sustainability

MDPI

Article

Spectral Heterogeneity Analysis and Soil Organic Matter Inversion across Differences in Soil Types and Organic Matter Content in Dryland Farmland in China

Qiang Shen [1], Shiwen Zhang [2,*] and Ke Xia [1]

[1] School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China; qiangshen@hhu.edu.cn (Q.S.); kexia@hhu.edu.cn (K.X.)
[2] School of Earth and Environmental, Anhui University of Science and Technology, Huainan 232001, China
* Correspondence: shwzhang@aust.edu.cn

Abstract: Soil organic matter (SOM) plays an important role in agricultural production and arable land quality improvement. Hyperspectral technology enables frequent surveys over large areas. In this study, we explored the spectral heterogeneity of differences in soil types and SOM content, and proposed a method for measuring SOM content in large areas using spectroscopy. The results indicate regional variations in factors affecting soil spectral absorption peaks, with noticeable latitudinal disparities. The first-order differential partial-least-squares method provided the best prediction for the SOM inversion. The coefficient of determination (R^2) for the SOM inversion model was 0.93, and the root mean square error (RMSE) was 3.42, with an 8.49 g/kg difference in the SOM content. When the difference in SOM content fell between 8 and 15 g/kg, the inversion effect model performed best. The optimal model R^2 exceeded 0.95, and the RMSE was less than 5. The comprehensive analysis showed that the organic matter content was an important factor affecting the SOM content estimate and must be considered in the real process. In addition, it is crucial to categorize soil samples on the basis of distinct soil types while maintaining a consistent range of SOM content within the same soil type, ideally between 8 and 15 g/kg. Subsequently, the first-order differential partial least squares method is applicable. These results are expected to contribute to the acquisition of high-quality information on variations in the SOM of complex large-scale areas.

Keywords: organic matter; VNIR-SWIR spectroscopy; spectral characteristic selection; agricultural soil; digital soil mapping

check for updates

Citation: Shen, Q.; Zhang, S.; Xia, K. Spectral Heterogeneity Analysis and Soil Organic Matter Inversion across Differences in Soil Types and Organic Matter Content in Dryland Farmland in China. *Sustainability* **2023**, *15*, 16310. https://doi.org/10.3390/su152316310

Academic Editor: Jeroen Meersmans

Received: 30 August 2023
Revised: 3 November 2023
Accepted: 23 November 2023
Published: 25 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Soil organic matter (SOM) is an important component of soil for preserving ecological equilibrium and agricultural sustainability [1]. Its impact on soil texture, structure, water retention, and its role in carbon and nutrient cycling is profound [2]. Consequently, a comprehensive understanding of the distribution and SOM content is imperative for effective land resource management, agricultural productivity, and food security preservation [3]. Although conventional soil sampling and analytical techniques have been valuable in research, they have certain limitations [4]. These methods typically require significant time and resources, mainly when collecting and analyzing soil samples across extensive areas [5]. Consequently, there is a pressing demand for a new approach to efficiently and accurately assess SOM content to advance agricultural modernization and enhance overall agricultural efficiency [6].

Numerous studies have provided evidence of the viability of hyperspectral techniques in the quantity monitoring of SOM content [7,8]. Bowers and Hanks demonstrated a marked correlation between organic matter content and the visible light band of soil [9]. In a separate study, observations indicated correlations between SOM content and specific bands, including the ultraviolet (UV) band at 376.8 nm, the visible band at 616.5 nm,

and the near-infrared (NIR) band at 724.1 nm [10]. Laamrani et al. conducted an assessment to determine the optimal hyperspectral bands for detecting organic matter content in agricultural soils using visible near infrared (VNIR) and shortwave infrared (SWIR) bands [11]. Despite the valuable insights gained from these findings regarding the hyperspectral inversion of SOM, it is important to acknowledge certain limitations. The first is the small area of monitoring. Many studies usually focus on smaller areas, and the inversion models developed are often localized [12,13]. Although these models may be effectively employed within particular limits, their applicability to other areas necessitates further investigation. Consequently, they cannot accurately predict the SOM content over a larger area. Furthermore, monitoring SOM through remote sensing is a complex process influenced by numerous factors. It is worth noting that several studies have overlooked the impact of significant soil variations resulting from diverse geographical regions on the spectral characteristics of organic matter [14]. In addition, most SOM inversion models are statistical models, and their performance is highly dependent on the selection of modeling samples [15]. Significant fluctuations in the SOM content can substantially influence the efficacy of model inversion. While scholars have examined the spectral characteristics of diverse SOM from various perspectives, they have yet to analyze the underlying causes of these phenomena [16]. No corresponding solutions have been proposed in the process of SOM inversion for some identified problems. Numerous scholars have also examined the use of multispectral data, such as Landsat8 and Sentinel-2, to predict SOM content [17,18]. Despite this approach's potential for extensive monitoring, multispectral remote sensing images possess limited bands and spectral resolution, which do not provide sufficiently detailed spectral information on SOM. Thus, SOM inversion is not ideal. A key problem in current research is how to realize hyperspectral remote sensing mapping of SOM in complex large regions while guaranteeing accuracy.

Different scholars have extensively studied the spectral characteristics of soil in complex large-scale areas [19,20]. Condit et al. conducted a study in which they carefully selected 160 soil samples from 36 states across the United States [21]. They determined the spectral characteristics of the near ultraviolet and visible bands and subsequently established an empirical regression equation to describe the reflectance characteristics of the soil. Stoner et al. conducted spectral reflectance measurements on 485 soil samples from the United States and Brazil. These samples were subsequently classified on the basis of the influence of their organic matter and iron elements [22]. Similarly, Wang Qianlong et al. collected 1661 soil samples from 13 provinces in China to develop a hyperspectral inversion model for determining soil nitrogen content. To validate their model, they tested it on 104 rice soil samples collected from Zhejiang Province [23]. Goge et al. used over 2000 sets of spectral data from various regions across France. They employed a local partial-least-squares model following neighborhood selection to estimate parameters such as organic carbon and cation exchange capacity [24]. However, there is currently a lack of research on the achievement of high-precision inversion results of SOM across a large geographical area encompassing differences in soil types and SOM content.

Given that the dryland farmland in China possesses fertile soils and serves as a crucial food production hub, it is imperative to promptly and accurately determine the distribution of organic matter within these areas. Sun et al. conducted a comprehensive examination of the primary determinants of organic carbon density in various spatial regions within dry farmland [25]. Huang et al. elucidated the spatial and temporal evolution of key indicators of cropland in dryland farming by rigorously defining the scope and content of this practice [26]. Dryland farmland in China encompasses two distinct climatic zones, and the spectral response of SOM exhibits noticeable latitudinal zonal variations. Consequently, the application of existing small-area modeling or multispectral imaging techniques for accurately monitoring organic matter content in the dryland farmland is challenging.

In view of the above problems, we provide an important reference for SOM mapping in complex large-scale areas. The objectives of this study were to (1) further clarify the spectral response characteristics of differences in soil types and SOM content, (2) establish

a methodology for hyperspectral inversion modeling of SOM in large-scale areas, and (3) analyze the spatial distribution pattern of SOM in dryland farmland in China.

2. Materials and Methods

2.1. Study Area

The delineation of dry farmland was established by employing a topographic gradient below 5° and a proportion of dryland and cultivated land exceeding 40% within a 1 km^2 grid. The dry farmland studied here is situated in the eastern region of China with an average elevation of 150 m above sea level, and the topography is relatively flat. The major climate types in the region are temperate and subtropical monsoon climates, typically involving four distinct seasons with rain and heat over the same period and sufficient light [27,28]. The study area was located from 112°33′54.8″ E to 135°3′55.953″ E longitude and from 32°8′24.506″ N to 48°27′21.127″ N latitude. The study was conducted in 7 provinces and 694 counties, covering a total land area of 57.68 million square kilometers. The main soil type in the northern part of the study area is black soil, which is dark in color, fertile, and high in organic matter. In the southern part, it is predominantly loess, with a low organic matter content and a loose texture.

2.2. Data Acquisition

In the study area, the sampling points were uniformly distributed, resulting in 399 sampling locations. The sampling depth ranged from 0 to 10 cm (Figure 1). The collected soil samples were placed in a dry and ventilated place for natural air drying and sieved to a size of <100 μm. After drying and sieving, the soil samples were divided into two parts to determine the SOM content and soil spectral reflectance. We analyzed the SOM content using the potassium dichromate volumetric method [29]. A FieldSpec4 portable object spectrometer was used to measure the spectral reflectance, and the measurement wavelengths ranged from 350 to 2500 nm. We chose 12 V, 50 W halogen lamps as a light source. The fiber optic probe was positioned 7 cm vertically above the soil sample, with the light source inclined at a 45-degree angle and a direct distance of 60 cm from the soil sample [30,31]. We completed a whiteboard calibration and experimented with the stabilization of the instrument before use. Each sample was sampled at 1 nm intervals, repeated 10 times, and the resulting average value was chosen as the sample's spectral reflectance.

To more accurately depict the impacts of differences in soil types and SOM content on spectral characteristics, we categorized soil samples within the research region into seven distinct soil types, as defined by the China Soil Classification System [32,33], published by the National Soil Census Office in 1993. Given the large geographical span of the study area according to its longitude and latitude, we stratified the samples into five grades (I–V) based on the disparity in SOM content between the highest and lowest levels. Comprehensive details are provided in Tables 1 and 2.

Table 1. SOM content of the soil types and sites tested. Given are the maximum, minimum, and mean content.

Soil Type	Main Provinces	Sample Points	SOM (g/kg)			
			Max	Min	Mean	Standard Deviation
Dark-brown earth	Heilongjiang	15	85.77	12.53	39.66	11.65
Fluvo-aquic soils	Heilongjiang, Hebei	197	77.13	4.96	20.42	9.75
Yellow-cinnamon soils	Henan	11	23.94	9.54	17.65	11.14
Cinnamon soils	Hebei, Henan	61	79.28	8.22	23.72	10.07
Chernozems	Heilongjiang, Jilin	32	44.48	5.72	25.49	11.98
Black soils	Heilongjiang, Jilin	68	71.15	9.47	28.51	12.42
Brown earth	Liaoning	15	27.51	6.09	16.60	5.68

Table 2. Classification of SOM content (Grade I to V) based on the difference between maximum and minimum content.

Grade	Content Difference (between Max and Min)	SOM (g/kg)			
		Max	Min	Mean	Standard Deviation
I	3.89	23.15	19.26	21.18	1.22
II	8.49	25.60	17.11	21.19	2.43
III	14.58	29.46	14.88	21.37	3.93
IV	23.11	35.26	12.15	21.73	5.82
V	80.81	85.77	4.96	23.21	11.65

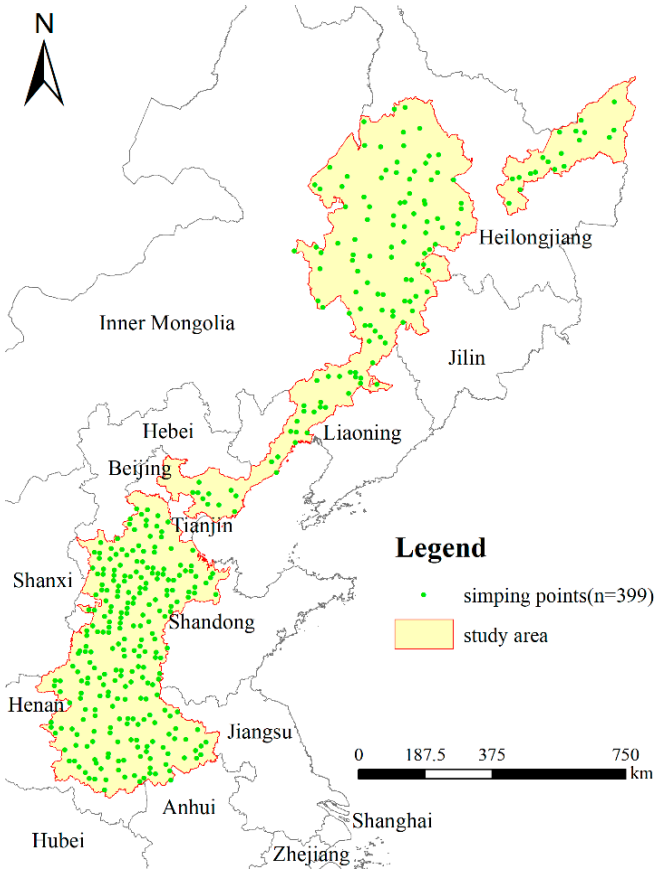

Figure 1. Study area location and sampling points distribution map.

2.3. Research Methods

We used the Kennard–Stone (K-S) algorithm to uniformly select samples with clear differences based on the Euclidean distance between spectral variables. This method ensures the rationality of the spatial distribution of the samples in the SOM inversion model [34,35]. The SOM content was inverted using the partial-least-squares regression (PLSR) and random forest regression (RFR) methods. The coefficient of determination (R^2) and root mean square error (RMSE) were introduced as precision indices to measure the degree of prediction of the model. The coefficient of determination reflects the fitting degree and stability of the model. The closer R^2 is to one, the greater the stability of the model. The RMSE reflects the ability of the model to estimate reality—that is, the smaller the value, the better the model's estimation [36,37].

2.3.1. Spectral Data Preprocessing

To reduce noise arising from the experimental environment and instrumentation, we excluded the spectral bands of 350–399 and 2401–2500 nm. Second, we applied the

multiplicative scatter correction (MSC) and Savitzky–Golay (S–G) convolution smoothing techniques to refine the original spectral data. The smoothed spectral data were transformed using first-order differential reflectance (FDR), second-order differential reflectance (SDR), and continuum removal (CR). The experiments revealed that the combination of the S-G convolution smoothing method and the FDR method yielded the most effective spectral transformation [38–40]. To mitigate data redundancy and minimize data storage requirements, we employed a wavelet neural network for spectral data compression [41,42].

2.3.2. Partial Least Squares Regression (PLSR)

Compared with conventional models, PLSR allows the use of fewer samples than variables and does not consider multivariate correlation [43,44]. Its model building method has been widely used in the field of spectral data processing.

To create the model, we established a spectral matrix X of $m \times n$ and the organic matter quantity measurement matrix Y of $n \times 1$, in which m is the number of spectral bands and n is the number of samples, as given in Equations (1) and (2):

$$X = TP^T + E \tag{1}$$

$$Y = UQ^T + F \tag{2}$$

where P and Q are the loads, E and F are the residual matrices from the PLSR simulation, and U and T are the score matrices.

Equation (3) provides the correlation coefficient matrix for B, the linear regression for U and T, and the prediction formula for the volume of organic matter as follows:

$$Y = TBQ = XP^TBQ \tag{3}$$

2.3.3. Random Forest Regression (RFR)

RFR is a machine learning method based on a classification and regression decision tree. Its main principle is that the bootstrap aggregation (bagging) integrated classifier, composed of weak classifiers and nodes, is divided into several decision trees using random classification technology [30,45]. The best classification results are selected through voting. The basic steps of random forest construction are as follows: k sample sets are extracted from the samples using bagging resampling technology, and K decision trees are constructed using the k sample sets, in which the k out-of-pocket data are composed of data that are not extracted each time. In the process of decision tree generation, a tree has n features, and m features ($m < n$) are randomly extracted at each node [46,47].

The final classification decision is given in Equation (4):

$$H(x) = \text{argmax}\sum_{i=1}^{k} I(h_i(x) = Y) \tag{4}$$

where $H(x)$ is a combined classification model system, h_i is a single decision tree, Y is the objective variable, and $I()$ is the demonstrative function.

The margin function is introduced to measure the error degree of the correct classification number that exceeds the average error number. The larger the margin, the more reliable is the classification, as given in Equation (5):

$$mg(X, Y) = av_i I(h_i(X) = Y) - \max av_i I(h_i(X) = j) \tag{5}$$

where $h_i(x)$ is a single decision tree classifier, (Y, X) is a randomly distributed dataset, and $I()$ is a demonstrative function.

2.4. SOM Mapping

To effectively predict the spatial variability of SOM in dry farmland, the optimal SOM inversion model was employed, in conjunction with the geostatistical method, to optimally

interpolate and visualize the spatial distribution of SOM. Additionally, the SOM content was classified according to China's soil-nutrient-grading standards.

3. Results

3.1. Soil Attributes Descriptive Statistics

The tables show that the distribution of different soil types in the study area is relatively scattered. The range of the SOM content was 4.96 to 85.77 g/kg. With regard to spatial distribution, the SOM content and soil types showed clear latitudinal zonal disparities. The amount of organic matter in the soils of the three northeastern provinces was higher than that in the other areas. The SOM content of the different soil types also varied considerably; dark-brown earth, black soils, and chernozems had a higher organic matter content, and these soils were distributed mainly in Heilongjiang and Jilin. The organic matter content in yellow-cinnamon soils and brown earth was the most deficient and was distributed mainly in the southern provinces of Henan.

We constructed the mean spectral reflectance curves for different soil types (Figure 2). The results showed that the general trend of the soil spectral reflectance curves was the same, but the values of soil spectral reflectance among the different soil types were different. The values for chernozems, black soils, and dark-brown earth, with a greater organic matter content, were typically lower, indicating that the quantity of spectral reflectance decreased as the amount of organic matter in the soil increased. The spectral curves of dark-brown earth, black soils, and chernozems were relatively smooth, whereas the other soil types had clear spectral absorption peaks.

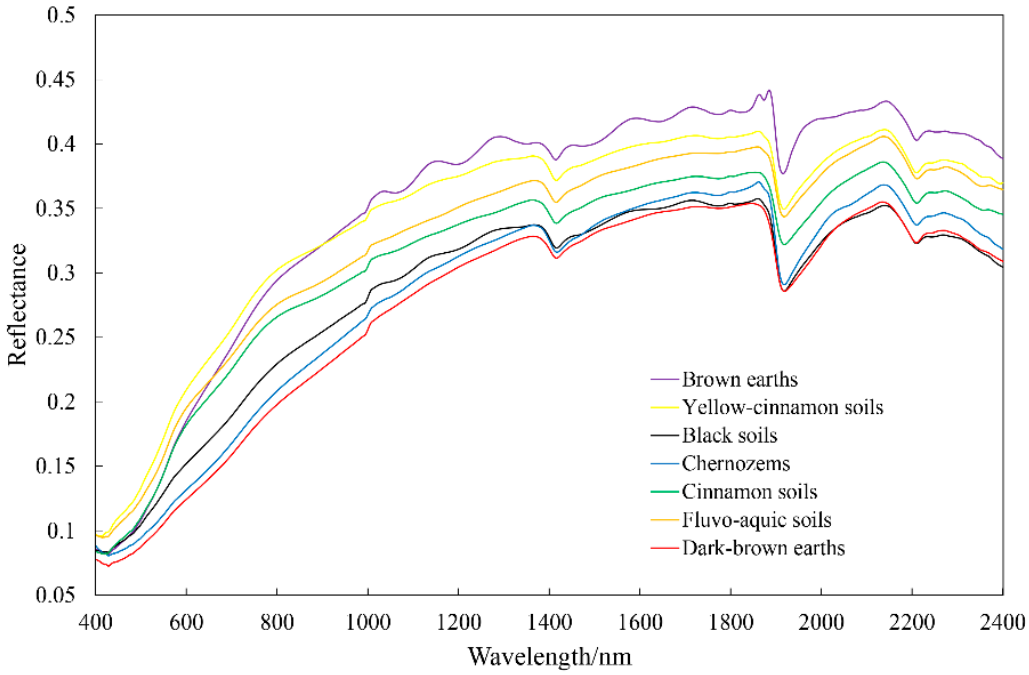

Figure 2. Comparison of the mean spectral curves of different soil types in the studied area.

3.2. Analysis of Soil Spectral Characteristics

Figure 3a,b show that the variation trend of the first-order differential curve of the spectrum was generally the same. The spectral absorption bands were around 800, 1000, 1400, 1900, and 2200 nm. With a decrease in latitude and a relative increase in the amount of organic matter, the reflectance numerical value of the first-order differential spectrum in the absorption peak band became more apparent, and the clearest reflectance absorption peaks were those of chernozems and black soils. Figure 3c shows that the correlation coefficient of the first-order differential curve of the different soil types exhibited significant fluctuations. The curve displayed minor fluctuations in the 400–900 nm range, with most

regions showing negative correlations. In the 900–2400 nm range, the curve exhibited more pronounced fluctuations, and the correlation coefficients varied between positive and negative values. Figure 3d shows a noticeable grading phenomenon in the correlation curve within the 400–800 nm range. As the difference in the organic material content increased, the degree of negative correlation also increased.

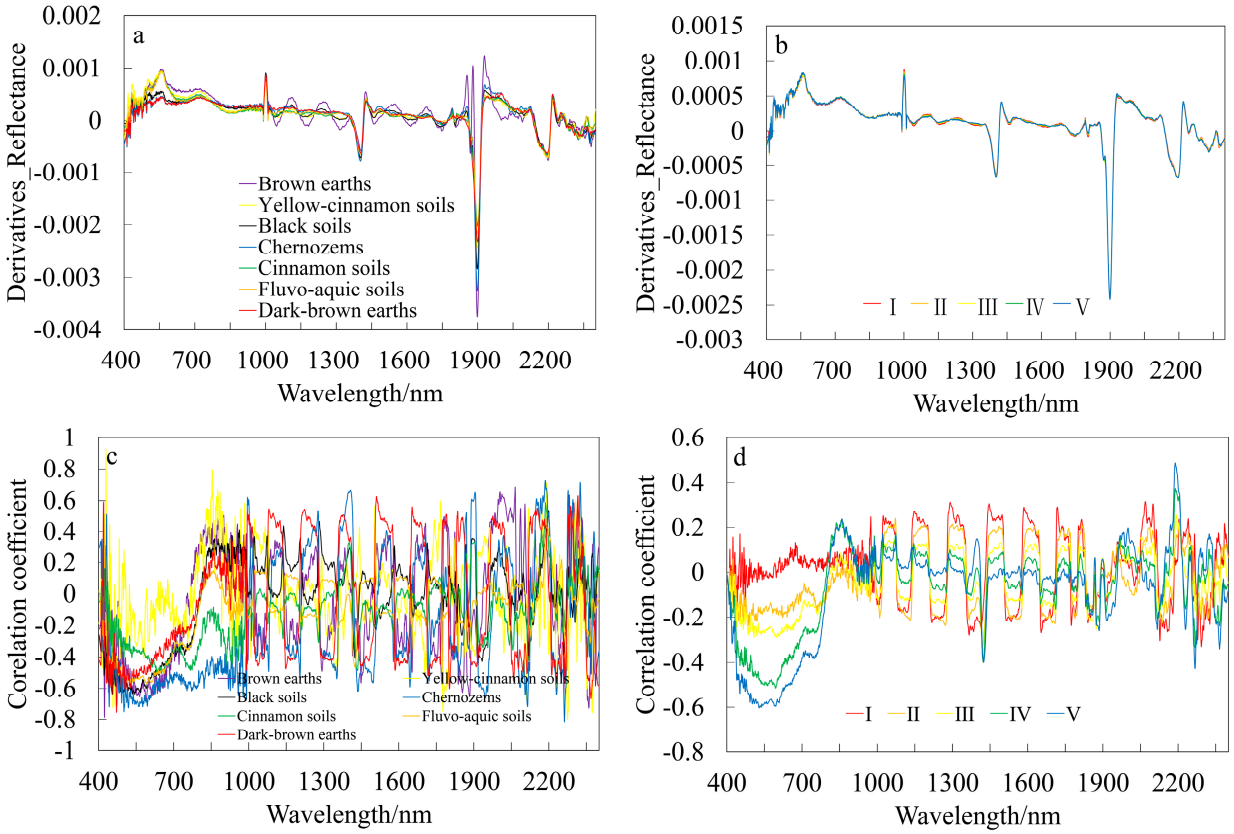

Figure 3. Curves for spectral first-order differential and correlation coefficient between SOM content and spectral reflectance: (**a**) FDR of different soil types; (**b**) FDR of difference in SOM content; (**c**) correlation coefficient curve of different soil types; and (**d**) correlation coefficient curve for grades I–V (difference in SOM content).

3.3. SOM Inversion Using Hyperspectral Data

3.3.1. Modeling and Analysis

The results of the validation using prediction models with the highest accuracy, according to the lowest RMSE and the highest R^2 values, are summarized in Tables 3 and 4. There was no significant difference between the two models for the full sample. For the soil types, except for chernozems, the PLSR method was superior to the RFR method, and the RMSE showed no significant difference, but R^2 was generally increased by more than 0.2. When compared with the model for the full sample, R^2 generally increased by more than 0.4; the inversion effect of SOM content in yellow-cinnamon soil was the best, and R^2 and RMSE were 0.98 and 0.52, respectively. The results in Figure 4a show that the influence of different soil types on the inversion effect should be considered when estimating the SOM content using hyperspectral techniques. Figure 4b shows that the difference in the organic matter content influenced the inversion effect of the two methods. For the RFR method, with an increase in the difference in quantity, model R^2 and RMSE gradually increased. For the PLSR method, the model worked best when we used the second-order difference for the SOM content; R^2 and RMSE were 0.93 and 3.42, respectively.

Table 3. SOM inversion results based on the RFR and PLSR models for different soil types.

Soil Types	RFR		PLSR	
	R^2	RMSE	R^2	RMSE
All samples	0.57	6.98	0.51	7.53
Dark-brown earth	0.24	10.91	0.79	8.14
Fluvo-aquic soils	0.76	6.09	0.95	6.66
Yellow-cinnamon soils	0.19	2.40	0.98	0.52
Cinnamon soils	0.74	5.48	0.97	5.54
Chernozems	0.94	6.26	0.92	6.29
Black soils	0.73	12.5	0.93	12.34
Brown earth	0.33	3.40	0.70	3.18

Table 4. SOM inversion results based on difference in SOM content (Grade I to V).

Grade	RFR		PLSR	
	R^2	RMSE	R^2	RMSE
I	0.24	1.14	0.51	1.21
II	0.30	3.00	0.93	3.42
III	0.30	4.57	0.91	5.38
IV	0.34	5.65	0.82	6.15
V	0.57	6.98	0.51	7.53

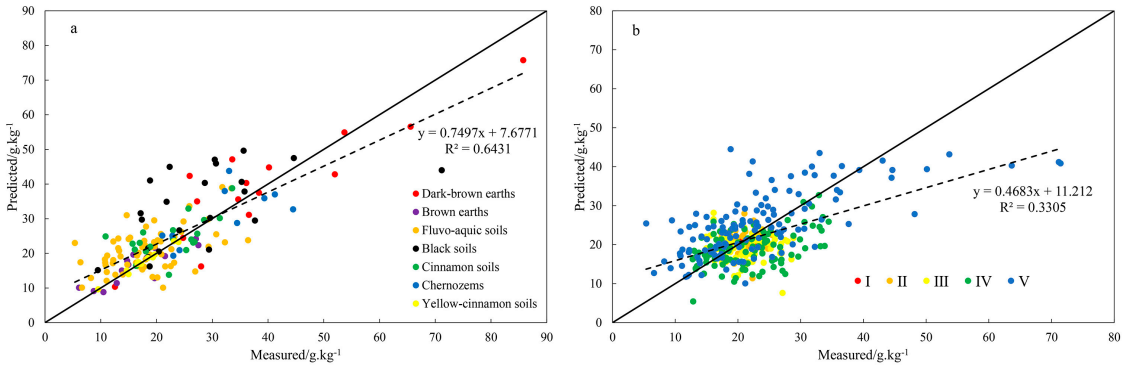

Figure 4. SOM inversion method verification (comparison results of actual measured values and spectral data predicted values) based on the PLSR model for (**a**) difference in soil types and (**b**) difference in SOM content (Grade I to V). (Black solid lines represent one–to–one bisectors. The black dotted line is the trend line for all soil samples.).

3.3.2. Effect of Content Differences on Inversion of SOM of Fluvo-Aquic Soil

To avoid the impact of varying soil types, tidal soil was used as an example to investigate the influence of difference in content on SOM inversion. The samples were divided into five categories (I–V), and the SOM inversion models were established using the first–order differential PLSR method (Table 5, Figure 5).

Table 5. Classification (Grades I–V) based on difference in SOM content of fluvo-aquic soil.

Grade	Content Difference (between Max and Min)	SOM (g/kg)			
		Max	Min	Mean	Standard Deviation
I	3.86	23.38	19.53	21.55	1.24
II	8.41	25.88	17.46	21.59	2.31
III	14.67	29.25	14.57	20.82	3.99
IV	23.71	35.95	12.24	20.82	5.61
V	72.17	77.13	4.96	20.42	9.76

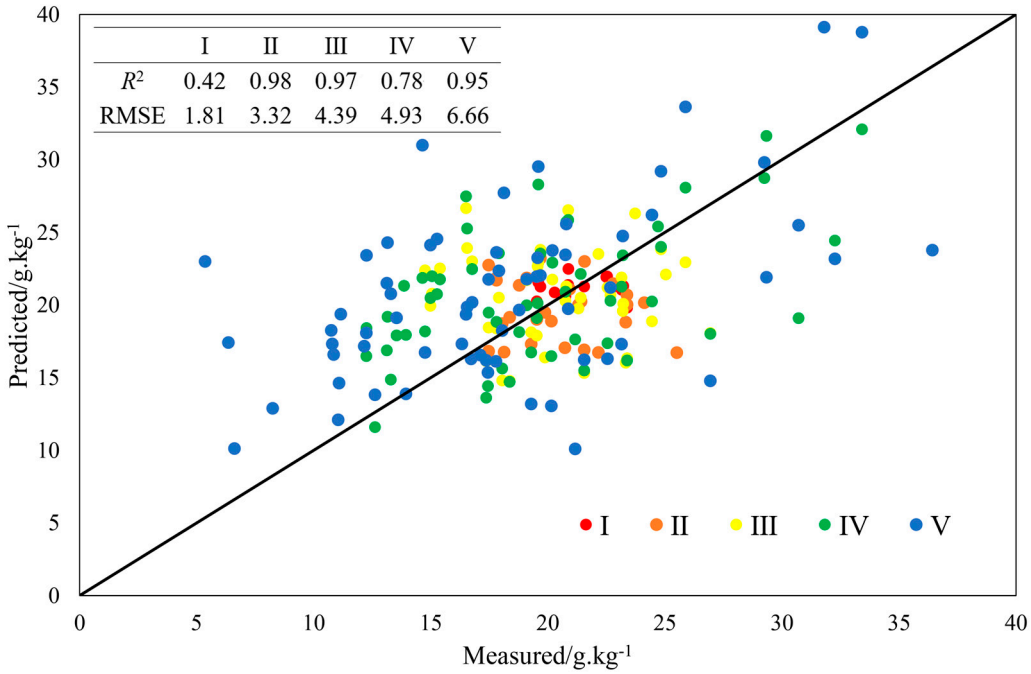

Figure 5. SOM estimation method verification (comparison results of actually measured values and predicted values from spectral data) of fluvo-aquic soil with difference in SOM content based on the PLSR method.

As shown in Figure 5, the inversion effect of the fluvo-aquic soil samples and all the other samples was the same for the difference in SOM content. The coefficient of determination showed a trend of initially increasing and then decreasing. The model precision of grades II to V was the highest; R^2 was greater than 0.95, and the R^2 of grade II was the highest, which was 0.98. The RMSE increased as the differences in the amount increased. The maximum RMSE was 6.66 for the fifth-order difference, and the prediction ability of the model was poor. We can see that the predictive ability and stability of the model reached a good range when the difference in SOM content was grade II. Generally, the effects of differences in soil types and SOM content should be fully considered when inverting SOM in complex large-scale areas.

3.4. Spatial Distribution of SOM in Dry Farmland

The optimal SOM inversion model was employed in conjunction with geostatistical techniques to interpolate and visually represent the spatial distribution of SOM in dry farmland (Figure 6). The SOM content in the dry farmland is generally high in the north and low in the south, and more than 70% of the area has an organic matter content of less than the medium level. The highest organic matter content was found in the northern part of the region, especially in Heilongjiang Province, where the organic matter content was above the medium level. The SOM contents in the central and southern regions are generally low, with the lowest organic matter content in the western hills of Liaoxi and the central Huanghuaihai Plain. The SOM inversion model based on hyperspectral analysis combined with the PLSR method of first-order differentiation can invert the spatial distribution trend and classification characteristics of SOM in complex large-scale areas.

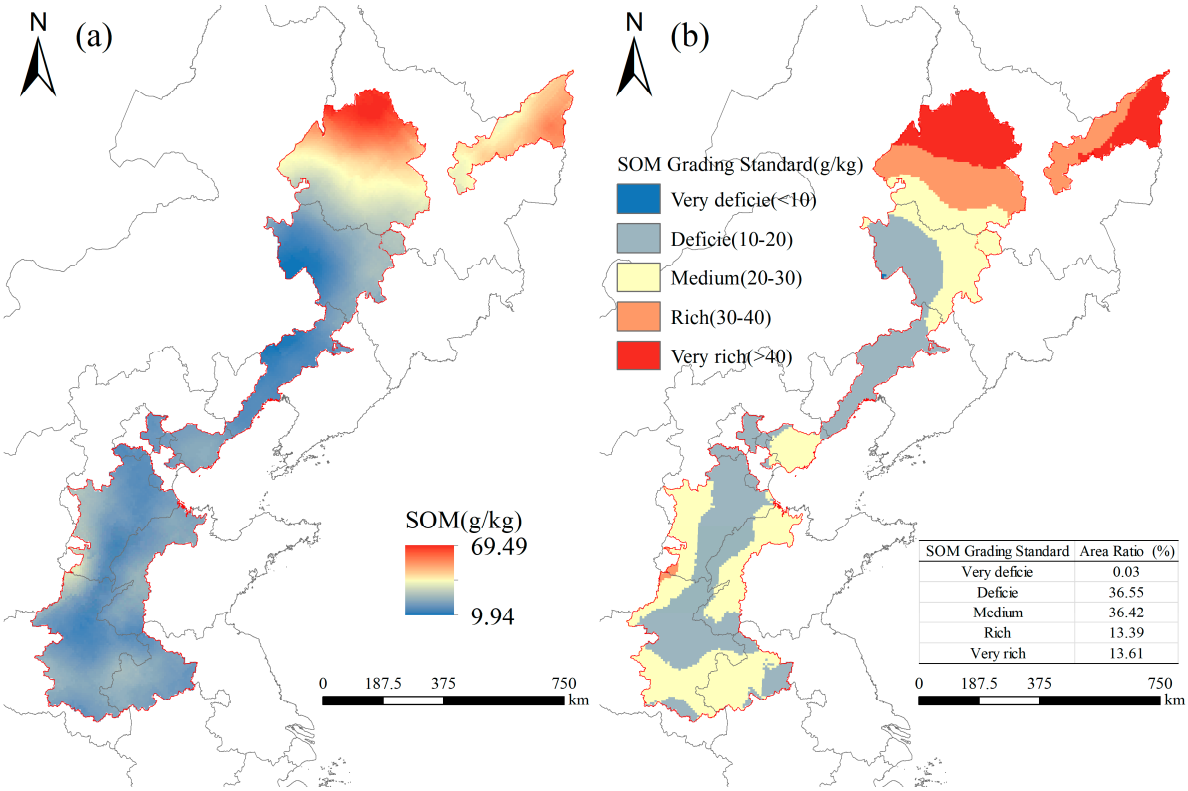

Figure 6. Spatial distribution of SOM in dryland farmland. (**a**) SOM content; (**b**) SOM grading stan–ard.

4. Discussion

4.1. Feasibility of Inverse SOM Using Hyperspectral Data

In this study, soil sampling sites were uniformly obtained from dry farmland in China, focusing on the effects of differences in soil types and SOM content due to latitudinal zonal disparities, to assess the feasibility of hyperspectral techniques in inverting SOM over complex large-scale areas. Some studies have shown that differential spectral transformation can effectively solve these problems [48,49]. The spectral bands sensitive to soil organic matter were isolated using various spectral transforms in conjunction with correlation analysis (Figure 3). The extracted correlation peaks aligned with the findings of Chang et al. [50]. Furthermore, the correlation coefficients showed a strong association between these bands and the SOM content. It has been observed that variations in soil composition and organic matter concentration do not exert any influence on the location of spectral absorption peaks, but they do have an impact on the correlation coefficients. This observation shows the existence of dissimilar spectral sensitivities of organic matter within different soil samples [37,51]. Inverse modeling of the SOM using hyperspectral data was promising (Tables 3 and 4). However, the evaluation of model performance, as measured using R^2 and RMSE, varies significantly across different studies because of variations in sample collection methods, spectral processing techniques, and model construction approaches. These variations can significantly impact the accuracy and reliability of spectral modeling [52,53]. The PLSR model can identify system information and noise and permits regression modeling provided that the number of sample points is less than the number of variables. This model has various applications in spectral inversion [54,55]. On the other hand, Sun et al. concluded that the sample segmentation method affects the estimation accuracy [56]. A similar performance can also be observed in the estimation of the SOM content with GA-PLSR. This study yielded similar findings, highlighting significant disparities in the inversion results derived from diverse modeling sample sets based on differences in soil types and SOM content. Regrettably, such discrepancies are frequently overlooked

in numerous studies. On the basis of the analysis of several soil samples, we concluded that the duplicated division of soil samples based on differences in soil types and SOM content is beneficial for reducing the chance of an abnormal estimation accuracy. Combined with the inversion effect, model accuracy, and other factors, the use of hyperspectral data is a promising method for inverting the SOM content of complex large-scale areas.

4.2. Influence of a Large Areas on SOM Hyperspectral Inversion

The results revealed that the accuracy of the SOM hyperspectral inversion models based on different samples was relatively diverse. Combined with the research results of other scholars, the main factors affecting the model inversion effect are soil Fe and Mn compounds, and SOM content [45,57]. First, this study showed clear absorption peaks for the spectral curves at 800, 1000, 1400, and 1900 nm. The results of previous studies showed that the characteristics of the spectral curves at 800–1000 nm were mainly due to the electron jump in metal ions (Fe^{2+}, Fe^{3+}, and Mn^{2+}) [58]. The values of dark-brown soil, black soil, and chernozems with a higher amount of organic matter were generally lower, indicating that the higher the amount of organic matter, the lower the hyperspectral reflectance. This agrees with the results of Qinhong Liao and Zhou Shi et al. [59]. The effects of Fe-Mn compounds on the hyperspectral model of SOM are mainly based on two aspects. On the one hand, Fe-Mn compounds are important dyes in soil that can change its color, and color is a decisive factor affecting the spectral reflectance of soil. In addition, Fe-Mn compounds have a strong masking effect on the expression of soil spectral characteristics, as has been proven in many studies [60,61].

From Table 6, we can see the difference in the SOM content and the Fe-Mn compounds of different soil types. Combined with the results of model inversion, it can be seen that soil with a high content of Fe-Mn compounds generally has a poor modeling effect. This is a good method for establishing an SOM content inversion model by classification according to soil type. Compared with sample-wide modeling, R^2 and RMSE increased by approximately 0.6 and 0.5, respectively. This difference not only manifested for different soil types but also showed a regional pattern. Because of the difference in soil types and the natural environment, the SOM content in the northern region was higher than that in the southern region (Figure 1, Table 1) [62]. Organic matter was the main factor in the northern part of the study area, whereas ferromanganese compounds were the main factor in the southern part, similar to the results of related research [63]. Based on this study, we recommend that when employing hyperspectral techniques to estimate SOM content, soil samples should be categorized according to different soil types while maintaining the variance in organic matter content within the same soil type within the range of 8–15. Further research is required to determine how to increase the precision of an SOM inversion model when the concentration of Fe-Mn compounds is high. In addition, many other factors affected the soil spectral properties. Lv et al. suggested that soil moisture is the main cause of pseudospectra and noise changes [64]. Meanwhile, due to the complexity of the soil formation process in the study area, the mineral composition of different soils may also be an important influencing factor [65]. Therefore, the influence and regulation of other factors on the spectral characteristics of soil and the comprehensive influence degree and regularity of various factors require further study.

Table 6. Statistics of average SOM, iron, and manganese in the different soil types tested.

	SOM (g/kg)	Fe (mg/kg)	Mn (mg/kg)
Dark-brown earth	39.66	60.48	38.64
Fluvo-aquic soils	20.42	13.72	16.04
Yellow-cinnamon soils	17.65	24.63	22.89
Cinnamon soils	23.72	15.48	18.91
Chernozems	25.49	26.47	20.61
Black soils	28.51	31.28	22.76
Brown earth	16.60	50.15	41.54

5. Conclusions

The expeditious acquisition of precise regional SOM maps is a crucial requirement for safeguarding land resources and devising pertinent policies. In this paper, we propose a reproducible workflow for realizing SOM inversion in complex large-scale areas using hyperspectral data. We further improved the accuracy of generating SOM maps of dry farmland in China. Our study revealed significant variations in the spectral characteristics of soils across various regions. Therefore, it is crucial to thoroughly account for the influences of differences in soil types and organic matter when conducting research in this field, as these factors can greatly impact the accuracy of inversion results. In the northern section of the dry farmland, SOM emerged as the prevailing determinant, whereas in the south-central region, iron, manganese, and other heavy metals were dominant. Differences in spectrally sensitive bands were observed among the various soil types, with correlation coefficients exceeding 0.8 at the 429 nm band for yellow-brown soils and 2262 nm for black calcium soils, corresponding to correlation coefficients of 0.91 and 0.82, respectively. When comparing inversion with difference in SOM content, the most optimal modeling was achieved with a content difference of 8.49, yielding an R^2 of 0.93 and an RMSE of 3.42. Our methodology and workflow can provide a reference for soil mapping in complex large-scale areas.

However, there remains a significant challenge in elucidating and quantifying the spectral characteristics of diverse soils through a more comprehensive delineation of factors that contribute to soil differentiation. Subsequent research endeavors should strive to enhance and streamline the process of SOM inversion and develop meticulous inversion techniques tailored to specific soil classifications, thereby enabling continuous and comprehensive monitoring of SOM over extended periods.

Author Contributions: Conceptualization, Q.S.; Methodology, Q.S.; Software, Q.S.; Investigation, K.X.; Resources, K.X.; Writing—original draft, Q.S.; Writing—review & editing, K.X.; Supervision, S.Z.; Project administration, S.Z.; Funding acquisition, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China [grant number 2016YFD0300801]; the National Natural Science Foundation of China [grant number 41471186]; and the Open Found of Shaanxi Key Laboratory of Land Consolidation [grant number 2018-ZD07].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to our lab's policies and confidentiality agreements.

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

References

1. Lavallee, J.M.; Soong, J.L.; Cotrufo, M.F. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. *Glob. Chang. Biol.* **2020**, *26*, 261–273. [CrossRef] [PubMed]
2. Augusto, L.; Boca, A. Tree functional traits, forest biomass, and tree species diversity interact with site properties to drive forest soil carbon. *Nat. Commun.* **2022**, *13*, 1097–1108. [CrossRef] [PubMed]
3. State Council of the PRC. *Circular of the State Council on the Issuance of the National Plan for Agricultural Modernization (2016–2020)*; State Council of the PRC: Beijing, China, 2016.
4. Wang, S.; Guan, K.Y.; Zhang, C.H.; Lee, D.; Margenot, A.J.; Ge, Y.F.; Peng, J.; Zhou, W.; Zhou, Q.; Huang, Y.Z. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: Assessing potential of airborne and spaceborne optical soil sensing. *Remote Sens. Environ.* **2022**, *271*, 112914. [CrossRef]
5. Liu, F.; Rossiter, D.G.; Zhang, G.L.; Li, D.C. A soil colour map of China. *Geoderma* **2020**, *379*, 114556. [CrossRef]
6. Gupta, N. Development of Staring Hyperspectral Imagers. In Proceedings of the 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 11–13 October 2011; pp. 1–8. [CrossRef]

7. Kamil, S.; Helena, Ł.; Rafał, P.; Andrzej, D.; Łukasz, G.; Mateusz, Ł.; Anna, R.P.; Jan, S.; Bogusław, U. Agricultural Drought Monitoring by MODIS Potential Evapotranspiration Remote Sens Data Application. *Remote Sens.* **2020**, *12*, 3411. [CrossRef]
8. Lu, X.; Yong, H.; Yu, W.; Long, G.; Tiezhu, S.; Yi, L.; Qinghu, J.; Teng, F.; Yaolin, L.; Abdul, M.M.; et al. Estimation of Organic Carbon in Anthropogenic Soil by VIS-NIR Spectroscopy: Effect of Variable Selection. *Remote Sens.* **2020**, *12*, 3394. [CrossRef]
9. Liu, L.; Shen, R.; Ding, G. Studies on the Estimation of SOM Content Based on Hyper-Spectrum. *Spectrosc. Spectr. Anal.* **2011**, *31*, 762–766. [CrossRef]
10. Bowers, S.A.; Hanks, R.J. Reflection of Radiant Energy from Soils. *Soil Sci.* **1965**, *100*, 130–138. [CrossRef]
11. Sha, J.M.; Chen, P.C.; Chen, S.L. Characteristics Analysis pf Soil Spectrum Response Resulted from Organic Material. *Res. Soil Water Conserv.* **2003**, *10*, 21–24. [CrossRef]
12. Qi, Y.B.; Qie, X.; Qin, Q.R.; Shukla, M.K. Prediction of soil calcium carbonate with soil visi-ble-near-infrared reflection (Vis-NIR) spectral in Shaanxi province, China: Soil groups vs. spectral groups. *Int. J. Remote Sens.* **2021**, *42*, 2502–2516. [CrossRef]
13. Wei, L.F.; Yuan, Z.R.; Wang, Z.X.; Zhao, L.Y.; Zhang, Y.X.; Lu, X.Y.; Cao, L.Q. Hyperspectral Inversion of Soil Organic Matter Content Based on a Combined Spectral Index Model. *Sensors* **2020**, *20*, 2777. [CrossRef] [PubMed]
14. Nayak, A.K.; Rahman, M.M.; Naidu, R.; Dhal, B.; Swain, C.K.; Nayak, A.D.; Tripathi, R.; Shahid, M.; Islam, M.R.; Pathak, H. Current and emerging methodologies for estimating carbon sequestration in agricultural soils: A review. *Sci. Total Environ.* **2019**, *665*, 890–912. [CrossRef] [PubMed]
15. Ustin, S.L.; Middleton, E.M. Current and near-term advances in Earth observation for ecological applications. *Ecol. Process.* **2021**, *10*, 1. [CrossRef] [PubMed]
16. Dou, X.; Wang, X.; Liu, H.J.; Zhang, X.L.; Meng, L.H.; Pan, Y.; Yu, Z.Y.; Cui, Y. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China. *Geoderma* **2019**, *356*, 113896. [CrossRef]
17. Vaudour, E.; Gomez, C.; Lagacherie, P.; Loiseau, T.; Baghdadi, N.; Urbina-Salazar, D.; Loubet, B.; Arrouays, D. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. *Int. J. Appl. Earch Obs. Geoinf.* **2021**, *96*, 102277. [CrossRef]
18. Xiao, Y.F.; Gao, X.H.; Li, G.W. Optimal Selection of Calibration Sample Sets for Predicting SOM Contents from Visible and Near Infrared Reflection Spectrum. *Soils* **2020**, *52*, 404–413. [CrossRef]
19. 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]
20. Wang, X.P.; Zhang, F.; Kung, H.T.; Johnson, V.C. New Methods for Improving the Remote Sensing Estimation of SOM Content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. *Remote Sens. Environ.* **2018**, *218*, 104–118. [CrossRef]
21. Gong, Z.N.; Wang, Q.W.; Zhang, C.; Guan, H.L. SOM Estimation Using Hyperspectral Remote Sensing Techniques in a Water-Level-Fluctuating Zone Around Guanting Reservoir, Beijing, China. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 6929–6932.
22. Shen, L.Z.; Gao, M.F.; Yan, J.W.; Li, Z.L.; Leng, P.; Yang, Q.; Duan, S.B. Hyperspectral Estimation of SOM Content using Different Spectral Preprocessing Techniques and PLSR Method. *Remote Sens.* **2020**, *12*, 1206. [CrossRef]
23. Du, C.; Zhou, J. Application of Infrared Photoacoustic Spectroscopy in Soil Analysis. *Appl. Spectrosc. Rev.* **2011**, *46*, 405–422. [CrossRef]
24. Soriano, D.; José, M.; Janik, L.J.; Viscarra, R.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]
25. Sun, Z.X.; Li, Y.; Zhao, Y.; Huang, Y.F.; Guo, X.L.; Cao, M. Analysis on Spatial Distribution Characteristics and Driving Forces of Soil Organic Carbon Density in Dry Farming Region. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 255–262. [CrossRef]
26. Huang, Y.F.; Zhang, S.W. *Spatial and Temporal Evolution of Cropland Quality and Charac-Terization of Fertile Tillage in Dryland Farming*; Science Press: Beijing, China, 2021.
27. Liu, W.; Zhao, Z.; Yuan, H.; Song, C.; Li, X. An Optimal Selection Method of Samples of Calibration Set and Validation Set for Spectral Multivariate Analysis. *Spectrosc. Spectr. Anal.* **2014**, *34*, 947–951. [CrossRef]
28. Wang, X.; Li, L.; Liu, H.J.; Song, K.S.; Wang, L.P.; Meng, X.T. Prediction of soil organic matter using VNIR spectral parameters extracted from shape characteristics. *Soil Tillage Res.* **2021**, *216*, 105241. [CrossRef]
29. Cheng, H.; Shen, R.; Chen, Y.; Wan, Q.; Shi, T.; Wang, J.; Wan, Y.; Hong, Y.; Li, X. Estimating Heavy Metal Concentrations in Suburban Soils with Reflectance Spectroscopy. *Geoderma* **2019**, *336*, 59–67. [CrossRef]
30. Nawar, S.; Buddenbaum, H.; Hill, J.; Kozak, J.; Mouazen, A. Estimating the Soil Clay Content and Organic Matter by Means of Different Calibration Methods of vis-NIR Diffuse Reflectance Spectroscopy. *Soil Tillage Res.* **2015**, *155*, 510–522. [CrossRef]
31. Xu, Z.; Chen, S.; Lu, P.; Wang, Z.; Li, A.; Zeng, Q.; Chen, L. Optimizing a Standard Spectral Measurement Protocol to Enhance the Quality of Soil Spectra: Exploration of Key Variables in Lab-Based VNIR-SWIR Spectral Measurement. *Remote Sens.* **2022**, *14*, 1558. [CrossRef]
32. Xu, L.; Li, Q.; Zhu, X.; Liu, S. Hyperspectral Inversion of Heavy Metal Content in Coal Gangue Filling Reclamation Land. *Spectrosc. Spectr. Anal.* **2017**, *37*, 3839–3844. [CrossRef]
33. Tang, H.; Meng, X.T.; Su, X.X.; Ma, T.; Liu, H.J.; Bao, Y.L.; Zhang, M.W.; Zhang, X.L.; Huo, H.Z. Hyperspectral prediction on soil organic matter of different types using CARS algorithm. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 105–113. [CrossRef]

34. Wang, C.; Wu, X.; Li, L.; Wang, Y.; Li, Z. Convolutional Neural Network Application in Prediction of Soil Moisture Content. *Spectrosc. Spectr. Anal.* **2018**, *38*, 42–47. [CrossRef]
35. National Soil Census Office. *Soil Classification System in China*; Agricultural Publishing House: Beijing, China, 1993.
36. Geng, Z. *Soil Science*; Science Press: Beijing, China, 2010.
37. Liao, Q.; Gu, X.; Li, C.; Chen, L.; Huang, W.; Du, S.; Fu, Y.; Wang, J. Estimation of fluvo-aquic SOM content from hyperspectral based on continuous wavelet transformation. *Trans. Chin. Soc. Agric. Eng.* **2012**, *28*, 132–139. [CrossRef]
38. Chen, W.Q.; Ran, H.F.; Cao, X.Y.; Wang, J.Z.; Teng, D.X.; Chen, J.; Zheng, X. Estimating PM2.5 with High-Resolution 1-km AOD Data and an Improved Machine Learning Model over Shenzhen, China. *Sci. Total Environ.* **2020**, *746*, 141093–141103. [CrossRef]
39. Shi, Y.; Jin, N.; Ma, X.L.; Wu, B.Y.; He, Q.S.; Yue, C.; Yu, Q. Attribution of climate and human activities to vegetation change in China using machine learning techniques. *Agric. For. Meteorol.* **2020**, *294*, 108146. [CrossRef]
40. Gonzalez, S.; Garcia, S.; Del, S.J.; Rokach, L.; Herrera, F. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Inf. Fusion* **2020**, *64*, 205–237. [CrossRef]
41. Li, Z.; Liu, M.; Ji, S. The Identification of the Origin of Chinese Wolfberry Based on Infrared Spectral Technology and the Artificial Neural Network. *Spectrosc. Spectr. Anal.* **2016**, *36*, 720–723. [CrossRef]
42. Li, P.; Zhu, J.; Ma, Y.; Wu, X. Near Infrared Spectral Analysis Based on Partial Least Squares. *J. Geomat.* **2015**, *40*, 53–57. [CrossRef]
43. Wang, X.M.; Yang, C.; Zhou, M.J. Partial Least Squares Improved Multivariate Adaptive Regression Splines for Visible and Near-Infrared-Based Soil Organic Matter Estimation Considering Spatial Heterogeneity. *Appl. Sci.* **2021**, *11*, 566. [CrossRef]
44. Novacoski, E.J.; Caetano, I.K.; Melquiades, F.L.; Aline, M.G.; Torres, Y.R.; Gonzalez-Borrero, P.P. Spectroscopic based partial least-squares models to estimate soil features. *Microchem. J.* **2022**, *180*, 107617. [CrossRef]
45. Wang, Q.; Chen, B.; Wang, F.; Song, Q.; Dou, Z.; Dai, J.; Xiao, Q.; Yang, X. Inversion for Spatial Distribution Pattern of SOM Based on HJ Image in Oasis Cotton Field. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 174–180. [CrossRef]
46. Xie, X.F.; Wu, T.; Zhu, M.; Jiang, G.J.; Xu, Y.; Wang, X.H.; Pu, L.J. Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. *Ecol. Indic.* **2021**, *120*, 106925. [CrossRef]
47. Azizi, K.; Ayoubi, S.; Nabiollahi, K.; Garosi, Y.; Gislum, R. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. *J. Geochem. Explor.* **2022**, *233*, 106921. [CrossRef]
48. Ewing, J.; Oommen, T.; Jayakumar, P.; Alger, R. Utilizing Hyperspectral Remote Sensing for Soil Gradation. *Remote Sens.* **2020**, *12*, 3312. [CrossRef]
49. Yu, L.; Hong, Y.S.; Zhou, Y.; Zhu, Q. Inversion of SOM Content Using Hyperspectral Data Based on Continuous Wavelet Transformation. *Spectrosc. Spectr. Anal.* **2016**, *36*, 1428–1433. [CrossRef]
50. Chang, R.; Chen, Z.; Wang, D.; Guo, K. Hyperspectral Remote Sensing Inversion and Monitoring of Organic Matter in Black Soil Based on Dynamic Fitness Iner-tia Weight Particle Swarm Optimization Neural Network. *Remote Sens.* **2022**, *14*, 4316. [CrossRef]
51. Du, P.; Kibbe, W.A.; Lin, S.M. Improved Peak Measurement in Mass Spectrum by Incorporating Continuous Wavelet Transform-Based Pattern Matching. *Bioinformatics* **2006**, *22*, 2059–2065. [CrossRef]
52. Guo, H.L.; Zhang, R.R.; Dai, W.H.; Zhou, X.W.; Zhang, D.J.; Yang, Y.H.; Cui, J. Mapping Soil Organic Matter Content Based on Feature Band Selection with ZY1-02D Hyperspectral Satellite Data in the Agricultural Region. *Agronomy* **2022**, *12*, 2111. [CrossRef]
53. Tong, Q.X.; Tian, G.L. *Analysis of Typical Chinese Surface Features and Their Characteristics*; Science Press: Beijing, China, 1999.
54. Silvero, N.E.Q.; Di Raimo, L.A.D.; Pereira, G.S.; de Magalhaes, L.P.; Terra, F.D.; Dassan, M.A.A.; Salazar, D.F.U.; Dematte, J.A.M. Effects of Water, Organic Matter, and Iron Forms in Mid-IR Spectra of Soils: Assessments from Laboratory to Satellite-Simulated Data. *Geoderma* **2020**, *375*, 114480. [CrossRef]
55. Wang, J.; He, T.; Li, Y.H. Studying on Extraction Methods for Land Quality Information Based on Hyperspectral Data. *J. Remote Sens.* **2005**, *9*, 438–445. [CrossRef]
56. Sun, W.; Liu, S.; Zhang, X.; Li, Y. Estimation of soil organic matter content using selected spectral subset of hyperspectral data. *Geoderma* **2022**, *409*, 115653. [CrossRef]
57. Gao, J.; Liang, C.; Shen, G.; Lv, J.; Wu, H. Spectral Characteristics of Dissolved Organic Matter in Various Agricultural Soils Throughout China. *Chemosphere* **2017**, *176*, 108–116. [CrossRef]
58. Wang, H.; Zhang, Z.; Karnieli, A.; Chen, J.; Han, W. Hyperspectral Estimation of Desert SOM Content Based on Gray Correlation-Ridge Regression Model. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 124–131. [CrossRef]
59. Wang, X. Spatial Variation of Organic Matter in Horizontal Zonal Soils in China. *Sci. Geogr. Sin.* **2001**, *21*, 19–23. [CrossRef]
60. Zhao, X.; Yang, M. Hyper-spectral Characteristics of Major Types of Soils in Red Soil Region of Jiangxi Province, China. *Acta Pedol. Sin.* **2018**, *55*, 31–42. [CrossRef]
61. Cheng, H.; Wan, Y.; Chen, Y.; Wan, Q.; Shi, T.; Shen, R.; Guo, K.; Hu, J. Study on the Characteristics and Mechanism of Visible and Near Inf rared Reflectance Spectra of Soil Heavy Metals. *Spectrosc. Spectr. Anal.* **2018**, *38*, 771–778. [CrossRef]
62. Shen, Q.; Zhang, S.; Ge, C.; Liu, H.; Zhou, Y.; Chen, Y.; Hu, Q.; Ye, H.; Huang, Y. Hyperspectral Inversion of Heavy Metal Content in Soils Reconstituted by Mining Wasteland. *Spectrosc. Spectr. Anal.* **2019**, *39*, 1214–1221. [CrossRef]
63. Chen, C.M.; Hall, S.J.; Coward, E.; Thompson, A. Iron-mediated organic matter decomposition in humid soils can counteract protection. *Nat. Commun.* **2020**, *11*, 2255. [CrossRef]

64. Lv, Z.Y.; Wang, F.J.; Cui, G.Q.; Benediktsson, J.A.; Lei, T.; Sun, W.W. Spatial-Spectral Attention Network Guided with Change Magnitude Image for Land Cover Change Detection Using Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
65. Kim, H.; Yu, J.; Wang, L.; Jeong, Y.; Kim, J. Variations in Spectral Signals of Heavy Metal Contamination in Mine Soils Controlled by Mineral Assemblages. *Remote Sens.* **2020**, *12*, 3273. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.