soil systems

MDPI

Article

Ensemble Modeling on Near-Infrared Spectra as Rapid Tool for Assessment of Soil Health Indicators for Sustainable Food Production Systems

John Walker Recha [1,*], Kennedy O. Olale [2], Andrew Sila [3], Gebermedihin Ambaw [1], Maren Radeny [1] and Dawit Solomon [1]

1 CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) East Africa, International Livestock Research Institute (ILRI), Nairobi P.O. Box 30709-00100, Kenya; g.ambaw@cgiar.org (G.A.); m.radeny@cgiar.org (M.R.); d.solomon@cgiar.org (D.S.)
2 Department of Chemistry, School of Pure and Applied Sciences, Kisii University, Kisii P.O. Box 408-40209, Kenya; kennedyolale@kisiiuniversity.ac.ke
3 World Agroforestry (ICRAF), United Nations Avenue, Nairobi P.O. Box 30677-00100, Kenya; a.sila@cgiar.org
* Correspondence: j.recha@cgiar.org

Abstract: A novel total ensemble (TE) algorithm was developed and compared with random forest optimization (RFO), gradient boosted machines (GBM), partial least squares (PLS), Cubist and Bayesian additive regression tree (BART) algorithms to predict numerous soil health indicators in soils with diverse climate-smart land uses at different soil depths. The study investigated how land-use practices affect several soil health indicators. Good predictions using the ensemble method were obtained for total carbon (R^2 = 0.87; RMSE = 0.39; RPIQ = 1.36 and RPD = 1.51), total nitrogen (R^2 = 0.82; RMSE = 0.03; RPIQ = 2.00 and RPD = 1.60), and exchangeable bases, m3. Cu, m3. Fe, m3. B, m3. Mn, exchangeable Na, Ca (R^2 > 0.70). The performances of algorithms were in order of TE > Cubist > BART > PLS > GBM > RFO. Soil properties differed significantly among land uses and between soil depths. In Kenya, however, soil pH was not significant, except at depths of 45–100 cm, while the Fe levels in Tanzanian grassland were significantly high at all depths. Ugandan agroforestry had a substantially high concentration of ExCa at 0–15 cm. The total ensemble method showed better predictions as compared to other algorithms. Climate-smart land-use practices to preserve soil quality can be adopted for sustainable food production systems.

Keywords: algorithms; climate-smart; soil quality; land use

Citation: Recha, J.W.; Olale, K.O.; Sila, A.; Ambaw, G.; Radeny, M.; Solomon, D. Ensemble Modeling on Near-Infrared Spectra as Rapid Tool for Assessment of Soil Health Indicators for Sustainable Food Production Systems. *Soil Syst.* **2021**, *5*, 69. https://doi.org/10.3390/soilsystems5040069

Academic Editor: Megharaj Mallavarapu

Received: 7 October 2021
Accepted: 6 November 2021
Published: 12 November 2021

1. Introduction

In sub-Saharan Africa, 62% of the rural population depends on agriculture as the main source of their livelihoods, and hence there is a close link between agriculture and soil health [1]. Agricultural sustainability necessitates a good understanding of soil characteristics which can inform farmers in making farming decisions and improve the practices that enhance soil quality [1,2]. Both the physical and chemical properties of soil have been used extensively to monitor soil health characteristics [3,4]; while these properties are important for farm productivity, they vary within fields and with land-use types [2,5]. If these soil properties are well-characterized, they should serve as indicators of soil health and be easy to measure using standardized methods [2]. The measurement of these soil health indicators faces significant technological difficulties due to the large number of properties involved [6]. Convectional analytical techniques such as wet chemical analysis have always been used for this purpose; however, these wet methods are time-consuming and expensive, prompting a need for a robust alternative method. Several authors have suggested near-infrared reflectance spectroscopy (12,500–4000 cm^{-1}; 800–2500 nm) as an alternative technique to wet chemical analysis [6–9]. Near-infrared absorption bands are overtones and combinations of fundamental vibrations of XH bonding, where X can be carbon, nitrogen, oxygen, or sulfur [10]. Near-infrared spectroscopy has the advantage of being rapid, non-destructive, inexpensive, precise, and can be used to estimate water-bearing minerals, such as clay minerals and organic matter, carbon and nitrogen, and cation exchange capacity [3], as well as micro-nutrients and exchangeable cations in soil samples [1,7,11]. Additionally, the technique has been applied in precision soil management as well as regular soil analysis [12]. Soriano-Disla et al. [8] reviewed soil spectroscopic models and published and listed several soil properties that could be determined by near-infrared spectroscopy; these properties include soil water content, clay, sand, soil organic carbon (SOC), CEC, exchangeable Ca and Mg, total N and pH. These spectroscopic models used different spectral preprocessing techniques such as wavelength range selection, the scatter correction method, mean normalization, baseline offset, and derivatives [9,13,14] to increase the robustness and predictability of the models. Additionally, modeling the relationship between near-infrared spectra with soil properties requires several multivariate procedures such as principal components regression, partial least squares regression (PLSR), stepwise multiple linear regression (SMLR), Fourier regression, locally weighted regression (LWR), and artificial neural networks. None of these multivariate procedures have gained widespread adoption since a model that works well for one application may be unsuitable for another. The search for an optimum algorithm for a specific NIR-based application is difficult since no single algorithm always performs well in any domain [15]. Fortunately, the growing 'ensemble' concept has prompted a fundamental shift in people's thinking [16]; rather than attempting to construct a single superb model, numerous simple models are used in tandem [15].

Ensemble modeling uses many PLS models with and without spectral preprocessing for prediction and combines several prediction models in order to improve the accuracy of weak models. Further, creating an ensemble entails two steps: (1) creating various models and (2) combining their estimates [17]. Ensembles are constructed using techniques such as Bayesian model averaging, boosting [18] and bagging [19]. The Bayesian model averages estimates from different models, weighted by their posterior evidence, while bagging bootstraps the training dataset and averages the estimates. Boosting builds models iteratively by varying case weights and employs the weighted sum of the sequence of model estimates [17]. The total ensemble algorithm technique has been little studied in situations that are common for calibration and prediction in chemistry [20]. This technique has gained increasing attention for the multivariate calibration of NIR spectra, by combining the results of multiple individual models to produce a single prediction [21]. The output of the total ensemble algorithm is computed by averaging the predicted values computed by its constituent learners [22]. The method's key assumption is that multiple models will detect and encode more features of the relationship between independent and dependent variables than a single model [23]. To obtain a good ensemble, it is generally believed that the member models should be as accurate and diverse as possible [24].

In this study, a new approach involving total ensemble modeling of NIR spectra was used to predict numerous soil health indicators from diverse climate-smart land uses at different soil depths. We compared the prediction accuracy and performance of the total ensemble method with other five machine learning algorithms: random forest optimization (RFO), gradient boosted machines (GBM), partial least squares (PLS), Bayesian additive regression trees (BART) and Cubist. Further, we used the predicted dataset to assess how land-use practices impact on selected soil health indicators.

2. Materials and Methods

2.1. Soil Samples and Types

A total of 315 samples were collected using soil coring auger at three depths, 0–15 cm, 15–45 cm, and 45–100 cm, from six climate-smart land-use types: agroforestry, community forest, cropland with soil and water conservation (SWC), crop land without SWC, grassland and control. These land-use types are in East African Climate-Smart Villages (CSV) in Lushoto (Tanzania), Hoima (Uganda), Nyando and Wote (Kenya). Collected soil samples were oven-dried at 105 °C and then finely ground to powder and passed through a 2 mm sieve. The soil types in Lushoto are Regosols, Lithic Leptosols, Cutanic Acrisols and Ferralic Cambisols [25], while in Hoima the soil types are Vertisols, which have 30% or more clay [26]. Nyando and Wote soil types are both fluviatile and lacustrine in origin and vary from colluviums to alluvium and lacustrine clays [27].

2.2. NIR Spectroscopy and Reference Laboratory Analysis

The finely ground soil samples were loaded into a glass Petri dish and then scanned in reflectance mode using a high-intensity contact probe attached to Fourier Transform Infrared Multi-Purpose Analyzer (FTIR MPA) from 8000 to 4000 cm^{-1}, 1250–2500 nm, at World Agroforestry Centre (ICRAF) in Nairobi, Kenya. For method validation, all the samples scanned using the NIR spectroscopy technique were taken for conventional methods for wet chemistry analysis. Selected physical and chemical properties such as pH, extractable P, K, Ca, Mg, Na, Mn, Fe, Cu, Zn, B, Mo, S, and Al, exchangeable bases (ExBas) (sum of Mehlich exch Ca, Mg, K, Na), exchangeable acidity (ExAc) and electrical conductivity (Ecd), as well as total N and C, were analyzed.

2.3. NIR Ensemble Modeling Using Spectroscopic Data

Ensemble modeling, unlike traditional single modeling methods, establishes many "weak" models then aggregates the predict results of each "weak" model through weighted average methods. In this study, we used regression modeling of the total ensemble algorithm plus other five machine learning algorithms, random forest optimization (RFO), gradient boosted machines (GBM), partial least squares (PLS), Bayesian additive regression trees (BART) and Cubist, to model the physical and chemical characteristics of soil. Using these modeling techniques, the processed spectral data were linked to laboratory-measured soil property data. RFO trains each tree separately using random sampling of the data, while GBM is a hybrid method that incorporates both boosting and bagging approaches [28]. BART is a nonparametric Bayesian regression method that employs dimensionally adaptable random basis elements to make inferences and estimate an unknown regression function [29], while the Cubist model includes boosting with training committees (typically more than one), which is comparable to the approach of "boosting" by generating a sequence of trees with successively adjusted weights [30].

2.4. Model Validation

To evaluate the accuracy of models, the coefficient of determination (R^2), root mean square error (RMSE), the ratio of performance to deviation RPD and ratio of performance to interquartile distance (RPIQ) were used. Calibration models having an $R^2 > 0.91$ are considered to be excellent, those with an R^2 between 0.82 and 0.90 are good, while an R^2 between 0.66 and 0.81 indicates satisfactory predictions [8]. RPD was calculated as the fraction of the standard deviation (SD) and the RMSEP (RPD = SD/RMSEP) [31], while RPIQ was calculated as a fraction of the interquartile range of the data (Q3–Q1) and the RMSEP (RPIQ = IQR/RMSEP) [32,33]. RPIQ > 1.03 indicates good predictions, 0.77–1.03 indicates reasonable prediction and <0.77 indicates non-reliable predictions. RPIQ is inversely related to R^2, and so was used in isolation to rank prediction performance. Larger values of RPIQ and smaller RMSE indicate better model performance [32].

2.5. Statistical Analysis

The reference ensemble method available in the R system for statistical computing version 3.1.0 via the "caretEnsemble" add-on package [34] was used as a modeling tool. Principal component analysis (PCA) was used to visualize the variability of soil spectral signatures in the whole dataset and to identify properties explaining the greatest variability in order to select the best indicators affecting soil health. Significant differences between land-use practices were tested by analysis of variance (ANOVA). Tukey's honest significance difference (HSD) tested the mean separation when analysis showed statistically significant differences ($p < 0.05$).

3. Results

3.1. Soil Properties across the Study Sites and Spectral Datasets

The laboratory-measured soil properties across the six land-use types as used for calibration are presented in Table 1. pH values range from slightly acidic to basic (4.43 to 9.08), while the average pH was 6.44 ± 1.34; this average suggests that the fields do not need lime to bring the pH to favorable levels. The total nitrogen and total carbon contents varied from 0.02% to 0.83% and 0.30–11.86%, respectively. The coefficients of variation (CV) were greater than 10% for all the elements analyzed (Table 1).

Table 1. Summary statistics for laboratory soil properties measured as potential indicators of soil health.

Soil Property	N	Min.	Median	Max.	Mean ± SD	Range	IQR	Skewness	CV%	Kurtosis
Total Nitrogen	315	0.02	0.11	0.83	0.13 ± 0.10	0.27	0.06	0.33	71.30	0.64
Total Carbon	315	0.30	1.46	11.86	1.79 ± 1.32	3.44	0.53	2.59	73.41	9.34
Sand	315	0.54	3.29	17.80	4.56 ± 3.71	17	2	1.91	81.36	3.49
Silt	315	1.68	6.22	56.39	9.67 ± 8.68	55	10	31.32	89.69	14.81
Clay	315	31.71	88.94	96.22	85.85 ± 10.72	65	10	−2.81	12.48	11.34
pH	315	4.43	6.35	9.08	6.44 ± 1.34	4.65	2.09	0.27	14.97	−0.83
m3.Al	315	456.00	951.00	2700.00	1006.97 ± 323.30	1274.00	479.00	0.45	32.11	−0.55
m3.B	315	0.00	0.65	4.18	0.77 ± 0.64	2.05	0.88	0.89	82.38	−0.26
m3.Cu	315	0.00	3.07	16.00	3.47 ± 2.63	7.34	2.08	1.25	75.76	1.44
m3.Fe	315	23.90	92.10	436.00	107.55 ± 66.52	238.10	31.10	2.30	61.85	8.20
m3.Mn	315	0.00	214.00	660.00	215.29 ± 155.74	390	186.40	0.07	72.34	−1.15
m3.P	315	0.00	1.91	166.00	6.98 ± 18.96	85.40	6.52	4.70	271.56	26.77
m3.S	315	0.00	3.29	226.00	9.24 ± 22.76	151.00	17.58	2.49	246.35	5.61
m3.Zn	315	0.00	1.01	32.30	2.14 ± 3.42	14.00	1.10	4.80	159.97	24.96
PSI	315	0.98	116.00	655.00	137.08 ± 87.23	332.00	132.05	0.87	63.64	−0.27
ExNa	315	0.00	0.05	11.70	0.66 ± 1.60	10.82	3.02	2.03	240.48	4.03
ExCa	315	0.31	8.60	44.05	12.46 ± 10.67	43.49	23.51	1.05	85.65	−0.49
ExMg	315	0.07	3.17	9.83	3.26 ± 1.77	6.50	1.82	1.10	54.36	0.09
ExK	315	0.00	0.28	5.17	0.72 ± 0.87	3.25	1.49	1.10	119.62	0.16
ExBas	315	0.49	12.25	58.26	17.11 ± 13.56	56.77	30.42	1.03	79.24	−0.58
ECd	315	0.01	0.05	0.77	0.08 ± 0.09	0.76	0.17	1.93	108.76	3.84
ExAc	315	0.00	0.00	8.75	0.27 ± 0.94	4.87	0.249	3.00	344.63	9.02

m3. = Mehlich 3 extractable; PSI = Phosphorus sorption index; ExNa = Exchangeable Na; ExCa = Exchangeable Ca; ExMg = Exchangeable Mg; ExK = Exchangeable K; ExBas = Exchangeable bases (sum of Mehlich exch Ca, Mg, K, Na); Ecd = Electrical conductivity; ExAc = Exchangeable Acidity; IQR = Interquartile Range; SD = standard deviation.

The PCA similarity maps of PC1 and PC2 made using the raw spectra show a clustering of soil samples among the spectra from various sampling counties in Kenya, Uganda and Tanzania (Figure 1). Further, there is a clear overlap of various spectra from sampling sites in Uganda. The clustering of soil samples could explain the effect of geographical origin of soil samples on NIR spectroscopy analysis when assessing both physical and chemical soil properties. Soil samples from the Mbuzii area in Tanzania were grouped away from Yamba soils samples using PC1, indicating that the largest contribution to variance in soil characteristics from Tanzania could be from these two regions from which the soil samples were taken. Although the Kenyan soil samples could not be discriminated using PC1, the samples were grouped separately according to the sampling site in PC2. There was no distinguishable variation profile amongst soil samples from Uganda to discriminate different sampling sites (Figure 1). These findings show that the primary sources of spectral variance are differences in soil types and that PCA may be used to distinguish sites but not land-use type.

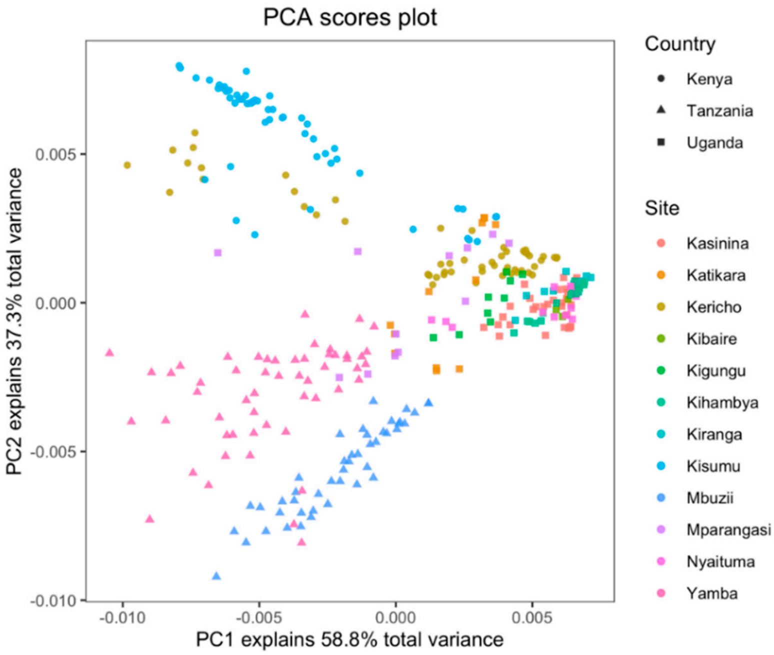

Figure 1. Biplot for the principal components explaining the largest variation.

3.2. Exploratory Analysis of Soils Near-Infrared Spectra

The spectra of all soils were similar in shape, with strong absorbance's around 5200 cm^{-1} and three distinct absorption peaks (at roughly 7060 cm^{-1}, 5200 cm^{-1}, and 4520 cm^{-1}) (Figure 2). The NIR raw spectral signature of scanned soil samples ranged from 8000 cm^{-1} to 4000 cm^{-1} regions. The NIR spectra reveal details of diverse soil constituents, including organic and inorganic components [2]. The band near 7060 cm^{-1} is caused by the overlapping of the first overtone of the O-H stretching vibration associated with water and hydroxyl groups in clay [2]. The band near 5200 cm^{-1} is related to the H-O-H bend and O-H stretch combination bands in water, while the band at around 4520 cm^{-1}, usually associated with clay minerals, results from metal–OH bending plus O-H stretching. Based on the recorded NIR spectra, the identified spectral regions and the vibrational groups are relevant in the generation of calibration models, the overtones of O-H and H-O-H stretch vibrations of free water, and the overtones and combinations of O-H stretching and metal-OH bends in the clay lattice, and are related to the well-defined absorption patterns in soil reflectance across the NIR [35]. NIR spectra with this shape have been recorded for a variety of soils [36,37].

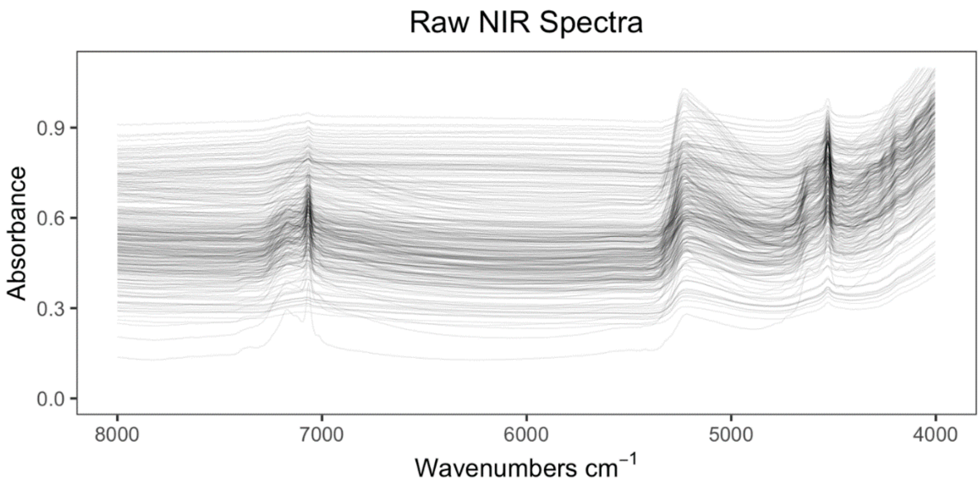

Figure 2. NIR raw spectral signatures for all soil samples.

3.3. Comparison of Machine Learning Algorithms for Prediction of Soil Properties

The descriptive regression statistics of the predicted vs. measured soil properties using total ensemble, random forest optimization, gradient boosted machines, partial least squares, Bayesian additive regression trees and Cubist methods are provided in Table 2 and the scatterplots of the predicted vs. measured soil properties from different models are illustrated in Figures 3 and 4. A good prediction was made for total carbon using the ensemble prediction (R^2 of 0.87; RMSE = 0.39; RPIQ = 1.36 and RPD = 1.51) and Cubist prediction (R^2 of 0.86; RMSE = 0.42; RPIQ = 1.26 and RPD = 1.40). Total nitrogen was also well-predicted using the total ensemble algorithm (R^2 of 0.82; RMSE = 0.03; RPIQ = 2.00 and RPD = 1.60) and Bayesian additive regression trees (R^2 of 0.78; RMSE = 0.04; RPIQ = 1.50 and RPD = 1.20). Carbon and nitrogen are two important constituents in soil organic matter, and their levels are highly connected [38]. Considering the classification by Zhao et al. [39] in terms of strong ($R^2 > 0.70$), moderate ($0.5 < R^2 < 0.7$), weak ($0.3 < R^2 < 0.5$) and very weak ($R^2 < 0.3$) prediction models, the total ensemble prediction method made: strong predictions ($R^2 > 0.70$) for exchangeable bases (sum of Mehlich exch Ca, Mg, K, Na), Mehlich 3 extractable Cu, Mehlich 3 extractable Fe, Mehlich 3 extractable B, Mehlich 3 extractable Mn, exchangeable Na and exchangeable Ca; moderate predictions ($0.5 < R^2 < 0.7$) for Mehlich 3 extractable Al, pH, the phosphorus sorption index (PSI) and exchangeable acidity (ExAc); weak predictions ($0.3 < R^2 < 0.5$) for Mehlich 3 extractable Zn, Mehlich 3 extractable P, and electrical conductivity(Ecd); and a very weak ($R^2 < 0.3$) prediction for Mehlich 3 extractable S (Table 2). The results show that total ensemble method consistently performs better than random forest optimization, gradient boosted machines, partial least squares, Bayesian additive regression trees and Cubist (lowest RMSE and highest RPIQ).

BART and Cubist made strong to moderate predictions, although cubist outperformed BART for total carbon (R^2 of 0.86; RMSE = 0.42; RPIQ = 1.26 and RPD = 1.40), pH (R^2 of 0.65; RMSE = 0.52; RPIQ = 4.02 and RPD = 2.58), Mehlich 3 extractable Al (R^2 of 0.56; RMSE = 185.55; RPIQ = 2.58 and RPD = 1.79), Mehlich 3 extractable B (R^2 of 0.71; RMSE = 0.34; RPIQ = 2.59 and RPD = 1.69), exchangeable Mg (R^2 of 0.66; RMSE = 1.01; RPIQ = 1.80 and RPD = 1.76), exchangeable bases (R^2 of 0.84; RMSE = 4.65; RPIQ = 6.54 and RPD = 3.94), Mehlich 3 extractable Fe (R^2 of 0.69; RMSE = 32.01; RPIQ = 0.97 and RPD = 1.248) and Mehlich 3 extractable P (R^2 of 0.41; RMSE = 16.79; RPIQ = 0.39 and RPD = 0.78) (Table 2). Exchangeable Ca and Mg have also been poorly predicted by other researchers [40,41], while EC and exchangeable Ca, Mg, K, and Na were poorly predicted with an RPD value < 1.3 by Pirie et al. [42]. Mehlich 3 extractable P, Mehlich 3 extractable Zn and electrical conductivity could not be predicted accurately with either model. Shepherd and Walsh [40] showed reasonable predictions for exchangeable Ca, Mg, and organic C in soils using NIR spectroscopy (R^2 = 0.78–0.88), while Udelhoven et al. [43] were unable to obtain suitable predictions for organic C, plant accessible N, K, and P on land-scape scaled experiment using NIR spectroscopy.

Table 2. Comparison of prediction accuracy from six machine learning algorithms using soil NIR spectra.

Soil Property	Method	R^2	RMSE	RPIQ	RPD	Soil Property	Method	R^2	RMSE	RPIQ	RPD
Total Carbon	RFO	0.83	0.46	1.15	1.28	m3.Cu	RFO	0.68	1.30	1.60	1.38
	GBM	0.75	0.59	0.90	1.00		GBM	0.59	1.51	1.38	1.19
	PLS	0.78	0.55	0.96	1.07		PLS	0.34	2.09	1.00	0.86
	BART	0.82	0.52	1.02	1.13		BART	0.72	1.22	1.70	1.47
	CUBIST	0.86	0.42	1.26	1.40		CUBIST	0.69	1.27	1.64	1.41
	ENS	0.87	0.39	1.36	1.51		ENS	0.73	1.2	1.73	1.50
Total Nitrogen	RFO	0.75	0.04	1.50	1.20	m3.Fe	RFO	0.63	34.96	0.89	1.14
	GBM	0.76	0.04	1.50	1.20		GBM	0.45	42.55	0.73	0.93
	PLS	0.72	0.04	1.50	1.20		PLS	0.53	40.98	0.76	0.97
	BART	0.78	0.04	1.50	1.20		BART	0.54	39.80	0.78	1.00
	CUBIST	0.67	0.05	1.20	0.96		CUBIST	0.69	32.01	0.97	1.24
	ENS	0.82	0.03	2.00	1.60		ENS	0.73	29.67	1.05	1.34
pH	RFO	0.56	0.58	3.60	2.31	m3.Mn	RFO	0.65	103.36	1.80	1.12
	GBM	0.56	0.60	3.48	2.23		GBM	0.49	125.62	1.48	0.92
	PLS	0.46	0.66	3.17	2.03		PLS	0.21	212.83	0.88	0.54
	BART	0.57	0.58	3.60	2.31		BART	0.72	92.26	2.02	1.25
	CUBIST	0.65	0.52	4.02	2.58		CUBIST	0.70	99.43	1.87	1.16
	ENS	0.66	0.51	4.10	2.63		ENS	0.75	85.30	2.19	1.35
m3.Al	RFO	0.49	201.58	2.38	1.65	m3.P	RFO	0.26	18.65	0.35	0.70
	GBM	0.41	212.01	2.26	1.56		GBM	0.17	19.66	0.33	0.67
	PLS	0.62	169.48	2.83	1.96		PLS	0.05	26.78	0.24	0.49
	BART	0.53	190.96	2.51	1.74		BART	0.16	19.84	0.33	0.66
	CUBIST	0.56	185.55	2.58	1.79		CUBIST	0.41	16.79	0.39	0.78
	ENS	0.68	157.12	3.05	2.11		ENS	0.41	16.58	0.39	0.79
m3.B	RFO	0.61	0.39	2.26	1.47	m3.S	RFO	0.03	15.5	1.13	2.23
	GBM	0.65	0.39	2.26	1.47		GBM	0.01	13.12	1.34	2.63
	PLS	0.52	0.48	1.83	1.19		PLS	0.11	12.68	1.39	2.73
	BART	0.62	0.38	2.32	1.51		BART	0.00	14.08	1.25	2.46
	CUBIST	0.71	0.34	2.59	1.69		CUBIST	0.02	12.94	1.36	2.67
	ENS	0.73	0.32	2.75	1.79		ENS	0.14	11.73	1.50	2.95
m3.Zn	RFO	0.33	2.59	0.42	0.86	ExNa	RFO	0.74	0.58	5.21	4.34
	GBM	0.40	2.50	0.44	0.89		GBM	0.57	0.74	4.08	3.40
	PLS	0.27	2.84	0.39	0.79		PLS	0.23	1.09	2.77	2.31
	BART	0.44	2.49	0.44	0.90		BART	0.72	0.6	5.03	4.19
	CUBIST	0.40	2.45	0.45	0.91		CUBIST	0.75	0.57	5.30	4.41
	ENS	0.49	2.24	0.49	1.00		ENS	0.81	0.50	6.04	5.03
PSI	RFO	0.32	63.64	2.07	1.46	ExCa	RFO	0.81	3.91	6.01	3.57
	GBM	0.38	60.69	2.18	1.53		GBM	0.79	4.29	5.48	3.25
	PLS	0.44	57.96	2.28	1.60		PLS	0.59	7.03	3.34	1.99
	BART	0.37	63.32	2.09	1.46		BART	0.80	3.97	5.92	3.52
	CUBIST	0.37	61.38	2.15	1.51		CUBIST	0.85	3.51	6.70	3.98
	ENS	0.52	52.61	2.51	1.76		ENS	0.85	3.47	6.78	4.02
ExMg	RFO	0.55	1.14	1.60	1.56	ExK	RFO	0.40	0.66	2.26	1.38
	GBM	0.50	1.22	1.49	1.46		GBM	0.33	0.71	2.10	1.28
	PLS	0.20	1.83	0.99	0.97		PLS	0.22	0.81	1.84	1.12
	BART	0.54	1.14	1.60	1.56		BART	0.47	0.62	2.40	1.47
	CUBIST	0.66	1.01	1.80	1.76		CUBIST	0.48	0.62	2.40	1.47
	ENS	0.67	0.96	1.90	1.85		ENS	0.51	0.60	2.48	1.52
ExBas	RFO	0.80	5.14	5.92	3.56	ECd	RFO	0.36	0.06	2.83	2.71
	GBM	0.77	5.74	5.30	3.19		GBM	0.37	0.05	3.40	3.25
	PLS	0.59	8.86	3.43	2.07		PLS	0.30	0.05	3.40	3.25
	BART	0.79	5.25	5.79	3.49		BART	0.23	0.07	2.43	2.32
	CUBIST	0.84	4.66	6.53	3.93		CUBIST	0.35	0.06	2.83	2.71
	ENS	0.84	4.65	6.54	3.94		ENS	0.40	0.05	3.40	1.38
ExAc	RFO	0.28	0.67	0.37	1.52						
	GBM	0.28	0.68	0.37	1.50						
	PLS	0.43	0.77	0.32	1.32						
	BART	0.38	0.66	0.38	1.54						
	CUBIST	0.34	0.64	0.39	1.59						
	ENS	0.52	0.55	0.45	1.85						

RFO = Random Forest Optimization; GBM = Gradient boosted machines; PLS = Partial least squares; BART = Bayesian additive regression trees; RPIQ = IQ/RMSE; RPD = SD/SEP.

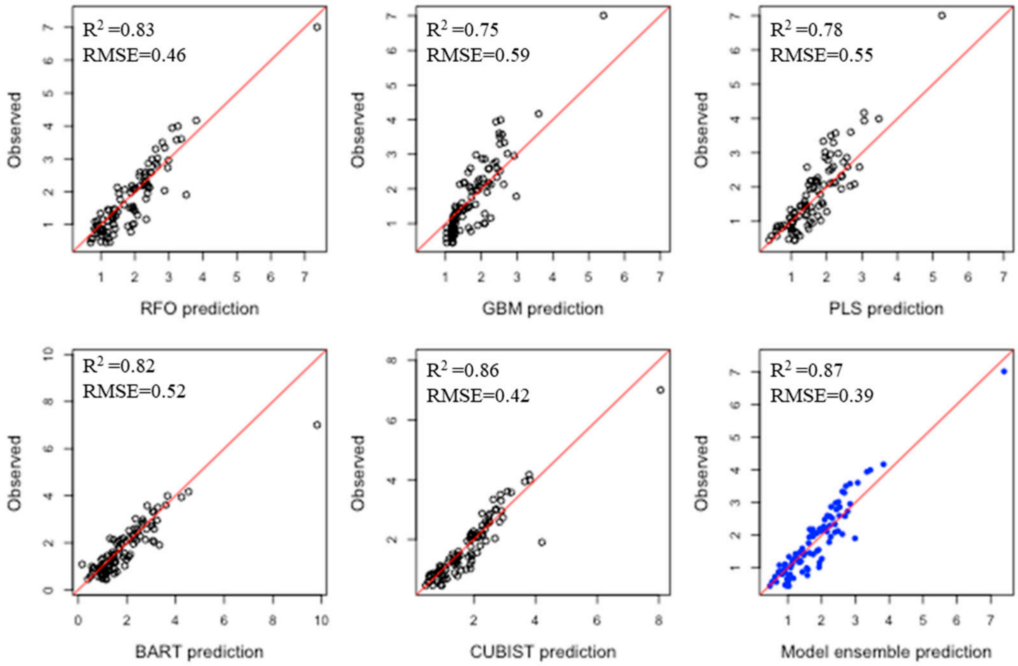

Figure 3. Scatter plot of observed and predicted total carbon content using RFO, GBM, PLS, BART, CUBIST and Total Ensemble.

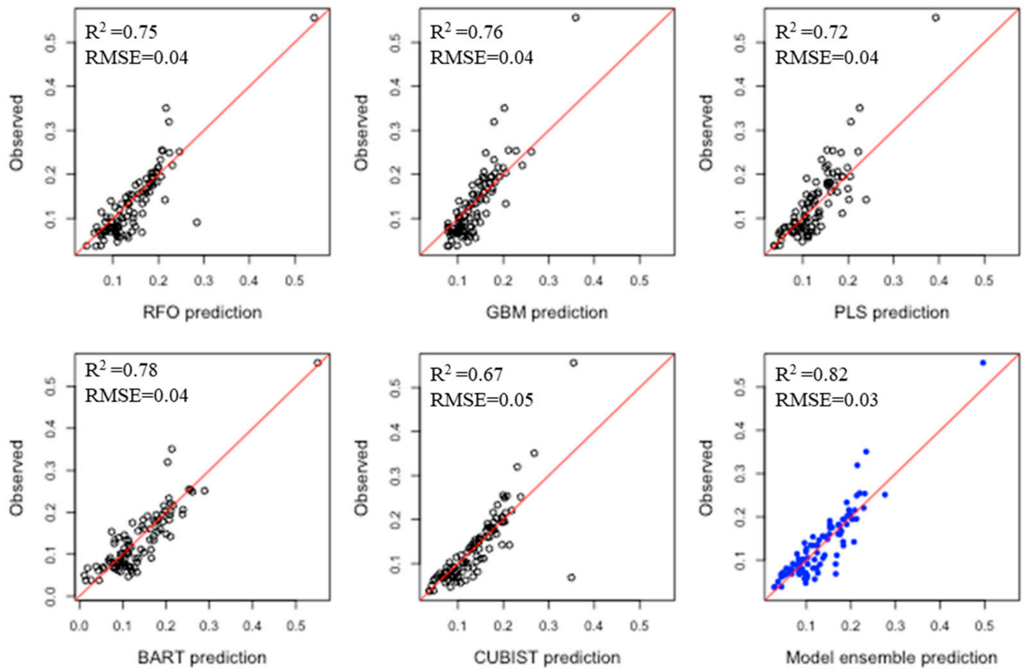

Figure 4. Scatter plot of observed and predicted total nitrogen content values using RFO, GBM, PLS, BART, CUBIST and total ensemble.

3.4. Soil Health Indicators of Different Land Uses as Predicted by Total Ensemble Algorithm

Soil properties in Kenya, Tanzania and Uganda differed significantly among land-use types and between soil depths (Table 3). In Kenya, the effect of land use at 0–15 cm soil depth was significant for TN, TC, Mn, P, Zn, PSI, exchangeable Na, K and electrical conductivity; soils collected from community forest (CF) at 0–15 cm and 15–45 cm soil depth had significantly higher TN and TC than the control. Soil pH, which is a measure of the soil acidity and alkalinity of soil with different land uses, was only significant at 45–100 cm depth ($p < 0.05$). The soil pH values of grassland (8.08 $\pm$ 0.59) and the control

(8.48 ± 0.30) were significantly higher than for cropland with soil and water conservation (CLSWC) (6.19 ± 0.65), community forest (6.416.94 ± 1.69) and agroforestry (6.41 ± 0.76; $p < 0.05$) (Table 3). The soils used for different land uses in Kenya are therefore suitable for agricultural activities (pH range of 6.62–7.46). Extremes in acidity or alkalinity of soil pH alter the available nutrients, resulting in imbalanced element uptake in plants [44]. Certain elements are more or less available at low soil acidic pH values (<5.5 units), thus resulting in nutrient constrains and unavailability in soils [45]. Maximum P availability occurs at soil pH range of 6.5–7.0; however, severe limitations with respect to P at 0–10 cm depth were observed in grassland (2.96 + 0.75 mg kg^{-1}), cropland without soil and water conservation (5.58 + 2.35 mg kg^{-1}) and community forest (5.50 + 3.02 mg kg^{-1}) as compared to the critical value of 8.5 mg kg^{-1} suggested for most crops [46]. The effect of land use and soil depth was non-significant for exchangeable Ca and exchangeable Mg at 0–15 cm, 15–45 cm and 45–100 cm, respectively. At 0–15 cm soil depth, cropland with soil and water conservation (CLSWC) had significantly high ExK (2.55 ± 0.57 cmolc kg^{-1}), compared to agroforestry (1.24 ± 0.47 cmolc kg^{-1}), and cropland without soil and water conservation (0.92 ± 0.53 cmolc kg^{-1}).

In Tanzania, the effect of land use on topsoil (0–15 cm) was significant for TN, TC, Mn, Al, Zn, PSI, exchangeable Ca, Mg, exbases and electrical conductivity (Table 3). TN was significantly high ($p < 0.05$) for the control (0.83 ± 0.03 %) than for cropland with soil and water conservation (0.18 ± 0.03 %) and other land uses at 0–15 cm soil depth. Soils from the community forest in Tanzania at 0–15 cm, 15–45 cm and 45–100 cm had significantly higher Fe contents than any other land use type, while community forest soils, collected at 0–15 cm and 15–45 cm soil depths had significantly higher Mn content than any other land-use type ($p < 0.05$; Table 3). Although Mn toxicity is ideal in acidic soils, soil pH across different land uses was not significant and ranged from 5.99 ± 0.46 to 6.47 ± 0.48, indicating that land uses have no significant effect on pH in Tanzania. Land use and soil depth were non-significant for exchangeable K and exchangeable Na, but remained significant for exchangeable Ca, exchangeable Mg at 0–15 cm, and 15–45 cm and 45–100 cm soil depths. Exchangeable K and exchangeable Ca in the soil also depends on the composition of parent rock materials [45].

In Uganda, land use and soil depth were significant for TN, TC, Mn, P, Zn, PSI, exchangeable Na and exchangeable K (Table 3). TN and TC were significantly high at the surface soil (0–15 cm) for five land use types compared to the control. Soil pH at the surface soil across different land uses were significant, with more acidic soils observed for the control (5.31 ± 0.59) and grasslands (5.79 ± 0.52). The highest soil pH (6.89 ± 0. 72) and the lowest soil pH (5.31 ± 0.59) were recorded for agroforestry and control land uses, respectively. Lower pH values for soils collected from grassland and the control could be as result of basic cation depletion in these soils. However, at 0–15 cm, 15–45 cm and 45–100 cm soil depths, soils from grassland recorded significantly higher Mn contents than the control (Table 3). The decline in soil pH across grassland could be due to encroachment into the grassland to create land for grazing and agricultural activities. Land use and soil depth did not affect exchangeable K and exchangeable Na, which could be a result of the removal of vegetation cover due to human and livestock interference [47]. A significantly high concentration of exchangeable Ca (18.21 ± 7.51 cmol/kg soil) for surface soils (0–15 cm) under agroforestry land use as compared to the control was probably due to decomposing litter that change the relative quantities of exchangeable base (Ca, Mg) and acid (Al, Fe) cations in soil [45]. Agroforestry through practices such as litter incorporation and manure application has been suggested to improve soil carbon sequestration [48].

Table 3. Soil quality indicators in different land uses in three countries at different depths as predicted by the ensemble algorithm on NIR spectra.

Country	Depth (cm)	Land Use (n)	TN %	T C %	pH Units	Al mg kg^{-1}	Cu mg kg^{-1}	Fe mg kg^{-1}	Mn mg kg^{-1}	P mg kg^{-1}	Zn mg kg^{-1}	PSI Units	ExNa cmolc kg^{-1}	ExCa cmolc kg^{-1}	ExMg cmolc kg^{-1}	ExK cmolc kg^{-1}	ExBas cmolc kg^{-1}	ECd cmolc kg^{-1}
KENYA	0–15	AF(6)	0.15 ± 0.06 ^{ab}	2.14 ± 0.89 ^{ab}	6.72 ± 0.89 ^{a}	676.33 ± 205.70 ^{a}	1.66 ± 0.99 ^{a}	168.4 ± 80.65 ^{a}	338.67 ± 75.46 ^{a}	61.74 ± 61.97 ^{ab}	8.71 ± 11.57 ^{ab}	61.66 ± 56.70 ^{b}	0.06 ± 0.05 ^{b}	18.34 ± 12.80 ^{a}	3.67 ± 1.62 ^{a}	1.24 ± 0.47 ^{bc}	23.31 ± 14.69 ^{a}	0.84 ± 0.05 ^{b}
		CF(6)	0.24 ± 0.16 ^{a}	3.88 ± 2.82 ^{a}	7.09 ± 0.73 ^{a}	818.83 ± 153.70 ^{a}	2.86 ± 2.02 ^{a}	153.00 ± 52.00 ^{a}	269.50 ± 56.87 ^{ab}	5.50 ± 3.02 ^{b}	4.73 ± 3.91 ^{ab}	136.00 ± 17.75 ^{a}	1.08 ± 1.36 ^{b}	28.78 ± 79.44 ^{a}	5.97 ± 0.80 ^{a}	2.03 ± 0.71 ^{ab}	37.85 ± 9.72 ^{a}	0.22 ± 0.14 ^{ab}
		CLSWC(6)	0.23 ± 0.05 ^{ab}	3.22 ± 0.62 ^{ab}	7.04 ± 0.73 ^{a}	843.83 ± 103.34 ^{a}	4.33 ± 0.29 ^{a}	145.00 ± 12.31 ^{a}	269.50 ± 56.87 ^{ab}	68.95 ± 53.30 ^{a}	12.97 ± 7.01 ^{ab}	89.55 ± 30.15 ^{ab}	0.04 ± 0.03 ^{b}	18.69 ± 6.06 ^{a}	4.33 ± 0.72 ^{a}	2.55 ± 0.57 ^{a}	25.62 ± 6.72 ^{a}	0.12 ± 0.05 ^{ab}
		CLNSWC(6)	0.13 ± 0.06 ^{ab}	1.78 ± 0.78 ^{ab}	6.62 ± 0.89 ^{a}	778.67 ± 48.32 ^{a}	2.13 ± 1.38 ^{a}	221.67 ± 78.57 ^{a}	301.00 ± 81.19 ^{b}	5.58 ± 2.35 ^{b}	2.33 ± 0.58 ^{b}	64.02 ± 32.80 ^{ab}	0.67 ± 0.26 ^{b}	17.32 ± 9.62 ^{a}	3.09 ± 1.92 ^{a}	0.92 ± 0.53 ^{c}	21.98 ± 12.11 ^{a}	0.06 ± 0.02 ^{ab}
		GL(6)	0.14 ± 0.06 ^{ab}	2.11 ± 0.47 ^{ab}	6.93 ± 1.08 ^{a}	765.00 ± 73.52 ^{a}	2.60 ± 1.86 ^{a}	217.88 ± 130.59 ^{a}	169.97 ± 75.45 ^{ab}	2.96 ± 0.75 ^{b}	1.71 ± 0.44 ^{b}	99.05 ± 20.06 ^{ab}	1.04 ± 0.38 ^{b}	22.92 ± 15.52 ^{a}	3.33 ± 2.17 ^{a}	1.59 ± 0.26 ^{c}	28.22 ± 18.34 ^{a}	0.76 ± 0.03 ^{b}
		C(6)	0.09 ± 0.02 ^{b}	1.27 ± 0.41 ^{b}	7.46 ± 0.89 ^{a}	804.17 ± 132.09 ^{a}	2.25 ± 1.68 ^{a}	145.40 ± 62.75 ^{a}	254.33 ± 91.44 ^{ab}	56.27 ± 6.32 ^{b}	1.16 ± 0.28 ^{b}	99.88 ± 24.82 ^{ab}	3.52 ± 2.22 ^{a}	21.91 ± 10.71 ^{a}	3.98 ± 2.06 ^{a}	1.54 ± 0.78 ^{bc}	34.52 ± 10.45 ^{a}	0.26 ± 0.17 ^{a}
	15–45	AF(6)	0.10 ± 0.02 ^{ab}	1.49 ± 0.42 ^{ab}	6.70 ± 0.99 ^{a}	789.67 ± 172.60 ^{ab}	1.79 ± 1.50 ^{a}	189.17 ± 101.92 ^{a}	255.17 ± 107.79 ^{a}	13.91 ± 11.28 ^{a}	2.99 ± 2.78 ^{a}	95.08 ± 46.17 ^{ab}	0.19 ± 0.17 ^{b}	15.37 ± 5.48 ^{a}	4.09 ± 1.24 ^{a}	1.43 ± 0.81 ^{abc}	21.07 ± 7.12 ^{a}	0.08 ± 0.06 ^{b}
		CF(6)	0.10 ± 0.04 ^{ab}	1.67 ± 0.60 ^{ab}	6.82 ± 1.53 ^{a}	941.17 ± 342.97 ^{ab}	2.83 ± 2.21 ^{a}	170.03 ± 92.27 ^{a}	301.17 ± 61.11 ^{a}	3.00 ± 3.01 ^{a}	2.35 ± 1.22 ^{a}	149.45 ± 50.47 ^{a}	1.73 ± 2.02 ^{b}	23.23 ± 15.02 ^{a}	4.79 ± 0.83 ^{a}	1.32 ± 0.58 ^{abc}	31.07 ± 17.57 ^{a}	0.12 ± 1.10 ^{ab}
		CLSWC(6)	0.15 ± 0.04 ^{a}	2.04 ± 0.65 ^{a}	6.55 ± 0.59 ^{a}	999.67 ± 198.04 ^{a}	4.07 ± 4.17 ^{a}	124.35 ± 35.66 ^{a}	191.60 ± 90.40 ^{b}	17.72 ± 30.79 ^{a}	4.33 ± 4.78 ^{a}	128.52 ± 42.29 ^{ab}	0.08 ± 0.05 ^{b}	13.70 ± 3.65 ^{a}	5.01 ± 1.85 ^{a}	1.86 ± 0.58 ^{ab}	20.66 ± 5.07 ^{a}	0.06 ± 0.02 ^{b}
		CLNSWC(6)	0.10 ± 0.05 ^{ab}	1.54 ± 0.61 ^{ab}	6.94 ± 0.85 ^{a}	694.67 ± 89.00 ^{ab}	2.00 ± 1.64 ^{a}	166.50 ± 34.64 ^{a}	105.85 ± 61.54 ^{ab}	2.37 ± 1.59 ^{a}	2.94 ± 1.95 ^{a}	68.67 ± 46.99 ^{b}	1.49 ± 1.13 ^{b}	18.15 ± 12.27 ^{a}	3.20 ± 2.44 ^{a}	0.88 ± 0.46 ^{bc}	23.71 ± 16.11 ^{a}	0.12 ± 0.10 ^{ab}
		GL(6)	0.07 ± 0.03 ^{b}	1.23 ± 0.17 ^{ab}	7.33 ± 1.13 ^{a}	605.50 ± 38.40 ^{b}	2.82 ± 2.52 ^{a}	153.03 ± 83.50 ^{a}	175.27 ± 98.26 ^{ab}	2.48 ± 1.58 ^{a}	4.78 ± 1.75 ^{a}	78.78 ± 27.82 ^{ab}	1.89 ± 1.43 ^{b}	23.48 ± 15.32 ^{a}	3.32 ± 2.64 ^{ab}	1.77 ± 0.35 ^{c}	29.46 ± 23.72 ^{a}	0.12 ± 0.07 ^{ab}
		C(6)	0.07 ± 0.01 ^{b}	1.07 ± 0.59 ^{b}	8.03 ± 0.60 ^{a}	774.17 ± 153.54 ^{ab}	2.73 ± 1.40 ^{a}	92.47 ± 925.57 ^{a}	265.50 ± 77.56 ^{ab}	12.24 ± 8.91 ^{a}	2.24 ± 1.16 ^{a}	83.65 ± 19.92 ^{ab}	4.64 ± 2.41 ^{a}	30.50 ± 6.07 ^{a}	4.37 ± 2.00 ^{a}	1.97 ± 0.58 ^{a}	41.47 ± 8.38 ^{a}	0.25 ± 0.09 ^{a}
	45–100	AF(6)	0.05 ± 0.01 ^{a}	0.74 ± 0.23 ^{a}	6.41 ± 0.76 ^{cd}	863.17 ± 316.21 ^{ab}	1.57 ± 1.64 ^{a}	163.83 ± 47.49 ^{a}	221.00 ± 41.69 ^{a}	15.52 ± 15.63 ^{a}	1.30 ± 0.75 ^{a}	111.23 ± 84.40 ^{ab}	0.29 ± 0.30 ^{b}	11.21 ± 6.80 ^{b}	3.67 ± 1.38 ^{a}	1.77 ± 0.74 ^{ab}	16.94 ± 7.98 ^{b}	0.06 ± 0.05 ^{b}
		CF(6)	0.07 ± 0.04 ^{a}	1.27 ± 0.62 ^{a}	6.94 ± 1.69 ^{bcd}	928.50 ± 362.77 ^{ab}	2.83 ± 2.07 ^{a}	153.90 ± 89.84 ^{ab}	304.17 ± 114.36 ^{a}	3.92 ± 4.61 ^{a}	1.63 ± 0.78 ^{a}	132.88 ± 60.88 ^{ab}	2.28 ± 2.28 ^{b}	23.40 ± 15.27 ^{ab}	4.48 ± 0.91 ^{b}	1.16 ± 0.26 ^{b}	31.31 ± 18.35 ^{ab}	0.12 ± 0.11 ^{ab}
		CLSWC(6)	0.09 ± 0.03 ^{a}	1.37 ± 0.67 ^{a}	6.19 ± 0.65 ^{d}	1124.33 ± 151.24 ^{cd}	3.08 ± 3.63 ^{a}	112.43 ± 20.22 ^{abc}	132.82 ± 40.16 ^{a}	2.28 ± 2.79 ^{a}	0.66 ± 0.33 ^{a}	165.23 ± 69.74 ^{a}	0.14 ± 0.13 ^{b}	10.24 ± 4.38 ^{b}	5.54 ± 2.49 ^{a}	1.28 ± 0.51 ^{ab}	17.20 ± 6.50 ^{b}	0.07 ± 0.08 ^{b}
		CLNSWC(6)	0.06 ± 0.01 ^{a}	1.02 ± 0.28 ^{a}	7.91 ± 0.48 ^{abc}	689.33 ± 157.26 ^{b}	2.38 ± 1.84 ^{a}	83.00 ± 23.44 ^{bc}	155.10 ± 146.93 ^{a}	5.71 ± 4.44 ^{a}	0.78 ± 0.27 ^{a}	63.40 ± 17.81 ^{b}	2.78 ± 1.59 ^{ab}	28.71 ± 9.43 ^{a}	4.82 ± 1.53 ^{a}	1.90 ± 0.61 ^{ab}	38.21 ± 11.82 ^{a}	0.21 ± 0.13 ^{ab}
		GL(6)	0.05 ± 0.01 ^{a}	0.88 ± 0.10 ^{a}	8.08 ± 0.59 ^{ab}	731.671 ± 203.55 ^{ab}	2.89 ± 2.44 ^{a}	80.00 ± 29.26 ^{bc}	190.98 ± 168.07 ^{a}	3.33 ± 1.97 ^{a}	0.85 ± 0.28 ^{a}	94.23 ± 34.45 ^{ab}	3.21 ± 1.52 ^{ab}	30.58 ± 9.84 ^{a}	4.92 ± 0.96 ^{a}	1.91 ± 0.75 ^{ab}	40.62 ± 11.14 ^{a}	0.17 ± 0.07 ^{ab}
		C(6)	0.06 ± 0.06 ^{a}	0.78 ± 0.28 ^{a}	8.48 ± 0.30 ^{a}	684.67 ± 186.27 ^{b}	2.89 ± 1.87 ^{a}	61.20 ± 3.44 ^{c}	250.00 ± 18.68 ^{a}	26.22 ± 29.61 ^{a}	3.50 ± 5.15 ^{a}	79.27 ± 7.05 ^{ab}	5.36 ± 2.90 ^{a}	38.14 ± 6.29 ^{a}	4.48 ± 1.71 ^{a}	2.32 ± 0.74 ^{a}	50.30 ± 7.06 ^{a}	0.31 ± 0.17 ^{a}
Tanzania	0–15	AF(6)	0.21 ± 0.04 ^{b}	2.64 ± 0.48 ^{a}	6.40 ± 0.25 ^{a}	908.38 ± 25.28 ^{ab}	7.44 ± 5.56 ^{a}	62.35 ± 10.23 ^{b}	306.33 ± 98.08 ^{ab}	2.39 ± 1.75 ^{a}	7.32 ± 2.79 ^{a}	121.67 ± 28.20 ^{a}	0.02 ± 0.01 ^{a}	10.56 ± 3.17 ^{a}	3.27 ± 0.12 ^{ab}	0.27 ± 0.18 ^{a}	14.11 ± 3.15 ^{a}	0.07 ± 0.01 ^{a}
		CF(3)	0.24 ± 0.05 ^{b}	2.76 ± 0.50 ^{a}	6.42 ± 0.22 ^{a}	832.67 ± 44.74 ^{b}	7.40 ± 1.23 ^{a}	157.33 ± 41.31 ^{a}	348.67 ± 87.27 ^{a}	2.53 ± 0.97 ^{a}	7.24 ± 1.65 ^{a}	94.47 ± 17.59 ^{a}	0.02 ± 0.00 ^{a}	11.92 ± 3.17 ^{a}	3.88 ± 065 ^{a}	0.13 ± 0.05 ^{a}	15.95 ± 3.86 ^{a}	0.09 ± 0.02 ^{a}
		CLSWC(6)	0.18 ± 0.03 ^{b}	2.23 ± 0.33 ^{a}	5.99 ± 0.46 ^{a}	995.17 ± 89.56 ^{a}	5.97 ± 3.07 ^{a}	78.00 ± 18.57 ^{b}	129.55 ± 97.35 ^{b}	5.80 ± 3.11 ^{a}	2.98 ± 2.40 ^{ab}	97.05 ± 6.71 ^{a}	0.02 ± 0.03 ^{a}	7.89 ± 2.49 ^{ab}	2.52 ± 0.92 ^{ab}	0.17 ± 0.14 ^{a}	10.61 ± 3.42 ^{ab}	0.04 ± 0.01 ^{a}
		CLNSWC(6)	0.22 ± 0.05 ^{b}	2.61 ± 0.61 ^{a}	6.47 ± 0.48 ^{a}	937.33 ± 86.31 ^{ab}	8.32 ± 4.70 ^{a}	81.40 ± 37.31 ^{b}	279.17 ± 124.67 ^{ab}	15.74 ± 22.08 ^{a}	5.85 ± 1.96 ^{a}	96.78 ± 30.02 ^{a}	0.02 ± 0.01 ^{a}	10.53 ± 3.48 ^{a}	3.30 ± 1.18 ^{ab}	0.31 ± 0.39 ^{a}	14.15 ± 5.00 ^{a}	0.08 ± 0.03 ^{a}
		GL(6)	0.21 ± 0.05 ^{b}	2.59 ± 0.43 ^{a}	6.10 ± 0.42 ^{a}	963.33 ± 764.66 ^{a}	6.46 ± 3.72 ^{a}	73.32 ± 18.43 ^{b}	159.37 ± 129.13 ^{ab}	4.49 ± 2.62 ^{a}	3.82 ± 3.27 ^{ab}	101.33 ± 14.60 ^{a}	0.01 ± 0.01 ^{a}	8.68 ± 3.21 ^{ab}	3.00 ± 0.50 ^{ab}	0.13 ± 0.09 ^{a}	11.82 ± 3.42 ^{ab}	0.06 ± 0.01 ^{ab}
		C(6)	0.83 ± 0.03 ^{a}	0.93 ± 0.23 ^{b}	6.28 ± 0.52 ^{a}	930.50 ± 64.20 ^{ab}	4.05 ± 2.01 ^{a}	60.28 ± 13.19 ^{b}	160.05 ± 113.86 ^{ab}	0.44 ± 0.49 ^{a}	0.47 ± 3.27 ^{b}	135.02 ± 39.39 ^{a}	0.02 ± 0.01 ^{a}	4.77 ± 1.90 ^{b}	2.05 ± 1.00 ^{b}	0.08 ± 0.10 ^{a}	9.92 ± 2.71 ^{b}	0.03 ± 0.00 ^{c}
	15–45	AF(6)	0.14 ± 0.05 ^{ab}	1.62 ± 0.64 ^{ab}	6.59 ± 0.34 ^{a}	879.50 ± 43.22 ^{a}	6.46 ± 5.34 ^{a}	53.97 ± 11.77 ^{b}	194.90 ± 127.46 ^{ab}	0.93 ± 1.08 ^{ab}	3.35 ± 3.31 ^{a}	133.92 ± 27.43 ^{a}	0.03 ± 0.03 ^{a}	8.23 ± 3.88 ^{a}	2.79 ± 0.66 ^{ab}	0.07 ± 0.05 ^{a}	11.11 ± 4.52 ^{a}	0.05 ± 0.01 ^{a}
		CF(3)	0.13 ± 0.01 ^{ab}	1.41 ± 0.11 ^{ab}	6.37 ± 0.12 ^{a}	844.00 ± 10.82 ^{a}	4.64 ± 0.07 ^{a}	116.67 ± 9.50 ^{a}	392.00 ± 124.90 ^{a}	0.00 ± 0.00 ^{b}	3.21 ± 1.35 ^{a}	107.67 ± 5.69 ^{a}	0.03 ± 0.00 ^{a}	7.97 ± 0.80 ^{a}	3.96 ± 0.22 ^{a}	0.08 ± 0.01 ^{a}	12.04 ± 1.00 ^{a}	0.05 ± 0.01 ^{a}
		CLSWC(6)	0.12 ± 0.05 ^{ab}	1.42 ± 0.62 ^{ab}	6.18 ± 0.52 ^{a}	949.83 ± 96.80 ^{a}	4.46 ± 1.83 ^{a}	54.52 ± 17.50 ^{b}	48.97 ± 30.32 ^{b}	1.42 ± 1.81 ^{ab}	0.56 ± 0.45 ^{aa}	115.00 ± 12.55 ^{a}	0.04 ± 0.04 ^{a}	6.63 ± 2.36 ^{a}	2.39 ± 0.95 ^{ab}	0.05 ± 0.02 ^{a}	9.11 ± 3.27 ^{a}	0.04 ± 0.02 ^{a}
		CLNSWC(6)	0.14 ± 0.05 ^{ab}	1.60 ± 0.55 ^{ab}	6.38 ± 0.40 ^{a}	916.67 ± 90.96 ^{a}	5.48 ± 2.34 ^{a}	58.82 ± 11.23 ^{b}	185.50 ± 127.83 ^{ab}	2.75 ± 3.11 ^{ab}	1.93 ± 1.43 ^{a}	107.27 ± 35.80 ^{a}	0.02 ± 0.01 ^{a}	7.52 ± 2.92 ^{a}	2.10 ± 0.59 ^{b}	0.08 ± 0.08 ^{a}	9.72 ± 3.51 ^{a}	0.04 ± 0.01 ^{a}
		GL(6)	0.20 ± 0.04 ^{a}	2.43 ± 0.60 ^{a}	5.98 ± 0.48 ^{a}	981.83 ± 91.02 ^{a}	5.94 ± 3.57 ^{a}	77.03 ± 16.88 ^{b}	134.30 ± 114.82 ^{b}	3.63 ± 2.22 ^{a}	2.79 ± 3.54 ^{a}	104.40 ± 25.07 ^{a}	0.02 ± 0.03 ^{a}	8.40 ± 2.56 ^{a}	2.28 ± 0.86 ^{b}	0.07 ± 0.06 ^{a}	10.77 ± 3.37 ^{a}	0.05 ± 0.01 ^{a}
		C(6)	0.06 ± 0.03 ^{b}	0.75 ± 0.28 ^{b}	6.27 ± 0.71 ^{a}	962.50 ± 100.41 ^{a}	3.45 ± 2.21 ^{a}	57.98 ± 32.52 ^{b}	110.69 ± 120.97 ^{b}	0.21 ± 0.34 ^{ab}	0.16 ± 0.28 ^{a}	141.88 ± 53.08 ^{a}	0.04 ± 0.02 ^{a}	4.29 ± 2.11 ^{a}	1.99 ± 1.16 ^{b}	0.02 ± 0.02 ^{a}	6.33 ± 3.00 ^{a}	0.03 ± 0.01 ^{a}
	45–100	AF(6)	0.05 ± 0.01 ^{b}	0.71 ± 0.24 ^{ab}	6.49 ± 0.59 ^{a}	877.17 ± 106.93 ^{a}	2.65 ± 1.14 ^{a}	42.85 ± 20.11 ^{b}	35.22 ± 43.35 ^{b}	0.05 ± 0.12 ^{a}	0.22 ± 0.27 ^{a}	143.93 ± 41.59 ^{a}	0.04 ± 0.01 ^{a}	4.99 ± 2.17 ^{a}	2.06 ± 0.75 ^{ab}	0.03 ± 0.03 ^{a}	7.11 ± 2.76 ^{a}	0.03 ± 0.01 ^{a}
		CF(3)	0.07 ± 0.01 ^{ab}	0.69 ± 0.08 ^{ab}	6.48 ± 0.17 ^{a}	888.00 ± 21.66 ^{a}	2.52 ± 0.25 ^{a}	91.30 ± 23.13 ^{a}	233.77 ± 141.97 ^{a}	0.00 ± 0.00 ^{a}	0.54 ± 0.63 ^{a}	142.67 ± 29.67 ^{a}	0.08 ± 0.01 ^{a}	5.46 ± 0.99 ^{a}	3.70 ± 0.58 ^{a}	0.06 ± 0.01 ^{a}	9.30 ± 1.56 ^{a}	0.04 ± 0.01 ^{a}
		CLSWC(6)	0.07 ± 0.04 ^{ab}	0.91 ± 0.52 ^{ab}	6.18 ± 0.47 ^{a}	1000.00 ± 105.50 ^{a}	3.69 ± 1.22 ^{a}	43.15 ± 16.81 ^{b}	21.55 ± 36.55 ^{b}	0.66 ± 0.84 ^{a}	0.04 ± 0.09 ^{a}	122.27 ± 20.81 ^{a}	0.05 ± 0.07 ^{a}	5.59 ± 2.62 ^{a}	2.29 ± 1.26 ^{ab}	0.02 ± 0.02 ^{a}	7,95 ± 3.87 ^{a}	0.04 ± 0.02 ^{a}
		CLNSWC(6)	0.06 ± 0.02 ^{ab}	0.81 ± 0.13 ^{ab}	6.49 ± 0.36 ^{a}	928.00 ± 13.44 ^{a}	4.11 ± 1.17 ^{a}	39.38 ± 9.40 ^{b}	59.62 ± 78.58 ^{b}	0.97 ± 1.09 ^{a}	0.19 ± 0.16 ^{a}	118.80 ± 25.23 ^{a}	0.04 ± 0.03 ^{a}	5.16 ± 0.99 ^{a}	1.57 ± 0.51 ^{a}	0.03 ± 0.03 ^{a}	6.79 ± 1.41 ^{a}	0.03 ± 0.01 ^{a}
		GL(6)	0.12 ± 0.05 ^{a}	1.45 ± 0.63 ^{a}	6.21 ± 0.40 ^{a}	962.67 ± 44.13 ^{a}	4.38 ± 2.66 ^{a}	56.42 ± 12.34 ^{b}	75.33 ± 54.92 ^{b}	0.78 ± 0.95 ^{a}	0.98 ± 1.21 ^{a}	123.17 ± 19.46 ^{a}	0.04 ± 0.03 ^{a}	6.65 ± 1.87 ^{a}	2.04 ± 0.81 ^{ab}	0.04 ± 0.02 ^{a}	8.77 ± 2.59 ^{a}	0.04 ± 0.02 ^{a}
		C(6)	0.05 ± 0.03 ^{b}	0.71 ± 0.34 ^{b}	6.29 ± 0.86 ^{a}	920.33 ± 178.02 ^{a}	3.33 ± 2.15 ^{a}	46.13 ± 18.50 ^{b}	73.62 ± 96.54 ^{b}	0.17 ± 0.19 ^{a}	0.24 ± 0.42 ^{a}	148.02 ± 56.09 ^{a}	0.04 ± 0.02 ^{a}	3.93 ± 2.66 ^{a}	1.84 ± 1.12 ^{b}	0.02 ± 0.02 ^{a}	5.84 ± 3.46 ^{a}	0.03 ± 0.01 ^{a}
Uganda	0–15	AF(6)	0.22 ± 0.03 ^{a}	3.07 ± 0.32 ^{ab}	6.89 ± 0. 72 ^{a}	937.17 ± 115.00 ^{b}	3.68 ± 0. 52 ^{b}	75.08 ± 9.67 ^{b}	468.83 ± 74.49 ^{a}	22.52 ± 45.35 ^{a}	3.25 ± 1.25 ^{a}	89.83 ± 17.08 ^{ab}	0.03 ± 0.01 ^{a}	18.21 ± 7.51 ^{a}	5.50 ± 0.79 ^{a}	0.91 ± 0.74 ^{a}	24.65 ± 8.67 ^{a}	0.10 ± 0.02 ^{a}
		CF(6)	0.36 ± 0.13 ^{ab}	4.50 ± 1.66 ^{a}	6.03 ± 0.87 ^{abc}	1031.50 ± 351.99 ^{ab}	3.40 ± 1.66 ^{ab}	193.50 ± 104.62 ^{a}	200.38 ± 154.89 ^{c}	14.45 ± 8.43 ^{a}	3.27 ± 0.59 ^{a}	94.95 ± 53.41 ^{ab}	0.13 ± 0.18 ^{a}	14.23 ± 7.23 ^{ab}	5.17 ± 2.13 ^{a}	0.61 ± 0.25 ^{a}	20.37 ± 9.25 ^{ab}	0.1140.02 ^{a}
		CLSWC(6)	0.20 ± 0.05 ^{ab}	2.68 ± 0.76 ^{ab}	6.46 ± 0.29 ^{ab}	922.50 ± 90.68 ^{b}	4.52 ± 0.32 ^{a}	100.02 ± 22.71 ^{b}	557.00 ± 62.53 ^{a}	6.08 ± 4.26 ^{a}	3.24 ± 0.59 ^{a}	64.78 ± 14.50 ^{b}	0.03 ± 0.02 ^{a}	12.15 ± 2.49 ^{abc}	3.24 ± 0.70 ^{ab}	1.17 ± 0.93 ^{a}	16.59 ± 3.66 ^{abc}	0.09 ± 0.03 ^{a}
		CLNSWC(6)	0.23 ± 0.04 ^{ab}	3.40 ± 0.32 ^{ab}	6.32 ± 0.25 ^{abc}	969.67 ± 127.33 ^{ab}	3.36 ± 0.60 ^{ab}	96.62 ± 20.24 ^{b}	398.83 ± 51.47 ^{ab}	2.57 ± 0.46 ^{a}	1.88 ± 1.41 ^{ab}	108.05 ± 42.91 ^{ab}	0.03 ± 0.02 ^{a}	14.41 ± 1.89 ^{ab}	4.13 ± 0.42 ^{ab}	0.56 ± 0.50 ^{a}	19.13 ± 2.07 ^{ab}	0.06 ± 0.02 ^{a}
		GL(6)	0.32 ± 0.25 ^{ab}	4.82 ± 3.47 ^{a}	5.79 ± 0.52 ^{bc}	1483.33 ± 483.47 ^{a}	2.19 ± 0.83 ^{ab}	141.38 ± 55.77 ^{ab}	104.23 ± 122.66 ^{c}	3.63 ± 0. 88 ^{a}	0.76 ± 0.35 ^{b}	220.35 ± 164.81 ^{a}	0.06 ± 0.03 ^{a}	7.96 ± 4.72 ^{bc}	3.11 ± 2.13 ^{ab}	0.40 ± 0.47 ^{a}	11.53 ± 6.74 ^{bc}	0.04 ± 0.02 ^{a}
		C(6)	0.13 ± 0.03 ^{b}	1.35 ± 0.73 ^{b}	5.31 ± 0.59 ^{c}	1391.67 ± 200.00 ^{ab}	2.16 ± 0.90 ^{b}	91.08 ± 29.77 ^{b}	252.66 ± 140.10 ^{bc}	4.91 ± 8. 09 ^{a}	0.76 ± 1.17 ^{b}	223.00 ± 104.40 ^{a}	0.26 ± 0.45 ^{a}	4.79 ± 3.12 ^{c}	2.13 ± 0.91 ^{b}	0.58 ± 0.57 ^{a}	7.73 ± 3.72 ^{c}	0.22 ± 0.27 ^{a}
	15–45	AF(6)	013. ± 0.04 ^{a}	1.76 ± 0.54 ^{a}	6.59 ± 0.84 ^{a}	1145.00 ± 165.62 ^{b}	3.54 ± 0.54 ^{a}	92.27 ± 9.40 ^{a}	390.83 ± 129.82 ^{a}	1.28 ± 1.03 ^{ab}	0.89 ± 0.52 ^{ab}	134.55 ± 50.11 ^{bc}	0.05 ± 0.03 ^{a}	10.17 ± 3.73 ^{a}	4.19 ± 1.16 ^{a}	0.82 ± 1.35 ^{a}	15.23 ± 4.77 ^{a}	0.06 ± 0.03 ^{b}
		CF(6)	0.16 ± 0.05 ^{a}	2.06 ± 0.73 ^{a}	5.71 ± 0.54 ^{ab}	1251.67 ± 252.62 ^{b}	3.37 ± 1.12 ^{a}	196.43 ± 136.02 ^{ab}	128.95 ± 119.65 ^{b}	3.34 ± 2.34 ^{a}	0.68 ± 0.32 ^{b}	240.35 ± 80.11 ^{bc}	0.78 ± 1.71 ^{a}	3.13 ± 4.42 ^{ab}	3.67 ± 1.70 ^{a}	047. ± 0.39 ^{a}	12.05 ± 5.82 ^{a}	0.05 ± 0.02 ^{ab}
		CLSWC(6)	0.13 ± 0.03 ^{a}	1.58 ± 0.64 ^{a}	6.46 ± 0.10 ^{a}	1024.17 ± 40.05 ^{b}	3.84 ± 0.90 ^{a}	120.40 ± 28.20 ^{ab}	519.17 ± 123.35 ^{a}	1.61 ± 2.50 ^{ab}	1.59 ± 0.83 ^{a}	101.97 ± 27.29 ^{c}	0.02 ± 0.01 ^{a}	9.68 ± 3.28 ^{a}	3.10 ± 0.82 ^{ab}	0.45 ± 0.35 ^{a}	13.25 ± 3.48 ^{a}	0.05 ± 0.02 ^{ab}
		CLNSWC(6)	0.15 ± 0.03 ^{a}	2.12 ± 0.20 ^{a}	6.17 ± 0.53 ^{a}	1144.50 ± 186.72 ^{b}	3.45 ± 0.50 ^{a}	109.78 ± 13.02 ^{ab}	358.00 ± 90.40 ^{a}	0.54 ± 0.34 ^{b}	0.69 ± 0.26 ^{b}	158.50 ± 53.09 ^{bc}	0.05 ± 0.02 ^{a}	10.33 ± 2.78 ^{a}	3.07 ± 0.31 ^{ab}	0.26 ± 0.23 ^{a}	13.71 ± 3.16 ^{a}	0.04 ± 0.01 ^{ab}
		GL(6)	0.23 ± 0.24 ^{a}	3.28 ± 20.95 ^{a}	5.13 ± 0.33 ^{b}	1863.33 ± 496.41 ^{a}	1.73 ± 0.80 ^{b}	103.38 ± 50.45 ^{a}	33.06 ± 41.91 ^{b}	1.29 ± 0.84 ^{ab}	0.39 ± 0.13 ^{b}	346.83 ± 172.25 ^{a}	0.07 ± 0.05 ^{a}	2.03 ± 2.16 ^{b}	0.83 ± 1.15 ^{c}	0.13 ± 0.06 ^{a}	3.06 ± 3.27 ^{b}	0.02 ± 0.01 ^{ab}
		C(6)	0.11 ± 0.09 ^{a}	1.26 ± 1.24 ^{a}	5.05 ± 0.62 ^{b}	1471.67 ± 190.62 ^{ab}	1.57 ± 0.88 ^{b}	70.35 ± 22.24 ^{b}	152.43 ± 106.79 ^{b}	0.23 ± 0.28 ^{b}	0.19 ± 0.15 ^{b}	275.33 ± 72.81 ^{ab}	0.05 ± 0.01 ^{a}	2.76 ± 1.68 ^{b}	1.47 ± 0.68 ^{bc}	0.15 ± 0.09 ^{a}	4.42 ± 2.22 ^{b}	0.07 ± 0.04 ^{a}
	45–100	AF(6)	0.09 ± 0.05 ^{a}	1.09 ± 0.87 ^{ab}	6.28 ± 0.74 ^{a}	1215.00 ± 70.92 ^{b}	2.60 ± 0.93 ^{ab}	103.15 ± 23.37 ^{ab}	331.83 ± 167.95 ^{a}	0.61 ± 0.53 ^{ab}	0.39 ± 0.23 ^{a}	170.63 ± 50.87 ^{b}	0.03 ± 0.01 ^{a}	5.72 ± 3.38 ^{a}	2.95 ± 1.03 ^{ab}	0.98 ± 2.05 ^{a}	9.68 ± 4.23 ^{a}	0.04 ± 0.02 ^{a}
		CF(6)	0.11 ± 0.03 ^{a}	1.18 ± 0.36 ^{ab}	5.84 ± 1.18 ^{a}	1330.83 ± 294.86 ^{b}	4.58 ± 2.86 ^{ab}	130.12 ± 69.57 ^{a}	99.99 ± 4.02 ^{bc}	1.16 ± 0.94 ^{a}	0.34 ± 0.17 ^{a}	172.52 ± 101.91 ^{b}	2,12 ± 4.71 ^{a}	5.16 ± 2.71 ^{a}	4.28 ± 3.11 ^{a}	0.32 ± 0.23 ^{a}	11.88 ± 9.48 ^{a}	0.06 ± 0.06 ^{a}
		CLSWC(6)	0.08 ± 0.01 ^{a}	0.75 ± 0.26 ^{ab}	6.11 ± 0.50 ^{a}	1118.33 ± 59.81 ^{b}	2.14 ± 0.66 ^{b}	118.45 ± 29.32 ^{ab}	392.00 ± 143.57 ^{a}	0.11 ± 0.28 ^{b}	0.49 ± 0.35 ^{a}	176.67 ± 48.27 ^{b}	0.02 ± 0.01 ^{a}	6.01 ± 2.13 ^{a}	2.83 ± 0.56 ^{ab}	0.23 ± 0.10 ^{a}	9.09 ± 2.16 ^{ab}	0.04 ± 0.01 ^{a}
		CLNSWC(6)	0.09 ± 0.01 ^{a}	0.96 ± 0.09 ^{ab}	5.88 ± 0.68 ^{a}	1278.33 ± 234.15 ^{b}	2.13 ± 0.76 ^{b}	100.60 ± 21.48 ^{ab}	275.00 ± 78.22 ^{ab}	0.00 ± 0.00 ^{b}	0.28 ± 0.04 ^{a}	253.67 ± 69.94 ^{b}	0.03 ± 0.01 ^{a}	5.43 ± 2.87 ^{a}	2.10 ± 1.11 ^{abc}	0.13 ± 0.08 ^{a}	7.69 ± 3.93 ^{b}	0.03 ± 0.01 ^{a}
		GL(6)	0.17 ± 0.17 ^{a}	2.15 ± 1.68 ^{ab}	5.06 ± 0.26 ^{a}	1928.33 ± 335.76 ^{a}	1.50 ± 0.76 ^{b}	76.25 ± 33.43 ^{ab}	18.06 ± 40.90 ^{c}	0.63 ± 0.47 ^{ab}	0.18 ± 0.20 ^{a}	391.33 ± 136.02 ^{b}	0.04 ± 0.02 ^{a}	0.61 ± 0.30 ^{a}	0.26 ± 0.23 ^{c}	0.09 ± 0.06 ^{a}	1.01 ± 0.45 ^{ab}	0.01 ± 0.00 ^{a}
		C(6)	0.07 ± 0.02 ^{a}	0.70 ± 0.22 ^{b}	5.11 ± 0.66 ^{a}	1455.00 ± 169.56 ^{b}	1.14 ± 0.63 ^{b}	57.25 ± 8.67 ^{b}	85.67 ± 140.90 ^{c}	0.15 ± 0.23 ^{b}	0.24 ± 0.25 ^{a}	290.33 ± 66.70 ^{ab}	0.06 ± 0.03 ^{a}	2.58 ± 1.83 ^{ab}	1.01 ± 0.60 ^{bc}	0.19 ± 0.23 ^{a}	3.84 ± 2.50 ^{ab}	0.06 ± 0.05 ^{a}

AF = Agroforestry; CF = Community forest; CLSWC = Cropland with soil and water conservation; CLNSWC = Cropland without soil and water conservation; GL = Grassland; C = Control. Means down the groups within a country followed by different letters are significantly different ($p < 0.05$) with respect to land use and soil depths; n = number of samples per land-use category.

4. Discussion

Total ensemble modeling was used as an alternative method for predicting soil quality indicators of different climate-smart land-use practices using soil NIR spectra. The ensemble modeling approaches have received little attention in contexts where calibration and prediction are common in chemistry [20]. The adoption of this method showed improved accuracy as compared to random forest optimization, gradient boosted machines, partial least squares, Bayesian additive regression trees and Cubist (Figures 3 and 4). A good prediction was obtained for total carbon (R^2 = 0.87), TN (R^2 = 0.82), exchangeable bases (sum of Mehlich exch Ca, Mg, K, Na), Mehlich 3 extractable Cu, Mehlich 3 extractable Fe, Mehlich 3 extractable B, Mehlich 3 extractable Mn, exchangeable Na and exchangeable Ca. Exchangeable Ca and Mg have also been satisfactorily predicted (R^2 > 0.70) by other authors [37,40,41]. Nevertheless, Chang et al. [38] found that some soil chemical or physical variables that do not have their primary response in the NIR region are poorly predicted, in comparison with biological indicators. However, the predictions of such properties can be achieved by correlations with biological properties [49]. Bian et al. [50] found that the linearity relationship of the total ensemble method was better than other modeling methods such as PLS.

Total ensemble modeling first generates diverse training subsets from the sample direction to build multiple sub-models [51], while gradient boosted machines (GBM) are hybrid methods that combines boosting and bagging techniques [28]. Cubist and random forests models have also been used to generate a more comprehensive predictive accuracy on NIR spectra [30,52]. The failure in calibration for Mehlich 3 extractable P and Mehlich 3 extractable Zn could probably be as a result of very narrow ranges, with extreme minimum values of 0.00 mg kg^{-1} for both Zn and P (Table 1). However, Minasny et al. [53] attributed the poor prediction of these properties to the fact that they are not related to the soil matrix or solid constituents. Additionally, electrical conductivity (R^2 = 0.40) and pH (R^2 = 0.65) do not exhibit a primary response in the NIR region and their prediction depends on their relationships with organic matter and clay contents [54]. Unsatisfactory predictions of Mehlich 3 extractable P and electrical conductivity (R^2 < 0.66 and RPD < 2) were also observed by other authors [42,54]. In general, electrolyte concentration-related parameters such as electrical conductivity (ECd) are not accurately predicted [55].

The concepts of soil health and soil quality are based on the ability of a given soil to sustain production and measure both positive and negative changes in soil health [46]. Soil characteristics and overall soil health are influenced by land-use systems and intrinsic features, with soil nutrient availability and soil metal concentrations accounting for the majority of the variation seen. Additionally, land-use patterns have an impact on the soil nutrient distribution and availability by affecting soil characteristics [56]. However, the different soil pH ranges for different climate-smart land-use types observed in this study suggest that soil pH is unlikely to have a negative impact on plant development and agricultural productivity. Some of the physical-chemical qualities of soil are static across time, while others are dynamic across a range of time scales. Some are resistant to change brought about by land-use types, whereas others are quickly influenced in both positive and negative ways by land use [57]. In most of the quality indicators assessed at different soil depths, agroforestry, community forest, cropland with soil and water conservation, cropland without soil and water conservation and grassland land uses resulted in better soil quality as compared to controlled land use. As a result, the findings of this study showed the potential of different climate-smart land-use types to improve or preserve soil quality. Soil quality changes are typically the result of increased erosion processes caused by plowing, burning, overgrazing, and other land-use practices that remove the protective vegetative cover [58]. Using a threshold value of 0.2% TN [25] to assess the agricultural productivity of soils at different depths for different land uses, the majority of topsoils (0–15 cm), mid depth (15–45 cm), and at a depth of 45–100 cm had >0.2% N. The results agree with other authors who concluded that climate-smart land use can reduce the degradation of soil quality. Adopting climate-smart land-use practices is a prerequisite to ensure soil health and crop production for sustainable food production systems. These practices, such as agroforestry, community forestry, cropland with soil and water conservation, and grasslands, can increase soil microbial activities and the cation exchange capacity [59]. The soil health indicators used for the assessment of any soil are based on the intended use of the soil. Hence, the set of soil parameters used to assess soil for sustainable rangeland may differ from those used to assess the soil for sustainable crop production, grasslands, and community forests. As a result, there are no universal soil health indicators that will work in all systems. Some soil health indicators, such as water-holding capacity and soil texture, do not change rapidly over time, while other soil health indicators fluctuate more throughout a field and over time according to soil management, crop variety, precipitation, and temperature.

5. Conclusions

This study highlighted the robustness and the predictive power of using total ensemble modeling on NIR spectra for predicting the soil chemical and physical properties as soil health indicators. The results are better when using the total ensemble algorithm as opposed to the random forest optimization, gradient boosted machines, partial least squares and Bayesian additive regression trees algorithms. Further, the developed models were able to predict the physical and chemical properties of soil collected from soils used for different land-use practices at different depths. The soils of the research locations are suitable for sustainable food production systems according to the current assessment of soil health indicators; however, land-use systems and sampling depth, as well as intrinsic traits, influence the overall soil health, with soil nutrient availability accounting for the majority of the variation observed. Thus, the findings of this study highlighted the land-use categories' potential for improving or sustaining soil quality. As a result, there is reason for adopting climate-smart agriculture (CSA) practices as a viable alternative.

Author Contributions: Conceptualization, J.W.R.; K.O.O.; A.S.; M.R.; and D.S.; methodology, J.W.R.; K.O.O.; and A.S.; software, J.W.R.; K.O.O.; and A.S.; validation, J.W.R.; K.O.O.; and A.S.; formal analysis, J.W.R.; K.O.O.; and A.S.; investigation, J.W.R.; K.O.O.; A.S.; and G.A.; resources, J.W.R.; K.O.O.; A.S.; M.R.; and D.S.; data curation, J.W.R.; K.O.O.; A.S.; and G.A.; writing—original draft preparation, J.W.R.; K.O.O.; and A.S.; writing—review and editing, J.W.R.; K.O.O.; A.S.; G.A.; M.R.; and D.S.; visualization, J.W.R.; K.O.O.; A.S.; G.A.; M.R.; and D.S.; supervision, J.W.R.; M.R; and D.S.; project administration, J.W.R.; M.R.; and D.S.; funding acquisition, J.W.R.; M.R.; and D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported through CCAFS-EC grant reference: 2000002575 for the project on Building Livelihoods and Resilience to Climate Change in East and West Africa: Agricultural Research for Development (AR4D) for large-scale implementation of Climate-Smart Agriculture. The funds are administered by the International Fund for Agricultural Development (IFAD), Rome, Italy, while the project was implemented by Alliance Bioversity-CIAT. The Article Processing Charges were paid through a grant from the World Bank "Accelerating Impacts of CGIAR Climate Research for Africa (AICCRA)" ESA regional project (Grant No. D7540) signed between CIAT and the International Development Association (IDA).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to confidentiality requirements of the funding organisations for the project.

Acknowledgments: The Kenya Agricultural and Livestock Research Organization (KALRO) and Vi Agroforestry are acknowledged for providing support during fieldwork in Nyando CSVs. Maseno University is acknowledged for the bulk density determination of Nyando CSVs soil samples. The Tanzania Agricultural Research Institute (TARI) is acknowledged for providing support during fieldwork as well as for bulk density determination of the Lushoto CSVs soil samples. The Uganda

National Agricultural Research Organisation (NARO) is acknowledged for providing support during fieldwork as well as for bulk density determination of the Hoima CSVs soil samples.

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

References

1. Segnini, A.; Posadas, A.; da Silva, W.T.L.; Milori, D.M.B.P.; Gavilan, C.; Claessens, L.; Quiroz, R. Quantifying soil carbon stocks and humification through spectroscopic methods: A scoping assessment in EMBU-Kenya. *J. Environ. Manag.* **2019**, *234*, 476–483. [CrossRef] [PubMed]
2. Comino, F.; Aranda, V.; García-Ruiz, R.; Ayora-Cañada, M.J.; Domínguez-Vidal, A. Infrared spectroscopy as a tool for the assessment of soil biological quality in agricultural soils under contrasting management practices. *Ecol. Indic.* **2018**, *87*, 117–126. [CrossRef]
3. Chodak, M.; Khanna, P.; Horvath, B.; Beese, F. Near infrared spectroscopy for determination of total and exchangeable cations in geologically heterogeneous forest soils. *J. Near Infrared Spectrosc.* **2004**, *12*, 315–324. [CrossRef]
4. Nosrati, K. Assessing soil quality indicator under different land use and soil erosion using multivariate statistical techniques. *Environ. Monit. Assess.* **2013**, *185*, 2895–2907. [CrossRef] [PubMed]
5. Butkute, B.; Šlepetiene, A. Application of near infrared reflectance spectroscopy for the assessment of soil quality in a long-term pasture. *Commun. Soil Sci. Plant Anal.* **2006**, *37*, 2389–2409. [CrossRef]
6. Cécillon, L.; Barthès, B.G.; Gomez, C.; Ertlen, D.; Genot, V.; Hedde, M.; Stevens, A.; Brun, J.J. Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *Eur. J. Soil Sci.* **2009**, *60*, 770–784. [CrossRef]
7. Ahmadi, A.; Emami, M.; Daccache, A.; He, L. Soil Properties Prediction for Precision Agriculture Using Visible and Near-Infrared Spectroscopy: A Systematic Review and Meta-Analysis. *Agronomy* **2021**, *11*, 433. [CrossRef]
8. Soriano-Disla, J.M.; Janik, L.J.; Rossel, R.A.V.; 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]
9. Pinheiro, É.F.M.; Ceddia, M.B.; 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]
10. Ozaki, Y.; Huck, C.W.; Beć, K.B. Near-IR Spectroscopy and Its Applications. *Mol. Laser Spectrosc. Adv. Appl.* **2018**, 11–38. [CrossRef]
11. De Mastro, F.; Cocozza, C.; Brunetti, G.; Traversa, A. Chemical and spectroscopic investigation of different soil fractions as affected by soil management. *Appl. Sci.* **2020**, *10*, 2571. [CrossRef]
12. Nduwamungu, C.; Ziadi, N.; Tremblay, G.F.; Parent, L.-É. Near-Infrared Reflectance Spectroscopy Prediction of Soil Properties: Effects of Sample Cups and Preparation. *Soil Sci. Soc. Am. J.* **2009**, *73*, 1896–1903. [CrossRef]
13. Askari, M.S.; O'Rourke, S.M.; Holden, N.M. Evaluation of soil quality for agricultural production using visible-near-infrared spectroscopy. *Geoderma* **2015**, *243–244*, 80–91. [CrossRef]
14. Kuang, B.; Mouazen, A.M. Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture. *Biosyst. Eng.* **2013**, *114*, 249–258. [CrossRef]
15. Chen, H.; Tan, C.; Lin, Z.; Wu, T. Classification of different liquid milk by near-infrared spectroscopy and ensemble modeling. *Spectrochim. Acta-Part A Mol. Biomol. Spectrosc.* **2021**, *251*, 119460. [CrossRef]
16. Dietterich, T.G. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Mach. Learn.* **1999**, *22*, 1–22.
17. Elder, J. The Apparent Paradox of Complexity in Ensemble Modeling. In *Handbook of Statistical Analysis and Data Mining Applications*; Academic Press: Cambridge, MA, USA, 2018; pp. 705–718, ISBN 978-0-12-416632-5.
18. Freund, Y.; Schapire, R.E. A short introduction to Boosting. *J. Jpn. Soc. Artif. Intell.* **1999**, *14*, 771–780.
19. Genuer, R.; Poggi, J.M.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [CrossRef]
20. Mevik, B.H.; Segtnan, V.H.; Næs, T. Ensemble methods and partial least squares regression. *J. Chemom.* **2004**, *18*, 498–507. [CrossRef]
21. Shao, X.; Bian, X.; Cai, W. An improved boosting partial least squares method for near-infrared spectroscopic quantitative analysis. *Anal. Chim. Acta* **2010**, *666*, 32–37. [CrossRef]
22. Reda, R.; Saffaj, T.; Ilham, B.; Saidi, O.; Issam, K.; Brahim, L.; El Hadrami, E.M. A comparative study between a new method and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy. *Chemom. Intell. Lab. Syst.* **2019**, *195*. [CrossRef]
23. Shao, X.; Bian, X.; Liu, J.; Zhang, M.; Cai, W. Multivariate calibration methods in near infrared spectroscopic analysis. *Anal. Methods* **2010**, *2*, 1662–1666. [CrossRef]
24. Chen, H.; Tan, C.; Lin, Z. Ensemble of extreme learning machines for multivariate calibration of near-infrared spectroscopy. *Spectrochim. Acta-Part A Mol. Biomol. Spectrosc.* **2020**, *229*, 117982. [CrossRef] [PubMed]

25. Winowiecki, L.; Vågen, T.G.; Massawe, B.; Jelinski, N.A.; Lyamchai, C.; Sayula, G.; Msoka, E. Landscape-scale variability of soil health indicators: Effects of cultivation on soil organic carbon in the Usambara Mountains of Tanzania. *Nutr. Cycl. Agroecosystems* **2016**, *105*, 263–274. [CrossRef]
26. Recha, J.; Radeny, M.; Kinyangi, J.; Kimeli, P.; Atakos, V.; Lyamchai, C.; Ngatoluwa, R.; Sayula, G. *Climate-Smart Villages and Progress in Achieving Household Food Security in Lushoto, Tanzania*; CCAFS Info Note; CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS): Copenhagen, Denmark, 2015.
27. Kinyangi, J.; Recha, J.; Kimeli, P.; Atakos, V. *Climate-Smart Villages and the Hope of Food Security in Kenya*; CCAFS Info Note; CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS): Copenhagen, Denmark, 2015.
28. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
29. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2012**, *6*, 266–298. [CrossRef]
30. Zhou, J.; Li, E.; Wei, H.; Li, C.; Qiao, Q.; Armaghani, D.J. Random forests and cubist algorithms for predicting shear strengths of rockfill materials. *Appl. Sci.* **2019**, *9*, 1621. [CrossRef]
31. Metzger, K.; Zhang, C.; Ward, M.; Daly, K. Mid-infrared spectroscopy as an alternative to laboratory extraction for the determination of lime requirement in tillage soils. *Geoderma* **2020**, *364*, 114171. [CrossRef]
32. 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]
33. O'Rourke, S.M.; Stockmann, U.; Holden, N.M.; McBratney, A.B.; Minasny, B. An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. *Geoderma* **2016**, *279*, 31–44. [CrossRef]
34. R Development Core Team. R: A Language and Environment for Statistical Computing. 2014, p. 2673. Available online: https://www.r-project.org/ (accessed on 6 October 2021).
35. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Dor, E.B.; 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. [CrossRef]
36. Ludwig, B.; Murugan, R.; Parama, V.R.R.; Vohland, M. Use of different chemometric approaches for an estimation of soil properties at field scale with near infrared spectroscopy. *J. Plant Nutr. Soil Sci.* **2018**, *181*, 704–713. [CrossRef]
37. Pudełko, A.; Chodak, M. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. *Geoderma* **2020**, *368*, 114306. [CrossRef]
38. Chang, C.; Laird, D.; Mausbach, M.J. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [CrossRef]
39. Zhao, D.; Arshad, M.; Li, N.; Triantafilis, J. Predicting soil physical and chemical properties using vis-NIR in Australian cotton areas. *Catena* **2021**, *196*, 104938. [CrossRef]
40. Shepherd, K.D.; Walsh, M.G. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988–998. [CrossRef]
41. 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]
42. Pirie, A.; Singh, B.; Islam, K. Spectroscopic Techniques to Predict Several Soil Properties. *Aust. J. Soil Res.* **2005**, *43*, 713–721. [CrossRef]
43. Udelhoven, T.; Emmerling, C.; Jarmer, T. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study Thomas. *Plant Soil* **2003**, *251*, 319–329. [CrossRef]
44. Zhao, J.; Dong, Y.; Xie, X.; Li, X.; Zhang, X.; Shen, X. Effect of annual variation in soil pH on available soil nutrients in pear orchards. *Acta Ecol. Sin.* **2011**, *31*, 212–216. [CrossRef]
45. Takoutsing, B.; Weber, J.C.; Tchoundjeu, Z.; Shepherd, K. Soil chemical properties dynamics as affected by land use change in the humid forest zone of Cameroon. *Agrofor. Syst.* **2016**, *90*, 1089–1102. [CrossRef]
46. Takoutsing, B.; Weber, J.; Aynekulu, E.; Rodríguez Martín, J.A.; Shepherd, K.; Sila, A.; Tchoundjeu, Z.; Diby, L. Assessment of soil health indicators for sustainable production of maize in smallholder farming systems in the highlands of Cameroon. *Geoderma* **2016**, *276*, 64–73. [CrossRef]
47. Moges, A.; Dagnachew, M.; Yimer, F. Land use effects on soil quality indicators: A case study of Abo-Wonsho Southern Ethiopia. *Appl. Environ. Soil Sci.* **2013**, *2013*, 1–9. [CrossRef]
48. Ambaw, G.; Recha, J.W.; Nigussie, A.; Solomon, D.; Radeny, M. Soil carbon sequestration potential of climate-smart villages in East African countries. *Climate* **2020**, *8*, 124. [CrossRef]
49. Sanderman, J.; Savage, K.; Dangal, S.R.S. Mid-infrared spectroscopy for prediction of soil health indicators in the United States. *Soil Sci. Soc. Am. J.* **2020**, *84*, 251–261. [CrossRef]
50. Bian, X.; Wang, K.; Tan, E.; Diwu, P.; Zhang, F.; Guo, Y. A selective ensemble preprocessing strategy for near-infrared spectral quantitative analysis of complex samples. *Chemom. Intell. Lab. Syst.* **2020**, *197*, 103916. [CrossRef]
51. Wang, K.; Bian, X.; Tan, X.; Wang, H.; Li, Y. A new ensemble modeling method for multivariate calibration of near infrared spectra. *Anal. Methods* **2021**, *13*, 1374–1380. [CrossRef]

52. Henaka Arachchi, M.P.N.K.; Field, D.J.; McBratney, A.B. Quantification of soil carbon from bulk soil samples to predict the aggregate-carbon fractions within using near- and mid-infrared spectroscopic techniques. *Geoderma* **2016**, *267*, 207–214. [CrossRef]
53. Minasny, B.; Tranter, G.; McBratney, A.B.; Brough, D.M.; Murphy, B.W. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* **2009**, *153*, 155–162. [CrossRef]
54. Zornoza, R.; Guerrero, C.; Mataix-Solera, J.; Scow, K.M.; Arcenegui, V.; Mataix-Beneyto, J. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol. Biochem.* **2008**, *40*, 1923–1930. [CrossRef]
55. Viscarrarossel, R.; Walvoort, D.; Mcbratney, A.; Janik, L.; Skjemstad, J. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]
56. Mulat, Y.; Kibret, K.; Bedadi, B.; Mohammed, M. Soil quality evaluation under different land use types in Kersa sub-watershed, eastern Ethiopia. *Environ. Syst. Res.* **2021**, *10*. [CrossRef]
57. Xia, Y.; Ugarte, C.M.; Guan, K.; Pentrak, M.; Wander, M.M.; Comino, F.; Aranda, V.; García-Ruiz, R.; Ayora-Cañada, M.J.; Domínguez-Vidal, A.; et al. Assessment of soil quality of arable soils in Hungary using DRIFT spectroscopy and chemometrics. *Soil Sci. Soc. Am. J.* **2018**, *11*, 80–91.
58. Arias, M.E.; González-Pérez, J.A.; González-Vila, F.J.; Ball, A.S. Soil health—A new challenge for microbiologists and chemists. *Int. Microbiol.* **2005**, *8*, 13–21. [CrossRef] [PubMed]
59. Tahat, M.M.; Alananbeh, K.M.; Othman, Y.A.; Leskovar, D.I. Soil health and sustainable agriculture. *Sustainability* **2020**, *12*, 4859. [CrossRef]