Review

From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review

Theodora Angelopoulou [1,2,*] **, Athanasios Balafoutis** [1] **, George Zalidis** [2,3] **and Dionysis Bochtis** [1]

1 Centre of Research and Technology-Hellas (CERTH), Institute for Bio-Economy and Agri-Technology (iBO), Thessaloniki, 57001 Thermi, Greece; a.balafoutis@certh.gr (A.B.); d.bochtis@certh.gr (D.B.)

2 Laboratory of Remote Sensing, Spectroscopy, and GIS, Department of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; zalidis@agro.auth.gr

3 Interbalkan Environment Center (i-BEC), 18 Loutron Str., 57200 Lagadas, Greece

* Correspondence: d.angelopoulou@certh.gr

check for updates

Abstract: Rapid and cost-effective soil properties estimations are considered imperative for the monitoring and recording of agricultural soil condition for the implementation of site-specific management practices. Conventional laboratory measurements are costly and time-consuming, and, therefore, cannot be considered appropriate for large datasets. This article reviews laboratory and proximal sensing spectroscopy in the visible and near infrared (VNIR)–short wave infrared (SWIR) wavelength region for soil organic carbon and soil organic matter estimation as an alternative to analytical chemistry measurements. The aim of this work is to report the progress made in the last decade on data preprocessing, calibration approaches, and system configurations used for VNIR-SWIR spectroscopy of soil organic carbon and soil organic matter estimation. We present and compare the results of over fifty selective studies and discuss the factors that affect the accuracy of spectroscopic measurements for both laboratory and in situ applications.

Keywords: reflectance spectroscopy; soil spectral libraries; VNIR-SWIR; soil organic matter; carbon sequestration

1. Introduction

Food production requires fertile soils that can be deteriorated by intensive agricultural practices [1]. In order to have sustainable agricultural systems, economic viability for both farmers and society should be considered, but environmental costs also need to be taken into account, avoiding the implementation of practices that could lead to irreversible soil degradation [2]. Sustainable land use management can be achieved by maintaining and enhancing ecosystem services that include environmental, physical, and socioeconomic aspects [3,4]. The consequences of soil degradation have been recognized by the scientific community, but it was not until 1987 that the United Nations Environment Programme (UNEP) made an agreement with the International Soil Reference and Information Centre (ISRIC) for the execution of the Global Assessment of Soil Degradation (GLASOD) project that produced a world map of human-induced soil degradation [5]. Although a potentially nonrenewable natural resource, the EU Thematic Strategy for Soil Protection was issued in 2006, identifying eight main threats to soils i.e., organic matter decline, erosion, compaction, decline of biodiversity, salinization, contamination, sealing, and landslides [6]. In addition to that, the Sustainable Development Goals, adopted by the United Nations on September 2015, identified the importance of preserving soil resources from degradation in order to achieve such goals [7,8]. A key indicator of soil fertility, and thus soil quality, is soil organic matter (SOM) content [9]. Soil organic carbon (SOC) as a component

of SOM affects soil fertility, but also has an impact on climate change, as soil represents one of the largest terrestrial carbon pools [10]. SOC losses could be induced from global warming that stimulates the rate of SOM decomposition, land use changes, and plant cover reduction [11,12]. To that effect, estimating SOC stocks is considered crucial for greenhouse gas (GHG) emission reports, which is not a very feasible task on large scales. Precision agriculture technologies could positively contribute to GHG emissions mitigation [13]. The various methods for SOC content determination can be categorized as analytical methods (i.e., dry and wet combustion), remote sensing based methods (i.e., space-borne and airborne), spectroscopy methods (i.e., visible near-infrared and short wave infrared, VNIR-SWIR, and mid-infrared, MIR spectroscopy), and laser-induced breakdown spectroscopy and inelastic neutron scattering methods [14]. Yet, there are still challenges that need to be addressed concerning data harmonization between different sources due to diversions in spatial and temporal data resolution, as well as a lack of the necessary information about soil sampling design and, in numerous cases, an absence of important auxiliary variable measurements for SOC stock estimation (i.e., bulk density) [15].

All analytical SOC measuring protocols have been thoroughly described in the literature [16], while remote sensing techniques for SOC estimation have been recently reviewed in detail [17]. However, there is a need to review the research work in laboratory and proximal sensing techniques in the VNIR-SWIR region, due to their significance on the development of nondestructive SOC estimation.

Therefore, this work aims to introduce the basic principles and approaches used for the analysis of spectroscopic measurements for SOC and/or SOM estimation, and not the humus fraction. Specifically, we present advances in soil reflectance spectroscopy from laboratory to proximal sensing applications, including a broad range of different approaches to inform the reader about progress made in the past decade. To that end, we cite selected papers that, to our knowledge, cover most of the methodologies used for modelling the acquired spectroscopic data, regarding multivariate statistical methods and preprocessing techniques, as well as the efforts made to address the effects of external factors on the accuracy of SOC/SOM predictions.

2. Laboratory VNIR-SWIR Spectroscopy

Over recent decades, the use of soil reflectance spectroscopy in laboratory conditions has gained much attention. Due to the principles of energy–matter interactions, a material can reflect, absorb, scatter, and emit electromagnetic radiation in a characteristic manner that depends on its molecular composition and shape, resulting in a unique spectral signature [18]. A sensor could measure the reflectance of an object at a wide area of wavelengths, providing information about its constituents.

Soil spectroscopy in the VNIR-SWIR region (400–2500 nm) has been evaluated as a possible alternative method for monitoring soil parameters, to address the need for continuous information about soil's condition, while reducing the cost of soil analysis [19–21]. The nondestructive nature of this technique enables simultaneous and repeatable measurements to be made, giving it a significant advantage over conventional laboratory measurements [22]. Exploiting the direct spectral responses of SOM—mainly due to overtones, as well as the bending and stretching of NH, OH, and CH groups [23]—has become one of the most studied and accurately measured soil properties regarding soil reflectance spectroscopy applications [23,24]. In addition to this, the presence of SOM affects the color of the soil, and hence, could be directly related to the visible region of the spectrum [25]. Research was initially focused on evaluating this technique in controlled laboratory conditions, which would make it less time consuming and expensive compared to conventional soil analysis [26]. In addition to that, the use of chemical reagents is not required, which is safe for the environment and provides safer working conditions [27]. Nevertheless, soil samples that are measured in laboratory conditions need the conventional preparation of drying and sieving prior to spectroscopic measurements [28].

One of the first studies that identified the potential to measure SOM content was made by Krishnan et al. [29], who observed that the slope of the spectral curve at around 800 nm increased with increasing SOM content. Dalal and Henry, on the other hand, observed that variations in soil moisture

content and soil texture were factors that affected the entire spectral signature of soil samples with the same OC content [30].

2.1. Preprocessing Techniques

Soil matrix consists of numerous parameters that are characterized by low concentrations and have overlapping absorptions that interfere with the spectroscopic measurements. For that reason, soil spectra in the VNIR-SWIR region are considered nonspecific, broad, and with low signal. To that end, spectral analysis requires the use of multivariate approaches to extract hidden information [31]. Commonly, prior to calibration, raw soil spectral data are subjected to preprocessing techniques to remove or minimize noise, enhance their signal, and select characteristic spectral features to improve the subsequent calibration models [32]. Preprocessing techniques are mathematical procedures that transform reflectance measurements and are able to remove variability from light scattering effects and enhance spectral features. These techniques can be categorized in scatter correction methods and spectral derivatives [33]; however, selecting the proper preprocessing technique is not a common procedure. Dotto et al. applied seven of the most common preprocessing techniques (i.e., Savitzky-Golay first derivative, normalization by range, standard normal variates (SNV), multiplicative scatter-correction, continuum removed reflectance (CRR), and the transformation to absorbance and application of a Savitzky-Golay first derivative with a first-order polynomial and a window size of 5 nm) and compared their performance within a wide range of multivariate calibration algorithms [34]. CRR was indicated as the most proper preprocessing technique for SOC prediction, irrespective of the multivariate method applied.

Another aspect of consideration is outlier detection and removal [35]. Though most studies remove extreme values from the modelling procedure, there is a notion that they should not be excluded from the dataset for a robust model to be able to detect such values [36]. An example is the work of De Santana et al. that aimed at providing adequate information about the methodology used for the evaluation of outliers [37], and proposed a new method for their detection using the random forest (RF) technique. It was found that RF generated a lower number of outliers from partial least square regression (PLSR).

2.2. Multivariate Calibrations

Spectral data analysis is performed with the use of multivariate statistical methods, the success of which is highly dependent on the selected calibration method [38]. PLSR is the most commonly-used linear method to describe the relationship between spectral data and soil properties due to its interpretability and low computation time [39,40]. Thus, it has been observed that the aforementioned relationship is not always linear, and that therefore PLSR could be considered insufficient for modelling soil properties [41]. Several studies have attempted to address this problem using machine learning techniques, in the hope that this would produce more accurate results than linear regression [27,28,42,43]. As Davis et al. [16] stated, the diverse and various methodologies for SOC estimation limit the researchers' ability of the to compare the results of their studies in order to generate conclusive results. Therefore, laboratory-based studies focus on finding the most accurate and effective approach for such a calibration. Hence, differences in these studies are mainly in the protocol used and the selection of the preprocessing technique and multivariate statistical method. The accuracy of a model can be assessed using a variety of different methods and metrics. The most common are (i) the coefficient of determination (R^2), which measures the percentage of variance of the dependent variable, as explained by the independent variable, (ii) the root mean square error (RMSE), (iii) the residual prediction deviation (RPD), which is the ratio of the standard deviation of the measured reference values to RMSE, and more recently, (iv) the ratio of performance to interquartile range (RPIQ).

For robust VNIR-SWIR soil spectral measurements, it is important to construct a database that is representative of soil variability for each studied area [44]. For the generation of the required data variability, Sithole et al. [45] acquired samples from various depths and different tillage systems that

produced a robust PLSR predictive model for SOC with $R^2 = 0.993$, RMSEP = 0.157, and RPD = 2.55, i.e., with very high accuracy in terms of its ability to predict SOC. The modified PLSR was utilized by Heinze et al. [46], as it was considered more accurate than PLSR. Soil samples were acquired from three long-term experimental sites that resulted in high diversity in SOC content (3.4–18 mg g $^{-1}$) and soil texture range. The results for SOC predictions were found to be very accurate, with R^2 above 0.96 for the two of the three sites, while poorer predictions were found for the site with a high sand content and smaller SOC range. Similarly, results were found by Stenberg et al. claiming that a high sand content reduces the accuracy of SOC estimation [39]. In contrast to previous studies, Bikindou et al. reported the ability of soil reflectance spectroscopy to accurately determine SOC in sandy soils with low SOM concentrations [47].

The accuracy and variability of results, depending on the multivariate statistical technique used for modelling SOC predictions, is evident in large number of studies that have evaluated and compared different algorithms. The available spectroscopic data have increased, and along with the use of machine learning techniques, created new opportunities for data analysis in the agricultural sector [48]. Spectroscopic models based on the M5 algorithm [49] were evaluated by Rossel et al. [50] using the Australian soil library. The library comprises over 10,000 samples with SOC measurements, providing an adequate sample size, which is an important parameter for the application of machine learning techniques. SOC models performed well, with an RPD of 2.17, and provided information about the important wavelengths used to partition the data and to predict the SOC content. Comparing PLSR, support vector machines regression (SVMR) and multivariate adaptive regression splines (MARS), in salt affected soils, Nawar et al. [32] reported different prediction accuracies that were dependent both on the multivariate technique and the preprocessing method. Overall, MARS, in combination with continuum removal for spectral preprocessing, showed better results ($R^2 = 0.81$), while PLSR performed better than SVMR. The study also showed a high level of correlation of SOM and raw spectra between the 550–680 nm regions, probably related to soil color. Due to the complex relationship between spectral data and soil parameters that is not always considered linear, de Santana et al. [37] compared RF to PLSR; the results were marginally better for RF, as they were able to identify outliers using a proximity matrix. In the same line, SVMR was found to perform better from linear multivariate methods (principal component analysis-PCA, PLSR) and back-propagation neural networks [51]. The selection of the most appropriate data pretreatment and calibration can be a laborious procedure due to the various combinations that can be applied. Gholizadeh et al. [52] compared some of the most common calibration techniques i.e., PLSR, RF, BRT, SVMR, and MBL with a new data-mining engine PARACUDA-II® [53]. PARACUDA-II® is based on the all possibilities approach (APA) and a conditional Latin hypercube sampling (cLHs) algorithm. It also has the capability of parallel programming, to assess all the possible combinations of eight different spectral preprocessing techniques against the original reflectance and the chemical data prior to model development. The performance was better compared to all other five multivariate techniques, with a $R^2 = 0.80$. The development of large soil spectral libraries has also given rise to the opportunity to explore the potential of deep learning in the analysis of such data. Padarian et al. [54] presented one of the first studies to evaluate convolutional neural networks (CNN), while introducing the concept of spectrograms to visualize soil spectra; it was found that CNN outperformed both the PLSR and Cubist models.

2.2.1. Feature Selection

Variable selection algorithms are applied to reduce the complexity of spectral data, and preselecting the useful features of the spectrum for calibration could increase the predictive ability of a model. Among the various techniques that could be used when selecting an optimum set of features, genetic algorithms have been shown to perform well for SOC prediction models [41,55,56]. Considering the complexity of genetic algorithms and the risk of overfitting when there are more than 200 variables, Peng et al. [42] evaluated a successive projection algorithm (SPA) combined with SVMR and PLSR to select the important wavelengths. This resulted in selecting 28 wavelength regions mainly found in the near

infrared region, which is highly correlated with SOC content. The best predictive model was the combination of SPA-SVMR ($R^2 = 0.73$) compared to SPA-PLSR ($R^2 = 0.62$). To make a step further, Chen et al. [57] proposed a combination of RF and a back propagation network to refine the selective informative wavelengths. It has been also suggested that selecting the inputs of a least squares SVMR model based on the regression coefficients obtained by PLS analysis could provide better prediction models [58]. Recently, Raj et al. [59] utilized a series of variable indicators (Pearson's correlation coefficient (r), biweight midcorrelation (bicor), mutual information based adjacency (AMI), variable importance in the projection (VIP)) combined with an ordered predictor selection (OPS) method to choose the optimum number of spectral variables. The best results were obtained with the use of the AMI indicator combined with SVMR ($R^2 = 0.68$) compared to the use of the full spectrum PLSR model. Bayer et al. [60] compared a physically-based model, where spectral feature analysis is coupled with multiple linear regression techniques (MLR), with a PLSR model. The results showed that model calibrations using significant spectral wavelengths could provide similar results with PLSR for SOC predictions.

2.2.2. Calibration Dataset

Apart from the proper preprocessing and calibration techniques, the stratification of local spectral libraries by homogeneity criteria of the spectral data regarding soil classes and land uses could lead to more accurate SOC predictions, as Moura-Bueno et al. [61] noted in their study. In the same context, Gupta et al. [36] applied a locally-weighted PLSR algorithm and selected an appropriate distance metric for calculating the similarity between a test sample and the calibration samples, and compared it to global calibrations with Lasso and ridge regression. Clingensmith et al. [62] evaluated four different calibration subsetting techniques coupled with PLSR, sparse partial least squares regression (SPLSR), and the heteroscedastic effects model (HEM). It was found that although the SPLSR and HEM algorithms can reduce the number of predictors, they yielded minor improvements in SOC predictions, probably due to the detritus nature of SOM.

Calibration sample size is also a factor that has been proven to affect model calibration; considering the use of soil reflectance spectroscopy as a cost-efficient method, it is important to determine the optimum number of samples for soil modelling. In general, a larger calibration sample size contributes to more robust models. Debaene et al. [63], after evaluating different sampling schemes on a field scale level, concluded that the use of K-means clustering provided better results over random sampling selection, analyte concentration, and PCA scores. However, it was observed that increasing the number of samples after a certain number did not provide proportional results in the decrease of RMSE, and thus, was not significant. In the same scope, Lucà et al. [64] found different levels of model performance variation with a selected calibration size that ranged from 14 to 144 samples and different modelling approaches, i.e., PCR, PLSR, and SVMR. Overall the SVMR model provided the best performance with 72 samples and an average RMSE of prediction of 1.10. It was also reported that creating a robust model depends on the representativeness of the data that can be selected based on the spectral characteristics and/or the physicochemical properties. Hence, the quality of the calibration dataset is affected by more factors than its size, but also by the spatial variability of the studied area.

2.3. Soil Moisture Effects

Since laboratory measurements are mainly conducted on air-dried and sieved soil samples, transferring such models to field conditions entails issues to be addressed, notably soil moisture that is known to affect the soils spectral signature [20,23,65,66]. Therefore, several studies have attempted to create prediction models for different levels of soil moisture [67–69]. The main absorption peaks of water content can be found near 1400, 1900, and 2200 nm due to the overtones and vibrations of the O–H group [70]. To quantify the effect of soil water content on the prediction models, Marakkala Manage et al. [71] used six different matrix potentials that could also provide information about the relationship between soil moisture, SOC, and soil texture. A typical decrease in reflectance was observed with

the increase of water content [72], though the most significant change in the spectral signature was found in the transition from dry to wet soils, where water retention forces change from adsorptive to capillary forces. The results indicated that for in situ measurements of SOC content, it is best for soils to be close to dry. Nocita et al. [73] proposed the normalized soil moisture index (NMSI) [74] to classify artificially-moistened soil samples to estimate SOC content from samples with unknown soil moisture contents. It was observed that there was an overlap in the mean spectra in the regions from 400 to 1000 nm at soil moisture levels above 0.15 gW gS^{-1}, while in the region of 1100–2500 nm, there were minor but detectable differences among samples. It was also found that calibration models developed from moist samples had better accuracy for SOC predictions than models created from dry samples. Hence, in the development and exploitation of national and global spectral libraries of dry soil samples, there is a need to remove the effect of soil moisture when moist samples are to be measured.

Several methods have been proposed to remove the effect of soil moisture from soil spectra [75]. The most common are direct standardization (DS) [76] and external parameter orthogonalization (EPO) [77]. Minasny et al. [78] reported that with increasing soil moisture, the accuracy of the PLSR model for SOC was decreasing, but with the use of the EPO algorithm, the accuracy of the wettest soil sample was comparable to that of a dry soil sample. Specifically, PLSR results ranged from R^2 = 0.52–0.78, while EPO-PLSR ranged from R^2 = 0.75–0.80. They concluded that it is preferable for the calibration sample size to be greater than 100 samples, and suggested further investigation to validate this method in open-field conditions. Similar results were reported in the study of de Santana [79]. Comparing DS and EPO, Roudier et al. [80] concluded that both approaches could mitigate the effect of soil moisture when using soil spectral libraries to develop calibration models, and that both performed similarly. It was also observed that the number of calibration samples affected the performance of DS and EPO, with the latter performing better with a limited number of samples. In addition to that, different soil types affected the prediction models (i.e., soils with uniform textures generated better results compared to soils formed from complex parent materials). Soil roughness is another factor that affects spectroscopic measurements, as it causes light scattering, and thus decreases reflectance. Rodionov et al. [81] aimed to classify the effects of soil moisture and roughness and use these finding for in situ applications. To do so, they created different datasets with disturbed and undisturbed soil samples. The disturbed samples were artificially wetted at six different soil moisture contents and sieved at seven different classes (the size of their aggregations did not exceed 30 mm). The undisturbed samples were air dried for two weeks; during that time, VNIR-SWIR spectra were recorded. They concluded that for sieved samples, it was possible to predict SOC at different soil moisture levels. Regarding soil roughness, an overestimation of SOC content was observed with an increase of aggregate size.

2.4. Soil Spectral Libraries for Local Calibrations

One of the first studies that tried to unlock the potential of soil spectral libraries (SSLs) and develop accurate calibration models was conducted by Shepherd and Walsh [82]. Consequently, several attempts have been made towards the creation of regional, continental, and global SSLs [20,83–87]. Considering the local character of predictions in most studies, Stevens et al. [83] evaluated the potential of using the LUCAS soil database [88] to possibly cover soil heterogeneity. It was observed that the accuracy of the predictions depended on soil classes (i.e., cropland, grassland, woodland mineral, and organic soils) and the use of auxiliary predictors (i.e., sand and clay). Better prediction models were developed for croplands, grasslands, and mineral soils, while sand content most affected the error of predictions when the SOC content was low. In a similar manner, Liu et al. [89] evaluated whether the classification of the Chinese SSL, according to the spectrally- and conventionally-derived soil type, could improve SOC estimations. Overall SOC prediction was improved by soil type stratification. The model's accuracy was slightly lower when soil types were spectrally derived, probably due to the misclassification of some soil types. From another perspective, the use of large-scale SSLs for accurate soil properties predictions on local scales could be challenging due to their diversity and heterogeneity. Lobsey et al. [90] developed a new method, named ReSampling-Local (RS-LOCAL), that selects a

subset from global SSL spiked with a small representative sample from a local site. They concluded that the proposed method could not only improve results, but also reduce the number of calibration samples, thus reducing the cost. In the same line, Gogé et al. [91] aimed to identify the best strategy to utilize a national SSL. They compared the models generated from a national soil database and the same database spiked with subsets from a local library using a fast Fourier transform local weighted method, and reported moderate improvements in accuracy.

2.5. Current Trends

Though VNIR-SWIR soil measurements are considered cost effective compared to analytical chemical analyses, purchasing a spectrometer is still expensive. Currently, there have been efforts to create smaller and cheaper spectrometers. The reduced cost is due to the smaller spectral range these sensors can cover. Therefore, it is imperative to specify the important wavelengths that are correlated with the estimation of a specific soil's properties. Evaluations of the accuracy of such devices is still at a very early stage, with Barthès et al. [92] comparing a JDSU MicroNIR 2200 spectrophotometer (Milpitas, CA, USA) equipped with an InGaAs array detector and a spectral range of 1151–2186 nm with a Foss NIRSystems 5000 (Laurel, MD, USA). The results were promising, though more research needs to be done.

2.6. Summary of VNIR-SWIR Spectroscopy Research Results on a Laboratory Scale

The work analyzed above on VNIR-SWIR spectroscopy under laboratory conditions for SOC/SOM estimations has been summarized in Table 1, in order to provide easy access to the methodology used and the results achieved in each case.

Table 1. Literature review comparing different multivariate methods for SOC/SOM estimations under laboratory conditions.

	Reference	Spectral Range (nm)	Multivariate Method	R^2	RMSE	RPD
1	Vohland et al. (2011) [41]	400–2500	PLSR GA-PLSR SVMR	0.10–0.72 0.13–0.68 0.73–0.81	0.37–0.65% 0.41–0.64% 0.38–0.86%	1.01–1.75 - -
2	Minasny et al. (2011) [78]	350–2500	PLSR EPO-PLSR	0.56–0.83 0.82–0.89	0.50–1.72 (log[g 100 g^{-1}]) 0.26–0.46 (log[g 100 g^{-1}])	- -
3	Rossel et al. (2012) [50]	350–2500	M5 algorithm	-	0.26 (log$_{10}$C)	2.16
4	Bikindou et al. (2012) [47]	1100–2500	PLSR	0.91	0.045%	-
5	Bayer et al. (2012) [60]	350–2500	PLSR MLR	0.69 0.74	0.45% 0.36%	1.53 1.93
6	Nocita et al. (2013) [73]	350–2500	PLSR	0.25–0.63	12.17–30.21 (g kg^{-1})	1.63–1.89
7	Stevens et al. (2013) [83]	400–2500	SVM Cubist	0.67–0.86 0.76–0.89	4.0–15 (g kg^{-1}) 6.4–50.6 (g kg^{-1})	1.74–2.62 1.99–2.88
8	Xuemei (2013) [58]	325–1075	PLSR SVM	0.77–0.80 0.83–0.86	3.32–3.91 (g kg^{-1}) 3.21–3.70 (g kg^{-1})	- -
9	Heinze et al. 2013 [46]	400–2500	MPLS	0.41–0.99	-	1.25–8.02
10	Debaene et al. 2014 [63]	350–2220	PLSR	0.42–0.72	0.12–0.27%	1.0–2.0
11	Rodionov et al. (2014) [81]	350–2500	PLSR	0.84–0.88	0.64–0.75 (g kg^{-1})	2.49–2.92
12	Rienzi et al. (2014) [67]	340–2220	PLSR	0.63–0.88	4.02–7.13 (g kg^{-1})	-

Table 1. *Cont.*

	Reference	Spectral Range (nm)	Multivariate Method	R²	RMSE	RPD
13	Gogé et al. (2014) [91]	400–2500	PLSR, FFT-LW	0.10–0.58	3.39–7.63 (g kg⁻¹)	0.57–1.37
14	Peng et al. (2014) [42]	350–2500	SVMR	0.72	2.83 (g kg⁻¹)	1.86
			SPA-PLSR	0.62	3.23 (g kg⁻¹)	1.63
			SPA-SVMR	0.73	2.78 (g kg⁻¹)	1.89
15	Wijewardane et al. (2016) [68]	350–2500	PLSR	0.40–0.71	0.73–1.01%	-
16	Morellos et al. (2016) [43]	305–2200	PCR	0.72	0.08%	1.89
			PLSR	0.71	0.08%	1.86
			LS-SVM	0.84	0.06%	2.25
			Cubist	0.79	0.07%	2.15
17	Nawar et al. (2016) [32]	350–2500	PLSR	0.50–0.79	0.28–0.42%	1.41–2.16
			SVR	0.51–0.75	0.26–0.37%	1.43–2.0
			MARS	0.66–0.81	0.22–0.33%	1.74–2.27
18	Roudier et al. (2017) [80]	350–2500	PLSR	-	0.93–1.60%	-
19	Lobsey et al. (2017) [90]	350–2500	PLSR,cubist	0.78–0.84	0.48–1.16%	-
20	Luca et al. (2017) [64]	350–2500	PCR	0.69	0.88%	-
			PLSR	0.79	0.71%	-
			SVMR	0.82	0.68%	-
21	Jiang et al. (2017) [93]	350–2500	PLSR	0.79–0.90	0.54–0.88%	-
22	Hong et al. (2018) [69]	350–2500	PLS-SVM	0.70–0.76	-	1.87–2.06
23	Xu et al. (2018) [51]	350–2500	PCR	0.81	6.01 (g g⁻¹)	2.31
			PLSR	0.85	5.48 (g g⁻¹)	2.54
			BPNN	0.86	5.16 (g g⁻¹)	2.69
			SVMR	0.88	4.85 (g g⁻¹)	2.84
24	de Santana et al. (2018) [37]	350–2500	RF	0.8	5.46 g/dm³	-
			PLSR	0.75	6.19 g/dm³	-
25	Raj et al. (2018) [59]	350–2500	SVM	0.68	0.62%	1.79
			PLSR	0.65	0.64%	1.72
26	Gholizadeh et al. (2018) [52]	350–2500	PLSR	0.63	0.29%	-
			RF	0.65	0.23%	-
			BRT	0.68	0.25%	-
			SVMR	0.71	0.20%	-
			MBL	0.78	0.20%	-
			PARACUDA	0.80	0.12%	-
27	Gupta et al. (2018) [36]	350–2500	PLSR	0.60–0.70	0.18–0.20%	-
28	Liu et al. (2018) [89]	350–2500	PLSR	0.51–0.82	1.63–3.18 (g kg⁻¹)	1.44–2.37
29	Sithole et al. (2018) [45]	450–2500	PLSR	0.99	0.16%	2.55
30	Ludwig et al. (2018) [55]	400–2500	GA-PLSR	-	0.04%	2.58
			SVMR	-	0.04%	2.67
			improved GA-PLSR	-	0.04%	2.89
31	Marakkala Manage et al. (2018) [71]	350–2500	PLSR	0.48–0.82	0.001–0.003 (kg kg⁻¹)	-
32	Vibhute et al. (2018) [40]	350–2500	PLSR	0.72–0.89	3.51–5.64 (g kg⁻¹)	-
33	de Santana et al. (2019) [79]	1150–2500	PLSR	0.86	1.85 (g/dm³)	2.59
			EPO-PLSR	0.81–0.85	2.15–2.16 (g/dm³)	2.02

Table 1. *Cont.*

	Reference	Spectral Range (nm)	Multivariate Method	R^2	RMSE	RPD
34	Padarian, et al. (2019) [54]	350–2500	PlS	0.35	130.5 (g kg^{-1})	-
			Cubist	0.79	43.75 (g kg^{-1})	-
			CNN	0.88	32.14 (g kg^{-1})	-
			CNN multi	0.69	16.81 (g kg^{-1})	-
35	Moura-Bueno et al. (2019) [61]	350–2500	PLSR	0.74	0.52%	-
			MLR	0.72	0.57%	-
			SVM	0.72	0.55%	-
			RF	0.72	0.56%	-
36	Clingensmith et al. (2019) [62]	350–2500	PLSR	0.42–0.53	0.32–0.48%	1.30–1.47
			SPLSR	0.45–0.65	0.31–0.42%	1.34–1.69
			HEM	0.44–0.63	0.31–0.43%	1.34–1.64
37	Barthes et al. (2019) [92]	1151–2186	PLSR	0.64–0.82	-	1.4–2.3

3. Proximal Soil Sensing

Regarding the need for continuous monitoring in the agricultural domain and for site specific applications, in situ soil sensors have been developed for rapid, cost-effective, and quasi real-time data acquisition that can provide high-resolution maps [94]. Proximal soil sensing is defined as the use of field-based sensors that are in close proximity with the ground, i.e., within a maximum distance of two meters [94]. To that end, proximal sensing directly in the field became a challenge, and gained interest due to its potential advantages [95]. These technologies concern either on-the-go sensors, mounted on agricultural vehicles, or hand-held instruments that can be used for site-specific management (e.g., variable rate applications) [96]. Due to the high sampling density these sensors provide, they are considered more effective in capturing field variability, hence addressing the problem of selecting the correct soil sampling strategy that will ensure representative soil samples. The accuracy of a sensor can be measured by means of measurement repeatability at the same time and place, and the correlation between reference measurements of soil properties [97]. Although promising, in situ soil reflectance spectroscopy applications require proper environmental conditions and various pretreatment methods to mitigate the effect of moisture content, soil roughness, and vegetation cover [98].

The use of proximal soil sensing techniques could increase the number of measured soil samples that are needed for an adequate characterization of soil heterogeneity on a field scale. To that end, researchers have focused on developing or updating already-developed sensors [96,99]. Achieving successful measurements would be of great benefit for agriculture, as they can assist on delineating management zones and control the inputs applied in the environment using precision agriculture techniques [100]. In this section, we present several prototypes that have been developed in research studies, as well as evaluations of commercially-used VNIR-SWIR sensors, and discuss the limitations and hindrances of their use.

3.1. Commercial Available In situ Soil Sensors

Bricklemyer and Brown [101] tested a commercially-available, on-the-go soil sensor (Veris Technologies Inc., Salina, KS, USA) in eight fields. Two approaches for model calibration were evaluated: (i) using measurements of the seven out of eight fields to independently predict SOC from the remaining field and continuing this procedure until each of the eight fields was predicted, (ii) adding nine randomly-selected samples from the remaining field to the calibration dataset. The accuracy was poor for both approaches with an RPD = 0.1, probably due to low SOC variability. Due to spectral correlation between soil properties, combining different sensors could potentially improve the accuracy of in situ measurements. To that end, Kweon et al. [102] utilized a dual-wavelength (660 and 940 nm), on-the-go soil optical sensor coupled with electrical conductivity measurements.

The relationship between these two wavelengths and SOM at levels below 5% was considered linear, and therefore, MLR was used for model calibration while slope, curvature, and elevation were also used as independent variables. The various combinations of the predictor values provided different results, with R^2 ranging from 0.55 to 0.94. These variations were attributed to the soil moisture content and temperature. In the same line, Knadel et al. [103] compared the predictive ability between a single sensor and a fusion of sensors (VNIR-SWIR, EC and temperature), and found that SOC estimation could be improved by the use of sensor fusion. Wetterind et al. [104] tested the P4000 drill rig-mounted spectrometer (Veris® Technologies, Salina, KS, USA) and also reported that predictions for SOM were marginally improved with a sensor combination. Similar results were reported by Pei et al. [105]. Utilizing the same sensor, Veum et al. [106] aimed to address the effect of soil moisture by applying the EPO transformation in conjunction with the Bayesian Lasso, along with additional covariate information. The accuracy for models trained on dry spectra and tested on field moist spectra was improved. Kuang et al. [107] compared the accuracy between PLSR and artificial neural networks (ANN) for online spectral measurements, and concluded that ANN provided better results in cross validation (R^2 = 0.83) compared to PLSR (R^2 = 0.71). Though using an independent dataset for calibration, the model's performance dropped significantly for both ANN and PLSR calibrations, with R^2 = 0.49 and R^2 = 0.46, respectively. Sorenson et al. [108] performed in situ spectral measurements with the same sensor as Wetterind et al. [104]. For model calibration, they evaluated different multivariate techniques, i.e., MARS, ANN, SVM, PLSR, RF, and Cubist. The lowest RMSE values were found with the use of Cubist and the highest with ANN. They also reported that a model's accuracy increases when it is developed from a homogeneous dataset, though this comes at the expense of model transferability.

3.2. Experimental Prototypes of In situ Soil Sensors

Apart from commercially-available equipment, researchers aimed to develop prototypes for in situ applications. To this end, Kodaira and Shibusawa [109] upgraded a prototype real-time soil sensor (RTSS) (SAS 1000, SHIBUYA MACHINERY Co., Ltd., Japan). The system was mounted on a tractor and comprised a sensor unit, a touch panel, a soil penetrator, and a set of probes. The sensor unit had two spectrophotometers, with spectral range from 350–1100 nm and 950–1700 nm respectively, and a differential global positioning system. The sensor was designed to acquire spectra at adjusted depths every 4 s which, based on the speed of the tractor (0.56 m s^{-1}), resulted in 144 measurements from 31.48 ha; this sampling density was considered adequate for high resolution mapping. For model calibration, the PLSR method with full cross calibration resulted in an accuracy of R^2 = 0.90 for SOM. Field conditions are highly variable and should be taken into consideration during measurements therefore, Kuang and Mouazen [110] utilized an online sensor developed by Mouazen and Ramon [111] that penetrated soil to a depth of 15 cm. Comparing the PLSR models generated for processed soils, fresh soils, and online measurements, it was observed that MC and increased clay content decreased the model's accuracy, while models created from dry samples with high clay content increased the accuracy. To address disturbing factors of in-field spectral measurements (i.e., illumination conditions), Rodionov et al. [112] developed a closed chamber with a commercial VNIR-SWIR spectrometer mounted on a tractor. The system was tested in a small-scale, bare soil field and at different spectral acquisition modes, i.e., stop-and-go and continuous. The modelling approach also accounted for soil moisture and roughness according to Rodionov et al. [81]. Regarding the continuous mode, spectral discontinuities were observed, and thus, the stop-and-go mode was considered more appropriate, providing an R^2 = 0.65. Although the results were significantly lower from laboratory measurements (R^2 = 0.94), they provided the potential for SOC pattern recognition. On-the-go spectral measurements are also affected by sensor movement and heterogeneous soil; hence, Franceschini et al. [113] aimed to mitigate these effects by comparing the results of utilizing the EPO, DS, and orthogonal signal correction (OSC) [114]. The in-situ measurements were performed in relatively dry soil conditions to minimize the effect of soil moisture. It was observed that the albedo of field measurements was generally lower than that of laboratory measurements, probably due to differences in the sensor configuration and illumination

conditions during acquisition. Although an improvement in the model's accuracy was observed with the use of OSC, the results were substantially inferior compared to those of laboratory-generated models, indicating that factors like soil-to-sensor distance and angle, gravels or straws, and changes in the illumination conditions should also be considered.

While proximal sensors have been largely developed and used to overcome the constraints of conventional laboratory chemical analyses, in some cases, the use of a single sensor is inadequate for the simultaneous estimation of several soil properties [115]. To address this, Rossel et al. [116] developed an integrated Soil Condition Analysis System (SCANS). SCANS is a fusion of proximal sensing technologies, analytic statistics, and smart engineering that can analyze a 1.20 m soil core after extracting it from the field. A VNIR-SWIR spectrometer, an active γ-ray attenuation sensor for bulk density estimation, a visible camera, and a Lepton long wave infrared camera mounted on the platform made it possible to estimate approximately thirteen physical, biochemical, and mineralogical soil properties. A total of 150 soil cores were derived and measured from a 600-ha cattle grazing farm. Model calibration was performed with the Cubist algorithm combined with the RS-LOCAL algorithm, providing a more efficient exploitation of the local spectral libraries. The EPO algorithm was also used to eliminate the effect of water in the soil spectra in order to achieve the transferability of the calibrated models with models from spectral libraries of dry soil. The results were promising for SOC estimation with a good level of accuracy ($R^2 = 0.83$) for a repeated 10-fold cross validation and $R^2 = 0.81$ for independent validation. The advantage of SCANS was that it provided accurate, fast, and cost-efficient predictions at high spatial resolution and from various depths.

3.3. Handheld Proximal Sensors

Apart from spectroradiometers mounted on vehicles, portable instruments equipped with a handheld contact probe are also used for direct in situ measurements. Cozzolino et al. [117] used the ASD FieldSpec III spectroradiometer with spectral analysis from 350–2500 nm (Analytical Spectral Devices, ASD, Boulder, CO, USA) and utilized the range 350–1850 nm. The spectral measurements were performed in 68 soil samples whose surface was flattened. The PLSR model calibration using only the NIR (950–1800 nm) wavelength region reported an RPD = 1.8 for SOC that suggests moderate accuracy. Gras et al. [118] evaluated seven different practices for acquiring spectra in the field using an ASDLabSpec 5000 spectrometer (Analytical Spectral Devices, Boulder, CO, USA). These procedures were applied either directly to the soil surface (in cores extracted with an auger) or to clods crumbled from the cores. No external validation was performed, as the aim was to select the most efficient approach for spectral acquisition. Different procedures gave different RPD values, and specifically for SOM estimations, RPD ranged from 2.1 for clods crumbled from cores after half a day, to 2.8 in raw core measurements. The same spectrophotometer was also used by Cambou et al. [119]; the spectral measurements were made on the outer core of samples collected with a manual auger. PLSR and seven different preprocessing techniques were used for model calibration and resulted in $R^2 = 0.75$ for SOC estimations and SOC. The poor results obtained could possibly be related to the fact that spectral measurements were not performed on the same samples that were conventionally analyzed.

3.4. Photosynthetic and Nonphotosynthetic Vegetation Affecting In situ Measurements

Although bare soil conditions are the most favorable for in situ measurements, in real field conditions, the presence of either green vegetation or straw cover is very common, and may lead to overestimations of SOC [120]. Rodionov et al. [121] aimed, in their study using the same spectra, to distinguish SOC from photosynthetic and nonphotosynthetic vegetation. They conducted an experiment under laboratory conditions to estimate the effect of vegetation fractional cover. Soil samples were placed in petri dishes with an increasing degree of plants and straws (24 soil samples with 13 coverage degrees). To characterize straw cover, they used the Cellulose Absorbance Index (CAI), and for green vegetation various, known indices were estimated, such as the Normalized Difference Vegetation Index (NDVI). The aforementioned indices were subtracted from field measurements to

prevent SOC overestimation. Increasing vegetation coverage resulted in higher reflectance in the whole spectral range and distinctive absorption peeks at 600–700 nm. Because the spectral response of straw was not very distinctive, both the use of CAI and NDVI was recommended. Specifically, CAI values ≥ 0.70 and NDVI values ≥ 0.80 were indicative of a SOC overestimation. Although this study confirmed the overestimation of SOC content with the presence of green vegetation and straw coverage, the model gave average predictions with R^2 = 0.58–0.66. The authors suggested the importance of estimating the vegetation coverage degree prior to SOC estimations.

3.5. Spiking Techniques

To address the lack of adequate local data for SOC and to better exploit existing SSls, the spiking technique is frequently evaluated. It is mostly used to augment an existing SSL with local spectra to improve the models accuracy [122–124]. Nawar et al. [125] showed that by using spiking for model calibration, the sample selection method and the number of selected samples also affects the model's accuracy. Specifically, a similarity analysis gave better predictive models with a minimum number of spiked samples compared to Kennard stone and random sampling. Guerrero et al. [126] suggested that an extra-weighted spiking subset could increase the prediction accuracy over spiked calibrations. Collecting field data and creating new SSLs for unprocessed field samples is a laborious procedure. To that end, Kühnel and Bogner [127] proposed the use of the synthetic minority oversampling technique (SMOTE) [128] to address the problem of limited field data. The technique resides in the principal of giving extra weight to existing soil samples to generate new synthetic spectra, hence augmenting existing datasets. This method aims to compensate for the loss of significant spectral data when using an EPO or DS algorithms to remove the effect of MC or for the time-consuming measurement of samples at different MC levels. It was observed that synthetic spectra gave more similar spectral responses with in situ spectra rather than spectra from air-dried, sieved samples. The augmentation of the spectral libraries with the synthetic spectra provided lower RMSE and higher RPD compared to spiking. However, the importance of selecting a representative site samples is significant.

3.6. Summary of VNIR-SWIR Soil Proximal Sensing Research Results in Field Scale

The work analyzed above on VNIR-SWIR soil proximal sensing under laboratory conditions for SOC/SOM estimations has been summarized in Table 2, in order to provide easy access to the methodology used and the results achieved in each case.

Table 2. Literature review comparing different multivariate methods for SOC/SOM estimations under field conditions.

	Reference	Spectral Range (nm)	Multivariate Method	R^2	RMSE	RPD
1	Bricklemyer and Brown (2010) [101]	350–2500	PLSR	0.00–0.42	-	1.0–1.3
2	Cozzolino et al. (2013) [117]	350–1850	PLSR	0.81	-	1.8
3	Kodaira and Shibusawa (2013) [109]	400–1700	PLSR	0.9	0.35%	2.9
4	Kweon et al. (2013) [102]	660 and 940	MLR	0.55–0.94	0.11–0.77%	1.50–4.27
5	Kuang and Mouazen, (2013) [110]	305–2200	PLSR	-	1.29–1.90 (g kg^{-1})	2.01–2.24
6	Gras et al. (2014) [118]	350–2500	MPLS	0.77–0.86	-	2.1–2.8

<div align="center">**Table 2.** *Cont.*</div>

	Reference	Spectral Range (nm)	Multivariate Method	R^2	RMSE	RPD
7	Knadel et al. (2015) [103]	350–2200	PLSR	0.57–0.94	-	1.4–3.9
8	Rodionov et al. (2015) [112]	350–2500	PLSR	0.65	-	-
9	Wetterlind et al. (2015) [104]					
10	Ji et al. (2015) [75]	350–2500	PLSR	0.63–0.70	0.21–0.27 (g kg^{-1})	1.44–1.79
11	Kuang et al. (2015) [107]	305–2200	PLSR ANN	0.37–0.81 0.39–0.90	1.46–3.88% 1.22–3.66%	1.15–2.29 1.22–3.01
12	Rodionov et al. (2016) [121]	350–2500	PLSR	0.84	0.73%	2.53
13	Cambou et al. (2016) [119]	350–2500	PLSR	0.75	-	2
14	Viscarra Rossel et al. (2017) [116]	350–2500	CUBIST	0.81	0.0041%	-
15	Kühnel and Bogner (2017) [127]	350–2500	smote/PLSR	0.40–0.86	1.90–16.63 (mg g^{-1})	-
16	Sorenson et al. (2017) [108]	350–2200	MARS ANN SVMR PLSR RF Cubist	0.76 0.01 0.75 0.54 0.78 0.8	0.66% 1.56% 0.67% 0.90% 0.62% 0.60%	2 0.9 2 1.5 2.1 2.2
17	Veum et al. (2018) [106]	350–2200	PLSR	0.23–0.82	0.19–0.96 g 100 g^{-1}	-
18	Nawar et al. (2018) [125]	305–2200	PLSR	0.74–0.78	0.16–0.18%	1.97–2.14
19	Nawar et al. (2019) [124]	305–2200	RF	0.12–0.75	0.17–0.33%	1.08–2.04
20	Pei et al. (2019) [105]	343–2222	PLSR NN RT RF	0.8 0.86 0.69 0.58	- - - -	- - - -

4. Discussion

Soil properties' monitoring and estimation are essential processes for site-specific management practices, though there is a lack of up-to-date and consistent information on their spatial distribution [129]. Considering the laborious and costly procedure of collecting a sufficient number of samples to develop high-resolution maps, more efficient methods have been evaluated as alternatives to support or, to some extent, replace time-consuming chemical soil analysis. The interaction of soil properties with electromagnetic radiation in the VNIR-SWIR spectral region (400–2500 nm) has provided the grounds by which to assess this method for soil properties estimations. The first attempts began in controlled laboratory conditions, and subsequently, in-field applications appeared. SOC, as one of the most important soil quality parameters, has gained much attention due to its characteristic interactions with electromagnetic radiation.

The procedure for spectroscopic measurement, modelling, and prediction of SOC requires well-defined steps, though the procedures in each step vary among studies. Firstly, soil samples from the studied area are analyzed with conventional laboratory measurements. Even these standardized procedures could differ in each study [32,119], as it was observed that laboratory measurements could also entail errors [130] that affect the predictive ability of the models, since they are dependent upon these specific laboratory reference measurements [131]. Laboratory soil spectroscopy has been evaluated for more than twenty years, and despite the development of soil spectral libraries on a global

scale, there is still not an agreed upon protocol for performing such measurements. That raises issues when it comes to sharing and comparing soil spectral data [132]. Although laboratory measurements are held in controlled conditions, each laboratory has a different measurement protocol (e.g., sample preparation, instrument configuration, and different instrument). Soil spectra are affected by systematic (i.e., different instruments) and nonsystematic effects (i.e., random noise and instabilities). To overcome such hindrances, Ben Dor et al. [132] proposed the internal soil standard and a simple protocol to measure soil properties under laboratory conditions; thanks to these efforts, soil spectral libraries from different sources could be merged. Prior to the multivariate analysis, preprocessing techniques are utilized to remove physical phenomena from spectra, enhance spectral features, and improve the model's performance [33]. The most efficient and commonly-used preprocessing techniques seem to be the first derivative [133], standard normal variate [134], and continuum removal [34].

Several studies have demonstrated that important wavelength selection could improve the accuracy of a model by removing unnecessary information and improving a model's interpretability [41,42,55,56,135]. Another factor that affects the accuracy of a model is the calibration size. At present, it is gaining much attention, because it could also prevent soil sampling from being a waste of resources [83]. However, finding the optimum calibration size is not an easy task [64]. It has been reported that with an increasing calibration size, the model's accuracy is increased up to a certain point were no significant improvements can be further observed. Moreover, deciding the calibration sampling strategy could define the representativeness of the calibration dataset. To that end, Kennard–Stone sampling, Fuzzy c-means sampling, and cLHS could provide better results than random sampling [63].

The use of local models for estimating soil properties of a different region could be proven insufficient due to underrepresented ranges of soil property values and differences in soil types in a given area [136]. To cover global soil diversity and variation, many countries have started initiatives to create national, regional, and global SSLs (Australia, European Union, Brazil, and China) [20,83–87], while there are also projects providing publicly-available data, containing metadata regarding soils' key properties and their spectral signature [88,137]. Building these large databases is more cost effective than using conventional analytical methods for the determination of soil properties. The increased size of the data generated by the creation of SSLs, along with improvements in computational resources, enabled the use of machine learning techniques for model calibration [138]. Machine learning and data mining techniques could generate information by exploring the relationships between existing data, and has the ability to learn without being strictly programmed. Using global or national models for local calibrations could also prove to be problematic because it usually captures general trends with a greater level of generalization. To address this shortcoming, several methods are now being successfully evaluated to increase the accuracy of such predictions, like spiking [93,126] and RS-LOCAL [90]. These techniques can augment an SSL with site-specific samples minimizing the required reference measurements while capturing the respective variability.

Although acquiring the spectroscopic data is a feasible and rapid procedure, information is still hidden in the data. The potential of VNIR-SWIR spectroscopy to predict SOC is highly dependent on the selection of the multivariate calibration technique. Apart from the most commonly-used PLSR method, large SSLs provide the opportunity for machine learning techniques (ANN, RF, MARS, SVM, Cubist) to be increasingly evaluated, and in most cases they outperformed PLSR. Specifically, PLSR and other linear approaches may fail to successfully represent the complex relationship between the measured reflectance spectra and soil attributes, while they are also prone to model noise. Compared to PLSR, SVM has demonstrated better performance, probably thanks to its tolerance in the presence of outliers and spectral noise [22]. SVM use kernel functions to project the data onto a new hyperspace where complex nonlinear patterns can be simply represented; therefore, it is capable of learning in a high-dimensional feature space with fewer training data [58]. MARS is a nonparametric method that is also used to estimate complex nonlinear relationships, and in most cases, has outperformed both PLSR and SVM. MARS is a generalization of recursive partitioning regression approaches which generates

piece-wise functions that are aggregated in terms of an additive model; therefore, the accuracy of the predictions is higher when the underlying function is continuous [27]. Random forests is also a nonparametric algorithm that is composed of several regression or classification trees that are a combination of bootstrap aggregation and random feature selection. By that means, it is able to rank controlling factors of SOM and provide the relative importance of spectral variables [139]. The Cubist is a rule-based model that is able to generate more comprehensive predictive accuracy than conventional statistical methods with minimized risk of over fitting. It can include various predictor variables (categorical or continuous), and the importance of a variable can be automatically obtained to interpret the variable contribution mechanism in the final predictor model [140]. Deep learning algorithms such as ANN are based on the ability to learn during the training procedure in which they are presented with inputs and a set of expected outputs. Although they have shown good accuracy, their use is limited due to factors such as the selection of the hidden layer size, the requirement of a large number of training data, and the tendency to overfit the data [141]. In a similar manner, CNN is a neural network that includes one or more convolutional layers in its architecture, and has the ability to exploit local correlations to make them more attractive than the fully-connected neural networks [54]. Despite the superiority shown in these techniques, there is a lack of consistency in the results; hence, PLSR is still utilized, as it provides sufficient accuracy [45]. Variations in the study areas and in the selection of calibration data and their size, numerous combinations of preprocessing, and multivariate methods are some of the reasons that no calibration method has achieved universal acceptance.

When it comes to the validation procedure, cross validation [117,123] is the most commonly-used method due to the small sample size of the dataset, though one should keep in mind that there is the risk of overfitting the model; therefore, testing its performance on independent datasets usually provides less accuracy [46].

Considering in situ soil measurements, they also share the same issues with laboratory soil spectroscopy when it comes to model calibration, but with the additional hindrance of external factors affecting the measurements. The most frequently-addressed parameter is soil moisture. Soil moisture has a distinct impact on soil reflectance by decreasing the albedo [113]. To that end, methods to eliminate this effect such as EPO and DS have been proposed with relatively good results [51,142]. Attempts to create samples with variable soil moisture content have also been made, though they are somewhat time consuming [73,80,143]. Additional factors that could interfere with in situ measurements are ambient conditions such as temperature, wind, and precipitation [144,145], soil roughness that could be altered by field operations, stones, plant residues and movement during spectral acquisition, i.e., vibrations [112,113,121]. Moreover, it should be noted that in-field measurements are usually performed in the top layer of the soil, with an exception being studies where the sensor penetrates the soil, which may impact the model's quality because laboratory reference measurements concern samples taken from a layer. Different systems with various configurations regarding the mode of spectral acquisition, the sensor distance from the ground, and illumination conditions vary among studies, increasing the obstacle of comparing the results from one study to another. Regarding the experimental stage of most studies, it is difficult to create a common protocol for in situ, on-the-go measurements.

Differences in evaluations of results are also observed among studies. The most frequent statistical indicators are RMSE, RPD, and R^2. It should be noted that the error should be referred to using the same units as those of the analyzed soil property [146]; most studies do not use the same evaluation methods, as shown in Tables 1 and 2. In addition to that, the normalization of the data due to skewness makes it even harder to compare studies.

5. Conclusions

The objective of this review was to summarize the progress made during the last decade using laboratory and proximal soil sensing in the VNIR-SWIR region for SOC estimations. Laboratory measurements are considered to be a well-established method for soil properties estimations, but there is not yet a commonly accepted universal model with broad application. The attention that soil

spectroscopy has gained is evident on the efforts of creating SSLs worldwide, albeit, with a nonagreed protocol. Research has also been made towards the use of proximal soil sensors for in situ applications that could estimate SOC in real time and on larger spatial scales. Overall, the results are promising for SOC estimations, and more research needs to be done in terms of selecting the proper spectral range of the sensor, the preprocessing methods, and the calibration techniques. Detrimental effects such as soil moisture, soil roughness, vegetation cover, and others that affect SOC spectral response need to be addressed for model transferability from laboratory to in field applications. Due to the various inconsistencies among studies, it is suggested that articles should include more information about the experimental design, the criteria used for the selection of the chemometric approach, and the pre- and post- processing procedures to facilitate comparisons of results among studies. The advent of much smaller and more affordable spectrometers could potentially provide even more rapid soil properties estimations to assist farmers in selecting the most appropriate management practices with respect to soil preservation.

Abbreviations

AMI	mutual information-based adjacency
ANN	artificial neural network
APA	all possibilities approach
BRT	boosted regression trees
CAI	cellulose absorbance index
cLHs	conditional Latin hypercube sampling
CNN	convolutional neural networks
CRR	continuum removed reflectance
DS	direct standardization
EC	electrical conductivity
EPO	external parameter orthogonalization
HEM	heteroscedastic effects model
MARS	multivariate adaptive regression splines
MBL	memory bases learning
MC	moisture content
MIR	mid infrared
MLR	multivariate linear regression
NDVI	normalized difference vegetation index
NMSI	normalized soil moisture index
OPS	ordered predictor selection
OSC	orthogonal signal correction
PCA	principal component analysis
PLSR	partial least square regression
R^2	coefficient of determination
RF	random forest
RMSE	root mean square error
RPD	residual prediction deviation
RS-LOCAL	re-sampling-local
SCANS	soil condition analysis system
SMOTE	synthetic minority oversampling technique

SNV	standard normal variate
SOC	soil organic carbon
SOM	soil organic matter
SPA	successive projection algorithm
SPLSR	sparse partial least squares regression
SSLs	soil spectral libraries
SVMR	support vector machines regression
SWIR	short wave infrared
VIP	variable importance in the projection
VNIR	visible near infrared

References

1. Johnston, A.E.; Poulton, P.R.; Coleman, K. Chapter 1 Soil Organic Matter: Its Importance in Sustainable Agriculture and Carbon Dioxide Fluxes. *Adv. Agron.* **2009**, *101*, 1–57.

2. 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]

3. Bogunovic, I.; Muñoz-Rojas, M.; Brevik, E.C. Soil ecosystem services, sustainability, valuation and management. *Curr. Opin. Environ. Sci. Heal.* **2018**, *5*, 7–13.

4. Lampridi, M.G.; Sørensen, C.G.; Bochtis, D. Agricultural Sustainability: A Review of Concepts and Methods. *Sustainability* **2019**, *11*, 5120. [CrossRef]

5. Oldeman, L.R.; Hakkeling, R.T.A.; Sombroek, W.G. *World Map of the Status of Human-Induced Soil Degradation: An Explanatory Note*, 2nd. rev. ed.; ISRIC: Wageningen, The Netherlands, 1991.

6. Panagos, P.; Montanarella, L. Soil Thematic Strategy: An important contribution to policy support, research, data development and raising the awareness. *Curr. Opin. Environ. Sci. Heal.* **2018**, *5*, 38–41. [CrossRef]

7. Keesstra, S.D.; Bouma, J.; Wallinga, J.; Tittonell, P.; Smith, P.; Cerdà, A.; Montanarella, L.; Quinton, J.N.; Pachepsky, Y.; Van Der Putten, W.H.; et al. The significance of soils and soil science towards realization of the United Nations sustainable development goals. *SOIL* **2016**, *2*, 111–128. [CrossRef]

8. Visser, S.; Keesstra, S.; Maas, G.; de Cleen, M.; Molenaar, C. Soil as a Basis to Create Enabling Conditions for Transitions Towards Sustainable Land Management as a Key to Achieve the SDGs by 2030. *Sustainability* **2019**, *11*, 6792. [CrossRef]

9. Carter, M.R. Soil quality for sustainable land management: Organic matter and aggregation interactions that maintain soil functions. *Agron. J.* **2002**, *94*, 38–47. [CrossRef]

10. Kirschbaum, M.U.F. Will changes in soil organic carbon act as a positive or negative feedback on global warming? *Biogeochemistry* **2000**, *48*, 21–51. [CrossRef]

11. Qiu, L.; Wei, X.; Zhang, X.; Cheng, J.; Gale, W.; Guo, C.; Long, T. Soil organic carbon losses due to land use change in a semiarid grassland. *Plant Soil* **2012**, *355*, 299–309. [CrossRef]

12. Martínez-Mena, M.; Alvarez Rogel, J.; Castillo, V.; Albaladejo, J. Organic carbon and nitrogen losses influenced by vegetation removal in a semiarid mediterranean soil. *Biogeochemistry* **2002**, *61*, 309–321. [CrossRef]

13. Balafoutis, A.; Beck, B.; Fountas, S.; Vangeyte, J.; Wal, T.; Soto, I.; Gómez-Barbero, M.; Barnes, A.; Eory, V. Precision Agriculture Technologies Positively Contributing to GHG Emissions Mitigation, Farm Productivity and Economics. *Sustainability* **2017**, *9*, 1339. [CrossRef]

14. FAO. Soil Organic Carbon the Hidden Potential. 2017. Available online: https://www.amazon.com/Soil-Organic-Carbon-Hidden-Potential/dp/9251096813 (accessed on 2 September 2019).

15. Jandl, R.; Rodeghiero, M.; Martinez, C.; Cotrufo, M.F.; Bampa, F.; van Wesemael, B.; Harrison, R.B.; Guerrini, I.A.; deB Richter, D., Jr.; Rustad, L.; et al. Current status, uncertainty and future needs in soil organic carbon monitoring. *Sci. Total Environ.* **2014**, *468–469*, 376–383. [CrossRef] [PubMed]

16. Davis, M.R.; Alves, B.J.R.; Karlen, D.L.; Kline, K.L.; Galdos, M.; Abulebdeh, D. Review of soil organic carbon measurement protocols: A US and Brazil comparison and recommendation. *Sustainability* **2017**, *10*, 53. [CrossRef]

17. Angelopoulou, T.; Tziolas, N.; Balafoutis, A.; Zalidis, G.; Bochtis, D. Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sens.* **2019**, *11*, 676. [CrossRef]

18. Shaw, G.A.; Burke, H.K. Spectral Imaging for Remote Sensing. *LINCOLN Lab. J.* **2003**, *14*.

19. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. *Chapter Four—Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring*; Sparks, D.L.B.T.-A., Ed.; Academic Press: San Diego, CA, USA, 2015; Volume 132, pp. 139–159. ISBN 0065-2113.

20. Brown, D.J.; Shepherd, K.D.; Walsh, M.G.; Mays, M.D.; Reinsch, T.G. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* **2006**, *132*, 273–290. [CrossRef]

21. Soriano-Disla, J.M.; Janik, L.J.; Viscarra Rossel, R.A.; Macdonald, L.M.; McLaughlin, M.J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Appl. Spectrosc. Rev.* **2014**, *49*, 139–186. [CrossRef]

22. Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. *Anal. Chim. Acta* **2018**, *1026*, 8–36. [CrossRef]

23. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]

24. 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. [CrossRef]

25. Baumgardner, M.F.; Kristof, S.; Johannsen, C.J.; Zachary, A. Effects of organic matter on the multispectral properties of soils. *Proc. Indian Acad. Sci.* **1970**, *79*, 413–422.

26. Cécillon, L.; Cassagne, N.; Czarnes, S.; Gros, R.; Vennetier, M.; Brun, J.J. Predicting soil quality indices with near infrared analysis in a wildfire chronosequence. *Sci. Total Environ.* **2009**, *407*, 1200–1205. [CrossRef] [PubMed]

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

28. Wetterlind, J.; Stenberg, B.; Rossel, R.A.V. Soil analysis using visible and near infrared spectroscopy. *Methods Mol. Biol.* **2013**, *953*, 95–107.

29. Krishnan, P.; Alexander, J.D.; Butler, B.J.; Hummel, J.W. Reflectance Technique for Predicting Soil Organic Matter1. *Soil Sci. Soc. Am. J.* **1980**, *44*, 1282. [CrossRef]

30. Dalal, R.C.; Henry, R.J. Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry1. *Soil Sci. Soc. Am. J.* **1986**, *50*, 120. [CrossRef]

31. Shi, T.; Guo, L.; Chen, Y.; Wang, W.; Shi, Z.; Li, Q.; Wu, G. Proximal and remote sensing techniques for mapping of soil contamination with heavy metals. *Appl. Spectrosc. Rev.* **2018**, *53*, 783–805. [CrossRef]

32. 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]

33. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]

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

35. 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]

36. Gupta, A.; Vasava, H.B.; Das, B.S.; Choubey, A.K. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. *Geoderma* **2018**, *325*, 59–71. [CrossRef]

37. de Santana, F.B.; de Souza, A.M.; Poppi, R.J. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *191*, 454–462. [CrossRef] [PubMed]

38. Geladi, P. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochim. Acta Part B At. Spectrosc.* **2003**, *58*, 767–782. [CrossRef]

39. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and near infrared spectroscopy in soil science. *Adv. Agron.* **2010**, *107*, 163–215.

40. Vibhute, A.D.; Kale, K.V.; Mehrotra, S.C.; Dhumal, R.K.; Nagne, A.D. Determination of soil physicochemical attributes in farming sites through visible, near-infrared diffuse reflectance spectroscopy and PLSR modeling. *Ecol. Process.* **2018**, *7*, 26. [CrossRef]

41. Vohland, M.; Besold, J.; Hill, J.; Fründ, H.-C. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma* **2011**, *166*, 198–205. [CrossRef]

42. 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]

43. Morellos, A.; Pantazi, X.E.; Moshou, D.; Alexandridis, T.; Whetton, R.; Tziotzios, G.; Wiebensohn, J.; Bill, R.; Mouazen, A.M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.* **2016**, *152*, 104–116. [CrossRef]

44. Stevens, A.; van Wesemael, B.; Bartholomeus, H.; Rosillon, D.; Tychon, B.; Ben-Dor, E. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma* **2008**, *144*, 395–404. [CrossRef]

45. Sithole, N.J.; Ncama, K.; Magwaza, L.S. Robust Vis-NIRS models for rapid assessment of soil organic carbon and nitrogen in Feralsols Haplic soils from different tillage management practices. *Comput. Electron. Agric.* **2018**, *153*, 295–301. [CrossRef]

46. Heinze, S.; Vohland, M.; Joergensen, R.; Ludwig, B. Usefulness of near-infrared spectroscopy for the prediction of chemical and biological soil properties in different long-term experiments. *J. Plant Nutr. Soil Sci.* **2013**, *176*, 520–528. [CrossRef]

47. Bikindou, F.D.A.; Gomat, H.Y.; Deleporte, P.; Bouillet, J.-P.; Moukini, R.; Mbedi, Y.; Ngouaka, E.; Brunet, D.; Sita, S.; Diazenza, J.-B.; et al. Are NIR spectra useful for predicting site indices in sandy soils under Eucalyptus stands in Republic of Congo? *For. Ecol. Manage.* **2012**, *266*, 126–137. [CrossRef]

48. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors (Switzerland)* **2018**, *18*, 2674. [CrossRef]

49. Quinlan, J.R. Learning with continuous classes. *Mach. Learn.* **1992**, *92*, 343–348.

50. Rossel, R.A.V.; Webster, R. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. *Eur. J. Soil Sci.* **2012**, *63*, 848–860. [CrossRef]

51. Xu, S.; Zhao, Y.; Wang, M.; Shi, X. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma* **2018**, *310*, 29–43. [CrossRef]

52. Gholizadeh, A.; Saberioon, M.; Carmon, N.; Boruvka, L.; Ben-Dor, E.; Gholizadeh, A.; Saberioon, M.; Carmon, N.; Boruvka, L.; Ben-Dor, E. Examining the Performance of PARACUDA-II Data-Mining Engine versus Selected Techniques to Model Soil Carbon from Reflectance Spectra. *Remote Sens.* **2018**, *10*, 1172. [CrossRef]

53. Carmon, N.; Ben-Dor, E. An Advanced Analytical Approach for Spectral-Based Modelling of Soil Properties. *Int. J. Emerg. Technol. Adv. Eng.* **2017**, *7*.

54. Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning to predict soil properties from regional spectral data. *Geoderma Reg.* **2019**, *16*, e00198. [CrossRef]

55. 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]

56. Shi, T.; Wang, J.; Chen, Y.; Wu, G. Improving the prediction of arsenic contents in agricultural soils by combining the reflectance spectroscopy of soils and rice plants. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 95–103. [CrossRef]

57. Chen, H.; Liu, X.; Jia, Z.; Liu, Z.; Shi, K.; Cai, K. A combination strategy of random forest and back propagation network for variable selection in spectral calibration. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 101–108. [CrossRef]

58. Xuemei, L.; Jianshe, L. Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration. *Meas. J. Int. Meas. Confed.* **2013**, *46*, 3808–3814. [CrossRef]

59. Raj, A.; Chakraborty, S.; Duda, B.M.; Weindorf, D.C.; Li, B.; Roy, S.; Sarathjith, M.C.; Das, B.S.; Paulette, L. Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach. *Geoderma* **2018**, *314*, 146–159. [CrossRef]

60. Bayer, A.; Bachmann, M.; Müller, A.; Kaufmann, H. A Comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African Ecosystem. *Appl. Environ. Soil Sci.* **2012**, *2012*. [CrossRef]

61. Moura-Bueno, J.M.; Dalmolin, R.S.D.; ten Caten, A.; Dotto, A.C.; Demattê, J.A.M. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. *Geoderma* **2019**, *337*, 565–581. [CrossRef]

62. Clingensmith, C.M.; Grunwald, S.; Wani, S.P. Evaluation of calibration subsetting and new chemometric methods on the spectral prediction of key soil properties in a data-limited environment. *Eur. J. Soil Sci.* **2019**, *70*, 107–126. [CrossRef]

63. Debaene, G.; Niedźwiecki, J.; Pecio, A.; Żurek, A. Effect of the number of calibration samples on the prediction of several soil properties at the farm-scale. *Geoderma* **2014**, *214–215*, 114–125. [CrossRef]

64. Lucà, F.; Conforti, M.; Castrignanò, A.; Matteucci, G.; Buttafuoco, G. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. *Geoderma* **2017**, *288*, 175–183. [CrossRef]

65. Vohland, M.; Ludwig, M.; Thiele-Bruhn, S.; Ludwig, B. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma* **2014**, *223–225*, 88–96. [CrossRef]

66. St. Luce, M.; Ziadi, N.; Zebarth, B.J.; Grant, C.A.; Tremblay, G.F.; Gregorich, E.G. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. *Geoderma* **2014**, *232–234*, 449–458. [CrossRef]

67. Rienzi, E.A.; Mijatovic, B.; Mueller, T.G.; Matocha, C.J.; Sikora, F.J.; Castrignanò, A. Prediction of Soil Organic Carbon under Varying Moisture Levels Using Reflectance Spectroscopy. *Soil Sci. Soc. Am. J.* **2014**, *78*, 958. [CrossRef]

68. Wijewardane, N.K.; Ge, Y.; Morgan, C.L.S. Prediction of soil organic and inorganic carbon at different moisture contents with dry ground VNIR: a comparative study of different approaches. *Eur. J. Soil Sci.* **2016**, *67*, 605–615. [CrossRef]

69. Hong, Y.; Yu, L.; Chen, Y.; Liu, Y.; Liu, Y.; Liu, Y.; Cheng, H. Prediction of soil organic matter by VIS-NIR spectroscopy using normalized soil moisture index as a proxy of soil moisture. *Remote Sens.* **2018**, *10*, 28. [CrossRef]

70. Ben-Dor, E. Quantitative remote sensing of soil properties. *Adv. Agron.* **2002**, *75*, 173–244.

71. Marakkala Manage, L.P.; Greve, M.H.; Knadel, M.; Moldrup, P.; de Jonge, L.W.; Katuwal, S. Visible-Near-Infrared Spectroscopy Prediction of Soil Characteristics as Affected by Soil-Water Content. *Soil Sci. Soc. Am. J.* **2018**, *82*, 1333. [CrossRef]

72. Lobell, D.B.; Asner, G.P. Moisture effects on soil reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66*, 722–727. [CrossRef]

73. Nocita, M.; Stevens, A.; Noon, C.; van Wesemael, B. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma* **2013**, *199*, 37–42. [CrossRef]

74. Haubrock, S.N.; Chabrillat, S.; Lemmnitz, C.; Kaufmann, H. Surface soil moisture quantification models from reflectance data under field conditions. *Int. J. Remote Sens.* **2008**, *29*, 3–29. [CrossRef]

75. Ji, W.; Viscarra Rossel, R.A.; Shi, Z. Accounting for the effects of water and the environment on proximally sensed vis-NIR soil spectra and their calibrations. *Eur. J. Soil Sci.* **2015**, *66*, 555–565. [CrossRef]

76. Wang, Y.; Veltkamp, D.J.; Kowalski, B.R. Multivariate Instrument Standardization. *J. Electroanal. Chem. Interfaclal Electro-chem.* **1991**, *63*, 2347–2352. [CrossRef]

77. Roger, J.-M.; Chauchard, F.; Bellon-Maurel, V. EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom. Intell. Lab. Syst.* **2003**, *66*, 191–204. [CrossRef]

78. Minasny, B.; McBratney, A.B.; Bellon-Maurel, V.; Roger, J.M.; Gobrecht, A.; Ferrand, L.; Joalland, S. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* **2011**, *167–168*, 118–124. [CrossRef]

79. de Santana, F.B.; de Giuseppe, L.O.; de Souza, A.M.; Poppi, R.J. Removing the moisture effect in soil organic matter determination using NIR spectroscopy and PLSR with external parameter orthogonalization. *Microchem. J.* **2019**, *145*, 1094–1101. [CrossRef]

80. Roudier, P.; Hedley, C.B.; Lobsey, C.R.; Viscarra Rossel, R.A.; Leroux, C. Evaluation of two methods to eliminate the effect of water from soil vis–NIR spectra for predictions of organic carbon. *Geoderma* **2017**, *296*, 98–107. [CrossRef]

81. Rodionov, A.; Pätzold, S.; Welp, G.; Pallares, R.C.; Damerow, L.; Amelung, W. Sensing of Soil Organic Carbon Using Visible and Near-Infrared Spectroscopy at Variable Moisture and Surface Roughness. *Soil Sci. Soc. Am. J.* **2014**, *78*, 949. [CrossRef]

82. 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]

83. 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*. [CrossRef]

84. Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Brown, D.J.; Demattê, J.A.M.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; et al. A global spectral library to characterize the world's soil. *Earth-Science Rev.* **2016**, *155*, 198–230. [CrossRef]

85. Shi, Z.; Wang, Q.L.; Peng, J.; Ji, W.J.; Liu, H.J.; Li, X.; Viscarra Rossel, R.A. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Sci. China Earth Sci.* **2014**, *57*, 1671–1680. [CrossRef]

86. Brodský, L.; Klement, A.; Penížek, V.; Kodešová, R.; Boruvka, L. Building soil spectral library of the czech soils for quantitative digital soil mapping. *Soil Water Res.* **2011**, *6*, 165–172. [CrossRef]

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

88. Tóth, G.; Jones, A.; Montanarella, L. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environ. Monit. Assess.* **2013**, *185*, 7409–7425. [CrossRef]

89. Liu, Y.; Shi, Z.; Zhang, G.; Chen, Y.; Li, S.; Hong, Y.; Shi, T.; Wang, J.; Liu, Y.; Liu, Y.; et al. Application of Spectrally Derived Soil Type as Ancillary Data to Improve the Estimation of Soil Organic Carbon by Using the Chinese Soil Vis-NIR Spectral Library. *Remote Sens.* **2018**, *10*, 1747. [CrossRef]

90. Lobsey, C.R.; Viscarra Rossel, R.A.; Roudier, P.; Hedley, C.B. Rs-Local Data-Mines Information From Spectral Libraries To Improve Local Calibrations. *Eur. J. Soil Sci.* **2017**, *68*, 840–852. [CrossRef]

91. Gogé, F.; Gomez, C.; Jolivet, C.; Joffre, R. Which strategy is best to predict soil properties of a local site from a national Vis-NIR database? *Geoderma* **2014**, *213*, 1–9. [CrossRef]

92. Barthès, B.G.; Kouakoua, E.; Clairotte, M.; Lallemand, J.; Chapuis-Lardy, L.; Rabenarivo, M.; Roussel, S. Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. *Geoderma* **2019**, *338*, 422–429. [CrossRef]

93. Jiang, Q.; Li, Q.; Wang, X.; Wu, Y.; Yang, X.; Liu, F. Estimation of soil organic carbon and total nitrogen in different soil layers using VNIR spectroscopy: Effects of spiking on model applicability. *Geoderma* **2017**, *293*, 54–63. [CrossRef]

94. Viscarra Rossel, R.A.; Adamchuk, V.I.; Sudduth, K.A.; McKenzie, N.J.; Lobsey, C. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. *Adv. Agron.* **2011**, *113*, 243–291.

95. Kuang, B.; Mahmood, H.S.; Quraishi, M.Z.; Hoogmoed, W.B.; Mouazen, A.M.; van Henten, E.J. Sensing Soil Properties in the Laboratory, in situ, and on-line. A review. Elsevier Inc., 2012, 1st ed. Vol. 114. Available online: https://www.sciencedirect.com/science/article/pii/B9780123942753000031 (accessed on 2 September 2019).

96. Christy, C.D. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Comput. Electron. Agric.* **2008**, *61*, 10–19. [CrossRef]

97. Sinfield, J.V.; Fagerman, D.; Colic, O. Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Comput. Electron. Agric.* **2010**, *70*, 1–18. [CrossRef]

98. Gehl, R.J.; Rice, C.W.; Gehl, R.J.; Rice, C.W. Emerging technologies for in situ measurement of soil carbon. *Clim. Change* **2007**, *80*, 43–54. [CrossRef]

99. Waiser, T.H.; Morgan, C.L.S.; Brown, D.J.; Hallmark, C.T. In Situ Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy. *Soil Sci. Soc. Am. J.* **2007**, *71*, 389. [CrossRef]

100. Sarkhot, D.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]

101. Bricklemyer, R.S.; Brown, D.J. On-the-go VisNIR: Potential and limitations for mapping soil clay and organic carbon. *Comput. Electron. Agric.* **2010**, *70*, 209–216. [CrossRef]

102. Kweon, G.; Lund, E.; Maxton, C. Soil organic matter and cation-exchange capacity sensing with on-the-go electrical conductivity and optical sensors. *Geoderma* **2013**, *199*, 80–89. [CrossRef]

103. Knadel, M.; Thomsen, A.; Schelde, K.; Greve, M.H. Soil organic carbon and particle sizes mapping using vis-NIR, EC and temperature mobile sensor platform. *Comput. Electron. Agric.* **2015**, *114*, 134–144. [CrossRef]

104. Wetterlind, J.; Piikki, K.; Stenberg, B.; Söderström, M. Exploring the predictability of soil texture and organic matter content with a commercial integrated soil profiling tool. *Eur. J. Soil Sci.* **2015**, *66*, 631–638. [CrossRef]

105. Pei, X.; Sudduth, K.; Veum, K.; Li, M. Improving In-Situ Estimation of Soil Profile Properties Using a Multi-Sensor Probe. *Sensors* **2019**, *19*, 1011. [CrossRef] [PubMed]

106. Veum, K.S.; Parker, P.A.; Sudduth, K.A.; Holan, S.H. Predicting Profile Soil Properties with Reflectance Spectra via Bayesian Covariate-Assisted External Parameter Orthogonalization. *Sensors* **2018**, *18*, 3869. [CrossRef] [PubMed]

107. Kuang, B.; Tekin, Y.; Mouazen, A.M. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. *Soil Tillage Res.* **2015**, *146*, 243–252. [CrossRef]

108. Sorenson, P.T.; Small, C.; Tappert, M.C.; Quideau, S.A.; Drozdowski, B.; Underwood, A.; Janz, A. Monitoring organic carbon, total nitrogen, and pH for reclaimed soils using field reflectance spectroscopy. *Can. J. Soil Sci.* **2017**, *97*, 241–248. [CrossRef]

109. Kodaira, M.; Shibusawa, S. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. *Geoderma* **2013**, *199*, 64–79. [CrossRef]

110. 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]

111. Mouazen, A.M.; Ramon, H. Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture content. *Soil Tillage Res.* **2006**, *86*, 218–229. [CrossRef]

112. Rodionov, A.; Welp, G.; Damerow, L.; Berg, T.; Amelung, W.; Pätzold, S. Towards on-the-go field assessment of soil organic carbon using Vis-NIR diffuse reflectance spectroscopy: Developing and testing a novel tractor-driven measuring chamber. *Soil Tillage Res.* **2015**, *145*, 93–102. [CrossRef]

113. Franceschini, M.H.D.; Demattê, J.A.M.; Kooistra, L.; Bartholomeus, H.; Rizzo, R.; Fongaro, C.T.; Molin, J.P. Effects of external factors on soil reflectance measured on-the-go and assessment of potential spectral correction through orthogonalisation and standardisation procedures. *Soil Tillage Res.* **2018**, *177*, 19–36. [CrossRef]

114. Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185. [CrossRef]

115. Ge, Y.; Thomasson, J.A.; Sui, R. Remote sensing of soil properties in precision agriculture: A review. *Front. Earth Sci.* **2011**, *5*, 229–238. [CrossRef]

116. Viscarra Rossel, R.A.; Lobsey, C.R.; Sharman, C.; Flick, P.; McLachlan, G. Novel Proximal Sensing for Monitoring Soil Organic C Stocks and Condition. *Environ. Sci. Technol.* **2017**, *51*, 5630–5641. [CrossRef] [PubMed]

117. Cozzolino, D.; Cynkar, W.U.; Dambergs, R.G.; Shah, N.; Smith, P. In Situ Measurement of Soil Chemical Composition by Near-Infrared Spectroscopy: A Tool Toward Sustainable Vineyard Management. *Commun. Soil Sci. Plant Anal.* **2013**, *44*, 1610–1619. [CrossRef]

118. Gras, J.-P.; Barthès, B.G.; Mahaut, B.; Trupin, S. Best practices for obtaining and processing field visible and near infrared (VNIR) spectra of topsoils. *Geoderma* **2014**, *214–215*, 126–134. [CrossRef]

119. Cambou, A.; Cardinael, R.; Kouakoua, E.; Villeneuve, M.; Durand, C.; Barthès, B.G. Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. *Geoderma* **2016**, *261*, 151–159. [CrossRef]

120. Bartholomeus, H.; Kooistra, L.; Stevens, A.; van Leeuwen, M.; van Wesemael, B.; Ben-Dor, E.; Tychon, B. Soil Organic Carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 81–88. [CrossRef]

121. Rodionov, A.; Pätzold, S.; Welp, G.; Pude, R.; Amelung, W. Proximal field Vis-NIR spectroscopy of soil organic carbon: A solution to clear obstacles related to vegetation and straw cover. *Soil Tillage Res.* **2016**, *163*, 89–98. [CrossRef]

122. Guerrero, C.; Zornoza, R.; Gómez, I.; Mataix-Beneyto, J. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. *Geoderma* **2010**, *158*, 66–77. [CrossRef]

123. Nawar, S.; Mouazen, A.M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *CATENA* **2017**, *151*, 118–129. [CrossRef]

124. Nawar, S.; Mouazen, A.M. On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. *Soil Tillage Res.* **2019**, *190*, 120–127. [CrossRef]

125. Nawar, S.; Mouazen, A.M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Comput. Electron. Agric.* **2018**, *151*, 469–477. [CrossRef]

126. Guerrero, C.; Stenberg, B.; Wetterlind, J.; Viscarra Rossel, R.A.; Maestre, F.T.; Mouazen, A.M.; Zornoza, R.; Ruiz-Sinoga, J.D.; Kuang, B. Assessment of soil organic carbon at local scale with spiked NIR calibrations: Effects of selection and extra-weighting on the spiking subset. *Eur. J. Soil Sci.* **2014**, *65*, 248–263. [CrossRef]

127. Kühnel, A.; Bogner, C. In-situ prediction of soil organic carbon by vis–NIR spectroscopy: An efficient use of limited field data. *Eur. J. Soil Sci.* **2017**, *68*, 689–702. [CrossRef]

128. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]

129. Kempen, B.; Dalsgaard, S.; Kaaya, A.K.; Chamuya, N.; Ruipérez-González, M.; Pekkarinen, A.; Walsh, M.G. Mapping topsoil organic carbon concentrations and stocks for Tanzania. *Geoderma* **2019**, *337*, 164–180. [CrossRef]

130. Demattê, J.A.M.; Dotto, A.C.; Bedin, L.G.; Sayão, V.M.; Souza, A.B. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. *Geoderma* **2019**, *337*, 111–121. [CrossRef]

131. Angelopoulou, T.; Dimitrakos, A.; Terzopoulou, E.; Zalidis, G.; Theocharis, J.; Stafilov, T.; Zouboulis, A. Reflectance Spectroscopy (Vis-NIR) for Assessing Soil Heavy Metals Concentrations Determined by two Different Analytical Protocols, Based on ISO 11466 and ISO 14869-1. *Water. Air. Soil Pollut.* **2017**, *228*. [CrossRef]

132. Ben Dor, E.; Ong, C.; Lau, I.C. Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma* **2015**, *245–246*, 112–124. [CrossRef]

133. Viscarra Rossel, R.A.; Brus, D.J.; Lobsey, C.; Shi, Z.; McLachlan, G. Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. *Geoderma* **2016**, *265*, 152–163. [CrossRef]

134. Adeline, K.R.M.; Gomez, C.; Gorretta, N.; Roger, J.M. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma* **2017**, *288*, 143–153. [CrossRef]

135. Ng, W.; Minasny, B.; Malone, B.P.; Sarathjith, M.C.; Das, B.S. Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling. *Comput. Electron. Agric.* **2019**, *158*, 201–210. [CrossRef]

136. Padarian, J.; Minasny, B.; McBratney, A.B. Transfer learning to localise a continental soil vis-NIR calibration model. *Geoderma* **2019**, *340*, 279–288. [CrossRef]

137. Tsakiridis, N.L.; Theocharis, J.B.; Zalidis, G.C. An evolutionary fuzzy rule-based system applied to real-world Big Data - The GEO-CRADLE and LUCAS soil spectral libraries. *IEEE Int. Conf. Fuzzy Syst.* **2018**, *2018-July*, 1–8.

138. Tsakiridis, N.L.; Theocharis, J.B.; Ben-Dor, E.; Zalidis, G.C. Using interpretable fuzzy rule-based models for the estimation of soil organic carbon from VNIR/SWIR spectra and soil texture. *Chemom. Intell. Lab. Syst.* **2019**, *189*, 39–55. [CrossRef]

139. Hong, Y.; Chen, S.; Liu, Y.; Zhang, Y.; Yu, L.; Chen, Y.; Liu, Y.; Cheng, H.; Liu, Y. Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. *CATENA* **2019**, *174*, 104–116. [CrossRef]

140. 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]

141. Gholizadeh, A.; Borůvka, L.; Saberioon, M.; Vašát, R. Visible, Near-Infrared, and Mid-Infrared Spectroscopy 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]

142. Ackerson, J.P.; Morgan, C.L.S.; Ge, Y. Penetrometer-mounted VisNIR spectroscopy: Application of EPO-PLS to in situ VisNIR spectra. *Geoderma* **2017**, *286*, 131–138. [CrossRef]

143. Yin, Z.; Lei, T.; Yan, Q.; Chen, Z.; Dong, Y. A near-infrared reflectance sensor for soil surface moisture measurement. *Comput. Electron. Agric.* **2013**, *99*, 101–107. [CrossRef]

144. Prudnikova, E.Y.; Savin, I.Y. Study of the optical properties of an exposed soil surface. *J. Opt. Technol.* **2016**, *83*, 642. [CrossRef]

145. Prudnikova, E.; Savin, I.; Vindeker, G.; Grubina, P.; Shishkonakova, E.; Sharychev, D. Influence of Soil Background on Spectral Reflectance of Winter Wheat Crop Canopy. *Remote Sens.* **2019**, *11*, 1932. [CrossRef]

146. England, J.R.; Rossel, R.A.V. Proximal sensing for soil carbon accounting. *SOIL* **2018**, *4*, 101–122. [CrossRef]