



Article

Mapping Soil Organic Matter and Analyzing the Prediction Accuracy of Typical Cropland Soil Types on the Northern Songnen Plain

Meiwei Zhang^{1,2}, Huanjun Liu^{1,2,*}, Meinan Zhang^{3,4}, Haoxuan Yang⁵, Yuanliang Jin⁶, Yu Han¹, Haitao Tang¹, Xiaohan Zhang¹ and Xinle Zhang¹

¹ Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China; zhangmw2019@gmail.com (M.Z.); daidaidegao123@gmail.com (Y.H.); tang1275715966@gmail.com (H.T.); zxh602742692@gmail.com (X.Z.); xinlezhang@neau.edu.cn (X.Z.)

² School of Public Administration and Law, Northeast Agricultural University, Harbin 150030, China

³ Department of Earth System Science, Tsinghua University, Beijing 100089, China; zhangmn18@mails.tsinghua.edu.cn

⁴ Institute of Forest Ecology, Environment and Nature Conservation, Chinese Academy of Forestry, Beijing 100091, China

⁵ College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China; 2011698@tongji.edu.cn

⁶ School of Environment, Tsinghua University, Beijing 100089, China; jiny118@mails.tsinghua.edu.cn

* Correspondence: liuhuanjun@neigae.ac.cn; Tel.: +86-137-960-751-27



Citation: Zhang, M.; Liu, H.; Zhang, M.; Yang, H.; Jin, Y.; Han, Y.; Tang, H.; Zhang, X.; Zhang, X. Mapping Soil Organic Matter and Analyzing the Prediction Accuracy of Typical Cropland Soil Types on the Northern Songnen Plain. *Remote Sens.* **2021**, *13*, 5162. <https://doi.org/10.3390/rs13245162>

Academic Editor: Eyal Ben-Dor

Received: 29 November 2021

Accepted: 15 December 2021

Published: 19 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Abstract: Soil organic matter (SOM) plays a critical role in agroecosystems and the terrestrial carbon cycle. Thus, accurately mapping SOM promotes sustainable agriculture and estimations of soil carbon pools. However, few studies have analyzed the changing trends in multi-period SOM prediction accuracies for single cropland soil types and mapped their spatial SOM patterns. Using time series 7 MOD09A1 images during the bare soil period, we combined the pixel dates of training samples and precipitation data to explore the variation in SOM accuracy for two typical cropland soil types. The advantage of using single soil type data versus the total dataset was evaluated, and SOM maps were drawn for the northern Songnen Plain. When almost no precipitation occurred on or near the optimal pixel date, the accuracies increased, and vice versa. SOM models of the two soil types achieved a lower root mean squared error ($RMSE = 0.55\%$, 0.79%) and mean absolute error ($MAE = 0.39\%$, 0.58%) and a higher coefficient of determination ($R^2 = 0.65$, 0.75) than the model using the total dataset and resulted in a mean relative improvement (RI) of 30.21%. The SOM decreased from northeast to southwest. The results provide reference data for the accurate management of cultivated soil and determining carbon sequestration.

Keywords: soil organic matter; northern Songnen Plain; single soil type; pixel date; prediction accuracy; MOD09A1

1. Introduction

Soil organic matter (SOM) is a vital component of the soil and contributes to the improvement of soil fertility status [1–4] and increasing grain yields [5–7]. As an important crop production and cultivation base, the northern Songnen Plain plays a valuable role in the sustainable development of China's national economy and food security. However, with the continuous development and utilization of cultivated land resources [8], a series of problems, such as serious soil erosion, degradation of cultivated land quality, and destruction of the agricultural ecological environment, have been caused by various natural factors and human production activities [9,10]. Therefore, accurate regional SOM predictions of target areas are essential for strengthening soil ecological protection and implementing management measures for precision agriculture.

Conventionally, a large number of soil samples are collected during field investigations to conduct analyses, and SOM prediction is both costly and difficult on a large regional scale [11–14]. Geostatistical methods (e.g., kriging and cokriging) [13,15–20] are gradually being applied to improve the prediction accuracy. However, due to the high spatial heterogeneity of soil properties, these methods also require numerous representative sample points to ensure prediction accuracy [19,21–24]. To overcome these pitfalls, SOM prediction using remote-sensing data is a cost-effective way to reduce sampling and analysis budgets in order to predict soil properties and categories over large areas [25,26]. Some regression-based statistical techniques for remote sensing-based SOM prediction have been developed, such as multiple linear regression [16,27–29], stepwise multiple regression [15,30–32], and mixed linear regression [6,33,34]. Therefore, regression models are widely used because of their simple operation mechanism, efficient calculation speed, and the interpretability of their prediction results and input variables [32]. Specifically, stepwise multiple regression provides the advantage of eliminating multicollinearity between input variables [35,36], and as such, it offers an improvement on traditional linear regression.

Extensive research has been performed on SOM prediction based on remote-sensing data, and most of the knowledge of soil spectroscopy has been generated through remote sensing [37]. The relationships between SOM and spectral characteristics have been increasingly studied, and previous studies have shown that SOM has a significant negative correlation with soil reflectance [38,39]. Simultaneously, some results have illustrated that the SOM content is sensitive to the visible near-infrared (VNIR) (400~1100 nm) and shortwave infrared (SWIR) (1100~2500 nm) spectral regions [39–42]; in particular, the VNIR spectral region provides a good alternative for predicting SOM [43–45]. The typical moisture absorption bands of the soil spectrum are located in the SWIR spectral region at 1400 nm, 1900 nm, and 2200 nm [46]. However, the prediction of soil properties using satellite remote-sensing data faces many challenges [37]. The SOM prediction accuracy is susceptible to being influenced by additional environmental factors [13,15,19,25,27,47–51], such as soil type (e.g., soil properties and the number of soil types in the sample dataset), seasonal conditions (e.g., precipitation and snowmelt time), and remote-sensing image features (e.g., pixel data). These factors will impact the soil reflection spectrum, and then the spectral index constructed by them as the model inputs will affect the prediction accuracy of SOM. Notably, one frequently applied approach in the use of remote-sensing images to analyze SOM prediction is to establish an analytic relationship between soil observation data and available variables related to factors that impact the prediction accuracy [26,52]. It is generally believed that the spatial distribution pattern of SOM is controlled by various environmental variables, and the use of auxiliary variables related to SOM spatial analysis can effectively enhance the accuracy [32,53]. Therefore, changes in SOM prediction accuracy may also be affected not only by one factor but also possibly by the combined actions of multiple factors, and effective approaches are needed to reveal the driving factors that impact and improve the SOM prediction accuracy and to map the spatial distribution of SOM more accurately.

With the continuous development of satellite remote-sensing technology, numerous satellite remote-sensing datasets have been applied to establish the prediction models of SOM and map its spatial distribution. The Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance product (MOD09A1) is an 8-day composite dataset, which has been widely used to perform the prediction of soil attributes and map their spatial distribution characteristics [51,54–56]. Because the temporal information from the MOD09A1 image at each pixel is inconsistent [30], each pixel may also have an inconsistent pixel date during the 8-day period. Hence, compared to less processed forms of remotely sensed imagery datasets, each image pixel of the MOD09A1 image contains detailed date information. However, the existing SOM prediction studies in the northern Songnen Plain generally used MOD09A1 images and SOM observation data from multiple soil types as the data source, and integrated precipitation to investigate the changes in SOM prediction accuracy [30,31,57,58]. For instance, Dou et al. [30] and Zhang et al. [31] performed regional

SOM prediction based on stepwise multiple regression and a soil sample dataset combined with multiple soil type data on the Songnen Plain and explored the impact of precipitation on SOM predictions. Zhang et al. [58] also used an SOM observation dataset composed of multiple soil type data and machine learning algorithms to predict the SOM content. However, it should be noted that the above studies both applied MOD09A1 images, but neither of them introduced pixel date or combined precipitation to further explore the comprehensive effects of these factors on SOM prediction. Hence, the MOD09A1 images were used mainly for the following reasons in our study. On the one hand, MODIS images have a high temporal resolution, which is suitable for exploring the trends of multi-period SOM prediction accuracies based on time series images, and these images have been widely used for the spatial prediction of soil properties and to accurately reveal their distribution patterns [30,51,54,55]. On the other hand, we can extract the specific date of the pixel where each training point is located according to the 8-day synthesis interval of each MOD09A1 image. This approach can be applied to comprehensively explore the impact of precipitation integrated with the pixel date of training samples on SOM prediction. Moreover, Meng et al. [59] and Bao et al. [60] separately applied hyperspectral experimental analysis and satellite data to build SOM prediction models in the northern part of the Songnen Plain, and simultaneously, the first study still applied an SOM dataset consisted of various soil type data. Therefore, few studies have focused on the combined impacts of precipitation and the pixel date of training samples on SOM prediction using time series remote sensing images. Meanwhile, the use of data on a single soil type, rather than data combining all soil types in a given region, to enhance the prediction accuracy has not been sufficiently explored, and its advantages need to be further verified.

Consequently, in this study, we consider the potential impact of the pixel date of training samples and precipitation data on SOM prediction accuracy by using time series 7 MOD09A1 images and SOM observation data of two typical cropland soil types over the bare soil period. Additionally, the SOM prediction accuracy of the model using data on a single soil type was compared with the results of the model using data on multiple soil types, and the SOM contents for cultivated lands were mapped in the study area. To compare our findings with other related studies of SOM prediction on the Songnen Plain based on the same algorithm and image conditions, we applied the stepwise multiple regression algorithm based on single soil type data and MOD09A1 images to further integrate the pixel date with precipitation data to analyze the changes in multi-period SOM prediction accuracies, and simultaneously, the advantages of using single soil type data would be revealed. We also compared the model performance with SOM prediction research using other algorithms or hyperspectral data. Our study provides comprehensive and systematic research on SOM prediction on the Songnen Plain, further clarifying the spatial distribution of SOM to help ensure the sustainable development of the soil ecological environment and promote national food security.

The specific objectives are to (1) comprehensively analyze the impacts of the pixel date of training samples and precipitation data on the accuracy of regional SOM prediction based on time series MOD09A1 images; (2) assess the predictive ability of SOM prediction models by using single soil type data; and (3) map the spatial distribution patterns of SOM content in the two typical cultivated lands and the study area of the northern Songnen Plain.

2. Materials and Methods

Figure 1 shows the framework of our study. The four major steps are indicated here: the green panels acquire and process time series MOD09A1 images and precipitation data, thus analyzing the impact of pixel date of training samples and precipitation on SOM prediction; the orange panels collect the SOM observation data and generate the training and validation sample datasets; the gray panels establish SOM prediction models using spectral indices and training datasets for two typical cropland soil types and the total dataset, thereby verifying the advantages of using single soil type data in SOM prediction;

and the purple panel selects the optimal models to map SOM content on the northern Songnen Plain. Notably, the three blue panels show the main objectives of our study.

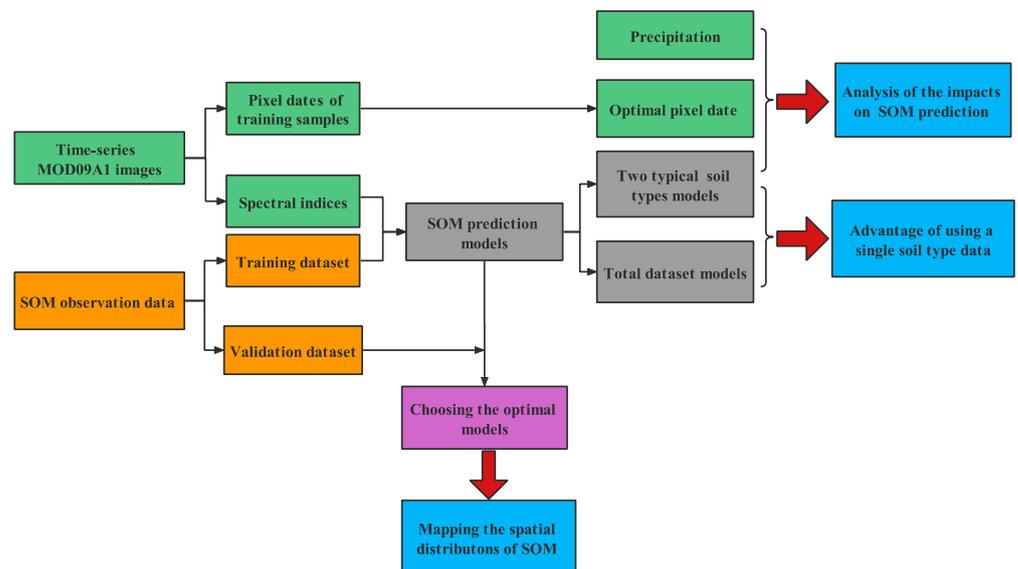


Figure 1. Workflow schematic for analyzing the prediction accuracy of regional soil organic matter (SOM) and mapping its spatial distribution. Notably, the models using two single soil types data (i.e., Arenosols and Phaeozems) are called the “two typical soil types models” and using the total dataset are called the “total dataset models”.

2.1. Study Area

As an important source of national food production, Northeast China is one of the regions with a large area of cropland in China. Our research area is located in the middle of the northern Songnen Plain in Heilongjiang Province; the longitude ranges from 122°E to 128°E, and the latitude ranges from 44°N to 49°N (Figure 2). The climate is dominated by temperate continental semi-arid and semi-humid monsoon conditions, and the annual mean precipitation varies from 400 to 600 mm [61]. The overall topography of the study area is relatively flat. The area also has fertile soils and is suitable for planting various crops, such as soybeans, maize, and rice.

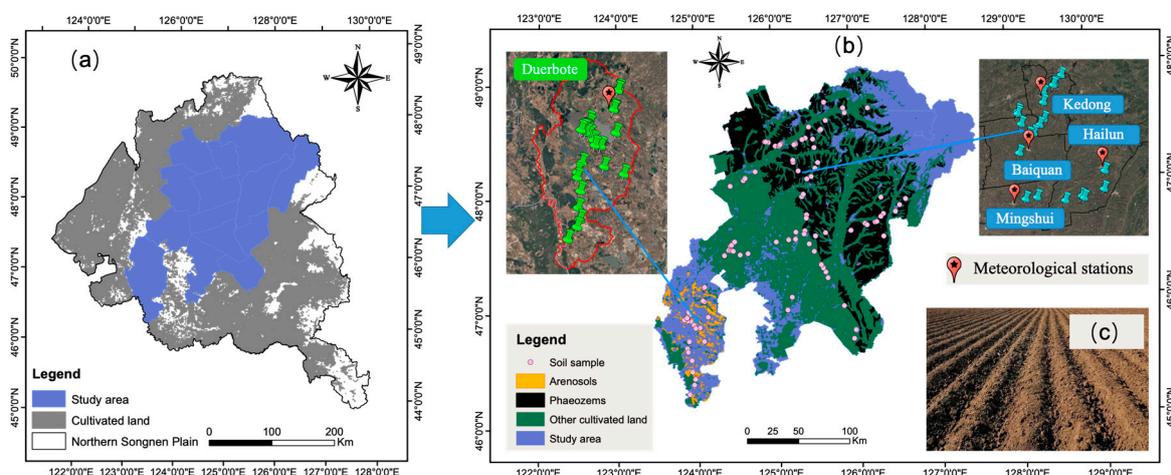


Figure 2. Overview of the northern Songnen Plain and study area (a), soil sampling locations of training samples, and main meteorological stations for Arenosols and Phaeozems (b). Photograph of the soil surface condition after plowing (c).

The northern Songnen Plain is commonly deemed to be an area with typical Mollicsols [62,63]. The two typical soils in the northeastern and southwestern parts of the study area are classified as black and aeolian soils, respectively, which are termed Phaeozems and Arenosols in the World Reference Base for Soil Resources (WRB) [64,65]. The Phaeozem region is a part of the world's three typical black soil belts with high SOM content and good water-holding capacity, and the Arenosols region has a relatively low SOM content and relatively poor water-holding capacity. The long-term management model of intensive agricultural production and the application of fertilizers and pesticides [8] have caused high spatial heterogeneity of soil nutrients [49], serious soil erosion, and poor soil fertility [9,10,66–68]. Thus, with the continuous decline in soil fertility in recent years, more studies should focus on assessing the regional changes in SOM and its influencing factors and on detecting the effectiveness of soil improvement measures. Therefore, the Phaeozem and Arenosols regions were selected as two typical research areas in our study, and the spatial distributions of SOM in these two regions were revealed with the hope of promoting an in-depth understanding of soil degradation and rational use and promoting reasonable agricultural protection measures in the target regions.

2.2. Data

2.2.1. Soil Sample Collection and Treatment

Field investigations were conducted in May in the cultivated areas during the bare soil period in the study area. Most farmers use the incineration method to eliminate crop straw, leading to almost no residue on the soil surface. This practice is usually conducted from late March to early April, and the soil surface remains exposed thereafter (Figure 2c). Hence, our study area has a unique period of bare soil between April and May; this period is characterized by no snow cover and almost no residues or vegetation on the cultivated soil surface and is regarded as the bare soil period [30,69]. Moreover, the bare soil period experiences less rainfall than that occurring from June to September during the rainy season [8,70], and simultaneously, the snow water remaining on the topsoil layer gradually evaporates. Therefore, the bare soil period not only has better soil surface conditions but also relatively dry topsoil.

Soil samples were collected from the 0–20 cm soil layer. To ensure that the SOM content of the sampling points was highly representative, a relatively wide plain cultivated area was selected for collection. We also confirmed in advance whether there would be precipitation on or near the sampling date to reduce the sampling error. The specific sample collection process was as follows. First, the Second National Soil Survey map was adopted, which is produced based on a large number of soil sample measured data, incorporating the advantages of the global soil classification system and with the specific distribution situation of soil types in China to assist in selecting the location of sample points. The map not only retains the information on soil genetics but also uses the diagnostic features of the diagnostic layer to classify the soils, formulating a relatively accurate soil classification system; thus, it is suitable for accurately collecting soil sample datasets of two typical cropland soil types in our research. To ensure that the sampling points could evenly and reasonably cover the entire study area and the areas of the target soil types, we also fully considered the spatial distributions of soil types in the study area and the heterogeneity of the soil surface. In particular, we made great efforts to ensure that each sampling site was not located at the junction of soil types. Then, each sampling point was obtained by mixing samples from five to six randomly selected subsample points within an area of 500 m by 500 m, which can characterize the average level of SOM content at a regional scale. Ultimately, the geographic locations of sampling points were recorded with a global positioning system (GPS, Beijing UniStrong Science and Technology Limited Company, China). In total, we obtained 160 soil samples from cultivated land types, including 40 Phaeozems and 39 Arenosols. The soil samples were air-dried and passed through sizes of ≤ 2 -mm mesh [71] and were then analyzed for SOM content using the potassium dichromate volumetric method [72].

2.2.2. Satellite Image Data Selection

Time series MOD09A1 images in 2018 over the bare soil period (Table 1) were selected from the Google Earth Engine (GEE) data pool (<https://code.earthengine.google.com/>, accessed on 27 September 2019). Using satellite images acquired over the bare soil period can effectively reduce negative effects [73,74], which is favorable for capturing more accurate spectral reflectance of the cultivated soil surface and offering better conditions to implement SOM predictions [30,59,60]. The MOD09A1 images provide 500-m resolution MODIS band 1~7 surface reflectance data, including ρ_1 (red band, 620~670 nm), ρ_2 (near-infrared band, 841~876 nm), ρ_3 (blue band, 459~479 nm), ρ_4 (green band, 545~565 nm), ρ_5 (mid-infrared band, 1230~1250 nm), ρ_6 (SWIR-1 band, 1628~1652 nm), and ρ_7 (SWIR-2 band, 2105~2155 nm). Here, the day of year (DOY) represents the beginning date of the 8-day period.

Table 1. Dates and periods of 7 MOD09A1 images.

Item	DOY						
	089	097	105	113	121	129	137
Image date	3/30	4/07	4/15	4/23	5/01	5/09	5/17
DOY period	089~096	097~104	105~112	113~120	121~128	129~136	137~144

2.2.3. Optimal Pixel Date and Precipitation Calculation

Since the 8-day synthesis interval of each image can provide the specific date of the pixel where each training point is located. Then, we calculated the statistics on the pixel dates of training samples within the 8-day period of each image. Ultimately, we defined the pixel date with the largest number of training samples as the optimal pixel date for each image period. As shown in Table 2, the optimal pixel dates were extracted from 7 image periods for the Phaeozems and Arenosols.

Table 2. Optimal pixel dates of seven image periods for the Phaeozems and Arenosols.

Item	Soil Type	
	Phaeozems	Arenosols
	Optimal Pixel Date	Optimal Pixel Date
DOY		
089	094 d	094 d
097	099 d	101 d
105	112 d	112 d
113	115 d	117 d
121	128 d	128 d
129	135 d	129 d
137	139 d	144 d

Meanwhile, according to the distribution regions of the training samples, we selected the Duerbote meteorological station for the Arenosols region and the Kedong, Baiquan, Mingshui, and Hailun stations for the Phaeozems region and then obtained precipitation data from the regional meteorological stations at 20~20 h daily [30,31]. Finally, we calculated the cumulative daily precipitation in the Phaeozems and Arenosols areas during the research period (see Figure 3 for details).

2.3. Construction of Spectral Indices

Studies have indicated that the spectral indicator can be applied to characterize SOM information [30–32,57] using band optimization algorithms [75,76], and the results confirmed that the difference index, ratio index, and normalized difference index are helpful to predict SOM accurately during the bare soil period [30,58,60,77,78]. Compared with the one-dimensional spectral band, the use of spectral indicators highlights the advantages

of enhancing the correlation between the SOM content and spectral indicators, providing more spectral information [79,80], and reducing the reflection spectrum error caused by terrain and atmospheric conditions [30,81,82], therefore improving the SOM prediction accuracy. Three types of spectral index were calculated, including the difference index (D_{xy}), ratio index (R_{xy}), and normalized difference index (ND_{xy}). The above indices were calculated based on the following equations:

$$D_{xy} = \rho_x - \rho_y \quad (1)$$

$$R_{xy} = \rho_x / \rho_y \quad (2)$$

$$ND_{xy} = (\rho_x - \rho_y) / (\rho_x + \rho_y) \quad (3)$$

where ρ_x , ρ_y and ρ_z represent the band x , y and z values, respectively; D_{xy} represents the difference in reflectance between ρ_x and ρ_y ; R_{xy} represents the ratio in reflectance between ρ_x and ρ_y ; and ND_{xy} represents the normalized difference in reflectance between ρ_x and ρ_y .

Ultimately, we applied the reflectance values of seven bands, ρ_1 , ρ_2 , ρ_3 , ρ_4 , ρ_5 , ρ_6 , and ρ_7 , as the base predictors and performed the above three mathematical transformations to calculate remote sensing indices in our study. The input variable dataset included seven bands $\rho_1 \sim \rho_7$, 21 difference indices, 21 ratio indices, and 21 normalized difference indices; thus, we constructed a total of 70 indices.

2.4. Prediction Method and Mapping of Soil Organic Matter (SOM)

Stepwise multiple regression is an effective statistical method for remote sensing-based SOM prediction and is usually based on the correlation coefficients between the input variables and SOM content to assess the importance of the input variables and test their statistical significance [30,83]. The method is applied to recheck the input variables from the previous steps at each step in the calculation process; that is, the variables previously entered into the model may become redundant variables during the later stages because of the relationships with other variables that are added to the model later [84,85], and the newly added variables would reduce the contribution value of previously determined input variables to the model [30]. The algorithm starts by selecting the variable with the highest correlation with SOM as the first input variable. According to the t -test for the regression coefficient, if the performance is significant, the variable is retained to construct the single-variable model. Then, the secondary input variable is selected based on the partial correlation coefficient to build the binary model [83]. Therefore, the stepwise multiple regression algorithm is similar to the implementing feature selection procedures because it selects only the most important variable to achieve maximum predictive power. To ensure that all the models are comparable, we adopted one input variable with the highest correlation with SOM content for each stepwise multiple regression model. Since the binary model generally obtains higher prediction accuracy than the single-variable model [30,31], we attempted to reserve two input variables for the models with the highest accuracy of two typical cropland soil types and the total dataset to improve the SOM prediction accuracy. Hence, single-variable and binary models were constructed by programming them in IBM SPSS Statistics 22 software.

To assess the model accuracy, we randomly selected the training and validation sample datasets at a 1:1 ratio [30,31] for the sample datasets. The training dataset was designated to train the models, and the validation sample dataset was reserved for validation purposes, which is consistent with previous literature [15,30,31,49,58–60,86,87]. We adopted the root mean square error ($RMSE$), mean absolute error (MAE) [51,88], and coefficient of determination (R^2) [88,89] to evaluate the model performance. The $RMSE$ and MAE assess the prediction accuracy, with a lower value denoting a higher prediction accuracy, and R^2 evaluates the model stability, with a higher value denoting higher stability. Moreover, we also applied the relative improvement (RI) [90,91] in $RMSE$ to measure the improvement

in prediction accuracy. These statistical indicators for assessing the model performance were expressed by the following equations:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{n}} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - o_i| \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (6)$$

$$RI_X = \frac{RMSE_{x1} - RMSE_{x2}}{RMSE_{x1}} \times 100\% \quad (7)$$

where p_i and o_i refer to the predicted and observed SOM values, respectively; n refers to the number of soil samples; \bar{o} refers to the mean of the observed SOM values; and $RMSE_{x1}$ and $RMSE_{x2}$ are the $RMSE$ values of the models using the total dataset and single soil type data, respectively.

Finally, the three optimal models with the highest prediction accuracy and adaptability for two typical cropland soil types and the total dataset were used to map SOM spatial distributions in the ArcGIS 10.2 platform.

3. Results and Analysis

3.1. Descriptive Statistics of the SOM Content

Table 3 shows the descriptive statistics on the SOM contents for the total, Phaeozem, and Arenosol datasets and their training and verification sample datasets by using IBM SPSS Statistics 22. The measured SOM content of the total dataset varied from 0.64 to 8.21%, and the mean content was 3.86%. The SD represents the standard deviation of the SOM contents. Notably, a larger SD means that the SOM content dataset is more discrete and unstable, and a smaller SD indicates that it is less discrete and closer to the mean. The results indicated that the Arenosol dataset had the lowest SD value, followed by the Phaeozem dataset, whereas the total dataset had the highest value. Moreover, the training and validation datasets of the three sample datasets separately had similar range, mean, and SD values; in particular, similar SD values indicate that SOM content datasets have similar dispersion and stability. Hence, the training and validation datasets randomly assigned for the three soil sample datasets are all highly representative.

Table 3. Descriptive statistics on the SOM contents for the total, Phaeozem, and Arenosol datasets and their training and validation sample datasets.

Soil Type	SOM				
	Minimum %	Maximum %	Range %	SD %	Mean %
Total dataset	0.64	8.21	7.57	1.48	3.86
Training dataset	0.64	8.21	7.57	1.53	3.93
Validation dataset	0.83	7.39	6.56	1.43	3.80
Phaeozems	3.46	8.21	4.75	1.35	5.40
Training dataset	3.46	8.21	4.75	1.47	5.20
Validation dataset	3.92	8.21	4.29	1.26	5.56
Arenosols	0.64	3.73	3.09	0.74	2.20
Training dataset	0.64	3.73	3.09	0.81	2.26
Validation dataset	0.82	3.55	2.73	0.70	2.16

3.2. SOM Prediction Using Single Soil Type Data

Tables 4 and 5 show the SOM prediction models for Phaeozems and Arenosols, respectively, based on 7 MOD09A1 images. Figure 3 presents the *RMSE* series corresponding to the optimal pixel dates of different DOYs and daily precipitation during the research period. We found that the variation trends in SOM prediction accuracy were similar for Phaeozems (Figure 3a) and Arenosols (Figure 3b), and the change trends in precipitation near the optimal pixel date were also similar. It is clearly seen that the variation trends in accuracies during the image period from DOY 97~121 are consistent for the two soil types. A large amount of continuous precipitation occurred near the optimal pixel date of DOY 105, the *RMSE* values increased, and the accuracies decreased significantly. Subsequently, a small amount of intermittent precipitation occurred during 112~128 d, and its date was not close to the optimal pixel dates of DOY 113~121; therefore, the soil moisture gradually decreased, the *RMSE* values decreased, and the accuracies showed an obvious increasing trend. There was a one-day interval (from 128 d to 129 d) for the optimal pixel dates of DOYs 121 and 129 for Arenosols, and the accuracies showed almost no change; however, the precipitation time series occurred during 128~139 d in the Phaeozem regions, and the accuracy decreased markedly. On DOY 137, continuous precipitation occurred near the optimal pixel date in these two regions, and the accuracy showed a continuous downward trend. Specifically, from the DOY 089 to 097 images, precipitation occurred before the optimal pixel date (101 d) in Arenosols, the *RMSE* value increased, and the accuracy decreased. However, almost no precipitation occurred during 089~099 d in the Phaeozem region, the accuracy showed an upward trend due to the strong water-holding capacity of the Phaeozems, and snow water remained in the topsoil layer at the beginning of April. As the snow water gradually evaporated, the soil moisture gradually decreased, and the *RMSE* value decreased from DOY 089~097 for the Phaeozems.

Table 4. Results of SOM prediction models for Phaeozems based on seven images.

Image DOY	Optimal Pixel Date	Input Variable	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²
089	094	ρ_3	1.07	0.81	0.54
097	099	R_{62}	0.98	0.88	0.60
105	112	R_{72}	1.12	1.00	0.45
113	115	D_{23}	1.07	0.81	0.65
121	128	ρ_3	0.86	0.69	0.65
129	135	ρ_3, R_{61}	0.79	0.58	0.75
137	139	R_{64}	0.99	0.74	0.38
		R_{61}	1.05	0.87	0.30

Table 5. Results of SOM prediction models for Arenosols based on seven images.

Image DOY	Optimal Pixel Date	Input Variable	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²
089	094	R_{43}	0.65	0.56	0.63
097	101	R_{23}	0.76	0.64	0.57
105	112	R_{63}	0.94	0.74	0.25
113	117	ρ_4	0.81	0.67	0.50
121	128	$ND\rho_{13}$	0.63	0.47	0.54
129	129	ρ_4	0.62	0.52	0.53
137	144	ρ_4, D_{52}	0.55	0.39	0.65
		R_{63}	0.90	0.49	0.63

With respect to the model input variables, when the optimal pixel date and its neighboring dates did not experience precipitation, the main input variables were generally VNIR bands or the spectral indices constructed by these bands and represented SOM information, such as on DOYs 113 and 121 for both soil types. However, when continuous precipitation occurred near the optimal pixel date, the inputs were mainly spectral indices composed of the SWIR bands and VNIR bands and characterized the impact of precipi-

tation, such as DOY 105 for both soil types and DOY 129 for the Phaeozems. In addition, the secondary input variables of binary models included the ρ_5 , ρ_6 , and ρ_7 bands, which primarily represented soil moisture information.

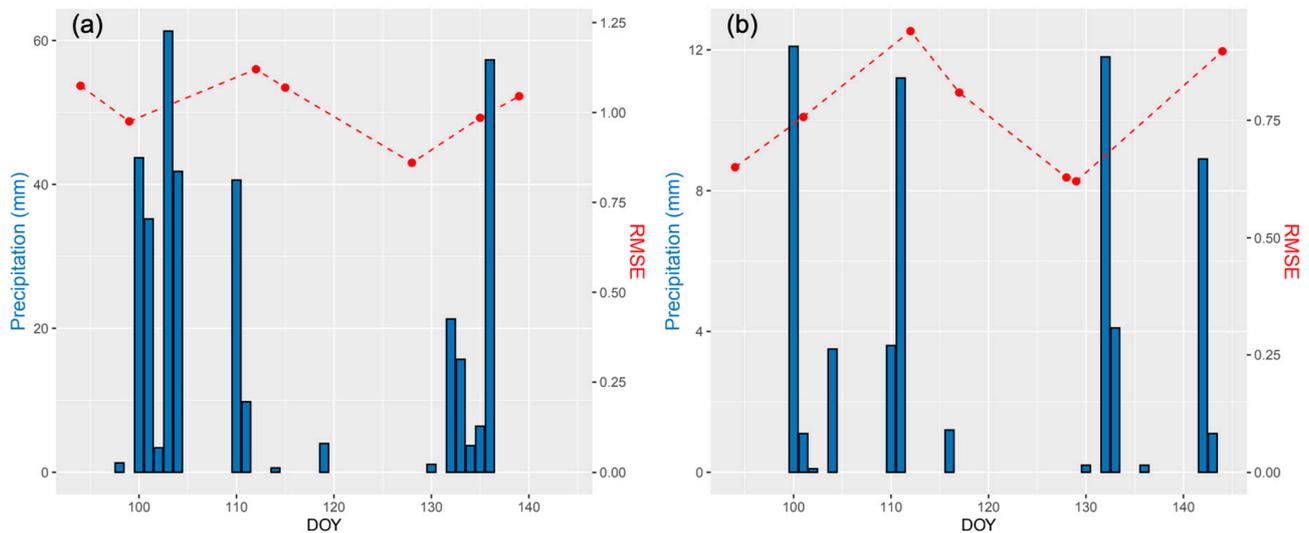


Figure 3. Time series of precipitation and root mean squared error (*RMSE*) on different days of the year (*DOYs*) for Phaeozems (a) and Arenosols (b). The *x*-axis values corresponding to the red dots represent the optimal pixel dates of training samples based on 7 images. The histogram values denote the precipitation on different dates. Notably, a smaller *RMSE* value indicates a higher prediction accuracy, whereas a larger *RMSE* value indicates a lower prediction accuracy.

3.3. SOM Prediction Using the Total Dataset

In comparison to the prediction accuracies using single soil type data, Table 6 shows the SOM prediction results using the total dataset mixed with multiple soil types. Taking the binary models of *DOYs* 121 and 129 as examples, which had the highest accuracies for Phaeozems and Arenosols, the *RMSE* values of the SOM prediction models were 0.79% and 0.55%, respectively; however, the *RMSE* for the total dataset on *DOY* 113 was 0.96%. Therefore, the model performance for single soil type performed better than that for the total dataset. Simultaneously, the application of single Phaeozem and Arenosol soil data resulted in *RIs* of 17.71% and 42.71%, respectively, with a mean value of 30.21%. In addition, Tables 4–6 show that the ratio spectral index was generally screened to characterize the impact of precipitation on SOM prediction, especially R_{61} , which was significantly correlated with SOM (Table 7).

Table 6. Results of SOM prediction using the total dataset based on 7 images.

<i>DOY</i>	Input Variables	<i>RMSE</i>	<i>MAE</i>	R^2
89	ρ_3	0.99	0.82	0.63
97	R_{61}	1.08	0.43	0.55
105	R_{61}	1.00	0.81	0.56
	ρ_1	1.00	0.76	0.49
113	ρ_1, R_{51}	0.96	0.77	0.62
121	D_{43}	1.10	0.77	0.39
129	R_{61}	1.04	0.79	0.51
137	R_{61}	1.17	0.88	0.47

3.4. Selecting the Optimal Models of SOM Prediction

According to the results of Sections 3.2 and 3.3, the binary models for *DOYs* 113, 121, and 129 (Table 8) obtained the highest prediction accuracy and adaptability for the total sample dataset, Phaeozem, and Arenosol, which were selected as three optimal

models to draw the spatial distribution patterns in the SOM contents. As shown in Table 8, we found that the main inputs of the three optimal models were all the VNIR bands, which characterized SOM content information, and the secondary input variables were the spectral indices composed of the VNIR and SWIR bands and further characterized and reduced the influence of soil moisture on SOM prediction.

Table 7. Correlation coefficients between SOM and main input variables.

DOY	Main Input Variables				
	R_{61}	ρ_3	ρ_1	D_{43}	R_{51}
089	0.70 **	−0.80 **	−0.78 **	−0.71 **	0.48 **
097	0.74 **	−0.72 **	−0.72 **	−0.64 **	0.34 **
105	0.75 **	−0.71 **	−0.69 **	−0.66 **	0.60 **
113	0.62 **	−0.70 **	−0.70 **	−0.66 **	0.65 **
121	0.48 **	−0.52 **	−0.58 **	−0.62 **	0.39 **
129	0.71 **	−0.61 **	−0.64 **	−0.63 **	0.52 **
137	0.68 **	−0.59 **	−0.59 **	−0.59 **	0.58 **

Note: “**” represents the level of statistical significance at $p < 0.01$.

Table 8. Optimal models for the total dataset, Phaeozem, and Arenosol.

Soil Type	Input Variables	Model
Total dataset	ρ_1, R_{51}	$SOM = 0.482 - 0.002 \times \rho_1 + 2.421 \times R_{51}$
Phaeozems	ρ_3, R_{61}	$SOM = -16.804 + 0.031 \times \rho_3 + 2.395 \times R_{61}$
Arenosols	ρ_4, D_{52}	$SOM = 6.448 - 0.005 \times \rho_4 + 0.002 \times D_{52}$

To test whether the established regression model is suitable for SOM mapping, the spatial distribution patterns of residual values were produced for three datasets (Figure 4). To observe the spatial distribution characteristics of different residual levels clearly, we divided the randomly distributed residual values into seven levels and assigned different colors to indicate the degrees of difference in the values. As shown in Figure 4, the residual points of different numerical levels showed random and irregular distributions in the study area; in particular, there was no aggregation distribution that was too high or too low in a specific area. Therefore, the three SOM prediction models we selected are suitable and can be used to relatively accurately map the spatial distribution patterns of SOM content in the study area.

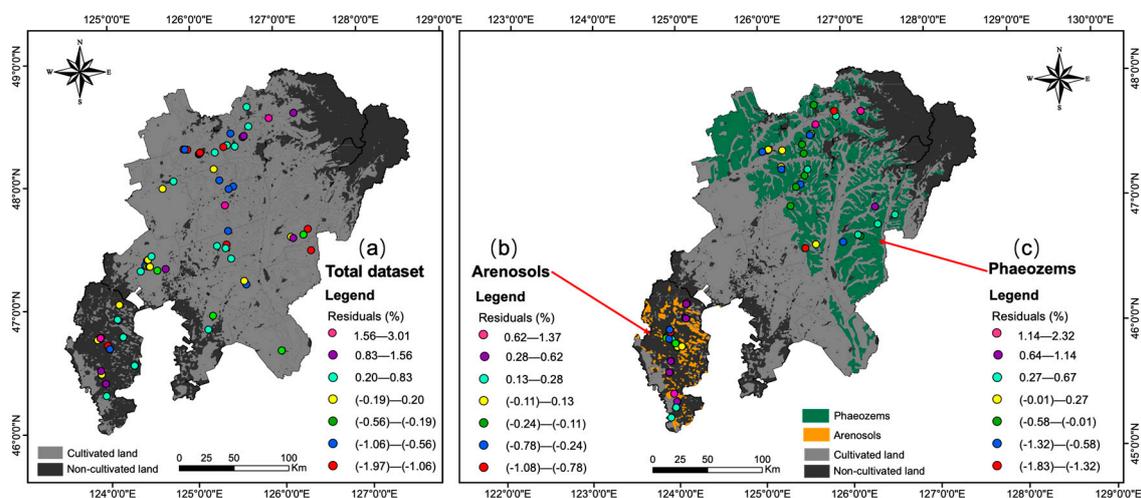


Figure 4. Spatial distribution patterns of residual values for the total (a), Arenosol (b), and Phaeozem (c) sample datasets.

3.5. Mapping the Spatial Distribution of SOM

We selected the three optimal models to map the SOM of the cultivated land in the study area (Figure 5a) and Arenosol and Phaeozem regions (Figure 5b) by using the total and two typical cropland soil type datasets, respectively. The Phaeozem areas were located in the northeastern part of the study area, and they were affected by a cold environment and higher latitudes characterized by rich SOM content, whereas Arenosols were distributed in the southwest with low SOM. Hence, consistent with previous SOM predictions for the Songnen Plain [30,31,43,58–60], the SOM content generally displayed a trend of being high in the northeast and low in the southwest. Simultaneously, comparing Figure 5a,b, we can find that the SOM spatial distribution characteristics of the corresponding Arenosol and Phaeozem regions in the two figures are highly consistent. Therefore, the results further prove the effectiveness of the optimal models based on our study, which can be applied to accurately reveal the spatial patterns of SOM content in the study area and the two typical soil regions.

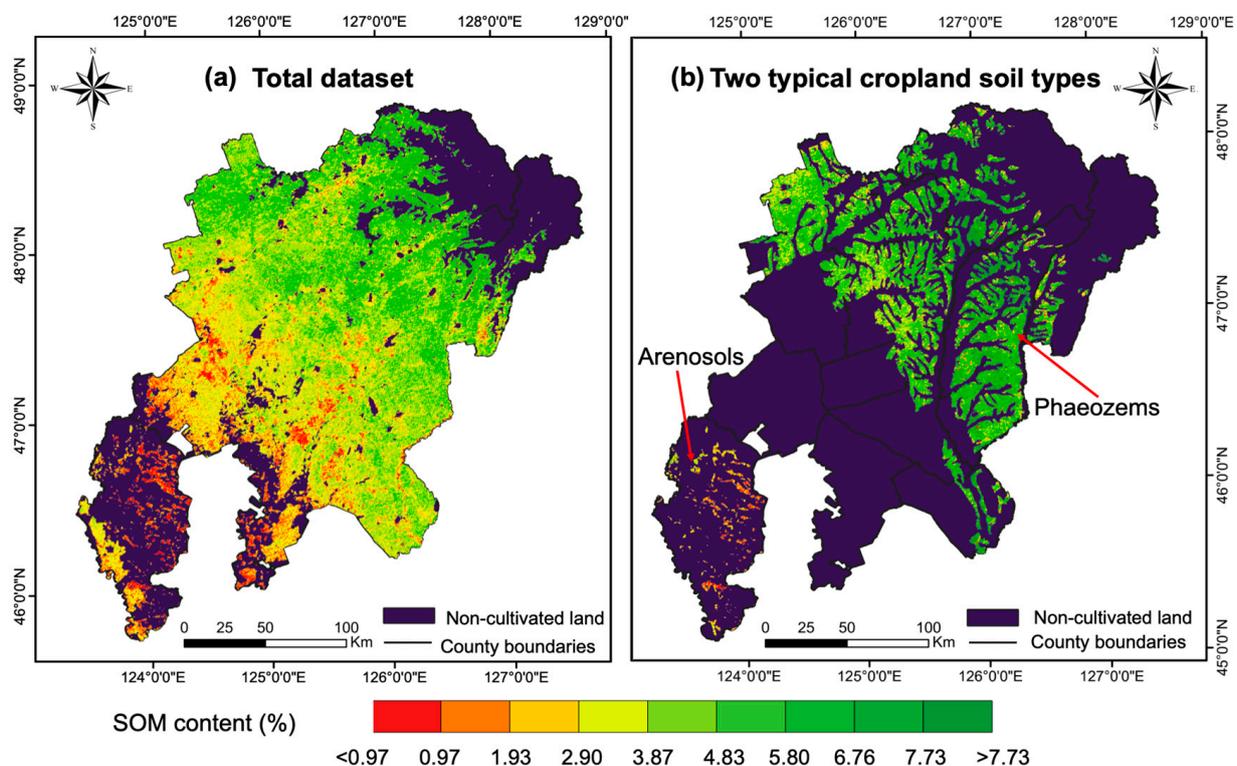


Figure 5. Maps of SOM content in the study area using the total dataset (a) and two typical cropland soil types (i.e., Arenosols and Phaeozems) (b).

4. Discussion

4.1. Impacts of Pixel Date and Precipitation on Prediction Accuracy

Our study comprehensively considered the impacts of pixel date and precipitation on the SOM prediction accuracy. Precipitation is an important source of soil moisture, and it can affect the soil surface water content and, in turn, impact the soil reflectance spectra [92–94]. Over the bare soil period in particular, precipitation is the main source of soil moisture for the topsoil layer. Many researchers have reported the influence of soil moisture on the VNIR reflectance spectrum [93,95–98]. Numerous studies also conducted with various remote-sensing data have demonstrated that the soil moisture content is the main limiting factor related to the soil when explaining the low performance of remote sensing-based SOM prediction models [30,31,99,100]. Therefore, the variability in soil moisture derived from the changes in precipitation can seriously impact the accuracy of

the SOM prediction model [30,101]. However, an important influencing factor that is rarely considered or discussed is the pixel date, which should receive more attention when analyzing the accuracy of SOM prediction. Our results showed that the variation trends in the SOM prediction accuracies of Phaeozems and Arenosols were similar due to the occurrence of similar variation trends in precipitation near the optimal pixel dates. The synergistic effect of the precipitation and pixel date primarily determined the model accuracy and its change trend. We detected that when almost no precipitation occurred on the optimal pixel date and its neighboring days, the prediction models generally achieved a higher accuracy, and the accuracies exhibited an increasing trend. However, continuous precipitation occurred near the optimal pixel date, especially on the optimal date itself, which resulted in a significant decrease in prediction accuracy, such as DOYs 121~129 for Phaeozems and DOYs 089~097 for Arenosols. Meanwhile, when a small amount of intermittent precipitation occurred after continuous precipitation and was not close to the optimal pixel dates, the accuracy gradually increased with a gradual decrease in soil moisture, and as such, the changes in accuracy presented an increasing trend from DOY 105 to DOY 121 for the Phaeozem and Arenosols. Moreover, due to the snowmelt time and good water-holding capacity, the topsoil surface of Phaeozems in early April was rich in residual snow water, which may lead to an increasing trend of SOM prediction accuracy between DOY 089 and 097 with the gradual evaporation of snow water.

For the model input variables, the spectral index formed by the ratio of the VNIR and SWIR bands can reduce the impact of precipitation on SOM prediction. In the three optimal models with highest model performance, the main input variables were the VNIR bands (i.e., SOM-sensitive bands), which were closely correlated with SOM content; the secondary input variables were the spectral indices formed by the mathematical transformation between the SWIR bands (i.e., moisture-sensitive bands) and SOM-sensitive bands and characterized soil moisture and decreased the influence of moisture changes on prediction accuracy. We also found that the model input variables underwent consistent changes with the impacts of the optimal pixel date and precipitation. When almost no precipitation occurred on or near the optimal pixel date, the main inputs were generally SOM-sensitive bands or spectral indices construed by these bands to characterize SOM information, such as DOYs 113 and 121 for both soil types. In contrast, the model inputs were mostly formed by moisture-sensitive and SOM-sensitive bands to characterize the effect of precipitation, such as DOYs 105 and 129 for Phaeozems and DOY 097 for Arenosols. Furthermore, soil texture affects many soil types of soil-forming processes and soil properties [15]. The soil texture of Arenosols is finer than that of Phaeozems. Previous research results showed that the heavier the soil texture, the higher the SOM content [15,102]; therefore, the SOM of Phaeozems is obviously higher than that of Arenosols. In particular, the soil moisture retention capacity is closely related to SOM depending on different soil textures [103–105]; that is, the higher the SOM content of soils, the stronger the water storage capacity [106,107], which makes it easier for the Phaeozems to store soil moisture. In our study, we detected that the inputs of Phaeozems more effectively characterized precipitation, whereas the Arenosols were more easily characterized by SOM due to the difference in water-holding capacity. For example, due to the poor water-holding capacity of Arenosols, the model input of DOY 097 primarily correlated SOM with bare soil characteristics; however, with the higher precipitation that occurred during the later stage, the input of DOY 113 characterized the impact of precipitation information. Moreover, the ratio spectral index, especially for R_{61} , was generally correlated with soil moisture and reduced the negative impact of precipitation on SOM prediction, which is consistent with the literature [30,31,43].

4.2. Comparison of Soil Type Impact on SOM Prediction

The model performance using single soil type data was compared with that using the total dataset, and the model using data on single soil type was better than that based on the total dataset. Although numerous scholars have studied how soil categorical factors can improve the accuracy of SOM prediction, soil type is considered less often [15,108]. Our

study used single soil type data to explore the possibility of improving the SOM prediction accuracy and obtained better model performance than has been reported in previous studies on the Songnen Plain. For example, Dou et al. [30] and Zhang et al. [31] used soil sample datasets that combined multiple soil types data and a MODIS image dataset as the data sources to predict regional SOM on the Songnen Plain; however, the *RMSE* value (0.81%) was higher than that of our research (0.55%). Compared with the results reported by Zhang et al. [58], who used the random forest algorithm and an SOM dataset composed of multiple soil type data, our research also achieved a slightly lower *RMSE* value using stepwise multiple regression and single soil type data. Simultaneously, the superiority of using single soil type data has also been verified in other SOM predictions. Bao et al. [59] used hyperspectral experimental analysis data and a competitive adaptive re-weighted sampling method to compare the accuracies of SOM prediction under various soil type grouping strategies and ultimately confirmed the advantage of soil classification prediction.

On the one hand, the spatial heterogeneity of soil properties can impact the prediction of soil attributes [109], and soil properties in different soil types play an important role in understanding soil moisture dynamics [110]. Especially in a large region with large spatial variation in soil moisture, soil moisture has a considerable influence on the model performance of SOM prediction. The total dataset was composed of complex soil types formed of different parent materials, resulting in a higher heterogeneity in soils and greater spatial variation in soil moisture. On the other hand, the distributions of the sampling points were relatively concentrated, and the *SD* values of the sampling dataset were small for single soil types; hence, the spatial variation in soil moisture resulting from differences in pixel temporal information was also smaller for two typical cropland soil types. However, the whole study area had a larger area and precipitation, and the sampling points of the total dataset were widely distributed, resulting in the greater spatial variability of pixel temporal information and soil moisture; therefore, its prediction accuracy was lower than that of single soil types. Furthermore, due to the long-term extensive utilization of cultivated land, which leads to serious soil erosion and complex and changeable features of surface soil, adjacent soil types are prone to proximity effects [111,112]. In particular, multiple soil types are cross-distributed without clear boundaries in our study area. For instance, meadow soils are widely interspersed among various soil types and readily form slightly convex topography conditions and complex features of surface soil under the action of external factors. As a result, the adjacent Arenosol and Phaeozem soils would have spectral characteristics similar to those of meadow soil. However, our research directly used a single soil type for soil sampling points and ensured that the sampling locations were not at the junction of soil types, allowing us to obtain more realistic soil spectral features of the Arenosols and Phaeozems and reduce the soil spectral similarity between adjacent soil types, thus more accurately mapping the SOM spatial distribution characteristics of the target areas.

Therefore, using sampling points of single soil type data to predict SOM for the two typical cropland soil types can effectively reduce the negative effects derived from the spatial heterogeneity in soil properties and the spatial variations in pixel temporal information and soil moisture and can avoid the proximity effect of adjacent soil types to achieve higher accuracy.

4.3. Limitations and Future Research

There are some limitations to our study. First, the bare soil period and MODIS imagery dataset were adopted as our research period and remote sensing data source. The bare soil period can reduce interference and promote real spectral features of the soil surface reflectance and more accurately map topsoil attributes [73,74]; however, some areas may not have bare soil periods, or the periods might be short due to the different seasonal tillage and continuous cropping systems. Thus, it remains to be verified whether the SOM prediction model established or the analysis methods introduced in our study can be applied to other areas. Simultaneously, the MOD09A1 image was used in our

study, which has the characteristics of high temporal resolution and spatially seamless and multitemporal series characteristics and has been widely applied to predict soil attributes [51,54–56]. However, MODIS is affected by the medium spatial resolution and heterogeneity of image pixels [113,114], thereby limiting the ability to enhance the precision of SOM predictions [58]. Hence, future researchers should attempt to apply remote sensing images with high spatial-temporal resolution (e.g., Sentinel-2 images) [58,60,89,115–117] or other composite and fusion images [51,74,118,119] to reduce the uncertainty of SOM spatial analysis. Second, our research is a simple case study of introducing the pixel date into SOM prediction. Further exploration can determine the pixel dates of the sampling points in advance to screen out the sample points with the same pixel date and ensure that no precipitation occurs on those dates, thus revealing the accuracy changes when the sample point data of the optimal pixel date are exclusively used to construct the SOM prediction model. In this way, the influence of soil moisture and spatial heterogeneity in pixel dates can be effectively reduced, and the SOM prediction accuracy may be significantly improved. Moreover, the stepwise multiple regression algorithm was used in our research. However, attempts should be made to use other algorithms to achieve better model performance for SOM prediction in future studies, such as partial least squares regression (PLSR) [1,120–122], machine learning algorithms (e.g., cubist algorithm) [87,91,123], and hybrid algorithms (e.g., random forest-kriging) [13,124–126]. These algorithms have been applied to predicting SOM content, and integrating different algorithms usually enhances SOM prediction accuracy. Finally, combining various promising algorithms, the spatial information on other environmental factors (e.g., soil surface roughness) [127–129] and agricultural management practices (e.g., no-tillage) [51,130,131] can be incorporated into SOM prediction to determine the decision variables and analyze the driving factors that lead to the differences in prediction accuracy, thereby enhancing the robustness and generalization ability of the model [50]. In summary, determining the method to weaken the negative effects, such as precipitation, on SOM prediction at a large scale to improve the prediction accuracy is the main topic for future research.

4.4. Research Innovations and Implications

First, our study is the first attempt to introduce pixel date using time series MOD09A1 images integrated with regional precipitation to analyze their impact on SOM prediction. The theoretical basis of this analysis method is the different effects of soil properties (e.g., SOM) and environmental factors (e.g., precipitation) on soil reflectivity. The spectral reflectance generally preserves a decreasing tendency with increasing SOM content [132,133], especially for the VNIR bands. The SOM content can be reflected by the differences in soil reflectance spectral characteristics [75,76,133]. However, the soil reflectivity usually increases with increasing soil moisture [92,100,134]. The three typical moisture absorption bands of the soil spectrum correspond to ρ_5 , ρ_6 , and ρ_7 , which can characterize soil moisture information derived from the changes in precipitation; thus, the variations in spectral curves in this spectral region will present a greater uncertainty with increasing SOM content [30,31]. Our results demonstrated that the analytical approach can clarify the changes in prediction accuracies and their trend over the bare soil period, which is conducive to revealing the driving factors that impact the accuracies of SOM prediction and to investigating reasonable approaches to enhancing the prediction accuracy. Second, our results illustrated that using single soil type data to establish SOM prediction models is an effective approach to improve model performance. The negative impacts derived from the spatial heterogeneity of different soil types and the greater spatial variation in soil moisture and pixel temporal information can be reduced. Therefore, our study provides a promising approach for improving the prediction accuracy and enhancing the robustness and practical application of regional-scale SOM prediction models.

Furthermore, our research results provide new ideas for digital soil mapping (DSM). The spatial distribution patterns of soil attributes are highly heterogeneous [49], leading to the current unsatisfactory accuracy of SOM prediction based on satellite remote-sensing

data, which is one of the reasons why the existing SOM mapping has considerable uncertainty. On the one hand, our research used remote-sensing data to map the spatial distribution of SOM content on the regional scale, which can overcome the limitations (e.g., labor-intensive, costly, and time-consuming) of traditional DSM techniques [135,136]; therefore, it does not require heavy field investigation work and a subsequent series of laboratory analyses. Simultaneously, the GEE cloud computing platform is used to obtain time-series remote sensing image datasets in our research, which has powerful computational and storage capacities including various remote-sensing images and geospatial datasets [137,138] and greatly enhances the work efficiency of scientific researchers [139]. More importantly, it also provides greater possibilities for conducting DSM on a large scale. On the other hand, the essence of DSM is the relationship between soil observation data and available environmental variables using predictive models to more accurately infer the temporal and spatial changes in soil properties, thereby enriching the soil information system [140,141]. The three main influencing factors (i.e., precipitation, pixel date, and soil type) on SOM prediction were considered in our study, and the mechanisms of their influence on SOM prediction accuracies were clarified from a theoretical perspective. Our results proved that using single soil type data can reduce the negative impacts and increase the accuracy of regional SOM prediction. Hence, our study provides data and methodological references to more accurately reveal the spatial distribution patterns of soil attributes, promote the development of DSM technology, and meet the needs of social development for soil database information. Finally, the spatial patterns of SOM in the Phaeozem and Arenosol regions and in the whole study area were accurately mapped based on the above favorable analysis conditions. It is essential to understand the soil fertility status and adopt active soil protection measures, such as adopting a reasonable increase in the amount of fertilization, developing an appropriate rotation or fallow system, and ultimately realizing the sustainable use of soil resources.

5. Conclusions

Our study focused on two typical cropland soil types on the northern Songnen Plain and attempted to incorporate the pixel dates of training samples and precipitation to evaluate their impacts on SOM prediction accuracy based on time series MOD09A1 images and SOM observation data from multiple soil types. We also demonstrated the advantages of using single soil type data to improve the SOM prediction accuracy and ultimately mapped the SOM content in the study area.

Our results showed that the pixel dates of the training samples and precipitation were the main factors controlling the model performance and inputs during the bare soil period. When precipitation did not occur on the optimal pixel date and its neighboring days, the model accuracies were high and generally showed an increasing trend, and the main inputs were generally SOM-sensitive bands or spectral indices constructed from these bands to characterize the SOM. Under the opposite conditions, the accuracies exhibited a decreasing trend, and inputs were generally constructed by SOM-sensitive and moisture-sensitive bands to represent the impact of precipitation. As anticipated, the ratio spectral index (e.g., R_{61}) was suitable for characterizing the impact of precipitation.

The SOM prediction accuracy for soil samples composed of single soil type outperformed those of multiple soil types. Models of single soil type can better reduce the spatial heterogeneity of soil properties, decrease the spatial variations in soil moisture and pixel temporal information, and prevent regional proximity effects between adjacent soil types. Moreover, compared to the total dataset, the RI of Arenosols displayed a greater improvement than that of Phaeozems due to the occurrence of less precipitation in the Arenosol region.

To conclude, our results indicated that integrating the pixel date with precipitation data can illustrate the variation in the prediction accuracy during the bare soil period, highlighting the advantage of using single soil type data and revealing that the SOM content shows a decreasing trend from northeast to southwest on the northern Songnen Plain. Our

study provides promising analytical approaches for investigating the driving factors that impact and improve SOM prediction accuracy. The research results are beneficial in that they serve as a data reference for improving the accuracy of remote-sensing models for soil physical and chemical parameters and guiding the implementation of more precise agricultural management measures.

Author Contributions: Conceptualization, M.Z. (Meiwei Zhang) and H.L.; methodology, M.Z. (Meiwei Zhang); software, M.Z. (Meiwei Zhang); validation, M.Z. (Meiwei Zhang); investigation, M.Z. (Meiwei Zhang) and H.L.; writing-original draft preparation, M.Z. (Meiwei Zhang); writing-review and editing, M.Z. (Meiwei Zhang), M.Z. (Meinan Zhang), H.Y., Y.J., H.L., Y.H., H.T., X.Z. (Xiaohan Zhang) and X.Z. (Xinle Zhang); visualization, M.Z. (Meiwei Zhang); supervision, H.L.; and funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2021YFD1500100) and the K. C. Wong Education Foundation.

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 privacy.

Acknowledgments: We thank AJE (<https://www.aje.com/>, accessed on 11 November 2021) for its linguistic assistance during the preparation of this manuscript. The authors are also very grateful to the editors and anonymous reviewers for their valuable and constructive suggestions that helped us to improve the manuscript.

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

References

1. Guo, L.; Zhao, C.; Zhang, H.; Chen, Y.; Linderman, M.; Zhang, Q.; Liu, Y. Comparisons of spatial and non-spatial models for predicting soil carbon content based on visible and near-infrared spectral technology. *Geoderma* **2017**, *285*, 280–292. [[CrossRef](#)]
2. 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](#)]
3. Huang, B.; Sun, W.; Zhao, Y.; Zhu, J.; Yang, R.; Zou, Z.; Ding, F.; Su, J. Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices. *Geoderma* **2007**, *139*, 336–345. [[CrossRef](#)]
4. Manlay, R.J.; Feller, C.; Swift, M.J. Historical evolution of soil organic matter concepts and their relationships with the fertility and sustainability of cropping systems. *Agric. Ecosyst. Environ.* **2007**, *119*, 217–233. [[CrossRef](#)]
5. Mishra, U.; Torn, M.S.; Masanet, E.; Ogle, S.M. Improving regional soil carbon inventories: Combining the IPCC carbon inventory method with regression kriging. *Geoderma* **2012**, *189*, 288–295. [[CrossRef](#)]
6. Oldfield, E.E.; Bradford, M.A.; Wood, S.A. Global meta-analysis of the relationship between soil organic matter and crop yields. *Soil* **2019**, *5*, 15–32. [[CrossRef](#)]
7. Liang, Z.; Chen, S.; Yang, Y.; Zhao, R.; Shi, Z.; Rossel, R.A.V. National digital soil map of organic matter in topsoil and its associated uncertainty in 1980's China. *Geoderma* **2019**, *335*, 47–56. [[CrossRef](#)]
8. Xu, X.; Xu, Y.; Chen, S.; Xu, S.; Zhang, H. Soil loss and conservation in the black soil region of Northeast China: A retrospective study. *Environ. Sci. Policy* **2010**, *13*, 793–800. [[CrossRef](#)]
9. Gao, X.; Hu, Y.; Sun, Q.; Du, L.; Duan, P.; Yao, L.; Guo, S. Erosion-induced carbon losses and CO₂ emissions from Loess and Black soil in China. *Catena* **2018**, *171*, 533–540. [[CrossRef](#)]
10. Li, H.; Zhu, H.; Qiu, L.; Wei, X.; Liu, B.; Shao, M. Response of soil OC, N and P to land-use change and erosion in the black soil region of the Northeast China. *Agric. Ecosyst. Environ.* **2020**, *302*, 107081. [[CrossRef](#)]
11. Kumar, S.; Lal, R.; Liu, D.; Rafiq, R. Estimating the spatial distribution of organic carbon density for the soils of Ohio, USA. *J. Geogr. Sci.* **2013**, *23*, 280–296. [[CrossRef](#)]
12. Pouladi, N.; Møller, A.B.; Tabatabai, S.; Greve, M.H. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma* **2019**, *342*, 85–92. [[CrossRef](#)]
13. Tziachris, P.; Aschonitis, V.; Chatzistathis, T.; Papadopoulou, M. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *Catena* **2019**, *174*, 206–216. [[CrossRef](#)]
14. Zeng, C.; Yang, L.; Zhu, A.-X.; Rossiter, D.G.; Liu, J.; Liu, J.; Qin, C.; Wang, D. Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method. *Geoderma* **2016**, *281*, 69–82. [[CrossRef](#)]
15. Zhang, S.; Huang, Y.; Shen, C.; Ye, H.; Du, Y. Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma* **2012**, *171*, 35–43. [[CrossRef](#)]

16. Meersmans, J.; De Ridder, F.; Canters, F.; De Baets, S.; Van Molle, M. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma* **2008**, *143*, 1–13. [[CrossRef](#)]
17. Schloeder, C.; Zimmerman, N.; Jacobs, M. Comparison of methods for interpolating soil properties using limited data. *Soil Sci. Soc. Am. J.* **2001**, *65*, 470–479. [[CrossRef](#)]
18. Wu, C.; Wu, J.; Luo, Y.; Zhang, L.; DeGloria, S.D. Spatial prediction of soil organic matter content using cokriging with remotely sensed data. *Soil Sci. Soc. Am. J.* **2009**, *73*, 1202–1208. [[CrossRef](#)]
19. Dai, F.; Zhou, Q.; Lv, Z.; Wang, X.; Liu, G. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. *Ecol. Indic.* **2014**, *45*, 184–194. [[CrossRef](#)]
20. Li, Y. Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? *Geoderma* **2010**, *159*, 63–75. [[CrossRef](#)]
21. Webster, R.; Oliver, M.A. Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* **1992**, *43*, 177–192. [[CrossRef](#)]
22. Heuvelink, G.; Bierkens, M. Combining soil maps with interpolations from point observations to predict quantitative soil properties. *Geoderma* **1992**, *55*, 1–15. [[CrossRef](#)]
23. McBratney, A.B.; Santos, M.M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
24. Zhao, Z.; Yang, Q.; Benoy, G.; Chow, T.L.; Xing, Z.; Rees, H.W.; Meng, F.-R. Using artificial neural network models to produce soil organic carbon content distribution maps across landscapes. *Can. J. Soil Sci.* **2010**, *90*, 75–87. [[CrossRef](#)]
25. Jeong, G.; Oeverdieck, H.; Park, S.J.; Huwe, B.; Liefß, M. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. *Catena* **2017**, *154*, 73–84. [[CrossRef](#)]
26. Zhou, T.; Geng, Y.; Chen, J.; Liu, M.; Haase, D.; Lausch, A. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. *Ecol. Indic.* **2020**, *114*, 106288. [[CrossRef](#)]
27. Takata, Y.; Funakawa, S.; Akshalov, K.; Ishida, N.; Kosaki, T. Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. *Soil Sci. Plant Nutr.* **2007**, *53*, 289–299. [[CrossRef](#)]
28. Abrougui, K.; Gabsi, K.; Mercatoris, B.; Khemis, C.; Amami, R.; Chehaibi, S. Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil Tillage Res.* **2019**, *190*, 202–208. [[CrossRef](#)]
29. Adhikari, K.; Hartemink, A.E. Digital mapping of topsoil carbon content and changes in the Driftless Area of Wisconsin, USA. *Soil Sci. Soc. Am. J.* **2015**, *79*, 155–164. [[CrossRef](#)]
30. Dou, X.; Wang, X.; Liu, H.; Zhang, X.; Meng, L.; Pan, Y.; Yu, Z.; Cui, Y. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China. *Geoderma* **2019**, *356*, 113896. [[CrossRef](#)]
31. Zhang, X.; Dou, X.; Xie, Y.; Liu, H.; Wang, N.; Wang, X.; Pan, Y. Remote sensing inversion model of soil organic matter in farmland by introducing temporal information. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 143–150.
32. Liu, S.; An, N.; Yang, J.; Dong, S.; Wang, C.; Yin, Y. Prediction of soil organic matter variability associated with different land use types in mountainous landscape in southwestern Yunnan province, China. *Catena* **2015**, *133*, 137–144. [[CrossRef](#)]
33. Doetterl, S.; Stevens, A.; Van Oost, K.; Quine, T.A.; Van WeseMAEL, B. Spatially-explicit regional-scale prediction of soil organic carbon stocks in cropland using environmental variables and mixed model approaches. *Geoderma* **2013**, *204*, 31–42. [[CrossRef](#)]
34. Rasmussen, C.; Heckman, K.; Wieder, W.R.; Keiluweit, M.; Lawrence, C.R.; Berhe, A.A.; Blankinship, J.C.; Crow, S.E.; Druhan, J.L.; Pries, C.E.H. Beyond clay: Towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* **2018**, *137*, 297–306. [[CrossRef](#)]
35. Goldberger, A.S.; Jochems, D.B. Note on stepwise least squares. *J. Am. Stat. Assoc.* **1961**, *56*, 105–110. [[CrossRef](#)]
36. Leigh, J.P. Assessing the importance of an independent variable in multiple regression: Is stepwise unwise? *J. Clin. Epidemiol.* **1988**, *41*, 669–677. [[CrossRef](#)]
37. Rogge, D.; Bauer, A.; Zeidler, J.; Mueller, A.; Esch, T.; Heiden, U. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sens. Environ.* **2018**, *205*, 1–17. [[CrossRef](#)]
38. Sudduth, K.; Hummel, J. Evaluation of reflectance methods for soil organic matter sensing. *Trans. ASAE* **1991**, *34*, 1900–1909. [[CrossRef](#)]
39. Behrens, T.; Schmidt, K.; Ramirez-Lopez, L.; Gallant, J.; Zhu, A.-X.; Scholten, T. Hyper-scale digital soil mapping and soil formation analysis. *Geoderma* **2014**, *213*, 578–588. [[CrossRef](#)]
40. Sullivan, D.G.; Shaw, J.; Rickman, D. IKONOS imagery to estimate surface soil property variability in two Alabama physiographies. *Soil Sci. Soc. Am. J.* **2005**, *69*, 1789–1798. [[CrossRef](#)]
41. Yanli, L.; Youlu, B.; Liping, Y.; Hongjuan, W. Hyperspectral extraction of soil organic matter content based on principal component regression. *N. Z. J. Agric. Res.* **2007**, *50*, 1169–1175. [[CrossRef](#)]
42. 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](#)]
43. Liu, Y.; Ding, X.; Liu, H.; Zhang, X.; Qu, C.; Hu, W.; Zhang, H. Quantitative analysis of reflectance spectrum of Black soil as affected by soil moisture for prediction of soil moisture in black soil. *Acta Pedol. Sin.* **2014**, *51*, 1021–1026.
44. Chen, F.; Kissel, D.E.; West, L.T.; Adkins, W. Field-scale mapping of surface soil organic carbon using remotely sensed imagery. *Soil Sci. Soc. Am. J.* **2000**, *64*, 746–753. [[CrossRef](#)]

45. Zheng, G.; Dongryeol, R.; Caixia, J.; Changqiao, H. Estimation of organic matter content in coastal soil using reflectance spectroscopy. *Pedosphere* **2016**, *26*, 130–136. [[CrossRef](#)]
46. Bower, S.A.; Hanks, R.J. Reflection of Radiant Energy from Soils. *Soil Sci.* **1965**, *100*, 130–138. [[CrossRef](#)]
47. 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](#)]
48. Piccini, C.; Marchetti, A.; Francaviglia, R. Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment. *Ecol. Indic.* **2014**, *36*, 301–314. [[CrossRef](#)]
49. Guo, P.-T.; Li, M.-F.; Luo, W.; Tang, Q.-F.; Liu, Z.-W.; Lin, Z.-M. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma* **2015**, *237*, 49–59. [[CrossRef](#)]
50. Zhao, Z.; Yang, Q.; Sun, D.; Ding, X.; Meng, F.-R. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. *Comput. Electron. Agric.* **2020**, *169*, 105172. [[CrossRef](#)]
51. Chen, D.; Chang, N.; Xiao, J.; Zhou, Q.; Wu, W. Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms. *Sci. Total Environ.* **2019**, *669*, 844–855. [[CrossRef](#)]
52. Muro, J.; Canty, M.; Conradsen, K.; Hüttich, C.; Nielsen, A.A.; Skriver, H.; Remy, F.; Strauch, A.; Thonfeld, F.; Menz, G. Short-term change detection in wetlands using Sentinel-1 time series. *Remote Sens.* **2016**, *8*, 795. [[CrossRef](#)]
53. Zhang, S.-w.; Shen, C.-y.; Chen, X.-y.; Ye, H.-c.; Huang, Y.-f.; Shuang, L. Spatial interpolation of soil texture using compositional kriging and regression kriging with consideration of the characteristics of compositional data and environment variables. *J. Integr. Agric.* **2013**, *12*, 1673–1683. [[CrossRef](#)]
54. Liu, F.; Geng, X.; Zhu, A.-X.; Fraser, W.; Waddell, A. Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS. *Geoderma* **2012**, *171*, 44–52. [[CrossRef](#)]
55. Xiao, W.; Chen, W.; He, T.; Ruan, L.; Guo, J. Multi-Temporal Mapping of Soil Total Nitrogen Using Google Earth Engine across the Shandong Province of China. *Sustainability* **2020**, *12*, 10274. [[CrossRef](#)]
56. Huang, N.; He, J.-S.; Niu, Z. Estimating the spatial pattern of soil respiration in Tibetan alpine grasslands using Landsat TM images and MODIS data. *Ecol. Indic.* **2013**, *26*, 117–125. [[CrossRef](#)]
57. Liu, H.; Pan, Y.; Dou, X.; Zhang, X.; Qiu, Z.; Xu, M.; Xie, Y.; Wang, N. Soil organic matter content inversion model with remote sensing image in field scale of blacksoil area. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 127–133.
58. Zhang, M.; Zhang, M.; Yang, H.; Jin, Y.; Zhang, X.; Liu, H. Mapping Regional Soil Organic Matter Based on Sentinel-2A and MODIS Imagery Using Machine Learning Algorithms and Google Earth Engine. *Remote Sens.* **2021**, *13*, 2934. [[CrossRef](#)]
59. Bao, Y.; Meng, X.; Ustin, S.; Wang, X.; Zhang, X.; Liu, H.; Tang, H. Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies. *Catena* **2020**, *195*, 104703. [[CrossRef](#)]
60. Meng, X.; Bao, Y.; Liu, J.; Liu, H.; Zhang, X.; Zhang, Y.; Wang, P.; Tang, H.; Kong, F. Regional soil organic carbon prediction model based on a discrete wavelet analysis of hyperspectral satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102111. [[CrossRef](#)]
61. Zhang, B.; Song, X.-F.; Zhang, Y.-H.; Han, D.-M.; Tang, C.-Y.; Lihu, Y.; Wang, Z.-L. The renewability and quality of shallow groundwater in Sanjiang and Songnen Plain, Northeast China. *J. Integr. Agric.* **2017**, *16*, 229–238. [[CrossRef](#)]
62. Duan, X.; Xie, Y.; Liu, G.; Gao, X.; Lu, H. Field capacity in black soil region, northeast China. *Chin. Geogr. Sci.* **2010**, *20*, 406–413. [[CrossRef](#)]
63. Song, X.-D.; Yang, F.; Ju, B.; Li, D.-C.; Zhao, Y.-G.; Yang, J.-L.; Zhang, G.-L. The influence of the conversion of grassland to cropland on changes in soil organic carbon and total nitrogen stocks in the Songnen Plain of Northeast China. *Catena* **2018**, *171*, 588–601. [[CrossRef](#)]
64. IUSS Working Group WRB. *World Reference Base for Soil Resources 2006*; World Soil Resources Reports No. 103; FAO: Rome, Italy, 2006.
65. Shi, X.; Yu, D.; Xu, S.; Warner, E.D.; Wang, H.; Sun, W.; Zhao, Y.; Gong, Z. Cross-reference for relating Genetic Soil Classification of China with WRB at different scales. *Geoderma* **2010**, *155*, 344–350. [[CrossRef](#)]
66. Zhao, P.; Li, S.; Wang, E.; Chen, X.; Deng, J.; Zhao, Y. Tillage erosion and its effect on spatial variations of soil organic carbon in the black soil region of China. *Soil Tillage Res.* **2018**, *178*, 72–81. [[CrossRef](#)]
67. Duan, X.; Xie, Y.; Liu, B.; Liu, G.; Feng, Y.; Gao, X. Soil loss tolerance in the black soil region of Northeast China. *J. Geogr. Sci.* **2012**, *22*, 737–751. [[CrossRef](#)]
68. Fang, H.; Sun, L.; Qi, D.; Cai, Q. Using ¹³⁷Cs technique to quantify soil erosion and deposition rates in an agricultural catchment in the black soil region, Northeast China. *Geomorphology* **2012**, *169*, 142–150. [[CrossRef](#)]
69. Yang, H.; Zhang, X.; Xu, M.; Shao, S.; Wang, X.; Liu, W.; Wu, D.; Ma, Y.; Bao, Y.; Zhang, X. Hyper-temporal remote sensing data in bare soil period and terrain attributes for digital soil mapping in the Black soil regions of China. *Catena* **2020**, *184*, 104259. [[CrossRef](#)]
70. Hui, L.; Feng, W.-t.; He, X.-h.; Ping, Z.; Gao, H.-j.; Nan, S.; Xu, M.-g. Chemical fertilizers could be completely replaced by manure to maintain high maize yield and soil organic carbon (SOC) when SOC reaches a threshold in the Northeast China Plain. *J. Integr. Agric.* **2017**, *16*, 937–946.
71. O’Kelly, B.C. Accurate determination of moisture content of organic soils using the oven drying method. *Dry. Technol.* **2004**, *22*, 1767–1776. [[CrossRef](#)]
72. Nelson, D.; Sommers, L.E. Total carbon, organic carbon, and organic matter. *Methods Soil Anal. Part 2 Chem. Microbiol. Prop.* **1983**, *9*, 539–579.

73. Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.* **2018**, *212*, 161–175. [[CrossRef](#)]
74. Gallo, B.C.; Demattê, J.A.; Rizzo, R.; Safanelli, J.L.; Mendes, W.d.S.; Lepsch, I.F.; Sato, M.V.; Romero, D.J.; Lacerda, M.P. Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. *Remote Sens.* **2018**, *10*, 1571. [[CrossRef](#)]
75. Bilgili, A.V.; Van Es, H.; Akbas, F.; Durak, A.; Hively, W. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *J. Arid Environ.* **2010**, *74*, 229–238. [[CrossRef](#)]
76. Summers, D.; Lewis, M.; Ostendorf, B.; Chittleborough, D. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecol. Indic.* **2011**, *11*, 123–131. [[CrossRef](#)]
77. Jin, X.; Du, J.; Liu, H.; Wang, Z.; Song, K. Remote estimation of soil organic matter content in the Sanjiang Plain, Northeast China: The optimal band algorithm versus the GRA-ANN model. *Agric. For. Meteorol.* **2016**, *218*, 250–260. [[CrossRef](#)]
78. Ge, Y.; Morgan, C.L.; Ackerson, J.P. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma* **2014**, *221*, 61–69. [[CrossRef](#)]
79. Noda, I. Progress in two-dimensional (2D) correlation spectroscopy. *J. Mol. Struct.* **2006**, *799*, 2–15. [[CrossRef](#)]
80. Zhang, Z.; Ding, J.; Wang, J.; Ge, X. Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices. *Catena* **2020**, *185*, 104257. [[CrossRef](#)]
81. Frazier, B.E.; Cheng, Y. Remote sensing of soils in the eastern Palouse region with Landsat Thematic Mapper. *Remote Sens. Environ.* **1989**, *28*, 317–325. [[CrossRef](#)]
82. Liu, H.J.; Ning, D.H.; Kang, R.; Jin, H.N.; Sheng, L. A Study on Predicting Model of Organic Matter Content Incorporating Soil Moisture Variation. *Spectrosc. Spectr. Anal.* **2017**, *37*, 566–570.
83. Zhan, X.; Liang, X.; Xu, G.; Zhou, L. Influence of plant root morphology and tissue composition on phenanthrene uptake: Stepwise multiple linear regression analysis. *Environ. Pollut.* **2013**, *179*, 294–300. [[CrossRef](#)] [[PubMed](#)]
84. Walker, E. Applied Regression Analysis and Other Multivariable Methods. *Technometrics* **1989**, *31*, 117–118. [[CrossRef](#)]
85. Kleinbaum, D.G.; Kupper, L.L.; Nizam, A.; Rosenberg, E.S. *Applied Regression Analysis and Other Multivariable Methods*; Cengage Learning: Boston, MA, USA, 2013.
86. Bao, Y.; Ustin, S.; Meng, X.; Zhang, X.; Guan, H.; Qi, B.; Liu, H. A regional-scale hyperspectral prediction model of soil organic carbon considering geomorphic features. *Geoderma* **2021**, *403*, 115263. [[CrossRef](#)]
87. Fernandes, M.M.H.; Coelho, A.P.; Fernandes, C.; Silva, M.F.D.; Marta, C.C.D. Estimation of soil organic matter content by modeling with artificial neural networks. *Geoderma* **2019**, *350*, 46–51. [[CrossRef](#)]
88. Minasny, B.; Setiawan, B.I.; Saptomo, S.K.; McBratney, A.B. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma* **2018**, *313*, 25–40.
89. Wang, B.; Waters, C.; Orgill, S.; Gray, J.; Cowie, A.; Clark, A.; Li Liu, D. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Sci. Total Environ.* **2018**, *630*, 367–378. [[CrossRef](#)]
90. Mishra, U.; Lal, R.; Liu, D.; Meirvenne, M.V. Predicting the Spatial Variation of the Soil Organic Carbon Pool at a Regional Scale. *Soil Sci. Soc. Am. J.* **2010**, *74*, 906–914. [[CrossRef](#)]
91. Qi-yong, Y.; Zhong-cheng, J.; Wen-jun, L.; Hui, L. Prediction of soil organic matter in peak-cluster depression region using kriging and terrain indices. *Soil Tillage Res.* **2014**, *144*, 126–132. [[CrossRef](#)]
92. Sadeghi, M.; Jones, S.B.; Philpot, W.D. A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. *Remote Sens. Environ.* **2015**, *164*, 66–76. [[CrossRef](#)]
93. Lobell, D.B.; Asner, G.P. Moisture effects on soil reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66*, 722–727. [[CrossRef](#)]
94. Weidong, L.; Baret, F.; Xingfa, G.; Qingxi, T.; Lanfen, Z.; Bing, Z. Relating soil surface moisture to reflectance. *Remote Sens. Environ.* **2002**, *81*, 238–246. [[CrossRef](#)]
95. Bricklemeyer, 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](#)]
96. Bogrekci, I.; Lee, W. Effects of soil moisture content on absorbance spectra of sandy soils in sensing phosphorus concentrations using UV-VIS-NIR spectroscopy. *Trans. ASABE* **2006**, *49*, 1175–1180. [[CrossRef](#)]
97. Minasny, B.; McBratney, A.B.; Pichon, L.; Sun, W.; Short, M.G. Evaluating near infrared spectroscopy for field prediction of soil properties. *Soil Res.* **2009**, *47*, 664–673. [[CrossRef](#)]
98. Sudduth, K.A.; Hummel, J. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. *Trans. ASAE* **1993**, *36*, 1571–1582. [[CrossRef](#)]
99. Prudnikova, E.; Savin, I. Some Peculiarities of Arable Soil Organic Matter Detection Using Optical Remote Sensing Data. *Remote Sens.* **2021**, *13*, 2313. [[CrossRef](#)]
100. 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*, 118–124. [[CrossRef](#)]
101. Huan-Jun, L.; Zhang, Y.-Z.; Zhang, X.-L.; Zhang, B.; Kai-Shan, S.; Zong-Ming, W.; Na, T. Quantitative analysis of moisture effect on black soil reflectance. *Pedosphere* **2009**, *19*, 532–540.
102. McGrath, D.; Zhang, C. Spatial distribution of soil organic carbon concentrations in grassland of Ireland. *Appl. Geochem.* **2003**, *18*, 1629–1639. [[CrossRef](#)]
103. Hudson, B.D. Soil organic matter and available water capacity. *J. Soil Water Conserv.* **1994**, *49*, 189–194.

104. Emerson, W.; McGarry, D. Organic carbon and soil porosity. *Soil Res.* **2003**, *41*, 107–118. [[CrossRef](#)]
105. Hamblin, A.; Davies, D. Influence of organic matter on the physical properties of some East Anglian soils of high silt content. *J. Soil Sci.* **1977**, *28*, 11–22. [[CrossRef](#)]
106. Bouyoucos, G.J. Effect of organic matter on the water-holding capacity and the wilting point of mineral soils. *Soil Sci.* **1939**, *47*, 377–384. [[CrossRef](#)]
107. Rawls, W.J.; Brakensiek, D.L.; Saxton, K. Estimation of soil water properties. *Trans. ASAE* **1982**, *25*, 1316–1320. [[CrossRef](#)]
108. Hengl, T.; Heuvelink, G.B.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93.
109. Tziolas, N.; Tsakiridis, N.; Ben-Dor, E.; Theocharis, J.; Zalidis, G. Employing a Multi-Input Deep Convolutional Neural Network to Derive Soil Clay Content from a Synergy of Multi-Temporal Optical and Radar Imagery Data. *Remote Sens.* **2020**, *12*, 1389. [[CrossRef](#)]
110. Martínez-Fernández, J.; González-Zamora, A.; Almendra-Martín, L. Soil moisture memory and soil properties: An analysis with the stored precipitation fraction. *J. Hydrol.* **2020**, *593*, 125622.
111. Liu, H.; Yang, H.; Xu, M.; Zhang, X.; Zhang, X.; Yu, Z.; Shao, S.; Li, H. Soil classification based on maximum likelihood method and features of multi-temporal remote sensing images in bare soil period. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 132–139.
112. Huanjun, L.; Xiaokang, Z.; Xinle, Z. Hyperspectral reflectance characteristics parameter extraction for soil classification model. *J. Remote Sens.* **2017**, *21*, 105–114.
113. Jain, M.; Mondal, P.; DeFries, R.S.; Small, C.; Galford, G.L. Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sens. Environ.* **2013**, *134*, 210–223. [[CrossRef](#)]
114. Liu, L.; Xiao, X.; Qin, Y.; Wang, J.; Xu, X.; Hu, Y.; Qiao, Z. Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine. *Remote Sens. Environ.* **2020**, *239*, 111624. [[CrossRef](#)]
115. O’rourke, S.; Holden, N. Determination of soil organic matter and carbon fractions in forest top soils using spectral data acquired from visible–near infrared hyperspectral images. *Soil Sci. Soc. Am. J.* **2012**, *76*, 586–596. [[CrossRef](#)]
116. Ataieyan, P.; Ahmadi Moghaddam, P.; Sepehr, E. Estimation of Soil Organic Carbon using Artificial Neural Network and Multiple Linear Regression Models based on Color Image Processing. *J. Agric. Mach.* **2018**, *8*, 137–148.
117. Selige, T.; Böhner, J.; Schmidhalter, U. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* **2006**, *136*, 235–244. [[CrossRef](#)]
118. Diek, S.; Fornallaz, F.; Schaepman, M.E.; De Jong, R. Barest pixel composite for agricultural areas using landsat time series. *Remote Sens.* **2017**, *9*, 1245. [[CrossRef](#)]
119. Blasch, G.; Spengler, D.; Itzerott, S.; Wessolek, G. Organic matter modeling at the landscape scale based on multitemporal soil pattern analysis using RapidEye data. *Remote Sens.* **2015**, *7*, 11125–11150. [[CrossRef](#)]
120. Conforti, M.; Castrignanò, A.; Robustelli, G.; Scarciglia, F.; Stelluti, M.; Buttafuoco, G. Laboratory-based Vis–NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content. *Catena* **2015**, *124*, 60–67. [[CrossRef](#)]
121. Shi, Z.; Ji, W.; Viscarra Rossel, R.; Chen, S.; Zhou, Y. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis–NIR spectral library. *Eur. J. Soil Sci.* **2015**, *66*, 679–687. [[CrossRef](#)]
122. Ostovari, Y.; Ghorbani-Dashtaki, S.; Bahrami, H.-A.; Abbasi, M.; Dematte, J.A.M.; Arthur, E.; Panagos, P. Towards prediction of soil erodibility, SOM and CaCO₃ using laboratory Vis-NIR spectra: A case study in a semi-arid region of Iran. *Geoderma* **2018**, *314*, 102–112. [[CrossRef](#)]
123. Ward, K.J.; Chabrilat, S.; Neumann, C.; Foerster, S. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered LUCAS soil database. *Geoderma* **2019**, *353*, 297–307. [[CrossRef](#)]
124. Shirzadi, A.; Shahabi, H.; Chapi, K.; Bui, D.T.; Pham, B.T.; Shahedi, K.; Ahmad, B.B. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena* **2017**, *157*, 213–226. [[CrossRef](#)]
125. Benke, K.K.; Norng, S.; Robinson, N.; Chia, K.; Rees, D.; Hopley, J. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. *Geoderma* **2020**, *366*, 114210. [[CrossRef](#)]
126. Rahmati, O.; Pourghasemi, H.R.; Melesse, A.M. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena* **2016**, *137*, 360–372. [[CrossRef](#)]
127. Cierniewski, J.; Verbrugge, M.; Marlewski, A. Effects of farming works on soil surface bidirectional reflectance measurements and modelling. *Int. J. Remote Sens.* **2002**, *23*, 1075–1094. [[CrossRef](#)]
128. Vaudour, E.; Gomez, C.; Loiseau, T.; Baghdadi, N.; Loubet, B.; Arrouays, D.; Ali, L.; Lagacherie, P. The impact of acquisition date on the prediction performance of topsoil organic carbon from Sentinel-2 for croplands. *Remote Sens.* **2019**, *11*, 2143. [[CrossRef](#)]
129. Denis, A.; Stevens, A.; Van WeseMAEL, B.; Udelhoven, T.; Tychon, B. Soil organic carbon assessment by field and airborne spectrometry in bare croplands: Accounting for soil surface roughness. *Geoderma* **2014**, *226*, 94–102. [[CrossRef](#)]
130. Huang, S.; Yan-Ni, S.; Wen-Yi, R.; Wu-Ren, L.; Zhang, W.-J. Long-term effect of no-tillage on soil organic carbon fractions in a continuous maize cropping system of Northeast China. *Pedosphere* **2010**, *20*, 285–292. [[CrossRef](#)]
131. Zhang, W.; Wang, X.; Xu, M.; Huang, S.; Liu, H.; Peng, C. Soil organic carbon dynamics under long-term fertilizations in arable land of northern China. *Biogeosciences* **2010**, *7*, 409–425. [[CrossRef](#)]
132. Ben-Dor, E.; Banin, A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci. Soc. Am. J.* **1995**, *59*, 364–372. [[CrossRef](#)]

133. McCarty, G.; Reeves, J.; Reeves, V.; Follett, R.; Kimble, J. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* **2002**, *66*, 640–646. [[CrossRef](#)]
134. Muller, E.; Decamps, H. Modeling soil moisture-reflectance. *Remote Sens. Environ.* **2001**, *76*, 173–180. [[CrossRef](#)]
135. Bouma, J. Using soil survey data for quantitative land evaluation. In *Advances in Soil Science*; Springer: Berlin/Heidelberg, Germany, 1989; pp. 177–213.
136. Grunwald, S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* **2009**, *152*, 195–207. [[CrossRef](#)]
137. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.; Goetz, S.J.; Loveland, T.R. High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *342*, 850–853. [[CrossRef](#)] [[PubMed](#)]
138. Huang, H.; Chen, Y.; Clinton, N.; Wang, J.; Wang, X.; Liu, C.; Gong, P.; Yang, J.; Bai, Y.; Zheng, Y. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sens. Environ.* **2017**, *202*, 166–176. [[CrossRef](#)]
139. Kumar, L.; Mutanga, O. Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sens.* **2018**, *10*, 1509. [[CrossRef](#)]
140. Lagacherie, P.; McBratney, A.; Voltz, M. *Digital Soil Mapping: An Introductory Perspective*; Elsevier: Amsterdam, The Netherlands, 2006.
141. Zhang, G.-L.; Feng, L.; Song, X.-d. Recent progress and future prospect of digital soil mapping: A review. *J. Integr. Agric.* **2017**, *16*, 2871–2885. [[CrossRef](#)]