Article

An Evaluation of Different NIR-Spectral Pre-Treatments to Derive the Soil Parameters C and N of a Humus-Clay-Rich Soil

Kurt Heil [1,2,*] **and Urs Schmidhalter** [1]

1   Chair of Plant Nutrition, Technical University Munich, Emil-Ramann-Str. 2, D-85350 Freising, Germany; schmidhalter@wzw.tum.de
2   Chair of Agricultural Systems Engineering, Technical University Munich, Dürnast 4, D-85354 Freising, Germany
*   Correspondence: kheil@wzw.tum.de; Tel.: +49-8161-71-3906

Abstract: Near-infrared reflectance spectroscopy (NIRS) was successfully used in this study to measure soil properties, mainly C and N, requiring spectral pre-treatments. Calculations in this evaluation were carried out using multivariate statistical procedures with preceding pre-treatment procedures of the spectral data. Such transformations could remove noise, highlight features, and extract essential wavelengths for quantitative predictions. This frequently significantly improved the predictions. Since selecting the appropriate transformation was not straightforward due to the large numbers of available methods, more comprehensive insight into choosing appropriate and optimized pre-treatments was required. Therefore, the objectives of this study were (i) to compare various pre-processing transformations of spectral data to determine their suitability for modeling soil C and N using NIR spectra (55 pre-treatment procedures were tested), and (ii) to determine which wavelengths were most important for the prediction of C and N. The investigations were carried out on an arable field in South Germany with a soil type of Calcaric Fluvic Relictigleyic Phaeozem (Epigeoabruptic and Pantoclayic), created in the flooding area of the Isar River. The best fit and highest model accuracy for the C (Ct, Corg, and Ccarb) and N models in the calibration and validation modes were achieved using derivations with Savitzky–Golay (SG). This enabled us to calculate the Ct, Corg, and N with an R^2 higher than 0.98/0.86 and an ratio of performance to the interquartile range (RPIQ) higher than 10.9/4.1 (calibration/validation).

Keywords: near-infrared spectroscopy (NIRS); partial least squares regression (PLSR); pre-treatments; soil properties (C, N); soil heterogeneity

1. Introduction

Near-infrared reflectance spectroscopy (NIRS) has been used successfully to characterize many soil properties, including C, N, cation exchange capacity, and soil texture. In comparison to conventional laboratory methods, NIRS is inexpensive, fast, non-destructive, produces no chemical substances, and requires minimal sample preparation, making it highly attractive. Several soil properties can also be determined with a single scan. Some researchers have noted that this technique can estimate primary soil properties (such as total C, N, and exchange capacity), as well as secondary soil properties (such as the respiration rate and potentially mineralizable N) at the same time [1–6]. Other authors have described the application of additional physical and soil chemical properties [7–12].

Near-infrared spectroscopy measures the reflection of radiation with wavelengths from 750 to 2500 nm. The NIR spectra contain broad bands that are produced by reflections with overlapping wavelengths. Reflections measured by this type of spectroscopy mainly correspond to combinations of vibrational modes with chemical C–H, O–H, overtones, and N–H bonds [13].

The recorded reflectance produces spectra that are unique for this sample. The collected reflectance lines include data on the molecules' properties in the sample and provide essential information about the sample composition.

However, the raw spectra also contain some information that complicates the evaluation. For example, background interference, overlapping absorption bands from other constituents, noise (caused by matrix/environmental effects, water absorption bands, instrumental noise, extraneous light sources), and weak deflections can make the mathematical derivations between light and particles more difficult or sometimes even impossible.

Therefore, it is common practice to carry out pre-processing and special calibration/validation procedures to reduce such effects and obtain valuable information about the soil samples' properties.

According to Gholizadeh et al. [4–6], the main challenge here is limiting the application of NIRS for the evaluation of soil properties. Finding suitable pre-treatments and identifying calibration/validation strategies is also complicated because there are many techniques available. The high number of methods renders an overview difficult.

Although many pre-processing transformations have been used in NIRS, the choice between them is somewhat arbitrary, and little is known about how this choice affects the final prediction of soil properties [14,15]. Indeed, finding an appropriate method needs more time than the main calculations. Table 1 gives an overview of the main applied pre-processing techniques used in the last ten years.

According to Gholizade et al. [4–6] and Xie et al. [16], these techniques can be divided into four main groups: Scaling, smoothing, baseline correction, normalization. Dotto et al. [17] grouped these methods into only two categories: Scatter correction and spectral derivation. Scatter corrections are represented by continuum distance, normalization by range, standard normal variables, and multiplicative dispersion correction. The pre-processing of spectral derivatives includes Savitzky–Golay and Norris–Williams derivatives [17]. In the following section, a basic overview of these procedures is given.

The smoothing procedure is used to reduce noise and is also referred to as a measurement error. Commonly used smoothing methods include the moving average, median filters, and Savitzky–Golay transformation [18]. In current soil-related studies, the moving average is the dominant smoothing practice.

A scaling procedure is applicable when the intensity of different spectra needs to be scaled so that the specifications can be compared. Scaling produces a spectrum with a mean of zero and a standard deviation of 1. Here, the procedure corrects light scattering by centering and scaling each spectrum.

Using baseline removal, background signals that are far from the zero lines are removed [4–6,19,20]. These signals must be removed, as they will cause peak heights and peak areas to be incorrectly calculated. The methods here are derivatives and are some of the best methods for eliminating baseline effects. According to Gholizade et al. [3,5,6], a different method to eliminate scattering is a multiplicative signal correction (MSC), a transformation method applied to equalize multiplicative and/or additive scattering effects [16].

Normalization generates new spectral data via the creation of shifted and scaled data. With these normalized values, scattering effects (e.g., changing slopes) are eliminated.

Some soil parameters are frequently derived with vis-NIR spectroscopy. The most apparent properties are soil organic carbon (SOC), total soil carbon, soil organic matter (humus content), and clay. However, achievements of the quality of carbon are rare. The applied methods for deriving these and other soil properties are listed in Vasques et al. [14,15]. Mainly referring to the last decade of research, Table 1 outlines the target variables and the used pre-processing techniques.

This table provides the results from a query of the ResearchGate database to select publications featuring the application of pre-processing methods to predict soil properties over the last ten years.

Table 1. Overview of applied evaluations with near-infrared reflectance (NIR) spectroscopy, mainly during the last decade, including their target parameters, spectral ranges, pre-processing techniques, and results.

References	Target Parameter	Range Target Parameter	Spectral Range [nm]	Area of Investigation		Soil Sampling Depth	Tested Pre-Processing Technique	RMSEPcal/R^2 of Best Pre-Processing Technique
					Soil Type			
[14,15]	Ct	0.0072 (0.0017–0.268) %	350–2500	USA, North-central Florida,		180 cm	SG 1st/2nd derivative using a 1st/2nd-order polynomial; normalization; search window 1 to 9	0.17%/0.86 Norris gap derivative with a search window of 7
					Ultisol, Spodosol, Entisol			
[21]	SOC	10.2 (0–55.9) g kg^{-1}	350–2500	USA Texas		105 cm	FD	2.8–6.5 g kg^{-1}/0.55–0.86 FD
					calcareous, hyperthermic Aridic Ustifluvents			
[22]	SOC	0.9 (0.0–2.7) %	125–2500	Mozambique, Limpopo National Park		2.5–5 cm	Original spectra, original spectra with 1st derivative smoothed 1st derivative, MSC, MSC smoothed, SNV, MSC 1st derivative, MSC smoothed 1st derivative	0.32%/0.83 1st derivative of MSC
					Eutric leptosol, Calcaric cambisol, Arenosols/haplic Luvisol, Ferralic arenosol			
[23]	SOC	1.96 (0.21–6.87) %	350–2500	Brazil, Santa Catarina State		200 cm	CR, NBR, SNV, MSC, ASG, SMO, SG 1st derivative, 1st order polynomial; search window 9	0.48%/0.82 NBR
					Oxisol			
[17]	SOC	1.84 (0.17–4.83) %	350–2500	Brazil, Santa Catarina State		200 cm	Smoothing SG 1st order polynomial; search window 5 CR, DT, BR	0.32%/0.90 CR
	Clay	59.56 (20.9–78.5) %						0.84%/0.62 DT
	Silt	32.94 (16.5–78.0) %						5.26%/0.56 CR
	Sand	7.51(1.0–35.5) %		Oxisol				6.0%/0.33 CR
[24,25]	Ct	10.75 (0.15–55.25) g kg^{-1}	400–6000	Main Hawaiin Islands		20 cm	Normalization, SG 1st derivative	2.28%/0.95 normalization, SG 1st derivative
					Andisol, Oxisol, Inceptisol, Ultisol			
[26]	SOC	13.53 (0.79–30.73) g kg^{-1}	350–2500	China, Yixing		20 cm	SNV, FD, MSC, WD, SD, MC	2.48 g kg^{-1}/0.72 FD, SD
					Different parent materials			
[27,28]	SOC	15.38 (0.79–30.73) g kg^{-1}	410–2450	China, Yixing		10 cm	SG smoothing+SG, FD with SG smoothing, SD with SG smoothing, SNV, MC, MSC	2.78/0.73 g kg^{-1} SG
					Zhongxiang Honghu Anthrosol Luvisol Leptosol Gleysol Planosol			

Table 1. *Cont.*

References	Target Parameter	Range Target Parameter	Spectral Range [nm]	Area of Investigation		Soil Sampling Depth	Tested Pre-Processing Technique	RMSEPcal/R^2 of Best Pre-Processing Technique
				Soil Type				
[29]	Cu	5.5–92.2 mg kg^{-1}	399–2459	Czech Republic		0–30 cm	SNV, MSC, SG smoothing with a second-order polynomial fit and 11 smoothing points, FD, SD CR	4.0 mg kg^{-1}/0.78 FD
	Pb	0.9–55.9 mg kg^{-1}		Vertisol, and partly also Chernozem				2.97 mg kg^{-1}/0.68 FD
	Mn	41.6–1027.6 mg kg^{-1}						97.2 mg kg^{-1}/0.6 FD
	Cd	0.0–0.73 mg kg^{-1}						0.04 mg kg^{-1}/0.80 CR
	Zn	6.6–213.1 mg kg^{-1}						13.7 mg kg^{-1}/0.77 FD
[30,31]	Ct,	32.0 (1.33–523.3) g kg^{-1}	2000–6000 nm	USA, Florida		6 cm	MSC-1st D, SG-Quad, SG-1st D, SG-1st D-Quad, log10(1/x), log10(1/x) SG-1st D	0.23/0.95 log g kg^{-1} SG-Quad
	SOC	31.54 (1.33–523.27) g kg^{-1}		Spodosol, Entisol, Ultisol, Alfisol, istosol				0.23/095 log g kg^{-1} SG
	RC	21.13 (0.67–502.07) g kg^{-1}						0.31/0.93 log g kg^{-1} SG
	HC	0.88 (0.05–19.24) g kg^{-1}						0.3/0.86 log g kg^{-1} SG-Quad
[32]	SOC	0.85 (0.01–2.3) %	350–2500	Egypt, Northwestern Sinai peninsula		-	Original spectra, SG smoothing, 1st derivative with SG smoothing, 2nd derivative with SG smoothing, CR, SNV with detrending, MSC, extended MSC	0.19%/0.85 CR
	Clay	27.22 (0.02–54.3) %		Entisol, Aridisol				5.32%/0.90 CR
[33]	Nt	1.36 (0.21–2.79) g kg^{-1}	340–2511	China, Guangdong Province, Conghua District		7 cm	SG smoothing search window 10 with FD, SG smoothing search window 10 with SD, SG smoothing search window 10 with RL	21.61 g kg^{-1}/0.82
	Pt	0.75 (0.13–3.15) g kg^{-1}						42.84 g kg^{-1}/0.79
	Kt	10.55 (0.62–30.39) g kg^{-1}		-				25.42 g kg^{-1}/0.90; all transformations were used for N, P, and K

Table 1. *Cont.*

References	Target Parameter	Range Target Parameter	Spectral Range [nm]	Area of Investigation	Soil Sampling Depth	Tested Pre-Processing Technique	RMSEPcal/R^2 of Best Pre-Processing Technique
				Soil Type			
[34]	Cd	0.0–1.0 mg kg^{-1}	400–2400	China, Sichuan Province	-	MSC with SG smoothing 2nd polynomial search window 7, FD with MSC with SG smoothing 2nd polynomial search window 7, SD MSC with SG smoothing 2nd polynomial search window 7, RL MSC with SG smoothing 2nd polynomial search window 7	1.5 mg kg^{-1}/0.77
	Cr	0.0–1000 mg g^{-1}					295.7 mg kg^{-1}/0.73
	Pb	0.0–1000 mg kg^{-1}		Pots texture between sand and loess			67.17 mg kg^{-1}/0.71

Ct (total carbon); SOC (organic carbon); RC (recalcitrant carbon; HC (hydrolyzable carbon); OM (organic matter); RMSEPcal/R2 (root mean square error of calibration/coefficient of determination of calibration); SG (Savitzky–Golay); CR (continuum removal); DT (detrend transformation); RL (reciprocal Logarithm); BR (band ratio pre-processing techniques); SNV (standard normal variate); FD (first derivative); MSC (multiplicative signal correction); WD (wavelet detrending); SD (second derivative); MC (mean centering); NBR (normalization by range); SMO (smoothed); MSC-1st D (MSC followed by a five-window smoothing SG filter applied to 1st derivative spectra); SNV (standard normal variate correction, SG = 5 windows smoothing SG filter); SG-Quad (5 windows quadratic smoothing SG filter); SG-1st D (1st derivative-5 window smoothing SG filter); SG-1st D-Quad (1st derivative-5 window quadratic smoothing SG filter); log10(1/x) (absorbance (where x is the reflectance value); log10(1/x) SG-1st D (transformation to absorbance and then application of the 1st derivative—5 window smoothing SG filter); ASG (absorbance and then the application of a Savitzky–Golay derivative).

Within this catalogue of transformations, some methods were more frequently used than others, including smoothing with different search windows, normalization, standardization, and derivation with Savitzky–Golay [14,15,19,23–25,27,28,30–32,35,36].

The selection of multivariate techniques was similarly difficult. Upon the first examination, regressions seemed to be an appropriate statistical method for building reliable calibration models. However, the multicollinearity of highly correlated wavelengths required additional methods. These methods were mainly principal component regression (PCR) and partial least squares regression (PLSR). Instead of using the original data, these methods develop orthogonal (uncorrelated) linear combinations from the spectral variables (components or factors). Other methods have also been established for analyzing spectrometer readings, including support vector machine (SVM), soft independent modeling of a class anthology (SIMCA), cluster analysis (CA), discriminant partial least square (DPLS), K-nearest neighbors (KNN), and linear discriminant analysis (LDA). A comprehensive review of the relevant qualitative methods and their applications can be found in Cen and He [37].

The derivation of a calibration model also involves validations. A commonly used validation method is internal or cross-validation, where the validation dataset is a part of the calibration dataset. In contrast, external validations use an independent data sample set (test set) and provide more reliable and relevant estimates. In general, in both procedures, it is necessary to select a representative sample set that provides the most considerable amount of information for the calculation, including as many variations as possible that may occur in future samples [20].

Among all multivariate statistics, PLSR has been applied the most frequently. Dotto et al. [17] searched the scientific citation database Scopus and found that from 2006 to 2016, publications with PLSR appeared with a frequency of around 65% for predicting soil properties with NIRS. This is the reason why this procedure is applied in the present evaluation.

The soil investigated in this work occupies a particular position among the various soils investigated, and its high clay content in combination with high C-content has not yet been described in the literature. The most frequently examined soil types are Ultisol, Spodosol, Entisol, Cambisol, Luvisol, Oxisol, and Alfisol, as shown in Table 1.

Therefore, the specific objectives were the following: (i) To compare various preprocessing transformations of spectral data to determine their suitability for modeling C (Ct, Corg, Ccarb) and N using NIR spectra with PLSR, and (ii) to evaluate which wavelengths are most important for the prediction of the selected parameters.

2. Materials and Methods

2.1. General Description, Soil, and Physiography

The study area is approximately 1.0 ha in size and is located in Freising, 30 km north of Munich, Germany ($44°78'977''$ E, $535°87'77''$ N), in a formerly flooded area of the Isar River.

The average annual temperature in the area is approximately 7.8 °C, and the average annual precipitation is 800 mm. Holocene sediments are the predominant soil material. The composition of this flat area is a consequence of repeated floods with periods of soil-forming processes. According to the German Soil Survey [38] and the WRB, Calcaric Fluvic Relictigleyic Phaeozem is the dominant soil type (see Figures 1–3). The soil texture of the field is Tl [38] with loamy clay. The mean clay content is about 65%, with 1% sand. The fine and middle fractions are the dominant texture classes of silt in the area (Table 2).

Figure 1. Aerial image and picture of the position of the field (Pulling 8) in the landscape.

1. eAh
2. IIreGo-rAp1
3. IIreGo-rAp2
4. IIreGo-Ah
5. IIIfeAh°rGo
6. IIIreGo1
7. IIIreGo2
8. IVfeAh°rGo1
9. IVfeAh°rGo2
10. IV feAh°Gr1
11. IV feAh°Gr2

Calcaric Fluvic
Relictigleyic
Phaeozem
(Epigeoabruptic,
Pantoclayic)
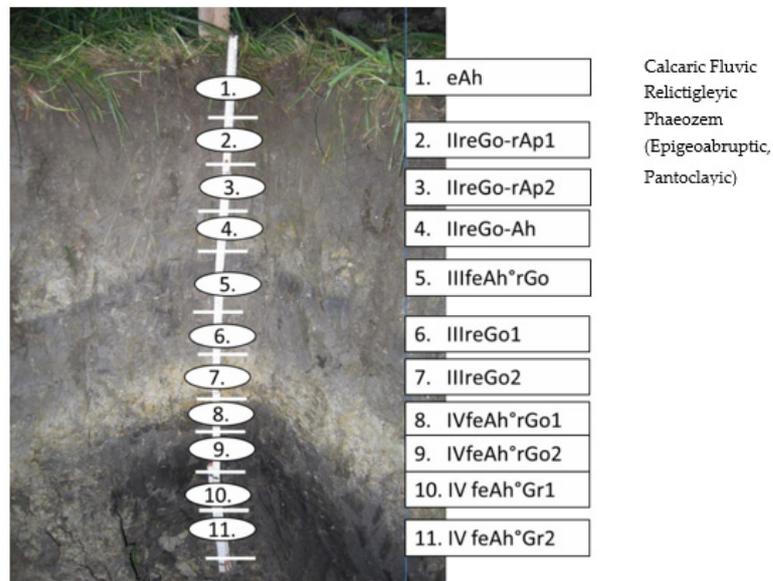
Figure 2. The soil profile of the investigated field Pulling 8 with a description of the names of the horizons and vertical discontinuities (abbreviations from [38]).
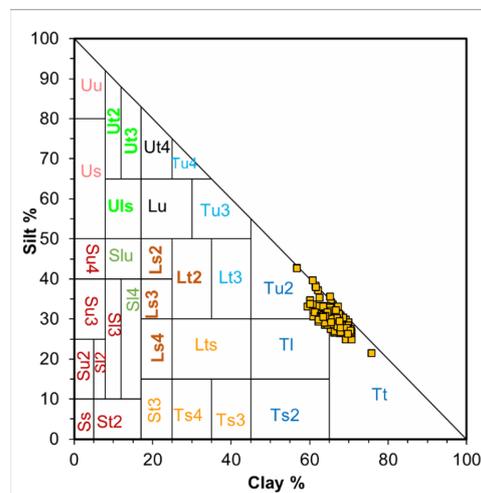

Figure 3. The particle size distribution of the whole soil profile Pulling 8.

Table 2. Chemical properties and particle size distribution of the soil profile Pulling 8.

	Horizons	Depth [cm]	Ct [%]	N [%]	Ccal [%]	Corg [%]	pH	Clay [%]	Sand [%]	Siltf/m [%]	Siltc [%]	Siltt [%]	Skeleton [%]	KA5	FAO/WRB
1	eAh	0–15	9.04	0.36	5.71	3.37	7.39	62.21	8.39	25.50	3.90	29.41	0.0	Lt3	C
2	II reGo-rAp1	15–20	8.00	0.21	6.03	1.98	7.42	66.75	4.68	26.36	2.22	28.58	0.0	Tt	C
3	II reGo-Ah2	20–30	7.66	0.21	5.69	1.97	7.47	67.28	2.57	27.81	2.33	30.15	0.0	Tt	C
4	II reGo-Ah	30–35	7.37	0.10	5.33	2.04	7.49	69.19	0.93	29.02	0.86	29.87	0.0	Tt	C
5	III feAh°rGo	35–45	8.16	0.12	6.87	1.28	7.51	65.77	0.54	32.62	1.06	33.68	0.0	Tt	C
6	III reGo1	45–54	9.33	0.07	8.65	0.68	7.53	62.08	0.71	34.10	3.12	37.22	0.0	Ts2	C
7	III reGo2	54–62	9.34	0.04	8.82	0.52	7.53	56.64	0.53	35.31	7.52	42.83	0.0	Tu2	C
8	IV feAh°Go1	60–70	9.03	0.08	8.05	0.99	7.53	61.33	0.21	35.19	3.26	38.45	0.0	Tu2	C
9	IV feAh°Go2	70–80	9.29	0.14	7.62	1.67	7.52	61.52	0.41	35.77	2.31	38.08	0.0	Tu2	C
10	IV feAh°Gr1	80–90	9.56	0.14	8.10	1.46	7.52	56.71	0.58	39.73	2.99	42.72	0.0	Tu2	C
11	IV feAh°Gr2	90–100	9.56	0.13	8.27	1.29	7.52	65.40	0.34	32.34	1.90	34.24	0.0	Tt	C

C_t-Carbon total; C_{cal}—C calcareous; $Silt_{f/m}$—Silt (fine+middle); $Silt_c$—Silt coarse; $Silt_t$—Sil$_t$ total.

2.2. Soil Sampling and Laboratory Analysis

A total of 120 soil samples were taken from 60 geo-referenced positions down to a depth of 1 m in April 2011 (Figure 1). Soil samples were collected using a regular grid of 10 × 25 m or following the experimental plots' borders.

Samples were air-dried at room temperature, cleaned of their visible plant residues, and then sieved using a 2-mm mesh sieve.

All samples were subjected to determination of their soil texture (clay, sand siltfine+ middle (2–20 μm), siltcoarse (20–63 μm), soil pH, C (Ct, Ccalc, Corg), and N). Soil carbon was analyzed by the dry combustion method using a Costech ECS 4010 Carbon–Nitrogen analyzer (Costech Analytical Technologies Inc., Valencia, CA, USA).

The soil texture was analyzed according to VDLUFA [39].

The soil samples' spectral reflectance was determined using an FT NIRS Bruker Vector 22/N (Ettlingen, Germany) with a spectral range of 833–2703 nm. The soil samples were distributed homogeneously in glass Petri dishes (9 cm diameter). The constant rotations captured large areas of the sample. The PbS detector scanned an area approximately 2 cm in diameter. A metal stamp (822 g) was used to compress the soil in the dishes to make the sample density uniform and avoid the influence of external light.

During the measurements, a gold standard plate was scanned every 40 min during calibration. Three replications were recorded from each sample. In this way, obviously wrong measurements were identified and eliminated during the measurements. Correct readings were averaged and used for the calculations.

2.3. Statistical Data Analysis

2.3.1. Pre-Procession Techniques for Spectral Data

Pre-processing techniques and multivariate statistical methods were applied using the Unscrambler 10.5 [40]. All combinations of pre-processing that were tested and compared are listed in Table 3.

Table 3. List of all applied combinations of pre-processing techniques for the spectral data.

Smoothing	Search Window	Derivation	Additional Technique	Abbreviation
None	0	Raw data	None	None-0-raw-none
			Standard normal variate	None-0-raw-SNV
			Multiplicative scatter correction	None-0-raw-MSC
			Standard normal variate and mean centering	None-0-raw-SNV+mc
			Standard normal variate and detrending, second-order polynomial	None-0-raw-SNV+det
Smoothing Savitzky–Golay derivative	3	Savitzky–Golay 1st derivative, 1st-order polynomial (within the command "Transform > Derivative > SG" in Unscrambler)	None	SG3-SG1-none
			Standard normal variate	SG3-SG1-SNV
			Multiplicative scatter correction	SG3-SG1-MSC
			Standard normal variate and mean centering	SG3-SG1-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG3-SG1-SNV+det
Smoothing Savitzky–Golay derivative	3	Savitzky–Golay 2nd derivative, 2nd-order polynomial (within the command "Transform > Derivative > SG" in Unscrambler)	None	SG3-SG2-none
			Standard normal variate	SG3-SG2-SNV
			Multiplicative scatter correction	SG3-SG2-MSC
			Standard normal variate and mean centering	SG3-SG2-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG3-SG2-SNV+det
Moving average	11	Raw data	None	MA11-raw-none
			Standard normal variate	MA11-raw-SNV
			Multiplicative scatter correction	MA11-raw-MSC
			Standard normal variate and mean centering	MA11-raw-SNV+mc
			Standard normal variate and detrending, second-order polynomial	MA11-raw-SNV+det
Moving average	25	Raw data	None	MA25-raw-none
			Standard normal variate	MA25-raw-SNV
			Multiplicative scatter correction	MA25-raw-MSC
			Standard normal variate and mean centering	MA25-raw-SNV+mc
			Standard normal variate and detrending, second-order polynomial	MA25-raw-SNV+det
Savitzky–Golay, 0-order polynomial (within the command "Transform > Smoothing > SG" in Unscrambler)	11	Raw data	None	SG11-raw-none
			Standard normal variate	SG11-raw-SNV
			Multiplicative scatter correction	SG11-raw-MSC
			Standard normal variate and mean centering	SG11-raw-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG11-raw-SNV+det
Savitzky–Golay, 0-order polynomial (within the command "Transform > Smoothing > SG" in Unscrambler)	25	Raw data	None	SG25-raw-none
			Standard normal variate	SG25-raw-SNV
			Multiplicative scatter correction	SG25-raw-MSC
			Standard normal variate and mean centering	SG25-raw-SNV+mc

Table 3. *Cont.*

Smoothing	Search Window	Derivation	Additional Technique	Abbreviation
			Standard normal variate and detrending, second-order polynomial	SG25-raw-SNV+det
Savitzky–Golay 1st derivative, 1st-order polynomial (within the command "Transform > Derivative > SG" in Unscrambler)			None	SG11-SG1-none
			Standard normal variate	SG11-SG1-SNV
			Multiplicative scatter correction	SG11-SG1-MSC
	11		Standard normal variate and mean centering	SG11-SG1-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG11-SG1-SNV+det
Savitzky–Golay 2nd derivative, 2nd-order polynomial (within the command "Transform > Derivative > SG" in Unscrambler)			None	SG11-SG2-none
			Standard normal variate	SG11-SG2-SNV
			Multiplicative scatter correction	SG11-SG2-MSC
	11		Standard normal variate and mean centering	SG11-SG2-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG11-SG2-SNV+det
Savitzky–Golay, 1st derivative, 1st-order polynomial (within the command "Transform > Derivative > SG" in Unscrambler)			None	SG25-SG1-none
			Standard normal variate	SG25-SG1-SNV
			Multiplicative scatter correction	SG25-SG1-MSC
	25		Standard normal variate and mean centering	SG25-SG1-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG25-SG1-SNV+det
Savitzky–Golay, 2nd derivative, 2nd-order polynomial (within the command "Transform > Derivative > SG" in Unscrambler)			None	SG25-SG2-none
			Standard normal variate	SG25-SG2-SNV
			Multiplicative scatter correction	SG25-SG2-MSC
	25		Standard normal variate and mean centering	SG25-SG2-SNV+mc
			Standard normal variate and detrending, second-order polynomial	SG25-SG2-SNV+det

SG3, SG11, SG25-Savitzky-Golay, search windows smoothing points 3, 11, and 25; SG1, SG2-Savitzky-Golay, derivative order 1 or 2, polynomial order: 1; SNV-standard normal variate; MSC-multiplicative scatter correction; det-detrending; mc-mean centering.

To determine the transformation that delivers the most accurate fit, 55 combinations were tested. All selected transformations are frequently cited in the literature [14,15,22–25,30–33,35,36,41].

The compilation of pre-processing procedures in Table 3 is structured into levels that are built upon each other, including smoothing, search window, derivation, and additional techniques.

For smoothing, the readings here were smoothed with different levels. In the first case, the search window was 3, and in the other case, the window was 11 or 25. The derivation techniques of the Savitzky–Golay first derivative, second-order polynomial, Savitzky–Golay second derivative, and second-order polynomial was chosen. The same additional procedure was always selected within each smoothing, search window, and derivation variation (raw data, mean centering, standard normal variate, multiplicative scatter correction, standard normal variate, and standard normal variate and detrending with a second-order polynomial). The first was applied as a control treatment, where the raw reflectance values were not pre-treated.

Such pre-treatment procedures have already been described in the literature [20].

Smoothing was used to reduce background or instrumental noise. The expression of 3 smoothing points indicated 1 point on the right and 1 point on the left of the smoothed point. Eleven smoothing points described 5 points on the left and 5 points on the right of the smoothed reading.

2.3.2. Partial Least Square Regression (PLSR)

After applying the pre-processing techniques, the next step was to build prediction/calibration models with the soil's physical and chemical properties as dependent target variables.

This PLSR extracted a set of components from the spectra relevant to the target variables [41]. The independent variables were the untreated and pre-treated readings of the wavelengths.

2.3.3. Indices for Evaluation PLSR

Calibration and validation models were assessed to determine their predictive qualities using different indices. Model calibration was performed based on the differences between observed and predicted values after fitting the model using the entire data set. Model validation and the performance of the pre-processing procedures and multivariate statistical methods were evaluated using randomly selected data subsets. For this purpose, the data set was randomly divided into 20 groups, with 15 observations in each group. Ninety percent of the groups were used for model training, and one group was the validation set (prediction set) for model testing. This division in cross-validation and prediction sets was replicated two times.

The applied indices were the coefficient of determination (R^2), the root mean square error (RMSE), the mean of the difference between the measured and calculated values (Bias), the standard error of prediction (SEP), the ratio of the performance deviation for predicted data (RPDpre), and the ratio of performance deviation for the measured data (RPDlab) [42,43]. R^2 (the coefficient of determination) is a commonly used measure for the goodness of fit. The RMSE is also commonly used to measure differences between the calculated and observed values from the measured model. The advantage of RMSE is its use of the same data units:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_l - \overline{y_l})^2}{\sum_{i=1}^{n} (y_i - \overline{y_l})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=t}^{n} (\hat{y}_i - y_i)^2}{n}}. \tag{2}$$

Bias is the mean difference between the measured and calculated values:

$$Bias = \frac{y_i - \hat{y}_i}{n} \tag{3}$$

This is defined as the difference between the predicted value and the observed value. A positive value means that the model has overestimated, while a negative value represents an underestimation.

$$SEP = \text{Standard error of prediction} \quad SEP = \sqrt{\sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{n}} = SD * (\hat{y}_i - y_i). \tag{4}$$

While the RMSE values indicate the accuracy of the calculation, the SEP index shows the prediction's precision. SEP squared is approximately equal to the RMSE squared minus the bias squared.

$$RPD_{lab} = \text{ratio of the performance deviation for the measured data} = RPD_{lab} = \frac{SD_{lab}}{RMSE} \tag{5}$$

$$RPD_{pred} = \text{ratio of the performance deviation for the predicted data} = RPD_{pred} = \frac{SD_{pred}}{RMSE} \tag{6}$$

Here, ȳ is the mean of the observed value, ŷ is the predicted value, y is the observed value, and n is the number of samples with i equal to 1, 2, ... , n. SD: Standard deviation observed (lab, measured) and predicted (pred).

While bias, RMSE, and SEP use the same measurement units as the component, the R^2 and RPD values are dimensionless so that they can be compared similarly between models for different components/properties. Higher values for the RPD indicate more accurate models. The RMSE should be lower than the SD, and the RPD should ideally be four or higher (Table 4). A large RMSE causes low RPD values compared to the SD or the low variability of the reference values.

Table 4. Levels of merit for the ratio of the performance deviation (RPD) for the application of NIR spectroscopy [42,44,45].

RPD Value	Classification	Application
0.0–1.9	Very poor	Not recommended
2.0–2.4	Poor	Rough screening
2.5–2.9	Fair	Screening
3.0–3.4	Good	Quality control
3.5–4.0	Very good	Process control
>4.1	Excellent	Any application

The ratio of performance to the interquartile range (RPIQ) is a less-often-used index. According to Bellon–Maurel et al. (2010), the soil samples often show skewed distributions instead of normal distributions. The quartiles in the RPIQ better represent the spread of the population.

$$RPIQ = (Q75–Q25)/RMSE \qquad (7)$$

RPIQ is the difference between the first and third quartiles (Q75–Q25), where Q25 is the value in 25% of the samples, and Q75 is the value in 75% of the samples.

2.3.4. Detection of Important Variables

Variables that have a significant effect are considered essential variables. In addition, the variables involved in important interactions are also important, even if their main effects are negligible.

In such cases, Martens' automatic uncertainty test indicates the significant variables in the multicomponent model [40].

3. Results and Discussion

3.1. Characteristics of the Spectral Reflectance Curves

The reflection curves are represented by the different soil properties, which differ in their position and composition. The shapes of the curves vary between 0.2 and 0.6. Changes of a darker and more whitish color with distinctly different absorbance characteristics [43] are observable here (Figure 4).

The same figure shows that the spectral reflectance increased gradually with wavelengths from 900–1400 nm. The reflectance reached its first reflection peak between 1390 and 1400 nm. Three more significant reflection peaks appeared at 1450, 1860, and 2120 nm. The two reflection peaks at 1420 and 1930 nm were also characteristic of soil curves. The reflection curves were affected by color, particle size, moisture, and mineral composition (especially the presence of iron). The soil reflectance curves in the NIR region for different soil types, along with the most critical parts for water and carbon and their responsible chemical groups, are shown in Vasques et al. [14,15].
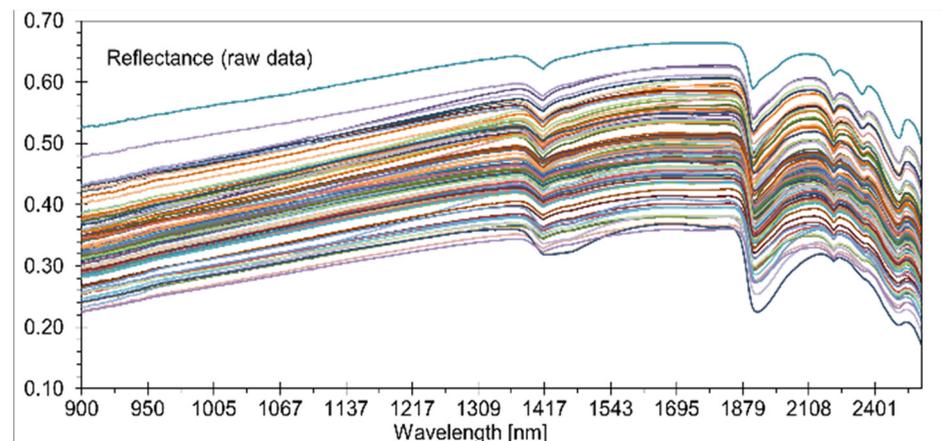
Figure 4. Spectral reflection curves for all soil samples.

According to Demattê et al. [46], Stenberg et al. [47,48], and Whiting et al. [49], the bands at 1400, 1900, and 2000 nm are produced by OH-groups and water molecule movements, and reflection at 2200 nm indicates the existence of kaolinite and other silicates [50]. Ben–Dor [51] described three major areas for clay minerals in general and smectite minerals, particularly around 1300–1400, 1800–1900, and 2200-2500 nm. According to Dotto et al. [17], the reflections around 1400–1900 and 2200–2500 nm for 1:1 minerals (kaolinite) are relatively strong, while the signal at 1900 nm is very weak. The soils from this investigation were formed from complex alluvial sedimentation interrupted by the building of A-layers. The spectral behavior of the soil horizons and vertical discontinuities varied mainly according to the Ct content, particle size distribution (mainly sand distribution and clay mineral composition), and the condition of Fe oxides (oxides and oxyhydroxides). Fe oxides exist independent of the degree of reduction in these gleyic soils. To clarify this context, in Figure 5, the soil profile is contrasted with the spectral signatures for the 11 soil horizons and vertical discontinuities. The following findings can be derived:

- Variation of the Corg content between the surface and subsurface parts caused a distinct differentiation in spectral shapes. Higher Corg content yielded reduced reflectance intensity along with the spectra, recognizable until horizon 4. A similar relationship was detectable at horizon 9 with an Ah-horizon and Corg content of nearly 1.7%;
- At horizons 10 and 11, higher values of Ccal produced increased spectral readings over the whole reflection shape. Horizons 5 and 8 indicated a more median position;
- The influence of Fe-oxides and -hydroxides was not detectable. Higher Fe content would produce lower reflection intensities. The strong influence of Corg (and also from Ccal, in part) were the main factors that hindered recognition of the interference between the spectrum and the possible occurrence of these minerals [52];
- The same conclusions can be drawn for the case of sand. The effect of quartz particles on spectral behavior was restricted by humus and carbonatic coatings;
- To summarise, Corg and Ccal dominated the spectral shapes. A sandier horizon, which can reflect more energy, was not recognizable. Increasing clay values with increasing depth was also not clearly detectable.
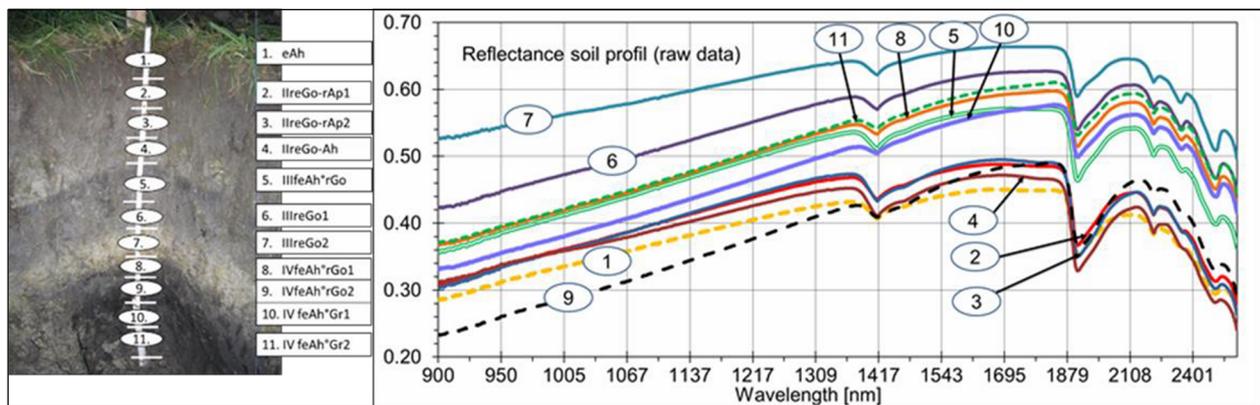
Figure 5. Spectral behavior of the different horizons and vertical discontinuities of the investigated soil profile.

3.2. Model Development

Multivariate calibrations were carried out with PLSR and leave-one-out cross-validation procedures for the 55 pre-processed spectra.

3.2.1. Influence of the Pre-Treatment Techniques on the Readings

Pre-processing techniques were used to improve the spectral characteristics and optimize the relationships with the soil properties of interest. As described before, these techniques were divided into the data treatments of smoothing, the degree of the search window, derivation, and additional techniques.

An evaluation of the relevant literature indicates that there is no single or combined pre-processing technique(s) that is optimal for all datasets (Table 1, [14,15,47,48]). Thus, the type and degree of pre-processing methods are data-specific. The same pre-processing was used for all spectra. Soil spectra were first reduced to 900–2659 nm to eliminate severe noise at the edges of each spectrum. The shapes of all pre-treatment procedures are shown in the Supplementary Materials (Figure S11).

Smoothing and degree of the search window: With smoothing, noise in the spectral signals is reduced, and the calibrations become more robust and more straightforward [53–55]. The first models were used as a 'control treatment', where the raw reflectance values were not smoothed (None-0-raw-) or only weakly smoothed (SG3-SG1- and SG3-SG2-). The next modeling stages included the smoothing technique "moving average", divided into calculations with search windows of 11 and 25 (MA11-raw- and MA25-raw-). The third modeling stage calculated smoothing according to the Savitzky–Golay method, divided into search windows of 11 and 25 points (SG11-raw- and SG25-raw-). While a "moving average" build involves values within a window of data points, Savitzky–Golay applies a polynomial to fit the data points within a window of points.

The unsmoothed raw data indicated (None-0-raw-, Figure S1, Supplementary Materials) small fluctuations with a maximum of ±0.01 nm, particularly in the lower reflectance ranges (up to about 1100 nm). It is assumed that most of these noise variations in this study are device-related issues caused by light scattering due to sample surface topology, particle size, and possible sample quantity variations in the sample containers. Such effects were described by Martens et al. [53].

Under smoothing transformations with increasing search windows, these variations were distinctly reduced. The number of the readings in the windows determines how many adjacent readings will be used to build an average value. Ranges of 5 to 9 nm have been frequently used in the literature [17,30,31] and a value of 101 nm [36]. Igne et al. [54] used window sizes of 25, as in this study. The smoothing effect was evident in all further transformations based on these smoothings (Supplementary Materials, pre-processing techniques MA11 and MA 25, Figures S4 and S5; SG11 and SG25, Figure S6–S11). Although

the reflectance spectra became much less complex, the variation in some of the spectra appeared to be exaggerated, and the others became more similar [53].

Derivation: Within every classification for the smoothing and search windows, the following derivation levels were selected: No derivation, Savitzky–Golay first derivative, first-order polynomial, Savitzky–Golay second derivative, and second-order polynomial.

Spectral derivatives were applied to eliminate the additive and multiplicative effects from the reflectance spectra [27,28]. The most basic derivation method is a slope calculation between two subsequent spectral measurement points; the second-order derivative is calculated by the difference between two successive points of the first-order derivative readings [20]. Calculation of the Savitzky–Golay derivative value at point i within the spectra was carried out using a polynomial. In the Unscrambler software, the selection of the polynomial degree is necessary.

However, Viscarra Rossel and Lark [55] noted that the application of derivations increased the fluctuations by calculating the gradients between adjacent readings. The results yielded strongly fluctuating values (Supplementary Materials, pre-processing techniques Figures S2 and S8). These fluctuations were also increased by the second derivation (Supplementary Materials, pre-processing techniques Figures S3 and S9). Therefore, not all the pre-processing methods were appropriate to improve the accuracy.

Consequently, a smoothing algorithm is often used together with the derivative [55]. This effect is particularly evident here in the first derivation with curve progressions, which indicate fluctuations. Rinnan et al. [20] used a smoothing procedure before calculating the derivatives to reduce the detrimental influence of conventional finite-difference derivatives on the signal-to-noise ratio.

In the current evaluation, the applications using derivations with low smoothing did not deliver the desired weakening of the peak values. The same applies to the second derivatives with prior smoothing, independent of the degree of smoothing. In both cases, the derivations produced a considerable amount of noise.

Additional techniques: Dotto et al. [17] classified multiplicative scatter correction (MSC), detrending, standard normal variate (SNV), and mean centering into the group of scatter-corrective pre-processing methods. In Figure S1, Supplementary Materials, the similarity between SNV and MSC's curve progression is observable, albeit with different reflection scales. The signal correction procedures are the same for SNV and MSC [56]. SNV is designed to center the underlying linear slope of each sample spectrum.

3.2.2. Selection of the Best Pre-Processing Technique

The predictive statistics of all soil properties models are shown in Figure 6 and Supplementary Materials, Table S1. The different pre-treatment techniques were evaluated by their potential to predict the single soil parameters. The results showed that the different pre-processing methods had considerable effects on the performance of the selected parameters.

(**a**)

(**b**)
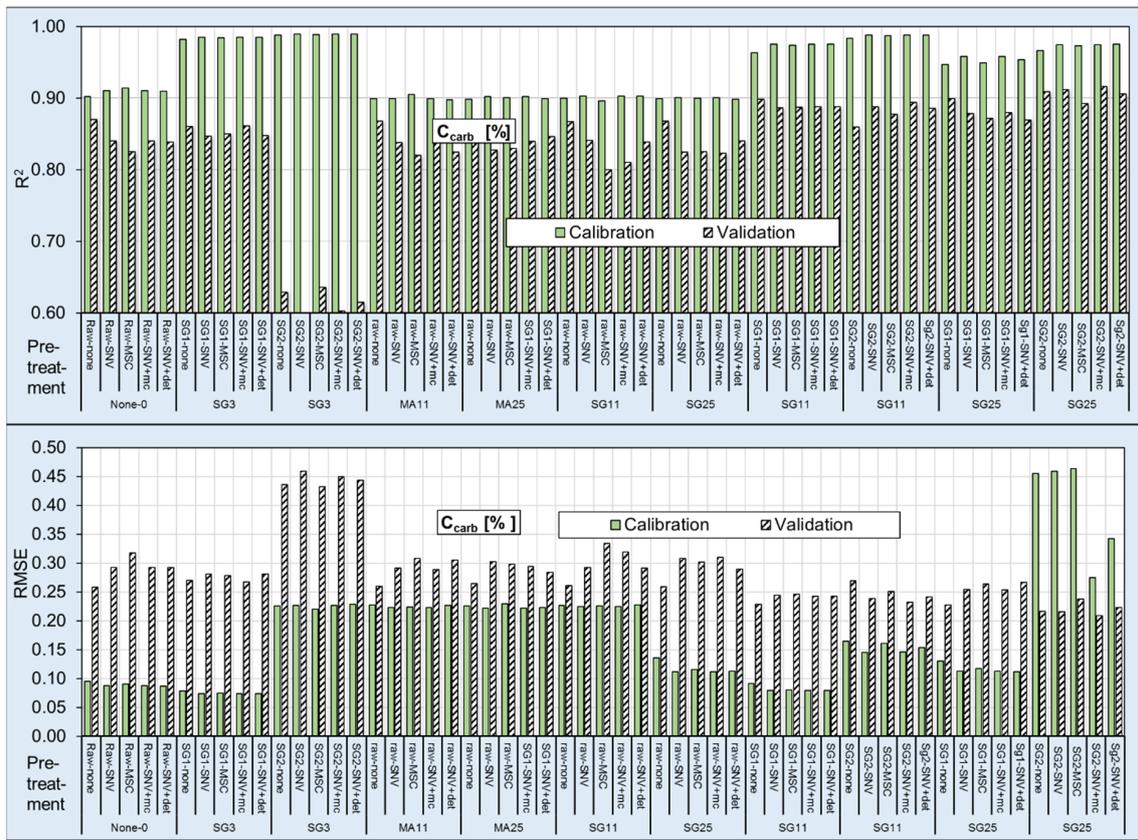
Figure 6. *Cont.*

(**c**)
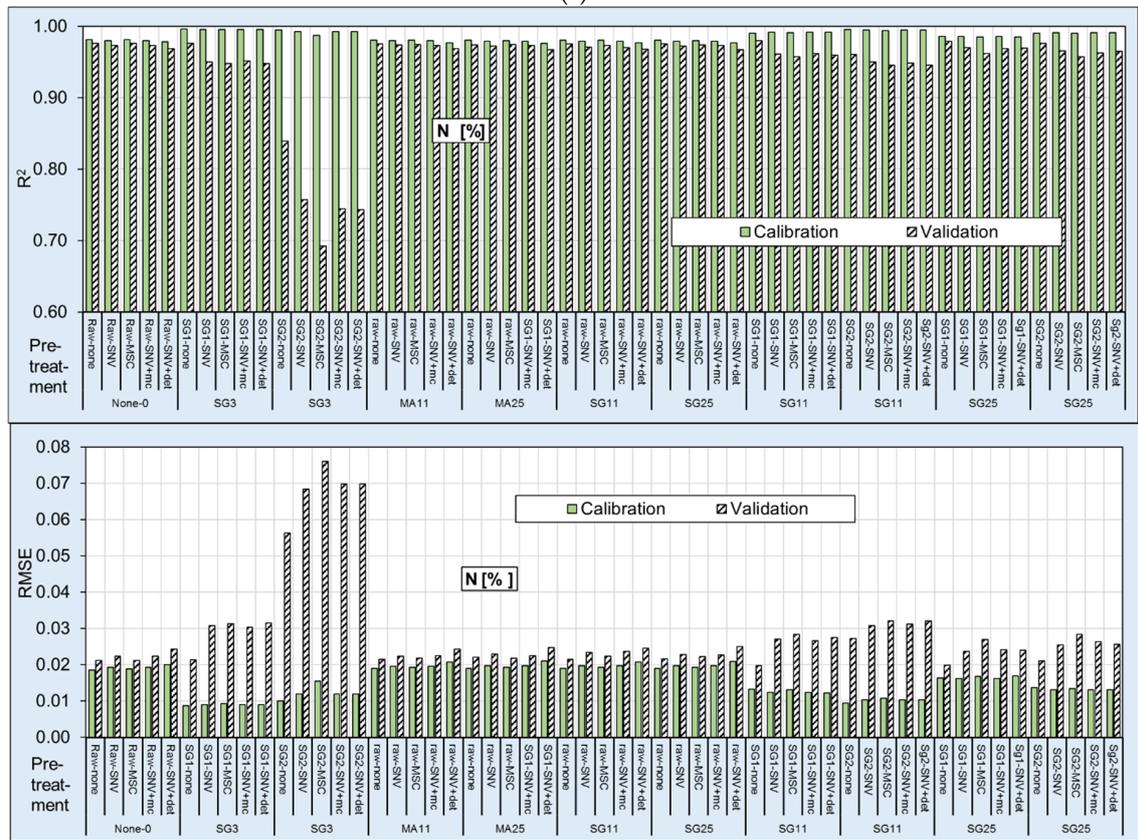
(**d**)

Figure 6. *Cont*.
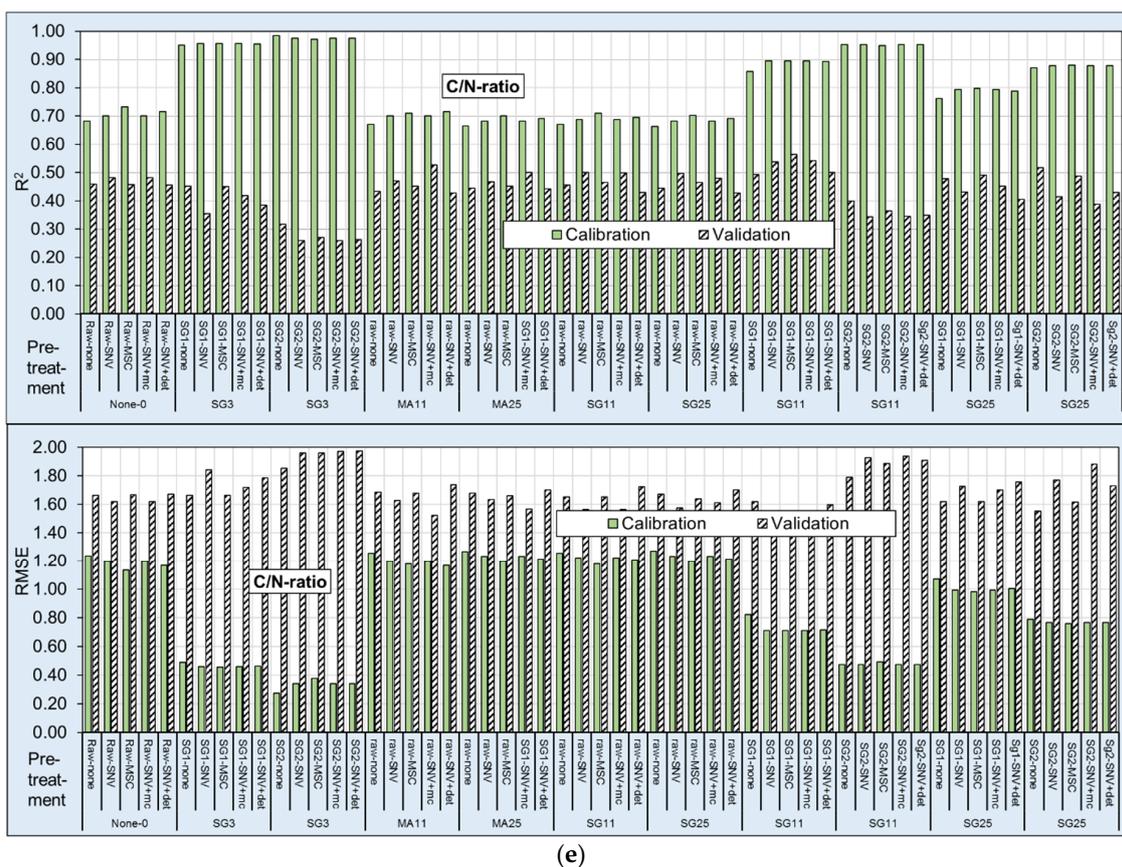
Figure 6. Indices for calculations with partial least squares regression (PLSR) R^2 and root mean square error (RMSE) for all applied pre-treatments; (**a**) Ct, (**b**) Corg, (**c**) Ccal, (**d**) N, (**e**) C/N. For abbreviations, see Table 3.

Ct: The pre-treatment trials showed that weak SG-smoothing combined with the Savitzky-Golay first derivative, first-order polynomial, and no additional technique (SG3-SG1-none) provided the best performance for Ct. With this pre-processing transformation, the PLSR models accurately predicted Ct, with R^2 values of 0.98/0.86 (calibration/validation) and RMSE values of 0.09%/0.25% (calibration/validation). The high RPD and RPIQ values of 7.16/10.92 and 2.58/4.08, respectively, for calibration/validation, were another indication of the model's quality. The results using the procedure of the Savitzky–Golay second derivative, second-order polynomial, and 11 smoothing segments without an additional technique (SG11-SG2-none) were slightly worse. These results compared well to previous work by other researchers. The studies by Vasques et al. [14,15], Sarkhot et al. [57,58], McDowell et al. [24,25], Knox et al. [30] and Pinheiro et al. [31] described Ct prediction models with lower R^2 and RPD values (R^2 0.86–0.95). A similar R^2 value (0.97) was also found in work by Reeves et al. [59].

Ccalc: For Ccalc, the calculation SG25-Sg2-SNV+det provided excellent calibration and validation accuracy, with an R^2 between 0.98 and 0.91 (calibration/validation) and an RMSE of 0.11 and 0.22. Both RPD and RPIQ presented levels with 6.3/7.83 and 3.23/4.11. Good model performance was also obtained using the treatments SG11-SG1, SG11-SG2, SG25-SG1, and SG25-SG2 with the other additional technique.

Corg: Here, the best performance was produced with the pre-treatment of SG3-SG1-none. The R^2 values of 0.99/0.94 (calibration/validation) and RMSE values of 0.09%/0.24% (calibration/validation) in combination with high RPD and RPIQ values of 11.59/22.41 and 4.19/8.42, respectively, for calibration and validation, indicated high quality. The results of the same procedure with the additional technique were slightly worse. The pre-treatment groups SG3-SG2 and SG11-SG2 also provided good predictions.

The models developed for organic C values in the literature provided slightly worse results [17,21–23,26–28,32,60]. These differences may have been influenced, at least in part, by the larger range of Corg values in this calculation (compare Table 1 and Figure 3).

Several studies reported that vis–NIR relates better to Corg than NIR alone (Viscarra Rossel et al. [2,3]. Islam et al. [61] noted better results for Australian soils, and Fystro [62] derived a similar result with Norwegian soils. However, in the literature, there are also contrary results. The inclusion of reflection values in the visible area (300–750 nm) delivered only a weak improvement in Swedish soils (Stenberg, 2010 [47,48]). Contrary results were also described by Dunn et al. [63] for soils from SE Australia and the USA. The observation that higher Corg content combines with a darker color is indeed correct, but it seems that this color is only reflected in the spectral signature if the soils come from the same geological substrate. Hummel et al. [64] noted that soil properties such as moisture, texture, and mineralogy could influence the brightness of soils and the same content of Corg.

The combined application of NIR with MIR has also often been discussed. According to Bellon–Maurel and McBratney [65], MIR generally provides slightly better results. The higher sensitivity of NIR than MIR to instrumental errors yielded better predictions with MIR.

N: Of all the parameters, N was the best for calculations. The R^2 values of the calibration were in no cases worse than 0.9. The best performance was reached with the procedure of SG3-SG1-none, with R^2 values of 0.99/0.98 (calibration/validation) and RMSE values of 0.01%/0.02% (calibration/validation). The high RPD and RPIQ values of 15.83/30.67 and 6.36/12.39, respectively, for calibration/validation also indicated the model's quality.

C/N: All results here were insufficient. The best results were calculated with the procedures of SG3-SG1 and SG3-SG2 followed by SG11-SG2. While the R^2 of the calibration was 0.95 combined with an RMSE of 0.3–0.5 and an RPD/RPIQ of 4.5–8.0 and 2.3–4.2, the results of the validations were weaker. A similar level between calibration and validation was not observed. Therefore, the calculations were not useable.

3.2.3. Description of the Prediction-Relevant Wavelengths

Figure 7 illustrates the most important wavelengths for the prediction of C and N. Notably, up to about 1400 nm, single wavelengths were most important. In contrast, above this value, the number of determining wavelengths/wavebands increased, and whole ranges became important. In the case of N and Corg, the regions of 1420–1730, 1840–2060, and 2160–2600 nm were nearly all (without any gaps) relevant for the predictions.

The prediction of Ccarb also indicted relevant wavelengths mainly above 1490 nm (1490–1730, 1850–1970, 2000–2100, 2218, and higher than 2280 nm). The number of selected wavelengths for the C/N ratio was lower than that for the other parameters. The ranges of 1660–1730, 1900–2050, and above 2170 corresponded to the important values of N and Corg.

Several authors have identified the most important single wavelengths and ranges of wavelengths for the prediction of Corg and N. Reflection values around 1100, 1600, 1700–1800, 2000, and 2200–2400 nm have been often described as important for Corg and N calibrations [47,48,64–69]; cited the range from 1650 to 2500 as the most relevant for measuring organic C. According to Mouazen et al. [70], the most significant wavelengths for Corg are in the VIS wavelength range (439, 490, and 661 nm) and the NIR region (1109, 1232, 1414, and 1522 nm). Other researchers have found the VIS range (400–700) to be important for the detection of Corg [10,71].

These values from the literature partly differed from those in the present evaluation. However, it was not easy to make specific assignments using the NIR spectrum because other different organic and inorganic molecules could be absorbed in these areas. This was especially true for wavelengths above 2000 nm. Kuang and Mouazen [72] provided other reasons why the important wavelengths can differ. For example, the concentration level and range influence the selected wavelengths. The wavelengths also depend largely on the

variability of different soil types in connection with the geological source substrate and type of cultivation [47,48,73].
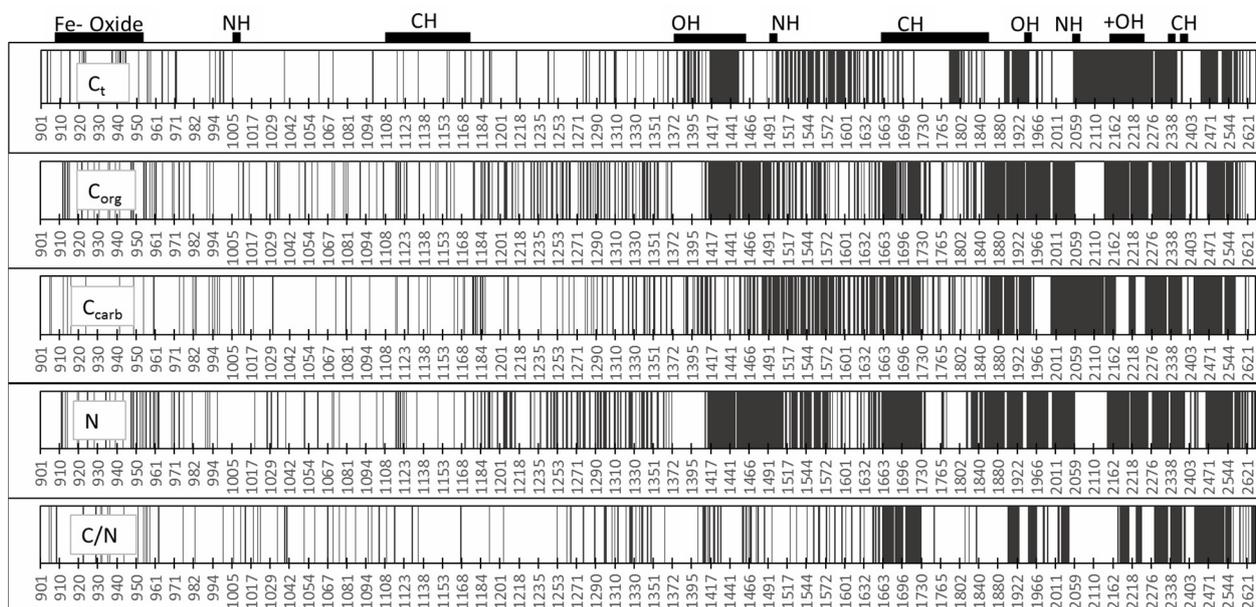

Figure 7. Important wavelengths for the prediction of Ct, Corg, Ccarb, and C/N. Specific reflection wavebands for the different areas for soil are specified on the top (modified from [73]).

In this context, Bellon–Maurel and McBratney [65] studied whether discrete variables are sufficient for the relevant calculations. The authors concluded that if the RPD values are over 2 and the R^2 values are over 0.75, the models are sufficient. An additional option here was the development of low-cost wavelength-based sensors to measure organic C.

4. Conclusions

The statistics obtained in this study for calibration, cross-validation, and validation showed that NIRS techniques could effectively provide fast and accurate predictions of Ct, Corg, Ccarb, and N.

The number of publications on the application of NIR spectroscopy in soil science has increased in recent years, as this technique has the advantage of a rapid application without intensive sample preparation. Altogether, 55 procedures of pre-treatments were calculated to determine the best soil sample model from a humus–clay-rich soil. In all calculations, the derivatives with Savitzky–Golay produced the lowest errors and enhanced the pre-treatments, such as detrending; MSC and SNV were not required. Derivatives using Savitzky–Golay calculated by differences were very popular to perform baseline corrections and enhance weak signals. However, in the current study, up to a wavelength of about 1400 nm, this procedure yielded severe fluctuations due to scattering in the original data. Thus, the condition of the measurement device, the sample presentation, the replications, and the accuracy during measurement have an enormous impact on the final results. Only an iterative process can help develop the best models. Smoothing can reduce these fluctuations, but up to a search window of 25, the effect was very low in the case of Ct, Corg, and N, while only Ccarb and C/N indicated an improvement. Important wavelengths for the predictions were mainly in a range of about 1400–2500 nm.

Further work will be undertaken to improve prediction accuracy by applying non-linear calibration techniques such as support vector machine, random forest.

Supplementary Materials: The following are available online at https://www.mdpi.com/1424-8220/21/4/1423/s1 Supplementary file "An Evaluation of Different NIR-Spectral Pre-Treatments to Derive the Soil Parameters C and N of a Humus-Clay-Rich Soil_Heil_2021", Figure S1: Spectra of

soils with different pre-processing methods: without smoothing–no derivation (None-0-Raw data); Figure S2: Spectra of soils with different pre-processing methods: smoothing Savitzky–Golay and a search window of 3; Savitzky–Golay 1st derivation and 1st order polynomial (SG3-SG1-); Figure S3: Spectra of soils with different pre-processing methods: smoothing Savitzky–Golay, a search window of 3, Savitzky–Golay 2nd derivation, and 2nd order polynomial; Figure S4: Spectra of soils with different pre-processing methods: a moving average search window of 11 segments and no derivation (MA11-raw-); Figure S5: Spectra of soils with different pre-processing methods: methods: a moving average search window of 25 segments and no derivation (MA25-raw-); Figure S6: Spectra of soils with different pre-processing methods: Savitzky–Golay, 0 order polynomial, a search window of 11 segments, and without derivation (SG11-raw-); Figure S7: Spectra of soils with different pre-processing methods: Savitzky–Golay, 0 order polynomial, a search window of 25 segments, and without derivation (SG25-raw-); Figure S8: Spectra of soils with different pre-processing methods: Savitzky-Golay, 1st derivation, 1st order polynomial, and a search window of 11 segments (SG11-SG1-); Figure S9: Spectra of soils with different pre-processing methods: Savitzky–Golay, 2nd derivation, 2nd order polynomial, and a search window of 11 segments (SG11-SG2-); Figure S10: Spectra of soils with different pre-processing methods: Savitzky–Golay, 1st derivation, 1st order polynomial, and a search window of 25 segments (SG25-SG1-); Figure S11: Spectra of soils with different pre-processing methods: Savitzky–Golay, 2nd derivation, 2nd order polynomial, and a search window of 25 segments (SG25-SG2-); Table S1: Pre-treatments applied to the reflectance spectra with statistics of calibration and validation for different soil properties.

Author Contributions: K.H.: Data collection, evaluation, calculations, writing, and visualization; U.S.: project administration and supervision. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, C.W.; Charles, R.; Hurburgh, J.R.; Laird, D.A.; Mausbach, M.J. Near-Infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [CrossRef]
2. Viscarra Rossel, R.A.; Walvoortb, T.D.J.J.; McBratney, A.B.; Janikc, 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]
3. Gomez, C.; Viscarra Rossel, R.A.; McBratney, A.B. Soil organic carbon prediction by hyperspectral remote sensing andfield vis-NIRspectroscopy: An Australian case study. *Geoderma* **2008**, *146*, 403–411. [CrossRef]
4. Boruvka, L.; Gholizade, A.; Saberioon, M.M.; Soom, M.A.M. Visible and near infrared reflectance spectroscopy to determine chemical properties of paddy soils. *J. Food Agric. Environ.* **2013**, *11*, 859–866.
5. Gholizadeh, A.; Borůvka, L.; Saberioon, M. Common Chemometric Indicators for Prediction of Soil Organic Matter Content and Quality from Soil Spectra: A Review and Research Perspectives. In Proceedings of the International Workshop "Soil Spectroscopy: The Present and Future of Soil Monitoring", Rome, Italy, 4–6 December 2013. [CrossRef]
6. Gholizadeh, A.; Boruvka, L.; Saberioon, M.M.; Vasat, R. Visible, near-infrared, and mid-infrared spec-troscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Appl. Spectrosc.* **2013**, *67*, 1349–1362. [CrossRef] [PubMed]
7. Shepherd, K.D.; Walsh, M.G. Development of reflectance spectra libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988–998. [CrossRef]
8. Reeves, J.B.; McCarty, G.W.; Meisinger, J.J. Near infrared reflectance spectroscopy for the analysis of agricultural soils. *J. Near Infrared Spectrosc.* **1999**, *7*, 179–193. [CrossRef]
9. Odlare, M.; Pell, M.; Svensson, K. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. *Geoderma* **2005**, *126*, 193–202. [CrossRef]
10. Cozzolino, D.; Morón, A. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil Till. Res.* **2006**, *85*, 78–85. [CrossRef]
11. Brown, D.J.; Mays, M.D.; Reinsch, T.G.; Wales, M.G. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* **2005**, *129*, 215–267. [CrossRef]

12. Bogrecki, I.; Lee, W.S. Spectral phosphorus mapping using diffuse reflectance of soils and grass. *Biosyst. Eng.* **2005**, *91*, 305–312. [CrossRef]

13. Aalhus, J.L.; Dugan, M.E.R.; Pawluczyk, O.; Prieto, N. A Review of the Principles and Applications of Near-Infrared Spectroscopy to Characterize Meat, Fat, and Meat Products. *Appl. Spectr.* **2017**, *71*, 1403–1426.

14. Vasques, G.M.; Grunwald, S.; Sickman, J.O. Comparison of multivariate methods for. inferential modeling of soil carbon using visible-near infrared spectra. *Geoderma* **2008**, *146*, 14–25. [CrossRef]

15. Vasques, G.M.; Grunwald, S.; Sickman, J.O. Modeling of Soil Organic Carbon Fractions Using Visible–Near-Infrared Spectroscopy. *Soil Sci. Soc. Am. J.* **2009**, *73*, 176–184. [CrossRef]

16. Xie, X.L.; Pan, X.Z.; Sun, B. Visible and Near-Infrared Diffuse Reflectance Spectroscopy for Prediction of Soil Properties near a Copper Smelter. *Pedosphere* **2012**, *22*, 351–366. [CrossRef]

17. Dotto, A.C.; Dalmolin, R.S.J.; ten Caten, A.; Grunwald, S. A systematic study on the application of scatter-corrective and spectralderivative pre-processing for multivariate. prediction of soil organic carbon by Vis-NIR spectra. *Geoderma* **2018**, *314*, 262–274. [CrossRef]

18. Savitzky, A.; Golay, M. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

19. Roberts, C.A.; Workman, J., Jr.; Reeves, J.B.; Duckworth, J. (Eds.) Mathematical Data Preprocessing. In *Near-Infrared Spectroscopy in Agriculture*; American Society of Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI, USA, 2004; Volume 44, pp. 115–132.

20. Rinnan, A.; Berg, F.; Van Den, E. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]

21. 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. [CrossRef]

22. Cambule, A.H.; Rossiter, D.G.; Smaling, E.M.A.; Stoorvogel, J.J. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. *Geoderma* **2012**, *183–184*, 41–48. [CrossRef]

23. Dotto, A.C.; Dalmolin, R.S.J.; Grunwald, S.; ten Caten, A.; Filho, W.F. Two pre-processing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopyd. *Soil Tillage Res.* **2017**, *172*, 59–68. [CrossRef]

24. McDowell, M.L.; Bruland, G.L.; Deenik, J.L.; Grunwald, S.; Knox, N.M. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma* **2012**, *189–190*, 312–320. [CrossRef]

25. McDowell, M.L.; Bruland, G.L.; Deenik, J.L.; Grunwald, S. Effects of subsetting bycarbon content, soil order, and spectral classification on prediction of soil total carbonwith diffuse reflectance spectroscopy. *J. Appl. Environ. Soil Sci.* **2012**, *2012*, 294121. [CrossRef]

26. Peng, X.; Shi, T.; Song, A.; Chen, Y.; Gao, W. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* **2014**, *6*, 2699–2717. [CrossRef]

27. Chen, S.; Xu, D.; Li, S. Monitoring soil organic carbon in alpine soils using in situ vis-NIR spectroscopy and a multilayer perceptron. *Land Degrad Dev.* **2020**, *31*, 1026–1038. [CrossRef]

28. Gholizadeh, A.; Borůvka, L.; Vašát, R.; Saberioon, M.; Klement, A.; Kratina, J. Estimation of Potentially Toxic Elements Contamination in Anthropogenic Soils on a Brown Coal Mining Dumpsite by Reflectance Spectroscopy:A Case Study. *PLoS ONE* **2015**, *10*, 1–14. [CrossRef] [PubMed]

29. Knox, N.M.; Grunwald, S.; McDowell, M.L.; Bruland, G.L.; Myers, D.B.; Harris, W.G. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma* **2015**, *239–240*, 229–239. [CrossRef]

30. Pinheiro, E.; Ceddia, M.; Clingensmith, C.M.; Grunwald, S.; Vasques, G.M. Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon. *Remote Sens.* **2017**, *9*, 293. [CrossRef]

31. Nawar, S.; Buddenbaum, H.; Hill, J.; Kozak, J.; Mouazen, A.M. Estimating the soil clay content and organic matter by means of different calibration methods of Vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res.* **2016**, *155*, 510–522. [CrossRef]

32. Peng, Y.; Zhao, L.; Hu, Y.; Wang, G.; Wang, l.; Liu, Z. Prediction of Soil Nutrient Contents Using Visible and Near-Infrared Reflectance Spectroscopy. *Isprs Int. J. Geo Inf.* **2019**, *8*, 437. [CrossRef]

33. Wang, X.; Shi, A.; Yaqing, X.; Huping, H.; Fuyao, C.; Yongjun, Y.; Shaoliang, Z.; Run, L. A Back Propagation Neural Network Model Optimized by Mind Evolutionary Algorithm for Estimating Cd, Cr, and Pb Concentrations in Soils Using Vis-NIR Diffuse Reflectance Spectroscopy. *Appl. Sci.* **2020**, *10*, 51. [CrossRef]

34. Muñoz, J.D.; Kravchenko, A. Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. *Geoderma* **2011**, *166*, 102–110. [CrossRef]

35. Stevens, A.; Nocita, M.; Tóth, G.; Montanarella, L.; van Wesemael, B. Prediction of soil organic carbon at the European scale by visible and near InfraRed reflectance spectroscopy. *PLoS ONE* **2013**, *8*, 1–20. [CrossRef] [PubMed]

36. Cen, H.; He, Y. Theory and Application of near Infrared Reflectance Spectroscopy in. Determination of Food Quality. *Trends Food Sci. Technol.* **2007**, *18*, 72–83. [CrossRef]

37. Bodenkundliche Kartieranleitung. 5. Verbesserte und 540 Erweiterte Auflage. Schweizerbart'sche Verlagsbuchhandlung: Hannover, Germany. 2005. Available online: https://www.schweizerbart.de/publications/detail/artno/184050681 (accessed on 18 February 2021).

38. VDLUFA. Methode C 2.2.1, Texturanalyse des Feinbodens Kombination von Nasssiebung und Pipettmethode nach KÖHN. Available online: https://www.vdlufa.de/Methodenbuch/index.php?option=com_content&view=article&id=7&Itemid=108&lang=de&lang=en (accessed on 18 February 2021).

39. CAMO. The Unscrambler User Manual Camo Software AS. Manual is up to Date for Version 9.6 of The Unscrambler. Available online: www.camo.com/downloads/products/The_Unscrambler.pdf (accessed on 10 October 2019).

40. Hong, Y.; Shen, R.; Cheng, H.; Chen, Y.; Zhang, Y.; Liu, Y.; Zhou, M.; Yu, L.; Liu, Y.; Liu, Y. Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. *Sci. Total Environ.* **2019**, *651*, 1969–1982. [CrossRef]

41. Westerhaus, M.; Workman, J.; Reeves, J.B.; Mark, H. Quantitative Analysis. In *Agronomy Monographs*; Roberts, C.A., Workman, J., Reeves, J.B., Eds.; American Society of Agronomy, Inc.: Madison, WI, USA; Crop Science Society of America, Inc.: Madison, WI, USA; Soil Science Society of America, Inc.: Madison, WI, USA, 2004; Chapter 7. [CrossRef]

42. Viscarra Rossel, R.A.; Jeon, Y.S.; Odeh, I.O.A.; McBratney, A.B. Using a legacy soil sample to develop a mid-IR spectral library. *Austral. J. Soil Res.* **2008**, *46*, 1–16. [CrossRef]

43. Williams, P. Tutorial: The RPD Statistic: A Tutorial Note. *NIR News* **2014**, *25*, 22–26. [CrossRef]

44. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **2010**, *29*, 1073–1081. [CrossRef]

45. Demattê José, A.M.; Ramirez-Lopez, L.; Marques, K.P.P.; Rodella, A.A. Chemometric soil analysis on the determination of specific bands for the detection of magnesium and potassium by spectroscopy. *Geoderma* **2017**, *288*, 8–22. [CrossRef]

46. Stenberg, B.; Viscarra Rossel, R.A. Diffuse reflectance spectroscopy for high resolution soil sensing. In *Proximal Soil Sensing*; Viscarra Rossel, R.A., McBratney, B.A., Minasny, B., Eds.; Springer Science + Business: Dordrecht, The Netherlands, 2010; pp. 29–47.

47. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Burlington, ON, Canada, 2010; Volume 107, pp. 163–215.

48. Whiting, M.L.; Li, L.; Ustin, S.L. Predicting water content using Gaussian model on soil spectra. *Remote Sens. Environ.* **2004**, *89*, 535–552. [CrossRef]

49. Clark, R.N. Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual Remote Sens.* **1999**, *3*, 3–58.

50. Ben-Dor, E. Characterization of soil properties using reflectance spectroscopy. In *Hyperspectral Remote Sensing of Vegetation*; Thenkabail, P.S., Lyon, J.G., Huete, A., Eds.; Crc Press: Boca Raton, FL, USA, 2011; pp. 513–557.

51. Henderson, T.L.; Szilagyi, A.; Baumgardner, M.F.; Chen, C.T.; Landgrebe, D. Spectral band selection from classification of soil organic matter content. *Soil Sci. Soc. Am. J.* **1989**, *53*, 1778–1784. [CrossRef]

52. Martens, H.; Nielsen, J.P.; Engelsen, S.B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. *Anal. Chem.* **2003**, *75*, 394–404. [CrossRef] [PubMed]

53. Igne, B.; Hurburgh, C.R. Local chemometrics for samples and variables: Optimizing calibration and standardizationprocesses. *J. Chemom.* **2010**, *24*, 75–86. [CrossRef]

54. Viscarra Rossel, R.A.; Lark, R.M. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *Eur. J. Soil Sci.* **2009**, *60*, 453–464. [CrossRef]

55. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard Normal Variate Transformation and De- Trending of near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777. [CrossRef]

56. Sarkhot, V.; Grunwald, S.; Ge, Y.; Morgan, C.L.S. Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. *Geoderma* **2011**, *164*, 22–32. [CrossRef]

57. Sarkhot, D.; Grunwald, S.; Ge, Y.; Morgan, C. Total and available soil carbon fractions under the perennial grass *Cynodon dactylon* (L.) Pers and the bioenergy crop *Arundo donax* L. *Biomass Bioenergy* **2012**, *41*, 122–130. [CrossRef]

58. Reeves III, J.B.; McCarty, G.W.; Mimmo, T. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environ. Pollut.* **2002**, *116*, 277–284. [CrossRef]

59. Viscarra Rossel, R.A.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]

60. Islam, K.; Singh, B.; McBratney, A. Simultaneous estimation of several soil properties by ultra- violet, visible, and near-infrared reflectance spectroscopy. *Soil Res.* **2003**, *41*, 1101–1114. [CrossRef]

61. Fystro, G. The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis–NIR spectroscopy and comparative methods. *Plant Soil* **2002**, *246*, 139–149. [CrossRef]

62. Dunn, B.W.; Beecher, H.G.; Batten, G.D.; 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*, 1–8. [CrossRef]

63. Hummel, J.W.; Sudduth, K.A.; Hollinger, S.E. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Comput. Electron. Agric.* **2001**, *32*, 149–165. [CrossRef]

64. Bellon-Maurel, V.; McBratney, A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils—Critical review and research perspectives. *Soil Biol. Biochem.* **2011**, *43*, 1398–1410. [CrossRef]

65. Ben-Dor, E.; Banin, A. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate. Several Soil Properties. *Soil Sci. Soc. Am. J.* **1995**, *59*, 364. [CrossRef]

66. Malley, D.; Lockhart, L.; Wilkinson, P. Determination of carbon, carbonate, nitrogen, and phosphorus in freshwater sediments by near-infrared reflectance spectroscopy: Rapid analysis and a check on conventional analytical methods. *J. Paleolimnol.* **2000**, *24*, 415–425. [CrossRef]

67. Morra, M.J.; Hall, M.H.; Freeborn, L.L. Carbon and Nitrogen Analysis of Soil Fractions Using. Near-Infrared Reflectance Spectroscopy. *Soil Sci. Soc. Am. J.* **1991**, *55*, 288–291. [CrossRef]

68. Morgan, C.L.S.; Waiser, T.H.; Brown, D.J.; Hallmark, C.T. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma* **2009**, *151*, 249–256. [CrossRef]

69. Mouazen, A.M.; Maleki, M.R.; De Baerdemaeker, J.; Ramon, H. On-line measurement of some selected soil properties using a VIS–NIR sensor. *Soil Tillage Res.* **2007**, *93*, 13–27. [CrossRef]

70. Kirshnan, P.; Alexander, J.D.; Butler, B.J.; Hummel, J.W. Reflectance technique for predicting soil organic matter. *Soil Sci. Soc. Am. J.* **1980**, *44*, 1282–1285. [CrossRef]

71. Kuang, B.; Mouazen, A.M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *Eur. J. Soil Sci.* **2011**, *62*, 629–636. [CrossRef]

72. Wang, J.; He, T.; Lv, C.; Chen, Y.; Jian, W. Mapping soil organic matter based on land 651 degradation spectral response units using Hyperion images. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *652*, 171–180. [CrossRef]

73. 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. [CrossRef]