remote sensing

MDPI

Article

A Step Forward in Hybrid Soil Laboratory Analysis: Merging Chemometric Corrections, Protocols and Data-Driven Methods

Bruno dos Anjos Bartsch [1], Nicolas Augusto Rosin [1], Uemeson José dos Santos [2], João Augusto Coblinski [3], Marcelo H. P. Pelegrino [1], Jorge Tadeu Fim Rosas [1], Raul Roberto Poppiel [1], Ednilson Batista Ortiz [4], Viviane Cristina Vivian Kochinki [4], Paulo Gallo [4], Eyal Ben Dor [5], Renan Falcioni [6], Marcos Rafael Nanni [6], João Vitor Ferreira Gonçalves [6], Caio Almeida de Oliveira [6], Nicole Ghinzelli Vedana Vedana [6], Renato Herrig Furlanetto [7,*] and José A. M. Demattê [1]

[1] Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Av. Pádua Dias 11, Piracicaba 13418-260, SP, Brazil; brunobartsch@usp.br (B.d.A.B.); narosin@usp.br (N.A.R.); marcelopelegrino@usp.br (M.H.P.P.); jorge.fimrosas@usp.br (J.T.F.R.); raulpoppiel@usp.br (R.R.P.); jamdemat@usp.br (J.A.M.D.)

[2] Federal Institute of Education, Science, and Technology of Pará, Campus Santarém, Av. Mal. Castelo Branco, Interventoria 621, Santarém 68020-820, PA, Brazil; uemeson.santos@ifpa.edu.br

[3] Institute of Soil Science and Plant Cultivation—State Research Institute, Czartoryskich 8, 24-100 Puławy, Poland; jcoblinski@iung.pulawy.pl

[4] ABC Foundation Laboratory, Castro 84165-250, PR, Brazil; ednilson@fundacaoabc.org (E.B.O.); viviane.vivian@fundacaoabc.org (V.C.V.K.); gallo@fundacaoabc.org (P.G.)

[5] Remote Sensing Laboratory, Geography Department, Porter School of the Environment and Earth Sciences, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 699780, Israel; bendor@tauex.tau.ac.il

[6] Department of Agronomy, State University of Maringá, Av. Colombo 5790, Maringá 87020-900, PR, Brazil; rfalcioni2@uem.br (R.F.); mrnanni@uem.br (M.R.N.); pg403805@uem.br (J.V.F.G.); pg55482@uem.br (C.A.d.O.); pg405864@uem.br (N.G.V.V.)

[7] Weed Science Laboratory, Gulf Coast Research and Education Center, University of Florida, Wimauma, FL 33598, USA

* Correspondence: re.herrigfurlane@ufl.edu; Tel.: +1-813-729-2184

check for updates

Citation: Bartsch, B.d.A.; Rosin, N.A.; dos Santos, U.J.; Coblinski, J.A.; Pelegrino, M.H.P.; Rosas, J.T.F.; Poppiel, R.R.; Ortiz, E.B.; Kochinki, V.C.V.; Gallo, P.; et al. A Step Forward in Hybrid Soil Laboratory Analysis: Merging Chemometric Corrections, Protocols and Data-Driven Methods. *Remote Sens.* **2024**, *16*, 4543. https://doi.org/10.3390/rs16234543

Academic Editor: Magaly Koch

Received: 22 October 2024
Revised: 21 November 2024
Accepted: 2 December 2024
Published: 4 December 2024

Abstract: The need to maintain soil health and produce more food worldwide has increased, and soil analysis is essential for its management. Although spectroscopy has emerged as an important tool, it is important to focus primarily on predictive modeling procedures rather than specific protocols. This article aims to contribute to a routine work sequence in a hybrid laboratory that seeks to provide the best data for its users. In this study, 18,730 soil samples from the state of Paraná, Brazil, were analyzed using three different laboratories, sensors and geometries for data acquisition. Thirty soil properties were analyzed, some using different chemical methodologies for comparison purposes. After a spectral reading, two literary protocols were applied, and the final prediction results were observed. We applied cubist models, which were the best for our population. The combination of different spectral analysis systems, with a standardized protocol using LB for the ISS detection of discrepant samples, was shown to significantly improve the accuracy of predictions for 21 of the 30 soil properties analyzed, highlighting the importance of choosing the extraction methodology and improving data quality, which have a significant impact on laboratory analyses, reaffirming spectroscopy as an essential tool for the efficient and sustainable management of soil resources.

Keywords: laboratory spectroscopy; pedometrics; soil quality; spectral standardization soil analysis

1. Introduction

Key soil performance indicators traditionally obtained from soil laboratory analysis are important tools to identify areas that need attention to improve soil management and health while increasing productivity. However, the increases in demand for soil properties and nutrient content information require time and expense. Thus, we should constrain the

number of samples to be analyzed. In this context, efforts have been made to make cheaper and faster auxiliary techniques for soil analysis [1,2].

The advent of visible and near-infrared spectroscopy (vis-NIR-SWIR, 400–2500 nm) has revolutionized soil analysis, offering rapid, economic, and environmentally friendly methods [3]. As a result, it is increasingly being used as an auxiliary to traditional soil analysis techniques in proxy models [4,5]. The main advantages of this approach include speed analysis, multiple sample scans, and a user-friendly operation. This permits the use of large sample sets at a lower cost compared to conventional methods [6,7]. Additionally, this technique is non-invasive, non-destructive, and "green", allowing the measurement of many samples for their several soil properties [8,9] in both field and laboratory conditions.

The global soil laboratory network GLOSOLAN [10] recently recognized the potential of vis-NIR-SWIR for soil analysis and added this technique to the vast list of soil analysis methods under initiative framework terms as GLSOLOAN-Spec. However, it also highlighted constraints such as the lack of common protocols, spectral libraries and skilled laborers. The diversity of equipment and lack of standards make data sharing difficult [11,12], mainly in proxy models where the absorption positions, intensity and shape of the spectrum are crucial [13]. Recently, to fill this gap, a new activity for soil standards and protocols was initiated in the framework of the IEEE Standard Association under the P4005 working group [14]

Generally, variance in reflectance results may be summarized into two main categories: systematic and non-systematic effects. Systematic effects are controlled factors, but they can vary from one laboratory to another depending on the inhouse protocol and instruments. To minimize these effects, the white reference sample (WR), measurement geometry, operator skills, sample preparation, and other factors must be kept constant and standardized. Non-systematic effects include noise and instabilities.

To address the challenge of aligning spectral measurements between laboratories, chemometric and data-driven approaches have been explored to reduce the reliance on traditional analytical methods. Pimstein et al. [15] proposed using an internal soil standard (ISS) to harmonize spectral data across different labs. Similarly, Jung et al. [16] suggested employing a common white reference to minimize spectrometer variations; however, this approach is impractical due to the inability to share a single reference among users and the potential for white reference deterioration over time. Ben Dor et al. [13] introduced a soil benchmark method using standard sand samples, characterized by their mineralogy and reflectance stability, to align spectra from various sources using two sand dunes from Southern Australia as soil standards (ISS), namely Lucky Bay (LB) and Wylie Bay (WB). Despite this, systematic factors must remain controlled for accurate standardization. In a more recent study, Jung et al. [16] applied unsupervised random forests to compute proximity matrices and clusters for spectrally similar soil samples, enabling internal quality control and effective outlier detection.

To advance soil spectroscopy techniques in laboratories, this study aims to (i) test spectral standardization using an internal benchmark, i.e., LB to generate correction factors and normalize variations; (ii) apply machine learning to the standardized spectra for segmentation, outlier detection, and removal; and (iii) use raw, standardized, and optimized spectra configurations in proxy models for predicting soil properties.

2. Materials and Methods

2.1. Soil Samples

In this study, we used 18,730 samples collected in the state of Paraná, Brazil (Figure 1). The samples came from the ABC Foundation [17] and Geotechnologies in the Soil Science Group [18]. The state of Paraná is an important Brazilian state regarding agricultural areas, parent material, including sandstone, and diabase. The main soils are Nitisols, Ferralsols, Arenosols and Litosols [19]. The climate is predominantly humid subtropical. According to the Köppen classification, the region is divided into a humid subtropical climate (Cfa) in

the north, center, west and southwest of the state, as well as the Ribeira River valley, and a humid temperate climate (Cfb) in the higher plateau lands [20].

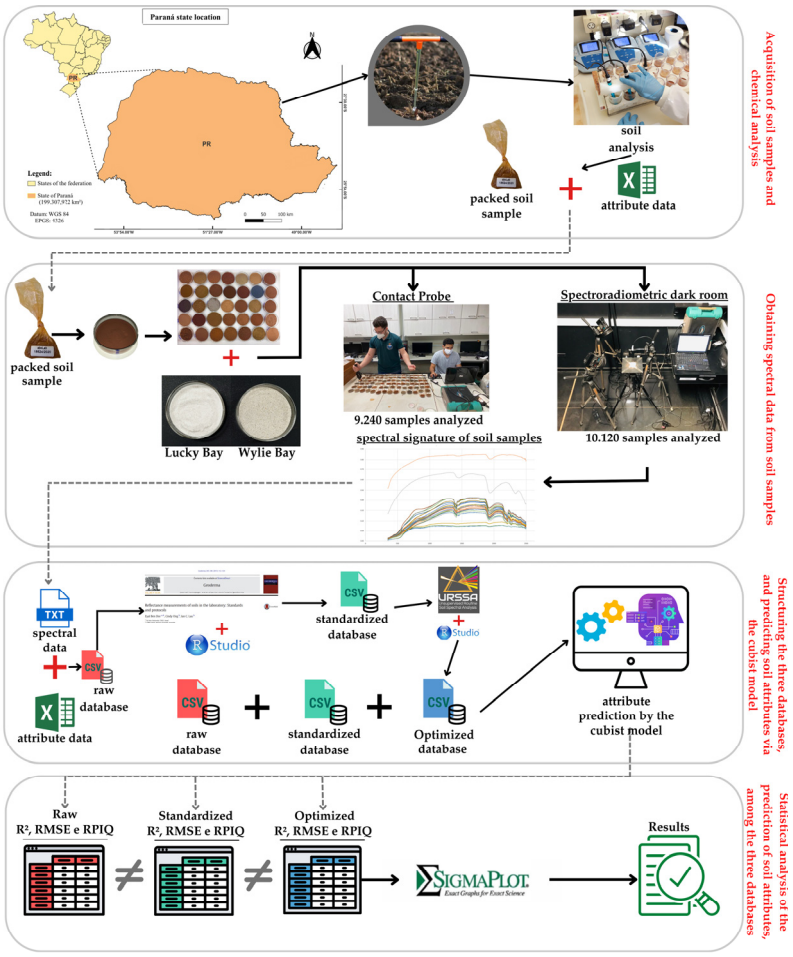

Figure 1. Schematic of the sequence from acquisition, sample preparation, readings, calibration, correction, and data segmentation for modeling.

2.2. Chemical and Physical Analysis

The soil samples were air-dried, grounded and sieved in a 2 mm mesh for physical and chemical determination (Appendix A) using traditional methods according to Teixeia et al. [21]. The chemical analyses involved the potential of hydrogen (pH) in SMP buffer solution and $CaCl_2$, organic matter (OM) (using the Walkey–Black method by colorimetry), and the following elements in Anion Exchange Resin (AER) and Mehlich-1 extractant (M): potential acidity (H + Al) and available aluminum (Al), phosphorus (P), potassium (K), calcium (Ca) and magnesium (Mg). Available micronutrients boron (B), copper (Cu), iron (Fe), manganese (Mn) and zinc (Zn) were determined in the DTPA extractant. Sand, silt and clay content were determined by the densimeter method, based on the sedimentation of soil constituents.

We also calculated the sum of the base (SB), Equation (2), cation exchange capacity (CEC), Equation (1), and base saturation (V%), Equation (3).

$$SB = Ca + Mg + K \quad (1)$$

$$CEC = SB + H + Al \quad (2)$$

$$V\% = (SB/CEC) \times 100 \quad (3)$$

2.3. Soil Spectroscopic Analysis

Each sample was placed in Petri dishes to reduce the influence of material roughness. Spectral data were acquired using spectroradiometers Fieldspec 3, 4 and Pro (Analytical Spectral Devices, ASD, Boulder, CO, USA). The three models cover the same spectral range from 350 to 2500 nm, but their resolutions differ: Fieldspec 3 has 3 nm resolution in the VIS/NIR (350–1000 nm) and 10 nm in the SWIR (1000–2500 nm); Fieldspec 4 offers higher resolution at 2 nm in the VIS/NIR and 8 nm in the SWIR; and Fieldspec Pro provides 1–3 nm in the VIS/NIR and 10 nm in the SWIR. All spectral analyses follow the standard spectral library analysis protocol [22], associated with the protocol by Ben Dor et al. [13] using a correction for the mother ISSs in CSIRO (LB and WB).

A total of 14,261 scans were performed with Fieldspec 3, 6918 with Fieldspec Pro, and 4469 with Fieldspec 4. The samples scanned with Fieldspec Pro were also shared with Fieldspec 3, resulting in a total of 18,730 samples scanned. We performed three replicates by rotating the sample to different positions, scanning it 100 times for each rotation and calculating the average spectrum. These three mean spectra were then averaged to obtain the final spectrum for the sample. The setup followed the long light geometry from Brazilian protocols [23], with a sensor positioned 8 cm away from the sample surface, capturing reflected light from a 2 cm^2 area. We corrected the splice points at 1000 and 1800 nm using the linear interpolation of 10 bands with the prospectr package [24] in R [25].

Modifications were employed to the experiments to emulate systematic effects, which included different environments and light (FieldSpec 3 and Pro were used in a dark room) and using fourteen different operators (experienced users and untrained students). And a standard white plate (Spectralon® (Labsphere, North Sutton, NH, USA)) was analyzed every 20 min to calibrate the device readings.

2.3.1. Standardized Dataset

To correct and standardize spectral data, protocols were used to employ the sand samples (LB and WB) proposed by Ben Dor et al. [13]. Given that the sensor calibration time lasts 20 min, spectral data collection was structured in batches of analyses. And in each lot, an analysis of two standard sands was performed, sands originating along the coast of Wylie Bay (WB) (90% quartz, 10% aragonite, and calcium carbonate) and Lucky Bay (LB) (99% quartz) in southwest Australia. Samples of WB and LB sands found in Western Australia are quite homogeneous and almost monomineral (quartz), consisting of a particle size and shape found in natural soils, being bright (LB) and semi-bright (WB) (covering a range large and dynamic spectrum) and having stable spectral characteristics. Moreover, these samples were chosen as spectral standards due to their spectral consistency, stability in space and time, mineralogical homogeneity, and representativeness in relation to the materials analyzed. These characteristics minimize error introduction and ensure that observed variations can be attributed to instrumental or operational factors, thereby facilitating data calibration and harmonization. The selection of these samples, which have a proven history as reliable standards in previous studies, strengthens the experimental design and allows for an independent replication and verification of the results obtained. Spectral correction Equation (4) is described below:

$$Rc\lambda = Ro\lambda \times (1 - (((S\rho\lambda - SBM\rho\lambda))/S\rho\lambda)) \quad (4)$$

where

$S\rho\lambda$ represents the WB and LB measured reflectance; $SBM\rho\lambda$ is the reflectance of the soil benchmark (SBM) ISS reference (the WB and LB sand samples); $Ro\lambda$ is the original soil sample reflectance; and $Rc\lambda$ is the corrected soil sample reflectance.

2.3.2. Optimized Dataset

As a step forward in spectrum correction, we additionally test a data-driven correction approach. To achieve this, the Unsupervised Routine Soil Spectral Analysis (URSSA) proposed by Poppiel et al. [26] was used, with the aim of removing outliers.

To reduce computational time, spectral data were resampled from 1 nm to 10 nm resolution using the prospectr package in R. Resampling is an important step to simplify the data volume while retaining the most relevant information. Next, the random uniform forest (RUF) technique was applied to the resampled data to identify spectral patterns and assess proximity between samples. In RUF, proximity is measured based on the frequency with which soil samples appear in the same terminal nodes (leaves) of decision trees. This proximity measure identifies samples that share similar characteristics in the spectral data.

After calculating the proximity matrix between samples, dimensionality reduction was performed through multidimensional scaling (MDS), which reduced the proximity matrix to two principal coordinates. This step aids in visualizing and separating samples within a two-dimensional space, facilitating the distinction of different patterns and potential outliers. With the principal coordinates obtained from MDS, Euclidean distances were calculated to identify each sample's nearest neighbors.

Spectral clustering was then conducted using the K-means algorithm, with a maximum of 10 clusters. Each cluster represents a group of spectrally similar samples. Outliers were identified by comparing the cluster results with traditional laboratory values. This allowed for the removal of samples that deviated significantly from expected laboratory values for the soil properties analyzed. The URSSA script, which integrates all these steps, was applied to 30 soil properties, ensuring a comprehensive and accurate correction. More details about this approach can be found in [26].

The technique of spectral resolution reduction decreases data volume while retaining the most relevant information. This simplification process facilitates analysis by eliminating spectral redundancies that may not significantly contribute to distinguishing soil characteristics. When combined with the random uniform forest (RUF) method, this approach enables the identification of relevant spectral patterns and the assessment of proximity between samples. RUF helps detect consistent spectral variations and selects the most informative spectral bands, which is particularly useful when dealing with data collected from different spectrometers and operators. By reducing the data to a subset of essential spectral characteristics, the impact of variations introduced by different operators and equipment is minimized.

Furthermore, by incorporating these resampled data into deep learning models, along with the use of the Internal Soil Standards (ISS) for standardization, it becomes possible to develop more robust models that are less sensitive to systematic errors. Resampling and feature selection make the models less dependent on noisy or redundant data, which are often introduced by differences in instruments and operators. Ultimately, this results in a more standardized and accurate spectral analysis, reducing systematic variability that would otherwise hinder data comparability across different measurement sets.

2.4. Predictive Modeling

Spectra in the Vis-NIR-SWIR range were used as input data for the models. For each soil property, the dataset was divided into training and validation through simple random sampling, where 70% of the samples were used for calibration and 30% for model validation. A simple 70–30% split for training and validation was chosen due to the use of the RUF algorithm for outlier removal, as this algorithm enhances model stability by reducing data variability and eliminating samples that could distort model fitting. This process allows the remaining data to be more representative of the overall dataset and, consequently, reduces the risk of overfitting.

The Cubist algorithm, provided by the Cubist package [27], frequently shows better results than other algorithms for soil spectroscopy [28,29] and is used in the R software [25]. It is a rule-based algorithm, derived from Quinlan's M5 model tree [30]. The Cubist algorithm

is particularly effective in soil spectroscopy as it combines the structure of decision trees with the robustness of multiple linear regression. Unlike other decision tree-based models that use the average of values at each terminal node, Cubist performs specific linear regressions at each node. This allows for a more precise modeling of complex relationships in spectral data, especially in cases with nonlinear interactions between spectral variables and soil properties. Additionally, Cubist incorporates the boosting technique, which adjusts the model through sequential iterations, reducing residual error and increasing accuracy [31].

3. Results

3.1. Analysis Description of Laboratory Data

The soil samples were analyzed using two extraction methods, Anion Exchange Resin (AER) and Mehlich-1 (M), to assess various physicochemical properties crucial for understanding soil fertility and composition. Figure 2 presents a comprehensive comparison of these methods through violin and box plots, revealing distinct patterns in nutrient availability and soil characteristics.

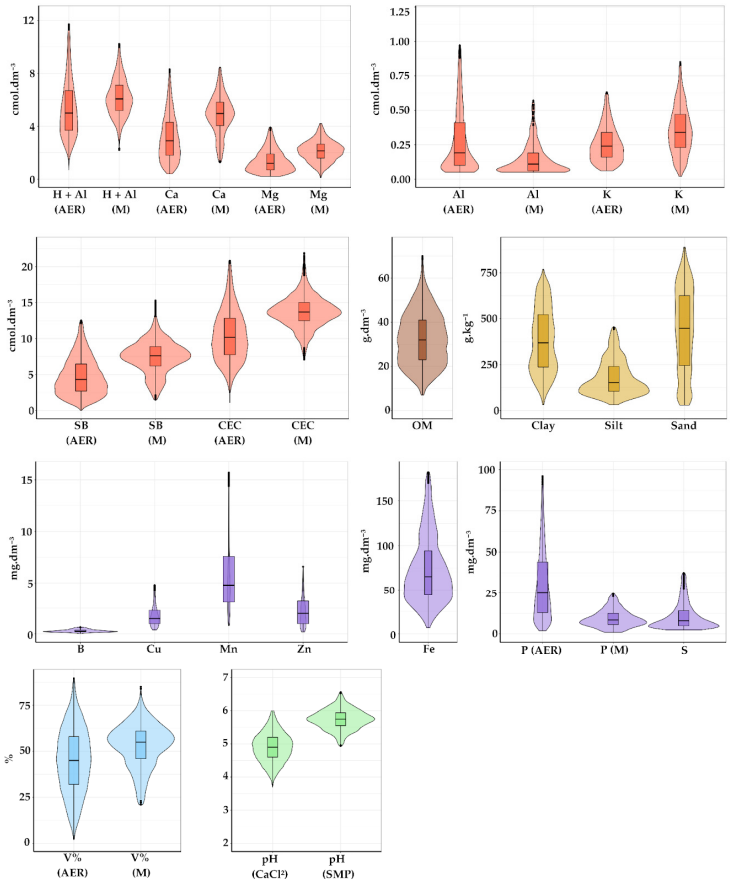

Figure 2. Violin plots show the full distribution of the data, while the box plots within them indicate the median, interquartile range, and potential outliers.

Organic matter content exhibited substantial variability or standard deviation (SD) (SD = 20.64 g.dm^{-3}), indicating diverse soil organic compositions across samples. Soil pH, measured in $CaCl_2$ and SMP, showed relatively low variability (SD = 0.47 and 0.32, respectively), with values generally ranging between 5 and 6, suggesting slightly acidic conditions. Notably, the AER method yielded higher variability in several properties compared to the M method, particularly for H + Al (SD_AER = 29.76 cmol$_c$.dm^{-3} vs. SD_M = 1.66 cmol$_c$.dm^{-3}) and exchangeable cations such as Ca (SD_AER = 22.46 cmol$_c$.dm^{-3} vs. SD_M = 1.59 cmol$_c$.dm^{-3}) and Mg (SD_AER = 11.39 cmol$_c$.dm^{-3} vs. SD_M = 0.82 cmol$_c$.dm^{-3}). Phosphorus availability differed markedly between methods, with AER extracting higher and more variable concentra-

tions (SD_AER = 36.66 mg.dm^{-3} vs. SD_M = 13.11 mg.dm^{-3}). The cation exchange capacity (CEC) showed high variability (SD_AER = 42.93 cmol$_c$.dm^{-3}, SD_M = 44 cmol$_c$.dm^{-3}), reflecting diverse soil textures, further evidenced by the considerable variation in clay (SD = 170.8 g.kg^{-1}), silt (SD = 103.34 g kg^{-1}), and sand (SD = 224.64 g.kg^{-1}) contents.

3.2. Correlation Between Soil Properties and Spectra of the Vis-Nir-Swir Region

To show the strength and direction of linear relationships between several soil properties and their medium spectrum, a Pearson correlation plot was used (Figure 3). In this plot, each soil property is represented along the Y axis, with colored cells indicating correlation values between the property and the spectrum in the X axis.

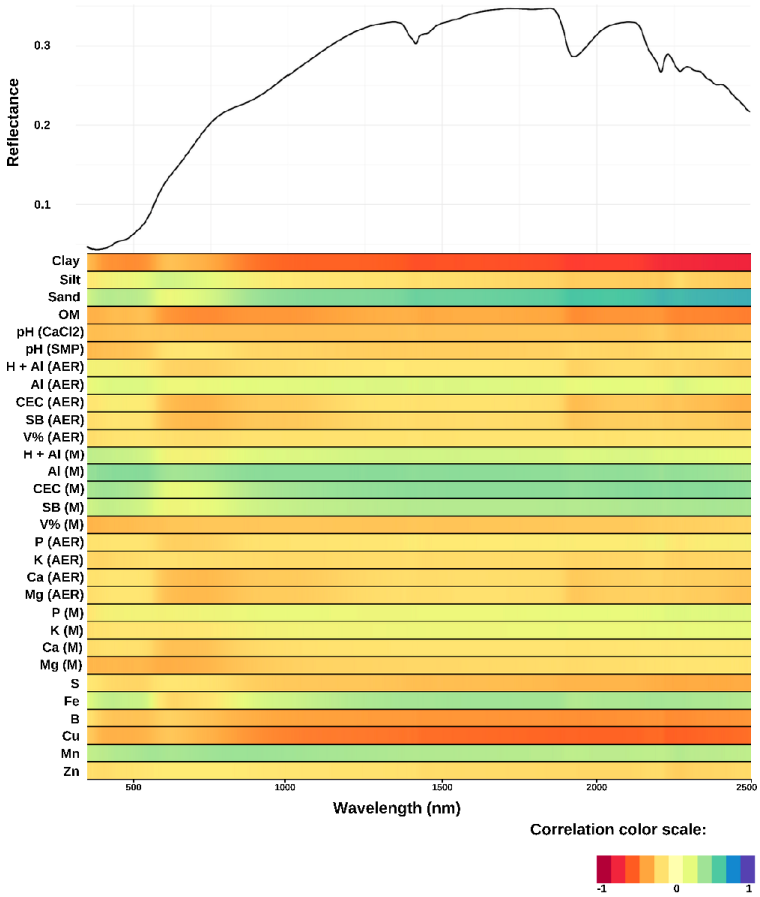

Figure 3. Pearson correlation plot of the linear relationships between soil properties and a medium spectrum.

The soil spectra had a significant correlation ($p < 0.01$) with clay ($-0.78 < r < -0.21$), sand ($0.08 < r < 0.64$), silt ($-0.20 < r < 0.18$), OM ($-0.45 < r < -0.24$), CEC (AER) ($-0.30 < r < -0.01$), CEC (M) ($0.09 < r < 0.41$) and V% (M) ($-0.28 < r < -0.14$). Higher levels of clay and organic matter in the soil are associated with the reduced reflection absorption of electromagnetic radiation, resulting in negative correlations. For micronutrients, B ($-0.41 < r < -0.08$) and Cu ($-0.54 < r < -0.17$) showed a negative correlation, likely due to their adsorption onto negatively charged clay surfaces and organic functional groups. In contrast, properties like H + Al, P, K, Ca, Mg, S, Mn and Zn showed no absorption in the VIS-NIR spectrum and low correlations with the spectrally detectable properties. However, due to their indirect correlations with properties absorbed in VIS-NIR (such as organic matter), it is possible to develop second-order regression models to estimate their values [32,33].

3.3. Standardized Database Predictive Models—ISS Approach

Figure 4 shows the mean and standard deviation of the VIS-NIR-SWIR spectra for the entire database, grouped into 10 clusters based on spectral similarity. While each cluster exhibits a unique spectral signature, similarities can be observed, particularly in the VIS and NIR regions. The spectra display typical soil absorptions, as noted in previous studies, at 400–600 nm, 1100 nm, 1400 nm, 1800–2000 nm, and 2200–2400 nm [30]. Absorptions in the visible range (400–780 nm) are linked to iron-containing minerals like hematite and goethite, though chromophores and organic matter can also cause absorption in this range [34].

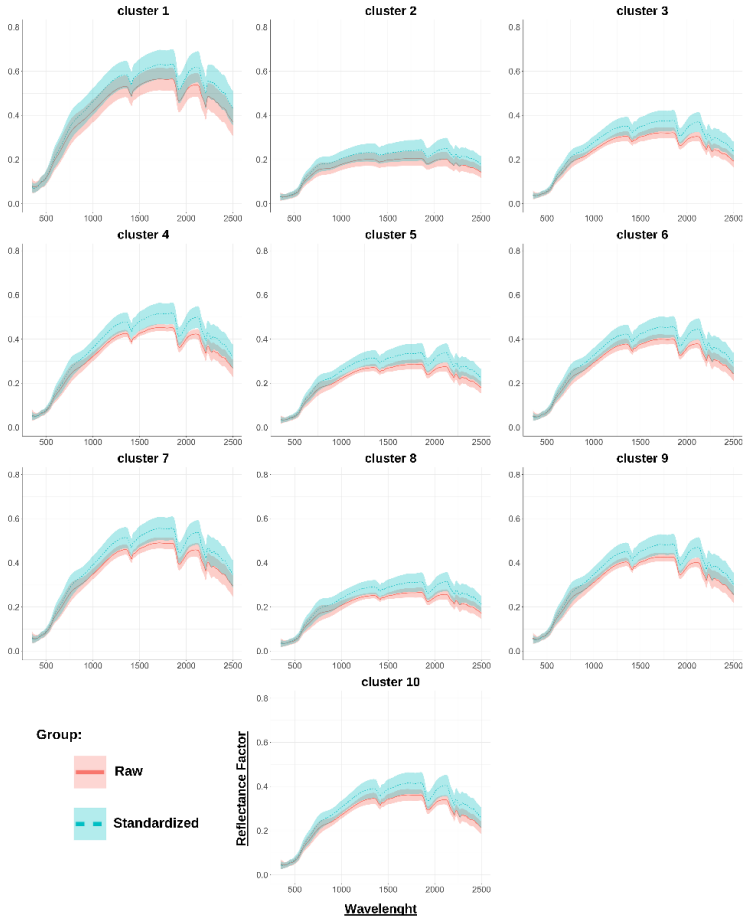

Figure 4. Cluster differences in the mean and standard deviation of both standardized (ISS) and raw spectra.

The correction did not affect spectral features tied to soil properties as the overall spectral shape remained consistent across all bands [35]. For soil properties with large sample sizes (over 9000 samples), such as OM, Al (AER), CEC (AER), SB (AER), and Ca (AER), standardized data outperformed raw data in RMSE and RPIQ metrics. These results underscore the importance of spectral standardization in integrating spectral libraries and improving model robustness [13].

3.4. Comparison of Results of Predictive Models Between Databases

All the results can be visualized in Table 1. Regarding chemical properties, for OM, the standardized treatment obtained the best model fit, with the mean value of R^2 reaching 0.87. However, the optimized treatment improved data dispersion with a mean RPIQ value of 1.68 and decreased the prediction error margin of 7.79 g.dm^{-3} in standardized treatment to 5.36 g.dm^{-3} in the optimized treatment. The improvement in RMSE and RPIQ metrics in the optimized treatment for predictive modeling can be attributed to a variance reduction [11], achieved by removing outliers. The same effect was observed

in the optimized treatment for the pH ($CaCl_2$), H + Al (AER) and Al (AER), with RMSE means equal to 0.33, 17.54 $mmol_c.dm^{-3}$ and 4.37 $mmol_c.dm^{-3}$, and RPIQ values of 0.9, 0.74 and 0.28 were obtained, respectively. Likewise, for the properties CEC (AER), SB (AER) and H + Al (M), there was an improvement in performance in all metrics in the optimized treatment, showing a mean R^2 value equal to 0.74, 0.52 and 0.38, and RPIQ values of 1.28, 0.85 and 0.74 were obtained, respectively. For the optimized treatment, the Al (M) property showed improved performance in predictive metrics, with an R^2 value of 0.38 and an RMSE of 0.4 $cmol_c.dm^{-3}$. In contrast, the raw treatment yielded better metrics for pH (SMP), V% (AER), and SB% (M), with mean R^2 values of 0.38, 0.45, and 0.63 and RPIQ values of 0.82, 0.99, and 0.11, respectively. For V% (M), the standardized treatment provided the best metrics, with an R^2 value of 0.38, an RMSE of 10.44%, and an RPIQ of 0.96. Finally, for CEC (M), the raw treatment yielded the highest mean R^2 value at 0.77 and an RPIQ of 0.07, although the optimized treatment showed a lower RMSE, with a mean value of 19.47 $cmol_c.dm^{-3}$.

Table 1. Accuracy parameters of the three proposed approaches, modeling by Cubist model.

				Mean Training			**Mean Validation**		
Soil Property	**Database**	**N**	**Outliers**	**R^2**	**RMSE**	**RPIQ**	**R^2**	**RMSE**	**RPIQ**
OM	Raw	18,716	-	0.93	5.33	1.70	0.84	8.19	1.1
	Std	18,716	-	0.93	5.21	1.74	0.87	7.79	1.16
	Opt	18,394	322	0.92	3.37	2.67	0.80	5.36	1.68
pH ($CaCl_2$)	Raw	18,731	-	0.79	0.22	1.39	0.43	0.35	0.85
	Std	18,731	-	0.78	0.22	1.37	0.43	0.35	0.85
	Opt	18,398	333	0.77	0.21	1.43	0.42	0.33	0.9
pH (SMP)	Raw	2363	-	0.77	0.16	1.28	0.38	0.24	0.82
	Std	2363	-	0.77	0.16	1.36	0.36	0.25	0.79
	Opt	2328	35	0.75	0.15	1.35	0.38	0.25	/
H + Al (AER)	Raw	16,368	-	0.82	12.69	1.10	0.54	20.33	0.69
	Std	16,368	-	0.82	12.59	1.11	0.53	20.44	0.69
	Opt	16,081	287	0.82	10.99	1.18	0.53	17.54	0.74
H + Al (M)	Raw	2363	-	0.78	0.77	1.16	0.36	1.34	0.66
	Std	2363	-	0.77	0.79	1.12	0.36	1.32	0.67
	Opt	2306	57	0.76	0.75	1.20	0.38	1.21	0.74
Al (AER)	Raw	9107	-	0.70	3.00	0.43	0.24	5.24	0.25
	Std	9107	-	0.71	3.11	0.42	0.23	4.77	0.27
	Opt	9068	39	0.72	2.74	0.44	0.21	4.37	0.28
Al (M)	Raw	565	-	0.70	0.25	0.21	0.31	0.42	0.12
	Std	565	-	0.72	0.25	0.20	0.31	0.38	0.13
	Opt	554	11	0.76	0.20	0.30	0.33	0.40	0.12
P (AER)	Raw	16,312	-	0.69	20.89	0.62	0.25	30.19	0.43
	Std	16,312	-	0.70	20.73	0.63	0.26	29.89	0.44
	Opt	16,148	164	0.74	13.65	0.88	0.35	22.20	0.54
P (M)	Raw	2489	-	0.56	9.13	0.36	0.08	9.80	0.31
	Std	2489	-	0.57	9.16	0.35	0.08	10.60	0.28
	Opt	2383	106	0.68	3.62	0.80	0.19	5.91	0.49
Ca (AER)	Raw	16,275	-	0.75	11.46	0.97	0.42	15.88	0.69
	Std	16,275	-	0.75	11.61	0.96	0.44	15.58	0.71
	Opt	16,146	126	0.81	8.09	1.36	0.52	12.80	0.86
Ca (M)	Raw	2363	-	0.80	0.71	1.32	0.48	1.15	0.82
	Std	2363	-	0.80	0.71	1.32	0.49	1.12	0.83
	Opt	2281	82	0.81	0.64	1.44	0.45	1.09	0.82
Mg (AER)	Raw	16,204	-	0.71	6.34	0.80	0.38	8.16	0.61
	Std	16,204	-	0.74	5.77	0.88	0.36	9.32	0.54
	Opt	16,125	79	0.79	4.25	1.18	0.47	6.82	0.73

Table 1. *Cont.*

				Mean Training			Mean Validation		
Soil Property	**Database**	**N**	**Outliers**	**R^2**	**RMSE**	**RPIQ**	**R^2**	**RMSE**	**RPIQ**
Mg (M)	Raw	2363	-	0.75	0.40	1.32	0.36	0.66	0.8
	Std	2363	-	0.75	0.41	1.30	0.37	0.63	0.84
	Opt	2325	38	0.75	0.37	1.45	0.37	0.62	0.87
K (AER)	Raw	15,946	-	0.70	0.89	0.91	0.24	1.41	0.57
	Std	15,946	-	0.71	0.84	0.95	0.25	1.49	0.54
	Opt	15,846	100	0.72	0.76	1.06	0.29	1.19	0.67
K (M)	Raw	2362	-	0.69	0.11	1.12	0.22	0.16	0.67
	Std	2362	-	0.70	0.10	1.18	0.24	0.18	0.68
	Opt	2351	11	0.70	0.10	1.12	0.23	0.16	0.7
SB (AER)	Raw	15,868	-	0.78	15.45	1.10	0.35	29.46	0.56
	Std	15,868	-	0.75	17.62	0.96	0.45	22.66	0.74
	Opt	15,736	132	0.82	11.99	1.38	0.52	19.46	0.85
SB (M)	Raw	2863	-	0.88	7.95	0.19	0.63	12.80	0.11
	Std	2863	-	0.86	8.27	0.18	0.60	13.58	0.11
	Opt	2583	254	0.85	7.31	0.18	0.57	12.96	0.1
V% (AER)	Raw	15,867	-	0.78	8.27	1.57	0.45	13.12	0.99
	Std	15,867	-	0.78	8.23	1.58	0.41	13.48	0.96
	Opt	15,845	22	0.78	8.29	1.57	0.43	13.15	0.99
V% (M)	Raw	2863	-	0.77	6.34	1.58	0.35	10.59	0.95
	Std	2863	-	0.75	6.57	1.52	0.38	10.44	0.96
	Opt	2784	79	0.75	6.27	1.44	0.29	10.75	0.84
CEC (AER)	Raw	15,868	-	0.89	13.48	1.83	0.55	32.16	0.76
	Std	15,868	-	0.86	16.44	1.52	0.66	24.45	1.01
	Opt	15,673	195	0.90	11.97	2.04	0.74	19.05	1.28
CEC (M)	Raw	2863	-	0.92	12.36	0.11	0.77	21.25	0.07
	Std	2863	-	0.91	13.03	0.11	0.76	22.27	0.06
	Opt	2583	280	0.91	11.12	0.11	0.75	19.47	0.06
S	Raw	5540	-	0.81	10.58	0.38	0.52	15.51	0.26
	Std	5540	-	0.82	10.05	0.40	0.47	17.78	0.23
	Opt	5345	195	0.73	7.54	0.53	0.34	12.19	0.33
Clay	Raw	2825	-	0.94	43.18	3.06	0.85	65.80	2.04
	Std	2825	-	0.94	41.32	3.20	0.82	73.33	1.8
	Opt	2792	33	0.94	42.04	3.12	0.82	71.18	1.85
Silt	Raw	2820	-	0.89	34.54	1.39	0.68	59.21	0.8
	Std	2820	-	0.89	35.20	1.37	0.67	57.61	0.84
	Opt	2780	40	0.89	34.87	1.38	0.69	56.34	0.85
Sand	Raw	2825	-	0.92	62.73	3.25	0.78	105.69	1.92
	Std	2825	-	0.92	63.34	3.21	0.80	102.07	1.98
	Opt	2820	5	0.92	62.56	3.24	0.79	103.77	1.96
B	Raw	2528	-	0.62	0.24	0.41	0.20	0.21	0.42
	Std	2528	-	0.55	0.26	0.38	0.17	0.25	0.34
	Opt	2438	90	0.76	0.08	0.90	0.44	0.12	0.59
Cu	Raw	2443	-	0.74	1.89	0.38	0.42	2.21	0.32
	Std	2443	-	0.72	1.91	0.38	0.43	2.36	0.3
	Opt	2342	101	0.88	0.86	0.82	0.73	1.24	0.56
Fe	Raw	2475	-	0.85	18.01	1.21	0.59	28.09	0.78
	Std	2475	-	0.85	17.46	1.30	0.58	29.99	0.7
	Opt	2406	69	0.86	16.31	1.35	0.59	27.33	0.77
Mn	Raw	2492	-	0.74	8.35	0.26	0.37	10.81	0.2
	Std	2492	-	0.77	6.69	0.32	0.20	16.63	0.13
	Opt	2278	214	0.79	4.93	0.39	0.50	7.53	0.26
Zn	Raw	2487	-	0.56	6.64	0.19	0.01	11.02	0.1
	Std	2487	-	0.49	8.51	0.15	0.05	7.81	0.17
	Opt	2329	158	0.57	4.88	0.26	0.01	15.01	0.08

N—number of samples; Std—standardized; Opt—optimized.

For the macronutrient properties, RMSE consistently showed better results in the optimized treatment across all nutrients. For P (AER), K (AER), Ca (AER), Mg (AER), P (M), and Mg (M), the optimized treatment significantly improved the R^2 value and RPIQ, with R^2 values between 0.19 and 0.52 and RPIQ values ranging from 0.49 to 0.87. For K (M) and P (M), the optimized treatment yielded similar effects as observed for OM, with a decrease in model fit but improvements in RMSE and RPIQ, where R^2 reached 0.23 and 0.34, and RPIQ was 0.70 and 0.33, respectively. Although P and K are not expected to show a direct spectral response in the Vis-NIR range, they can still be predicted with notable precision, as supported by the existing literature [36]. The only macronutrient property for which the optimized treatment did not yield the best potential was Ca (M). In this case, the standardized treatment showed the highest potential, with an R^2 value of 0.49 and an RPIQ of 0.83.

The optimized data treatment showed the best prediction for micronutrients B, Cu, Fe and Mn, with mean RMSE values ranging from 27.33 to 0.12 mg.dm^{-3} and the RPIQ ranging from 0.26 to 0.77. Alternatively, although low, the standardized treatment obtained the best result for Zn, with mean R^2 values of 0.05 and an RPIQ of 0.17. Notably, the model fit for B and Cu improved significantly, with R^2 values from 0.2 and 0.42 in the raw database to 0.44 and 0.73, respectively, in the optimized database. Micronutrients typically occur in soil at concentrations too low to be directly detected in soil spectra, even when they exhibit inherently strong spectral peaks. Consequently, predicting micronutrient levels relies largely on their correlations with other soil properties. For instance, elements like Cu, Mn, and Zn often correlate with carbonate minerals and, more prominently, with Fe-bearing minerals [37].

Our results indicate that combining different spectral analysis systems can overcome the limitations of each individual method. Although each system did not achieve optimal performance in isolation, the integration of a system that standardizes spectral signatures with another that automatically detects discrepant samples resulted in significant improvements in the accuracy of soil attribute predictions.

Of the 30 properties analyzed, 21 showed significantly better responses, demonstrating the effectiveness of the combined approach. Furthermore, the superiority of the Mehlich-1 extraction method over the Anion Exchange Resin method suggests that the choice of chemical analysis methodology is crucial for obtaining reliable results. This reinforces the importance of spectroscopy as a valuable tool for soil analysis, promoting a more efficient and sustainable management of soil resources, which is key to meeting the growing demand for food and preserving soil health.

4. Discussion

The literature indicates that the successful prediction of P and K using Vis-NIR-SWIR spectroscopy largely depends on co-variation with other spectrally active soil constituents, such as organic matter, iron oxides, clay mineralogy, and moisture [38]. For instance, models for total P often leverage correlations with IR-active compounds, while K models are influenced by the absorbance patterns of illite clay minerals in the SWIR range [39]. However, this co-variation varies across datasets, leading to fluctuations in prediction accuracy. In our study, the prediction of P and K remained imprecise, even after optimization. Similar challenges have been reported in previous studies, where P prediction has yielded low R^2 values (0.23–0.47), reflecting the complexity of P availability, which is influenced by sorption dynamics and soil pH [40–42].

Conversely, Ca prediction models have demonstrated higher accuracy. This is likely due to Ca's association with strongly IR-active materials, which provide distinct spectral features. For instance, [37] and [43] reported satisfactory Ca prediction models with R^2 values of 0.86–0.90 for untreated samples.

Organic matter (OM) also shows strong correlations with spectral reflectance, mainly in the VIS range, largely because of its influence on soil color. Elements such as Fe and Cu enhance red to brown hues, Mn darkens soils, and Zn lightens them, directly affecting

spectral signals [44] in this region as well. Despite this, micronutrient prediction using spectroscopy remains challenging. For example, Johnson et al. [43] reported unsatisfactory predictions for Mehlich-3-extractable Fe, Mn, Cu, and Zn, with R^2 values of 0.38, 0.32, 0.43, and 0.44, respectively, in soils from sub-Saharan Africa. Similarly, Terra et al. [33] observed low prediction accuracy for extractable Fe (R^2 = 0.39) and Zn (R^2 = 0.26) in Brazilian soils, though Mn showed moderate accuracy (R^2 = 0.54), and Cu achieved better accuracy (R^2 = 0.69) in the Vis-NIR range.

The spectral regions linked to micronutrient prediction have been explored extensively. Morón and Cozzolino [45] identified significant spectral bands for Fe, Cu, and Zn in the 400–800 nm range, around 1100 and 1400 nm, and between 1900 and 2500 nm in Uruguayan soils. They attributed the strong correlations in these bands to the adsorption characteristics of micronutrient oxides and the secondary adsorption of clay minerals. Similarly, Singh et al. [46] noted that absorption around 550 and 850 nm is related to the electronic transitions of Fe^{2+} and Fe^{3+}.

These findings underscore the variability in prediction accuracy for soil nutrients and the importance of considering both co-variations and specific spectral features to enhance model performance. To that end, the optimization enabled a better performance of the spectral-based prediction models.

5. Creation Sequence, Advantages and Limitations of a Hybrid Laboratory

Creating a low-environmental-impact hybrid laboratory for soil analysis involves integrating traditional laboratory methods (chemical) with advanced spectroscopy techniques. This procedure aims to optimize the analysis of soil properties, reducing the use of harmful chemical products and improving the efficiency of analytical processes. The hybrid analysis laboratory works as follows: Upon arrival at the laboratory (after sieving and air-drying), all soil samples (100%) undergo initial spectroscopy analysis and standardization using the ISS samples. The resulting spectra are then processed using a chemometric approach, which segments the population into subpopulations based on specific characteristics. The number of subpopulations depends on the samples received and can be determined using clustering or other appropriate techniques. A representative subset of these subpopulations is then selected for wet chemistry analysis, which serves as the basis for calibrating soil property prediction models. This hybrid approach enhances efficiency, with the recommended percentage of samples used for calibration varying according to the analytical context and methods applied. However, a recent study [1] suggests that a hybrid approach, where 80% of soil samples are analyzed by spectroscopic techniques and 20% by traditional ones, can optimize the quality of analytical control. This ratio of between 20 and 80% is an estimate and will depend on other factors, such as the total number of samples, the heterogeneity of the population, the total cost of the analyses and the accuracy of the prediction model [1]. The use of spectroscopic techniques is not limited to quantification alone. After completing the wet analysis, its quality must be evaluated, and any outliers should be reanalyzed before proceeding.

At this stage, choosing the right sensors for wet chemistry is crucial. These sensors must be able to measure multiple soil properties in a single reading with high precision, such as pH, organic matter and cation exchange capacity (CEC). At this stage, laboratories have a range of strategies they can adopt, with spectrometers operating in the visible, near-infrared and short-wave infrared (VIS-NIR-SWIR: 400–2500 nm) and mid-infrared (MIR: 2500–25,000 nm) ranges. VIS-NIR-SWIR spectroscopy allows multiple soil properties to be analyzed in a single spectral reading, speeding up the analysis process. Sensors can be used to analyze large volumes of samples quickly, which is ideal for applications in precision agriculture, large-scale monitoring and resource management [47]. In addition, the sensors are generally portable and easy to operate, allowing the analysis to be carried out directly in the field without the need for sophisticated laboratories. It is worth saying that the spectral measurement and analysis render the utilization of toxic reagents preserving the environment and can be termed as “green”.

It is essential to establish protocols for quality control and staff training, which include regular checks on the precision and accuracy of the sensors. This can be conducted by continuously comparing sensor results with the data obtained by traditional methods [13,15]. The use of reference samples such as the LB and WB as ISS for spectroscopy and repeated analysis are recommended practices. The same practices, however, should be implemented in wet chemistry analyses, which is still not considered in the VIS-NIR-SWIR analysis. Laboratory staff must be trained to operate the sensors, interpret the data generated and understand the calibration methodology [23]. This includes knowledge of chemometrics, which is essential for analyzing and interpreting spectral data. These aspects are in line with recent initiatives such as the ISO soil standard and protocol for VIS-NIR-SWIR measurement within the IEEE Standards Association (https://standards.ieee.org/, accessed on 30 August 2024) and the Global Soil Partnership in the World Soil Laboratory Network, using wet and spectral data (http://www.fao.org/global-soil-partnership/glosolan/en/, accessed on 30 August 2024), which aim to guarantee compliance with standards and the effectiveness of spectroscopic analyses, requiring laboratories to be certified by competent bodies, as well as demanding detailed reports and the creation of an accessible database for comparing results, in order to improve the quality of analyses, ensure environmental safety and increase public confidence in the information generated [15].

An important takeaway from this work is that the proposed standards and protocols for spectral analyses not only enable the better harmonization and follow-up of VIS-NIR-SWIR analyses using multiple soil spectral libraries, but they can also be used to assess the performance of individual laboratories from one measurement session to another. This approach facilitates the quality control of spectral data within each measurement scheme, ensuring consistency and reliability. It is also worth mentioning that a similar level of harmonization is strongly required for the same procedures.

Finally, sensor data should be integrated into a data management system that enables real-time analysis and visualization. This may involve software that interprets spectral readings and provides recommendations based on predictive models. Implementing a feedback system is highly recommended to support continuous improvement in analytical processes, allowing for regular result reviews, calibration model updates, and adaptations in laboratory practices. What follows is a film illustration of the system: https://youtu.be/kv0ATo-ddoc (accessed on 1 December 2024) and https://youtu.be/Gz4_FGVcEEw (accessed on 1 December 2024).

Challenges and Limitations

Despite the many advantages that laboratories offer, some challenges and limitations are still noticeable when implementing this routine. The first concerns the dependence on reference data, since the effectiveness of spectroscopic models depends on the accuracy of the data obtained by traditional methods [13,48]. To that end, it is important that all laboratories that generate SSLs use the same ISS such as LB and WB in order to allow the future harmonization of different users. If the reference data are inaccurate, this can compromise the quality of the predictions. Allied to this is the variability in results since different sensors can show variations in measurements, which require strict protocols for standardization, and our work has focused on overcoming this challenge [15,23]. Acquiring high-quality sensor infrastructure needed to integrate them into existing processes can be expensive, and this includes not only the cost of the equipment but also the need for specialized software for data analysis and maintaining a routine spectral and radiance calibration of the sensors. In addition, the results obtained by proximal sensors can show variability, which can be a challenge for data acceptance. The accuracy and reliability of sensors depend on factors such as environmental conditions and soil characteristics, which can lead to discrepancies with traditional methods [22]. Although proximal sensors are promising, they have limitations, such as the inability to detect certain soil properties with the same precision as traditional methods. For instance, this can limit their applicability in analyses that require high precision.

An important bottleneck is the construction of datasets. This is the main concern and doubt of users. How does one construct a dataset? Novais et al. [49] has indicated some strategies on how to do this. In summary, each wet laboratory should initiate this library and thus have its own one to continue the work. In the meantime, open access systems such as the Brazilian Soil Spectral Services (Braspecs) [50] and other upcoming ones will give support to this understanding.

Implementing a hybrid laboratory requires continuous research to optimize methods and guarantee the accuracy of predictive models. The change from traditional methods to more modern technologies can be met with resistance from laboratory staff and managers. This resistance can be based on a preference for established methods or a lack of confidence in new technologies. And finally, despite its advantages, spectroscopy cannot completely replace traditional methods, especially in analyses that require extreme precision or in specific soil conditions. Overcoming these obstacles requires careful planning, investments in training and technology, and a commitment to the continuous improvement of analytical processes. The successful integration of proximal sensors can lead to a more efficient and sustainable laboratory, capable of meeting the growing demands for soil analysis.

6. Conclusions

The adoption of a methodology for data standardization, as recommended by Ben Dor et al. [13], was essential to balance data from different sensors. In addition, the detection of outliers by the method proposed by Jung et al. [16] contributed to the construction of a more robust and reliable database for predicting soil properties.

We can state that the prediction of Vis-NIR-SWIR reflectance data on physicochemical properties, macronutrients and micronutrients was promising after applying these methodologies, where, of the 30 properties analyzed, 21 presented superior statistical models, offering cost and time efficiency.

The protocol sequence presented in this work is innovative and can serve as a basis for the development of a robust sequence of methods for the estimation of soil analysis, contributing to environmental health and reducing dependence on chemical products. As mentioned, this does not substitute the wet laboratory analysis but is integrated in the system. The bottleneck of discussions regarding chemistry analysis (Ca, K, Mg and others) showed here that it is possible to use this protocol but depends on the population of soil property distribution, number of samples, spectroscopic technique, and acceptable error of the laboratory.

The research community can achieve the development of this fantastic technique in the near future but needs to direct efforts towards the indicated bottlenecks. With the insertion of artificial intelligence, it is expected to reach improvements in discerning many soil properties.

Author Contributions: Conceptualization, J.A.M.D.; Methodology, B.d.A.B., N.A.R. and J.A.M.D.; Software, B.d.A.B. and N.A.R.; Validation, N.A.R., M.H.P.P., E.B.O., V.C.V.K., P.G., E.B.D., R.F., M.R.N., R.H.F. and J.A.M.D.; Formal analysis, N.A.R., U.J.d.S., J.A.C., M.H.P.P., J.T.F.R., R.R.P., E.B.O., V.C.V.K., P.G., E.B.D., R.F., M.R.N., J.V.F.G. and C.A.d.O.; Investigation, B.d.A.B., N.A.R., U.J.d.S., J.A.C., M.H.P.P., J.T.F.R., R.R.P., E.B.O., V.C.V.K., P.G., E.B.D., R.F., M.R.N., J.V.F.G., C.A.d.O. and N.G.V.V.; Data curation, B.d.A.B., U.J.d.S., J.T.F.R. and R.R.P.; Writing—original draft, B.d.A.B., N.A.R. and J.A.C.; Writing—review & editing, R.F., R.H.F. and J.A.M.D.; Visualization, R.H.F.; Supervision, J.A.M.D.; Project administration, J.A.M.D.; Funding acquisition, J.A.M.D. All authors have read and agreed to the published version of the manuscript.

Funding: The National Council for Scientific and Technological Development: Programa de Apoio à Fixação de Jovens Doutores no Brasil 168180/2022-7; Fundação Araucária: CP 19/2022—Jovens Doutores; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior: 001; Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP pn 2021/05129-8 for sensor financial support; and CEAGRE—Centro de Excelência em Agricultura Exponencial for financial support.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors gratefully acknowledge the ABC Foundation for providing soil samples and data essential for this study. We also thank the São Paulo Research Foundation (FAPESP) for their financial support through grants 2014/22262-8 and 2021/05129-8. Additional support from the Coordination for the Improvement of Higher Education Personnel (CAPES), the Araucária Foundation, the National Council for Scientific and Technological Development (CNPq) and Geotechnology in the Soil Science Group at ESALQ/USP (http://esalqgeocis.wixsite.com/english, accessed on 28 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the study design, collection, analyses, interpretation of the data, writing of the manuscript, or decision to publish the results.

Appendix A

Table A1. Descriptive Statistics of Soil Properties.

Soil Property	Number of Samples	Units	1° Q	Median	3° Q	Min	Max	Mean	Std Dev
OM	18,716	$g.dm^{-3}$	23	32	42	7	70	34.53	20.64
pH ($CaCl_2$)	18,731	-	4.6	4.9	5.2	3.7	6.1	4.92	0.47
pH (SMP)	2363	-	5.54	5.74	5.95	4.9	6.56	5.74	0.32
H + Al (AER)	16,368	$mmol_c.dm^{-3}$	38	52	70	7	117	58.16	29.76
H + Al (M)	2363	$cmol_c.dm^{-3}$	5.22	6.12	7.23	2.2	10.2	6.33	1.66
Al (AER)	9107	$mmol_c.dm^{-3}$	1.1	2.4	1.1	0.5	12.1	4.44	5.64
Al (M)	565	$cmol_c.dm^{-3}$	0.07	0.12	0.27	0.1	0.57	0.3	0.46
P (AER)	16,312	$mg.dm^{-3}$	14	27	47	2	47	35.88	36.66
P (M)	2489	$mg.dm^{-3}$	5.9	8.9	13.4	1	24.4	11.05	13.11
Ca (AER)	16,272	$cmol_c.dm^{-3}$	18	29	44	4	83	32.94	22.46
Ca (M)	2363	$cmol_c.dm^{-3}$	4.02	4.97	5.85	1.3	8.47	4.9	1.59
Mg (AER)	16,204	$cmol_c.dm^{-3}$	7	12	20	2	39	14.78	11.39
Mg (M)	2413	$cmol_c.dm^{-3}$	1.61	2.15	2.66	0.1	4.2	2.16	0.82
K (AER)	15,946	$cmol_c.dm^{-3}$	1.6	2.4	3.5	0.6	6.3	2.71	1.62
K (M)	2362	$cmol_c.dm^{-3}$	0.23	0.34	0.48	0	0.85	0.36	0.19
SB (AER)	15,868	$mmol_c.dm^{-3}$	27.2	43.8	66.4	0.3	125	49.87	33.82
SB (M)	2863	$cmol_c.dm^{-3}$	6.6	8.1	10.2	1.5	15.3	15.86	22.08
V% (AER)	15,867	%	32	45	58	2	97	45.06	17.54
V% (M)	2863	%	45	54	61	21	85	52.22	13.17
CEC (AER)	15,868	$mmol_c.dm^{-3}$	78.3	102.9	130.4	11	208	107.9	42.93
CEC (M)	2863	$cmol_c.dm^{-3}$	12.8	14.2	16.6	7.1	21.9	32.11	44
S	5540	$mg.dm^{-3}$	5	9	18	2.5	37.4	16.69	23.7
Clay	2825	$g.Kg^{-1}$	237	369	522	32	769	381.5	170.8
Silt	2820	$g.Kg^{-1}$	106	154	245.8	32	455	184.5	103.34
Sand	2825	$g.Kg^{-1}$	245	448	626	29	890	434.1	224.64
B	2528	$mg.dm^{-3}$	0.21	0.29	0.41	0.1	0.71	0.34	0.37
Cu	2443	$mg.dm^{-3}$	1.1	1.8	3.3	0.4	6.6	2.89	3.49
Fe	2475	$mg.dm^{-3}$	44	67	100	7	184	77.78	45.42
Mn	2492	$mg.dm^{-3}$	3.4	5.5	10.38	0.8	20.8	10.34	15.54
Zn	2487	$mg.dm^{-3}$	1.1	2.3	3.8	0.2	7.8	3.88	10.24

References

1. Demattê, J.A.M.; Dotto, A.C.; Paiva, A.F.S.; Sato, M.V.; Dalmolin, R.S.D.; de Araújo, M.D.S.B.; da Silva, E.B.; Nanni, M.R.; ten Caten, A.; Noronha, N.C.; et al. The Brazilian Soil Spectral Library (BSSL): A General View, Application and Challenges. *Geoderma* **2019**, *354*, 113793. [CrossRef]
2. Rossel, R.A.V.; Behrens, T. Using Data Mining to Model and Interpret Soil Diffuse Reflectance Spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
3. Li, S.; Viscarra Rossel, R.A.; Webster, R. The Cost-effectiveness of Reflectance Spectroscopy for Estimating Soil Organic Carbon. *Eur. J. Soil Sci.* **2022**, *73*, e13202. [CrossRef]
4. Coblinski, J.A.; Giasson, É.; Demattê, J.A.M.; Dotto, A.C.; Costa, J.J.F.; Vašát, R. Prediction of Soil Texture Classes through Different Wavelength Regions of Reflectance Spectroscopy at Various Soil Depths. *Catena* **2020**, *189*, 104485. [CrossRef]
5. Coblinski, J.A.; Inda, A.V.; Demattê, J.A.M.; Dotto, A.C.; Gholizadeh, A.; Giasson, É. Identification of Minerals in Subtropical Soils with Different Textural Classes by VIS–NIR–SWIR Reflectance Spectroscopy. *Catena* **2021**, *203*, 105334. [CrossRef]

6. Mohanty, B.; Gupta, A.; Das, B.S. Estimation of Weathering Indices Using Spectral Reflectance over Visible to Mid-Infrared Region. *Geoderma* **2016**, *266*, 111–119. [CrossRef]
7. 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. [CrossRef]
8. Wang, S.; Li, W.; Li, J.; Liu, X. Prediction of Soil Texture Using FT-NIR Spectroscopy and PXRF Spectrometry With Data Fusion. *Soil Sci.* **2013**, *178*, 626–638. [CrossRef]
9. Wight, J.P.; Ashworth, A.J.; Allen, F.L. Organic Substrate, Clay Type, Texture, and Water Influence on NIR Carbon Measurements. *Geoderma* **2016**, *261*, 36–43. [CrossRef]
10. FAO Spectroscopy. Available online: https://www.fao.org/global-soil-partnership/glosolan/en/ (accessed on 9 October 2024).
11. Demattê, J.A.M.; Bellinaso, H.; Araújo, S.R.; Rizzo, R.; Souza, A.B. Spectral Regionalization of Tropical Soils in the Estimation of Soil Attributes. *Revista Ciência Agronômica* **2016**, *47*, 589–598. [CrossRef]
12. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Adv. Agron.* **2015**, *132*, 139–159.
13. Ben Dor, E.; Ong, C.; Lau, I.C. Reflectance Measurements of Soils in the Laboratory: Standards and Protocols. *Geoderma* **2015**, *245–246*, 112–124. [CrossRef]
14. Karyotis, K.; Chabrillat, S.; Ben-Dor, E. P4005: The IEEE SA Standard and Protocol Scheme for Soil Spectral Measurement in Both Laboratory and Field. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 2493–2495.
15. Pimstein, A.; Notesco, G.; Ben-Dor, E. Performance of Three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions. *Soil Sci. Soc. Am. J.* **2011**, *75*, 746–759. [CrossRef]
16. Jung, A.; Götze, C.; Glässer, C. White-Reference Based Post-Correction Method for Multi-Source Spectral Libraries. *Photogramm. Fernerkund. Geoinf.* **2010**, *2010*, 363–369. [CrossRef]
17. Fundação ABC. Fundação ABC—Pesquisa e Desenvolvimento Agropecuário. Available online: https://fundacaoabc.org/ (accessed on 9 October 2024).
18. HOME GeoCiS GeoCiS. Available online: https://esalqgeocis.wixsite.com/english (accessed on 9 October 2024).
19. Schad, P. World Reference Base for Soil Resources—Its Fourth Edition and Its History. *J. Plant Nutr. Soil Sci.* **2023**, *186*, 151–163. [CrossRef]
20. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; de Moraes Gonçalves, J.L.; Sparovek, G. Köppen's Climate Classification Map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [CrossRef]
21. Teixeira, P.C.; Donagemma, G.K.; Fontana, A.; Teixeira, W.G. *Manual de Métodos de Análise de Solo*, 3rd ed.; Embrapa: Brasilia, DF, Brazil, 2017.
22. Demattê, J.A.M.; Dotto, A.C.; Bedin, L.G.; Sayão, V.M.; e Souza, A.B. Soil Analytical Quality Control by Traditional and Spectroscopy Techniques: Constructing the Future of a Hybrid Laboratory for Low Environmental Impact. *Geoderma* **2019**, *337*, 111–121. [CrossRef]
23. Romero, D.J.; Ben-Dor, E.; Demattê, J.A.M.; e Souza, A.B.; Vicente, L.E.; Tavares, T.R.; Martello, M.; Strabeli, T.F.; da Silva Barros, P.P.; Fiorio, P.R.; et al. Internal Soil Standard Method for the Brazilian Soil Spectral Library: Performance and Proximate Analysis. *Geoderma* **2018**, *312*, 95–103. [CrossRef]
24. Stevens, A.; Ramirez-Lopez, L. *Prospectr: Processing and Sample Selection for VisNIR Spectral Data*; Massachusetts Institute of Technology (MIT): Cambridge, MA, USA, 2013.
25. R Core Team R. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
26. Poppiel, R.R.; Paiva, A.F.D.S.; Demattê, J.A.M. Bridging the Gap between Soil Spectroscopy and Traditional Laboratory: Insights for Routine Implementation. *Geoderma* **2022**, *425*, 116029. [CrossRef]
27. Kuhn, M.; Quinlan, R. *Cubist: Rule—And Instance-Based Regression Modeling*; Rulequest Research Pty Ltd.: Empire Bay, NSW, Australia, 2021.
28. Chen, S.; Xu, H.; Xu, D.; Ji, W.; Li, S.; Yang, M.; Hu, B.; Zhou, Y.; Wang, N.; Arrouays, D.; et al. Evaluating Validation Strategies on the Performance of Soil Property Prediction from Regional to Continental Spectral Data. *Geoderma* **2021**, *400*, 115159. [CrossRef]
29. Moura-Bueno, J.M.; Dalmolin, R.S.D.; ten Caten, A.; Dotto, A.C.; Demattê, J.A.M. Stratification of a Local VIS-NIR-SWIR Spectral Library by Homogeneity Criteria Yields More Accurate Soil Organic Carbon Predictions. *Geoderma* **2019**, *337*, 565–581. [CrossRef]
30. Quinlan, J.R. AI '92: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (AI'92), Hobart, Tasmania, 16–18 November 1992; pp. 343–348.
31. Khaledian, Y.; Miller, B.A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. [CrossRef]
32. de Santana, F.B.; de Souza, A.M.; Poppi, R.J. Visible and near Infrared Spectroscopy Coupled to Random Forest to Quantify Some Soil Quality Parameters. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018**, *191*, 454–462. [CrossRef] [PubMed]
33. Terra, F.S.; Demattê, J.A.M.; Viscarra Rossel, R.A. Spectral Libraries for Quantitative Analyses of Tropical Brazilian Soils: Comparing Vis–NIR and Mid-IR Reflectance Data. *Geoderma* **2015**, *255–256*, 81–93. [CrossRef]
34. 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.

35. Gholizadeh, A.; Neumann, C.; Chabrillat, S.; van Wesemael, B.; Castaldi, F.; Borůvka, L.; Sanderman, J.; Klement, A.; Hohmann, C. Soil Organic Carbon Estimation Using VNIR–SWIR Spectroscopy: The Effect of Multiple Sensors and Scanning Conditions. *Soil Tillage Res.* **2021**, *211*, 105017. [CrossRef]
36. Xu, S.; Zhao, Y.; Wang, M.; Shi, X. Comparison of Multivariate Methods for Estimating Selected Soil Properties from Intact Soil Cores of Paddy Fields by Vis–NIR Spectroscopy. *Geoderma* **2018**, *310*, 29–43. [CrossRef]
37. 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. [CrossRef]
38. Mouazen, A.M.; Kuang, B.; De Baerdemaeker, J.; Ramon, H. Comparison among Principal Component, Partial Least Squares and Back Propagation Neural Network Analyses for Accuracy of Measurement of Selected Soil Properties with Visible and near Infrared Spectroscopy. *Geoderma* **2010**, *158*, 23–31. [CrossRef]
39. Kusuma, C.G.; Bhoomika, S.A.; Dharumarajan, S. Prediction of Soil Nutrients Using Visible-near-Infrared Reflectance Laboratory Spectroscopy. In *Remote Sensing of Soils*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 493–502.
40. Confalonieri, M.; Fornasier, F.; Ursino, A.; Boccardi, F.; Pintus, B.; Odoardi, M. The Potential of near Infrared Reflectance Spectroscopy as a Tool for the Chemical Characterisation of Agricultural Soils. *J. Near Infrared Spectrosc.* **2001**, *9*, 123–131. [CrossRef]
41. He, Y.; Huang, M.; García, A.; Hernández, A.; Song, H. Prediction of Soil Macronutrients Content Using Near-Infrared Spectroscopy. *Comput. Electron. Agric.* **2007**, *58*, 144–153. [CrossRef]
42. Reeves, J.B.; Smith, D.B. The Potential of Mid- and near-Infrared Diffuse Reflectance Spectroscopy for Determining Major- and Trace-Element Concentrations in Soils from a Geochemical Survey of North America. *Appl. Geochem.* **2009**, *24*, 1472–1481. [CrossRef]
43. Dunn, B.W.; Batten, G.D.; Beecher, H.G.; Ciavarella, S. The Potential of Near-Infrared Reflectance Spectroscopy for Soil Analysis—A Case Study from the Riverine Plain of South-Eastern Australia. *Aust. J. Exp. Agric.* **2002**, *42*, 607. [CrossRef]
44. Mozaffari, H.; Moosavi, A.A.; Ostovari, Y.; Nematollahi, M.A.; Rezaei, M. Developing Spectrotransfer Functions (STFs) to Predict Basic Physical and Chemical Properties of Calcareous Soils. *Geoderma* **2022**, *428*, 116174. [CrossRef]
45. Moron, A.; Cozzolino, D. Exploring the Use of near Infrared Reflectance Spectroscopy to Study Physical Properties and Microelements in Soils. *J. Near Infrared Spectrosc.* **2003**, *11*, 145–154. [CrossRef]
46. Singh, K.; Majeed, I.; Panigrahi, N.; Vasava, H.B.; Fidelis, C.; Karunaratne, S.; Bapiwai, P.; Yinil, D.; Sanderson, T.; Snoeck, D.; et al. Near Infrared Diffuse Reflectance Spectroscopy for Rapid and Comprehensive Soil Condition Assessment in Smallholder Cacao Farming Systems of Papua New Guinea. *Catena* **2019**, *183*, 104185. [CrossRef]
47. Demattê, J.A.M.; da Silva Terra, F. Spectral Pedology: A New Perspective on Evaluation of Soils along Pedogenetic Alterations. *Geoderma* **2014**, *217–218*, 190–200. [CrossRef]
48. Angelopoulou, T.; Dimitrakos, A.; Terzopoulou, E.; Zalidis, G.; Theocharis, J.; Stafilov, T.; Zouboulis, A. Reflectance Spectroscopy (Vis-NIR) for Assessing Soil Heavy Metals Concentrations Determined by Two Different Analytical Protocols, Based on ISO 11466 and ISO 14869-1. *Water Air Soil Pollut.* **2017**, *228*, 436. [CrossRef]
49. Novais, J.J.M.; Rosin, N.A.; Rosas, J.T.F.; Poppiel, R.R.; Dotto, A.C.; Paiva, A.F.S.; Bellinaso, H.; Albarracín, H.S.R.; Amorim, M.T.A.; Bartsch, B.D.A.; et al. The Brazilian Soil Spectral Library Data Opening. *Dokuchaev Soil Bull.* **2024**, *119*, 261–305. [CrossRef]
50. Demattê, J.A.M.; Paiva, A.F.D.S.; Poppiel, R.R.; Rosin, N.A.; Ruiz, L.F.C.; Mello, F.A.D.O.; 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. [CrossRef]