

## Article

# Inversion Estimation of Soil Organic Matter in Songnen Plain Based on Multispectral Analysis

Siyu Tang, Chong Du \* and Tangzhe Nie

School of Hydraulic and Electric Power, Heilongjiang University, Harbin 150080, China; 2201873@s.hlju.edu.cn (S.T.); 2019036@hlju.edu.cn (T.N.)

\* Correspondence: duchong@hlju.edu.cn

**Abstract:** Sentinel-2A multi-spectral remote sensing image data underwent high-efficiency differential processing to extract spectral information, which was then matched to soil organic matter (SOM) laboratory test values from field samples. From this, multiple-linear stepwise regression (MLSR) and partial least square (PLSR) models were established based on a differential algorithm for surface SOM modeling. The original spectra were subjected to basic transformations with first- and second-derivative processing. MLSR and PLSR models were established based on these methods and the measured values, respectively. The results show that Sentinel-2A remote sensing imagery and SOM content correlated in some bands. The correlation between the spectral value and SOM content was significantly improved after mathematical transformation, especially square-root transformation. After differential processing, the multi-band model had better predictive ability (based on fitting accuracy) than single-band and unprocessed multi-band models. The MLSR and PLSR models of SOM had good prediction functionality. The reciprocal logarithm first-order differential MLSR regression model had the best prediction and inversion results (i.e., most consistent with the real-world data). The MLSR model is more stable and reliable for monitoring SOM content, and provides a feasible method and reference for SOM content-mapping of the study area.

**Keywords:** soil organic matter; Sentinel-2A; remote sensing; differential algorithm; multispectral modeling; PLSR



**Citation:** Tang, S.; Du, C.; Nie, T. Inversion Estimation of Soil Organic Matter in Songnen Plain Based on Multispectral Analysis. *Land* **2022**, *11*, 608. <https://doi.org/10.3390/land11050608>

Academic Editors: Chiara Piccini and Rosa Francaviglia

Received: 31 March 2022

Accepted: 19 April 2022

Published: 21 April 2022

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



**Copyright:** © 2022 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), as an extensive component of soil, is an important indicator to measure the fertility, status and degradation degree of cultivated soil [1]. It plays a role in increasing moisture retention and, consequently, the drought tolerance of crops [2,3]. SOM also constitutes a huge organic carbon pool in terrestrial ecosystems [4,5]. It is of great practical significance to estimate the soil organic pool by mastering the spatial distribution information of large-scale soil organic matter content instantaneously [6]. Estimating soil organic matter pools has a significant impact on ecology and sustainable land use in the long term. Precision agriculture and long-term regional land development are aided by timely monitoring of SOM data [7]. Traditional biochemical analysis methods are time-consuming and labor-intensive, in addition to being ineffective and unsuitable for gathering information such as soil organic matter content over a large expanse [8,9].

Soil-reflection spectroscopy has successfully enabled rapid and cost-effective SOM estimation, assisting in fulfilling regional to global soil evaluation and monitoring requirements [10]. The majority of research has concentrated on soil spectral studies in controlled indoor environments. There are concerns with test parameters, including light-source power, light-source distance, and irradiation angle during the test. The majority of soil samples used in the tests are stabilized soils. As a result, the spectral data is more unclear, making it harder to share the results of known inverse models of soil characteristics. The advantages of remote sensing technology include a vast coverage area for ground object

information, as well as periodicity, currency, precision, and reliability [11]. The quantitative description of soil organic matter by remote sensing technology has always been a research hotspot of many scholars. According to research, the spectral characteristics of SOM are primarily reflected in the absorption of incident light energy by organic matter, and soil reflectance decreases as organic matter content increases [12,13].

There are still image factors, such as mixed pixels, water content, and spectral resolution in remote-sensing data, that must be taken into consideration for soil organic matter analysis. Therefore, the theoretical underpinning of using existing remote-sensing data for SOM mapping involves extracting adequate information from soil spectral data and generating a soil spectral index [14]. Researchers have made some progress in related fields. The application of traditional spectral index inversion theory was used to improve the estimation model's accuracy. Spectral indices could quantify the interrelationships between the SOM's characteristic bands utilizing spectral indices, enhancing weak band connections and reducing model complexity. Preprocessing transformations used to remove image-specific reflectance include soil moisture and particle size to transform soil spectral data, remove signal noise, and highlight features for quantitative model estimation [15]. The derivative algorithm is one of the common preprocessing transformations that reduces spectrum interference by eliminating baseline drift and improving spectral resolution, resulting in increased separation of overlapping peaks and less spectral interference [16]. Although a large number of studies have been carried out in related fields using differential spectral technology [17], relatively few studies have explored the predictive ability in monitoring soil nutrient content.

Previous studies have demonstrated that spectral preprocessing is an important component of multivariate modeling analysis and would improve the predictive performance of models [18–21]. A prediction model based on soil spectral information can effectively and rapidly estimate soil physical and chemical parameters. MLSR (multiple linear stepwise regression) has been developed on the basis of multiple linear regression. Considering the advantage of avoiding collinearity, MLSR has been used to develop models that estimate soil properties [22,23]. The regression equation was introduced using stepwise regression based on the effect, significance or contribution rate of global independent variables on the dependent variable. A linear regression model generates predictions about the dependent variable by eliminating independent variables that are not important to the dependent variable. PLS (partial least squares) is a widely used linear multivariate regression method in the field of soil spectroscopy [18,24,25]. PLS was more accurate than principal component regression or multiple linear regression in predictions of soil salinization using soil conductance in the semi-arid region of Brazil [26]. In PLS, the correlation between principal components is relatively insignificant, while the correlation with the dependent variable is the largest. At the same time, PLS can overcome the strong interpretation of independent variables by principal component analysis. It can effectively extract the comprehensive variables with substantial explanatory power to the system and improve the estimation ability of the model. Therefore, mathematical models are conducive to relate reflectance spectra to SOM content to predict soil nutrients [27,28].

Sentinel-2A is a high-resolution multispectral imaging satellite that covers 13 spectral bands. The bands vary in wavelength from 433 to 2280 nm, including ten bands in the visible near-infrared spectrum and three in the short infrared band. Sentinel-2A has an imaging bandwidth of 290 km, a spectral resolution of 15–180 nm and spatial resolutions including 10, 20 and 60 m. Compared with Landsat TM and other remote-sensing images, Sentinel-2A remote-sensing imaging has higher spectral and spatial resolutions and a shorter revisit cycle (the cycle is five days); it is primarily utilized in global ecological environment monitoring [29]. Morteza Sadeghi investigated soil moisture approaches using Sentinel-2 and Landsat-8 satellites and discovered that Sentinel-2 was more appropriate for the task [30]. Sentinel-2 is suitable for monitoring and mapping soil organic matter, but not soil texture (clay, silt, and sand content) [31]. Qi Gao presented two methodologies for the

retrieval of soil moisture from remotely sensed SAR images, with a spatial resolution of 100 m [32]. However, few studies have used Sentinel-2A imagery to monitor soil nutrients.

Given the enormous range of remote-sensing imagery available due to periodic updates, a comprehensive grasp of image characteristics is critical when selecting which image to employ. Different researchers come to different conclusions in terms of the reflectance band and estimation model used to calculate organic matter content [33,34]. In order to comprehensively understand the prediction ability and feasibility of differential spectroscopy in soil nutrient content, the trial used differential processing of the high-resolution Sentinel-2A spectral data based on mathematical transformations to develop models to predict soil organic matter content in the study area.

On the basis of the above, the research aims to construct a SOM evaluation model based on spectral indices and compare the prediction accuracy of different methods in the study region, using Sentinel-2A remote-sensing images as the data source and measured test data of SOM content as analysis data. The objective of this research is: (1) to investigate the use of Sentinel-2A remote-sensing images as a reference for estimating soil organic matter; (2) to analyze the correlation of mathematical transformation (reciprocal, reciprocal logarithm, square root, and square and cubic transformation) with the first- and second-order differential of reflectance and SOM; (3) to construct single-band and multi-band MLRSR and PLSR inversion models and evaluate the spectral indices and model performance in SOM estimation.

## 2. Materials and Methods

### 2.1. Experimental Site

The study area, Daqing, is located in the southwest of Heilongjiang Province in northeast China ( $45^{\circ}46'–46^{\circ}55'$  N and  $124^{\circ}19'–125^{\circ}12'$  E). The region is located in the middle of the Songnen Plain, a Mesozoic subsidence area with a flat, slightly undulating, small ground slope. The landform in the region gradually declines from north to south and is generally plain, with a relative height difference of 10 to 35 m. It connects with the Suihua area in the east, faces Jilin Province (Songhua River) in the south, and borders the city of Qiqihar in the west and north. Winters are cold and snowy, whereas spring and autumn monsoons are more humid. The frost-free season lasts only a few weeks each year. With mean annual precipitation of 427.5 mm, a mean annual temperature of  $4.2^{\circ}\text{C}$ , and a mean evaporation amount of 1635 mm; rain and heat are in the same season, which is beneficial for crop and forage grass growth. The cultivated land area of Daqing accounts for roughly 20% of the whole area and consists of an established farming industry (Figure 1).

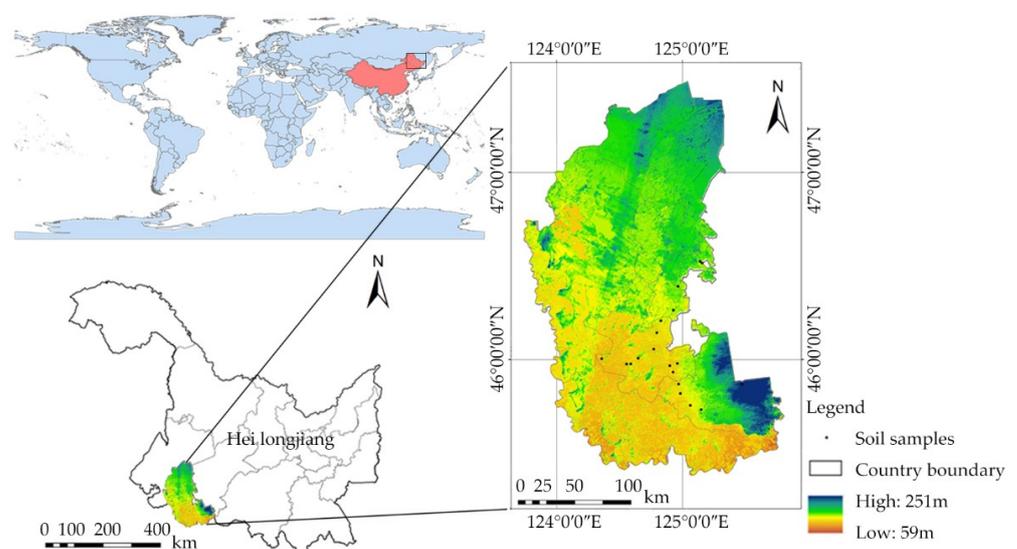


Figure 1. Distribution area diagram of the soil sampling area.

## 2.2. Soil Sampling

In this study, soil sampling was conducted in Daqing in July 2021, and 19 soil samples were randomly obtained. Five surface soil (soil depth of 20 cm) samples were collected and mixed within a 1 m radius of a specific sampling location, and approximately 500 g of soil per sampling site from the mixed models was used for chemical analysis. The actual longitudes and latitudes of the samples were recorded using a global positioning system (GPS) at the time of field sampling in order to obtain the reflectivity of the sampling point in the remote-sensing image. In the laboratory, all samples were air-dried and ground to pass through a 2-mm sieve to remove impurities such as gravel and animal and plant residues.

Chemical analysis was used to determine the SOM content. The concentrations of all soil samples from each sample point were measured through the potassium dichromate method [35]. The SOM content varied from 13.42 to 22.04 g kg<sup>-1</sup>. The coefficient of variation (CV) was 0.15, indicating that SOM showed medium variability across all samples (i.e., 0.1 CV 1.0). To ensure the rationality of model establishment and validation, the data were randomly divided into 14 prediction sites and 5 validation sites. (Table 1).

**Table 1.** Descriptive statistics of SOM (g/kg).

Sample Set	Max.	Min.	Range	Mean	SD	CV
Calibration	22.02	13.42	8.60	17.59	2.70	15.4%
Validation	17.33	14.48	2.85	15.49	0.75	4.9%
Total	22.02	13.42	8.60	17.04	2.57	15.1%

Notes: SD, standard deviation; CV, coefficient of variation.

## 2.3. Remote-Sensing Image Processing

Based on region size, we selected six Sentinel-2A images in July 2021 for the experiment. Ten visible bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12), which are near-infrared and short-wave infrared bands with a resolution of 10 m, were selected from the images (Table 2). The preprocessing of remote-sensing images mainly included radiometric correction, atmospheric correction, geometric correction, image mosaic, and image clipping. Sentinel-2A remote-sensing images were processed with Sen2cor software for radiometric and atmospheric correction, an ESA plug-in dedicated to the creation of L2A level data that is used to reduce radiometric inaccuracies caused by atmospheric influence and to invert the true surface reflectance of objects. Compared with the typical atmospheric correction software (SMAC and 6S), Sen2cor operation is more straightforward, without human input parameters. The geometric correction of Sentinel-2A remote-sensing imaging was completed with ENVI software, and the geometric correction error was less than 1 pixel. The boundary of the research area was classified in ArcGIS software, and remote-sensing images were clipped and arranged as a mosaic. Remote-sensing images of the original research region were obtained by clipping remote-sensing image data of six scenes with vector boundary data from the research area.

To properly manipulate the data from the sample points, a vector map of the administrative divisions of Daqing was obtained using the BIGEMAP map loader. Furthermore, ArcMap and ArcGIS software were used to complete the longitude and latitude distribution map of sampling points by clicking add data and the directory window. Then ArcMap and ENVI were used in combination to extract Sentinel-2A images for each sampling point corresponding to a DN (digital number) value of each band.

**Table 2.** The relevant parameters of Sentinel-2A.

Sentinel-2A				
Band	Band Name	Central Wavelength/nm	Spectral Width/nm	Spatial Resolution/m
1	Coastal Aerosol	433	20	60
2	Blue	490	65	10
3	Green	560	35	10
4	Red	665	30	10
5	Vegetation Red edge	705	15	20
6	Vegetation Red edge	740	15	20
7	Vegetation Red edge	783	30	20
8	NIR	842	115	10
8A	Narrow NIR	865	20	20
9	Water Vapour	945	20	60
10	SWIR-Cirrus	1375	30	60
11	SWIR	1610	90	20
12	SWIR	2190	180	20

#### 2.4. Statistical Modeling

##### 2.4.1. Differential Algorithm

Differential spectral technology, which is a common spectrum processing approach, can effectively dig spectral effective information and provide better resolution than the original spectral reflectance. It also improves the correlation between spectral data and soil parameters, allowing for better monitoring of progress in soil nutrient content research and improved prediction accuracy. The reference formula is as follows [36].

The first derivative can be described as:

$$FDR_{(\lambda)} = \left[ R_{(\lambda_{i+1})} - R_{(\lambda_i)} \right] / [\lambda_{i+1} - \lambda_i] \quad (1)$$

The second derivative (SDR) can be described as:

$$SDR_{(\lambda)} = \left[ R'_{(\lambda_{i+1})} - R'_{(\lambda_i)} \right] / [\lambda_{i+1} - \lambda_i] \quad (2)$$

where  $\lambda_i$  is the wavelength of the  $i$ -th band,  $R_{(\lambda_{i+1})}$ ,  $R_{(\lambda_i)}$  are the reflectance at bands  $\lambda_{i+1}$ ,  $\lambda_i$ , and  $R'_{(\lambda_{i+1})}$ ,  $R'_{(\lambda_i)}$  are the first derivative at bands  $\lambda_{i+1}$ ,  $\lambda_i$ , respectively.

##### 2.4.2. Multiple Linear Stepwise Regression

MLSR is mainly a comparative analysis of all independent variables according to influence or contribution size to all dependent variables through the F test [23]. Variables significant by the sum of squares are selected for the regression equation. Only one variable is introduced for each step. When a variable is introduced, the partial regression sum of squares of each variable is then tested. If the introduced variable is found to be insignificant, it is removed from the partial regression equation. If more than two variables are introduced in successive steps, it is determined whether or not any existing variables can be removed. Further, when no independent variables can be eliminated, a new independent variable with significant influence is selected for evaluation. This process is repeated until none of the introduced variables can be removed. The original independent variable is also tested, and the gradual regression equation ends.

The formula of the gradual regression equation is:

$$SOM = a_0 + \sum_{i=1}^n a_i R_{\lambda_i} \quad (3)$$

where  $a_0$ ,  $a_1 = 1$ ,  $n$  is the regression coefficient,  $i$  is the number of bands used for modeling,  $\lambda_i$  is the wavelength of the  $i$ th modeling band, and  $R_{\lambda_i}$  is the reflectance value at wavelength  $\lambda_i$ .

### 2.4.3. Partial Least Squares Regression

PLSR adopts the idea of extracting principal components from principal component analysis, which can simplify the data structure [25]. There are  $p$  dependent variables and  $m$  independent variables considered. The basic practice is to extract the first component  $x_i$  in the independent variable set and the first component  $u_i$  in the dependent variable set, requiring maximum correlation between  $x_i$  and  $u_i$ . The regression of the dependent variable with  $x_i$  is then established, and the algorithm is terminated until the equation reaches satisfactory accuracy. Otherwise, the extraction of the second pair component continues to achieve satisfactory accuracy. If the  $n$  components are finally extracted from the independent variable set from the independent variable set, the partial least squares regression will establish the regression equation between the dependent variable and  $x_1, x_2, \dots, x_n$ . This represents the regression equation between the dependent variable and the original independent variable: the partial least squares regression equation. In PLS calibration, significant wavelengths can be assessed on the basis of variable important in projection (VIP). If the VIP score of a specific wavelength exceeds 1, then the wavelength is considered important [37,38].

### 2.5. Construction of Spectral Indexes

SOM exhibits unique spectral response properties in visible and near-infrared bands, and the soil spectral reflectivity and SOM content are generally significantly negatively correlated [39,40]. The increase and decrease of SOM content can be reflected from the soil reflection spectrum to a certain extent. The determination of soil spectral reflectance becomes a novel approach to assessing SOM content due to the particular response relationship. Furthermore, the soil spectrum and SOM content show a nonlinear variation caused by the interaction of soil structure and the spectral measurement environment, making the absorption belt and reflection belt of the spectral curve not visible. On the other hand, low-order (first-order, second-order) differential transformation of the spectrum is less sensitive to noise, eliminating some of the background and noise influence and improving the correlation between spectral data and organic matter content.

Therefore, the spectral data are processed by conventional mathematical transformation and differential processing to increase sensitivity to the SOM content of the spectral index. The original spectrum was subjected to six different types of traditional mathematical transformations and respective first and second derivatives. Spectral characteristic indicators were screened using the Pearson correlation analysis method. SOM content measured in the laboratory is the dependent variable of the function, with the characteristic spectral index as the independent variable. The model was constructed between SOM content and the transformed spectral data of the reflection spectrum. The correlation between SOM content and the reflectance of remotely sensed images was analyzed in SPSS. The correlation coefficient was calculated by Formula (4):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

where  $r$  is the correlation coefficient of SOM and reflectance, and  $x_i$  and  $\bar{x}$  are the measured value and mean value of reflectivity, respectively;  $y_i$  and  $\bar{y}$  are the measured value and mean value of SOM content, respectively. When  $r > 0$ , reflectivity is positively correlated with SOM, and when  $r < 0$ , reflectivity is negatively correlated with SOM. The closer  $r$  is to 1, the more stable the model is and the better the fit is [41].

The prediction of SOM model stability is measured by the determination coefficient  $R^2$ ; the larger the  $R^2$ , the more stable the model; the accuracy is tested by RMSE. The smaller the RMSE, the higher the model accuracy [42,43].

The calculation formula is shown in (5) and (6):

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$RMSE = 1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{6}$$

where  $\hat{y}_i$  indicates the values estimated by the model;  $y_i$  indicates the measured values;  $\bar{y}$  indicates the average of the measured values; and  $n$  is the number of observations of the variable to be modelled.

2.6. Flow Chart

Figure 2 shows the flowchart of the research to estimate the model between SOM content and spectral index with differential transformations.

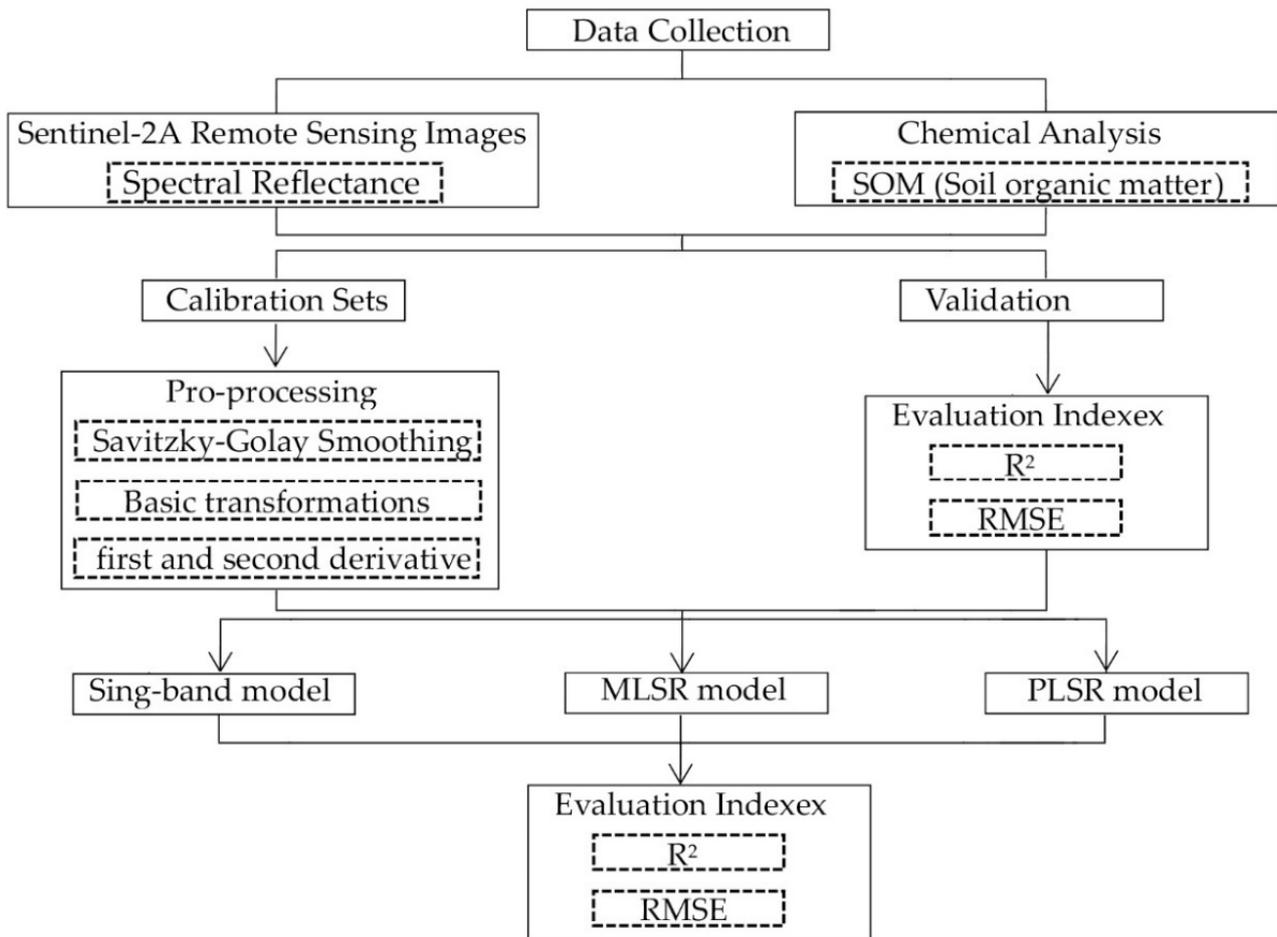
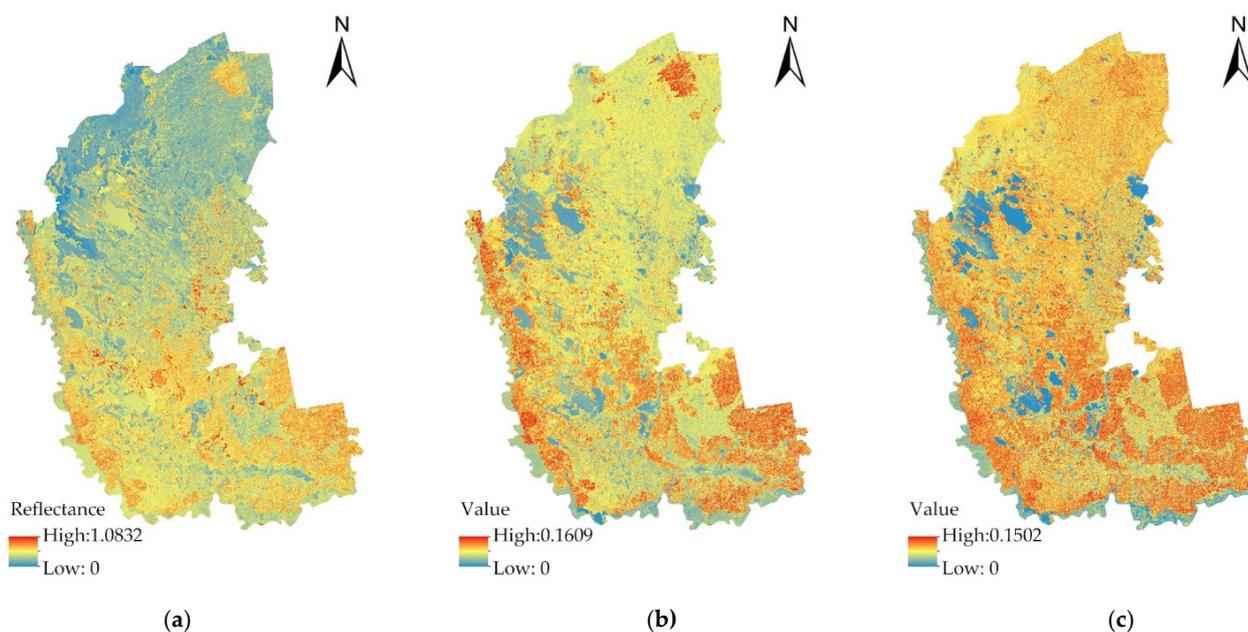


Figure 2. Conducting SOM prediction models.

3. Results and Analysis

3.1. Differential Analysis of the Multispectral Data

The first-order differential and second-order differential are processed by IDL software, with a remote-sensing third band image as an example (Figure 3). The image can better express the real situation of the object, and the first-order differential image better distinguishes the water body from the soil. Raw remote-sensing images contain much information, including noise, which can be excluded by differential image processing of remote-sensing images. However, the meaning of the information in the differential processing image cannot be seen directly, so it needs to be further analyzed based on actual data.



**Figure 3.** Derivative processing of remote image. (a) Original remote-sensing image; (b) First derivative processing of remote-sensing image; (c) Second derivative processing of remote-sensing image.

### 3.2. Correlation between SOM Content and Spectral Metrics

The remote-sensing estimation and inversion of site parameters are based on the relationship between remote-sensing data and site parameters. The correlation between multispectral reflectance data and measured SOM data was analyzed in SPSS to clarify the relationship and to find the spectral information sensitive to SOM content. All original remote-sensing bands exhibited a high degree of correlation (Table 3). According to the first-derivative image data, B3' and B4' have a higher correlation to SOM (Table 4), whereas the overall correlation was relatively low in the second-derivative image data (Table 5). In general, many bands have a high correlation with the original remote-sensing data, and the correlation value of the band is relatively large.

**Table 3.** The correlation between original image data and SOM.

	B3	B4	B5	B6	B7	B8	B8A	B11	SOM
B3	1	0.993 **	0.934 **	0.708 **	0.670 **	0.636 *	0.485 *	0.459	−0.738 **
B4		1	0.960 **	0.822 **	0.705 **	0.674 **	0.531	0.514	−0.779 **
B5			1	0.947 **	0.866 **	0.843 **	0.730 **	0.706 **	−0.852 **
B6				1	0.972 **	0.964 **	0.895 **	0.868 **	−0.854 **
B7					1	0.995 **	0.953 **	0.905 **	−0.763 **
B8						1	0.960 **	0.919 **	−0.762 **
B8A							1	0.982 **	−0.635 *
B11								1	−0.640 **

Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

Through the correlation table of each dataset and SOM content, the correlation of each band of original image data is more significant than 0.6, with B5 and B6 reaching 0.8. However, the first-derivative image data was less-associated with SOM content, with only 0.7 in the B3 band. No sensitive band exists with the SOM after the second-derivative image data. The correlation was significantly reduced compared to the original data in the differential image. According to the aforementioned relationship, the differential processing single-band model results in significant spectral information loss, and the relationship between multispectral data and SOM analysis is not ideal.

**Table 4.** The correlation between first-derivative image data and SOM.

	B3'	B4'	B5'	B6'	B7'	B8'	B8A'	B11'	SOM
B3'	1	0.428	−0.364	−0.016	0.144	0.184	−0.108	0.193	−0.770 **
B4'		1	0.286	0.210	−0.313	0.364	0.420	0.252	−0.595 **
B5'			1	0.847 **	0.142	0.633 *	0.687 **	0.066	−0.058
B6'				1	0.400	0.753 **	0.589 *	0.113	−0.354
B7'					1	0.721	0.880	0.685	−0.177
B8'						1	0.570 *	−0.038	−0.209
B8A'							1	0.643 *	−0.095
B11'								1	−0.263

Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

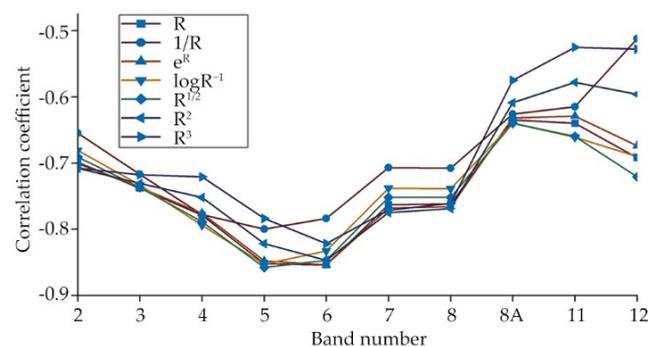
**Table 5.** The correlation between second-derivative image data and SOM.

	B3''	B4''	B5''	B6''	B7''	B8''	B8A''	B11''	SOM
B3''	1	−0.847 **	−0.336	0.695 **	−0.276	0.156	−0.627 **	0.758 **	−0.485
B4''		1	0.077	−0.617 *	0.132	0.053	0.432	−0.626 *	0.504
B5''			1	−0.742 **	0.087	−0.135	0.355	−0.517	−0.292
B6''				1	−0.335	0.153	−0.530	0.749 **	−0.088
B7''					1	−0.911 **	0.490	−0.237	0.394
B8''						1	−0.494	0.139	−0.202
B8A''							1	−0.873 **	0.089
B11''								1	−0.109

Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

### 3.3. Single-Band Inversion Model

The spectral reflectance of the different variation processing was correlated with SOM content to determine the sensitive bands according to the magnitude of the correlation coefficient using SPSS software (Figure 4). SOM exhibited significant spectral response properties in the visible and near-infrared wavelengths and was negatively correlated with spectral reflectance in Sentinel-2A remote-sensing images. The correlation coefficient of the original spectral reflectance peaked at around 740 nm ( $r = -0.854$ ,  $p < 0.001$ ). The fifth and sixth bands had remarkable correlation coefficients for the transformed converted spectral index. The correlation between the spectral index of the first-order differential ( $R'$ ) and SOM content was significant in the third waveband, with the weakest correlation coefficient ( $r = -0.770$ ,  $p < 0.05$ ). Differential processing significantly reduces the correlation compared to the other forms of the spectral index, which were linked considerably with SOM ( $|r| > 0.800$ ,  $p < 0.001$ ). Square root processing ( $R^{1/2}$ ) showed the most-significant correlation occurring at 705 nm ( $r = -0.858$ ,  $p < 0.001$ ).



**Figure 4.** Correlation statistics of organic matter content and reflectivity ( $R$  represents reflectance spectra,  $1/R$  represents inverted transformation,  $e^R$  is exponential transformation,  $\log R^{-1}$  is logarithm the reciprocal of reflectance,  $R^{1/2}$  is square-root transformation,  $R^2$  is square transformation,  $R^3$  is cubic transformation).

The bands with a correlation  $r > 0.5$  with SOM were employed as independent variables to develop separate SOM prediction models, while the measured SOM contents were used as the dependent variable (Table 6). The correlation coefficient of calibration determination ( $R_c^2$ ) and root mean square error of calibration (RMSEC) were used as an assessment indicator. The prediction coefficient of determination ( $R_p^2$ ) and the root mean square error of the validation (RMSEP) set were used to assess accuracy of the final model. The single-band model based on the original spectrum's sensitive band reflectance and SOM produced satisfactory accuracy. In addition, the model with square root processing ( $R^{1/2}$ ) had the best modeling efficiency ( $R_c^2 = 0.74$ , RMSEC = 1.50), but in validation sets performed poorly ( $R_p^2 = 0.69$ , RMSEP = 1.31). The single-band model based on the  $R'$  and  $1/R$  processed spectra with SOM had inferior modeling ( $R_c^2 = 0.60$  and  $0.61$ , RMSEP = 1.86 and 1.75, respectively). The other five models ( $1/R$ ,  $\log R^{-1}$ ,  $R^{1/2}$ ,  $R^2$ ,  $R^3$ ) all showed stronger modeling ( $R_c^2 > 0.68$ , RMSEP < 1.66). Compared to the original reflectivity, the  $R'$  transformation showed the best prediction, with an increase in  $R^2$  of 0.03 ( $R_p^2 = 0.82$ ), but showed extreme uncertainty (RMSECP = 3.72). The modeling of the organic matter single-band model was enhanced, but not dramatically, by spectral processing. Preprocessing of the original spectrum was used in the best suitable model utilizing the single-band ( $R_p^2 = 0.79$ , RMSEP = 2.18).

**Table 6.** Single-band inversion model of soil organic matter content based on spectral index.

Spectral Index	Sensitive Band	Correlation Coefficient	Inversion Model	Calibration		Validation	
				$R_c^2$	RMSEC	$R_p^2$	RMSEP
$R$	6	−0.854 **	$Y = -34.206R + 25.651$	0.71 **	1.52	0.79 **	2.18
$1/R$	5	−0.800 **	$Y = 0.832/R + 12.289$	0.61 **	1.75	0.74 *	1.66
$\log R^{-1}$	5	−0.853 **	$Y = 14.507 \log R^{-1} + 6.375$	0.70 *	1.53	0.71	1.35
$R^{1/2}$	5	−0.858 **	$Y = 30.729 - 31.491R^{1/2}$	0.74 **	1.50	0.69 *	1.31
$R^2$	6	−0.847 **	$Y = -67.546R + 79.856R^2 + 26.838$	0.72 **	1.55	0.77 *	2.40
$R^3$	6	−0.822 **	$Y = -12.333R - 261.662R^2 + 642.675R^3 + 24.151$	0.68 *	1.66	0.75 *	2.66
$R'$	3	−0.770 **	$Y = -149.981R + 18.636$	0.60 **	1.86	0.82	3.72

Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

### 3.4. Performance of Multiple Linear Stepwise Regression

MLSR models were utilized to evaluate the correlation between SOM content and spectral index, referring to the results of the single-band correlation analysis. Table 7 showed the sensitive band combinations used in the regression analysis. A variable variance significance level of 0.05 was set as the criterion for variable selection and exclusion. The maximum variance invasion factor (VIF) of each spectral band was less than 10, indicating no multicollinearity between bands. By comparing model accuracy, six better models were selected.

The band reflectance in the basic mathematical transformation was excluded as an opt-in variable to conduct the MLSR model, except for the raw spectra. However, the modeling is well-based on the differential treatment of mathematical transformations (Figure 5). The multi-band model demonstrated better predictive results than the single-band model in terms of accuracy and stability. Predictions of raw spectral data under MLSR models outperformed all single-band models ( $R_c^2 = 0.78$ , RMSEC = 1.32) in the calibration set. The raw and first-order differential processing ( $R'$ ) of the spectral indices resulted in a significantly improved model, but the effect of validation was poor and unstable ( $R_p^2 = 0.16$  and  $0.55$ , respectively). The MLSR models based on  $(1/R)'$  and  $(1/R)''$  showed better performance than the original spectrum. The inverse first-order differential  $(1/R)'$  verification set, on the other hand, was poor at predicting and hence was not taken into account ( $R_p^2 = 0.37$ ). The MLSR model constructed by  $(\log R^{-1})'$  and  $(\log R^{-1})''$  were slightly less effective than the original spectrum modeling in terms of modeling performance

( $R_c^2 = 0.71$  and  $0.69$ , RMSEC = 1.51 and 1.71, respectively), but the accuracy and stability of validation sets were significantly improved ( $R_p^2 = 0.93$  and  $0.76$ ). Overall, the second-order differential  $(1/R)''$  model performed the best ( $R_c^2 = 0.91$ , RMSEPC = 1.54). However, the validation model performed a little worse ( $R_p^2 = 0.84$ , RMSEP = 1.23), but was considered to be unsuccessful at prediction. The validation of the model based on  $(\log R^{-1})'$  performed incredibly well, with the  $R^2$  improving by 0.09 and the RMSE reducing by 0.43 compared to the  $(1/R)''$  transformation, even though it did not get the best match and model stability ( $R_c^2 = 0.71$ , RMSEC = 1.51,  $R_p^2 = 0.93$ , RMSEP = 1.11). The pre-processing method of the reciprocal logarithm first-order differential spectrum was used in the best suitable model utilizing the MLSR.

**Table 7.** Multi-band inversion model of soil organic matter content based on spectral index.

Sensitive Band	Spectral Index	Inversion Model	Calibration		Validation	
			$R_c^2$	RMSEC	$R_p^2$	RMSEP
$R_6, R_{8A}$	$R$	$Y = -34.206a - 57.592b + 25.651$	0.78 *	1.32	0.16	2.55
$R_3, R_6, R_8, R_4$	$R'$	$Y = -146.835a - 75.734b + 184.192c - 110.819d + 24$	0.89 **	0.92	0.55 *	2.04
$R_6, R_{11}$	$(1/R)'$	$Y = -3.179a + 0.159b + 13.273$	0.80 **	1.27	0.37	1.29
$R_{8A}, R_3, R_4$	$(1/R)''$	$Y = 1.53a - 1.138b - 1.157c + 14.502$	0.91 **	0.87	0.84 **	1.23
$R_{8A}, R_7$	$(\log R^{-1})'$	$Y = 46.497a - 69.465b + 11.418$	0.71 *	1.51	0.93 **	1.11
$R_{8A}, R_4$	$(\log R^{-1})''$	$Y = 47.412a - 43.495b + 13.005$	0.69 *	1.71	0.76 *	1.86

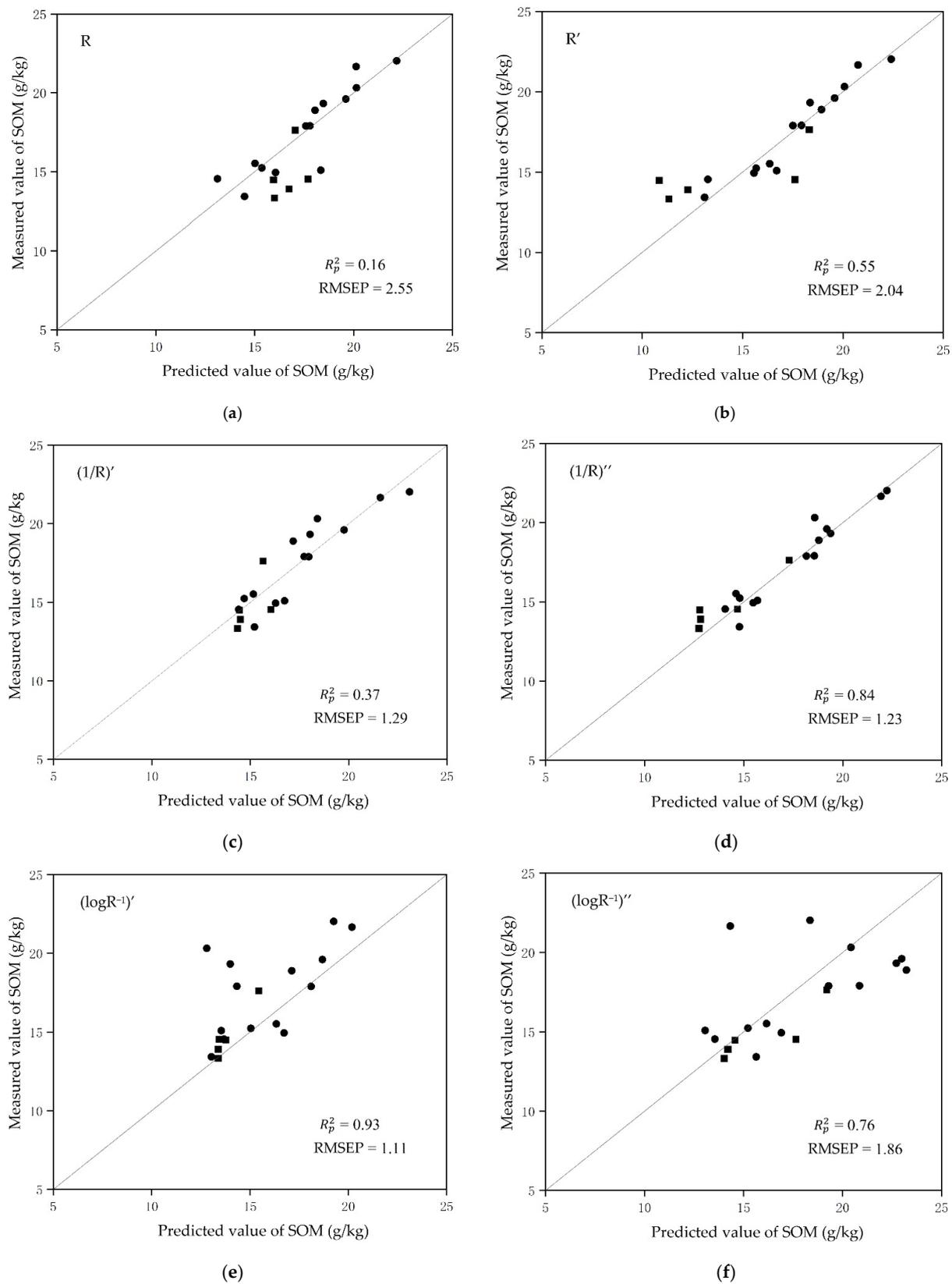
Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

### 3.5. Performance of Partial Least Square Regression

PLSR was established for the eighteen spectral transformations using the Unscrambler software. The results showed that the number of factors obtained by PLSR analysis varies considerably. We performed a full cross-validation before establishing a predictive model. Cross-validation is a method of predicting how well a model will fit the hypothesis validation set. The number of PLSR factors based on the original reflectance was three, increasing to nine after reciprocal transformation. The correlation coefficient of cross-validation determination ( $R_{cv}^2$ ) and root mean square error of cross-validation (RMSECV) were used as assessment metrics to optimize various spectrum post approaches.  $R_p^2$  and RMSEP were used to assess the final effect [22,44]. PLSR regression models based on mathematical transformation with differentiation gave disappointing outcomes.

Among the basic processing, the  $R^2$  model showed the most prediction accuracy in cross-validation ( $R_{cv}^2 = 0.69$ , RMSECV = 1.62) and in independent validation (RMSEP = 11.93), in which the prediction stability was not reliable (Table 8). Similarly, the PLSR model based on  $1/R$  method also performed well in cross-validation ( $R_{cv}^2 = 0.68$ , RMSECV = 1.65) but had poor accuracy in independent validation ( $R_p^2 = 0.51$ ). Prediction was worst using the  $R^3$  method ( $R_{cv}^2 = 0.50$ , RMSECV = 2.06). A PLSR model based on  $\log R^{-1}$  yielded the best prediction results ( $R_{cv}^2 = 0.66$ , RMSECV = 1.70,  $R_p^2 = 0.79$ , RMSEP = 1.55).

For the first-order derivative processing method, the accuracy and stability were reduced to varying degrees compared to the basic processing (Table 9). The prediction when using the  $(1/R)'$  method was best in cross-validation ( $R_{cv}^2 = 0.65$ , RMSECV = 1.8) and in independent validation (though with poor prediction stability) ( $R_p^2 = 0.67$ , RMSEP = 0.84). In terms of the prediction results of the six calibration models, the  $(\log R^{-1})'$  model validations performed the best ( $R_{cv}^2 = 0.62$ , RMSECV = 1.80,  $R_p^2 = 0.90$ , RMSEP = 0.51). Furthermore, PLSR regression models based on second-order derivative processing against SOM content had no practical significance ( $R_{cv}^2 < 0.60$ ). The model based on  $(\log R^{-1})''$  performed best among the processed sets ( $R_{cv}^2 = 0.57$ , RMSECV = 1.92,  $R_p^2 = 0.83$ , RMSEP = 1.28), suggesting that the PLSR model was not suitable for estimating the organic matter content of the region (Table 10).



**Figure 5.** Model validation. (a) Original multi-band measured value; (b) First derivative of the multi-band measured value; (c) First derivative of reciprocal multi-band measured value; (d) Second derivative of reciprocal multi-band measured value; (e) First derivative of the reciprocal logarithm multi-band measured value; (f) Second derivative of the reciprocal logarithm multi-band measured value.

**Table 8.** Results of partial least squares regression analysis of the original spectral data and soil organic matter content.

Spectral Index	PLR Factors	Cross-Validation		Validation	
		$R_{cv}^2$	RMSECV	$R_p^2$	RMSEP
$R$		0.63 *	1.78	0.78 *	1.81
$1/R$		0.68 **	1.65	0.51	1.65
$\log R^{-1}$		0.66	1.70	0.79 *	1.55
$R^{1/2}$		0.63	1.76	0.78 **	1.63
$R^2$		0.69 *	1.62	0.87 *	11.93
$R^3$		0.50	2.06	0.61	2.98

Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

**Table 9.** Results of partial least squares regression analysis of first-derivate transformation and soil organic matter content.

Spectral Index	PLR Factors	Cross-Validation		Validation	
		$R_{cv}^2$	RMSECV	$R_p^2$	RMSEP
$R'$		0.56	1.93	0.91 *	2.57
$(1/R)'$		0.65 **	1.72	0.67 *	0.84
$(\log R^{-1})'$		0.62 *	1.79	0.90 **	0.51
$(R^{1/2})'$		0.33	2.38	0.89 *	2.58
$(R^2)'$		0.64 *	1.75	0.22	3.14
$(R^3)'$		0.54 *	1.98	0.05	3.66

Notes: \* and \*\* represent the confidence level at 0.05 and 0.001, respectively.

**Table 10.** Results of partial least squares regression analysis of second-derivate transformation and soil organic matter content.

Spectral Index	PLR Factors	Calibration		Validation	
		$R_{cv}^2$	RMSE	$R_p^2$	RMSEP
$R''$	—				2.26
$(1/R)''$		0.25	2.52	1.07	0.90
$(\log R^{-1})''$		0.57 *	1.92	1.41	0.83
$(R^{1/2})''$		0.26	2.50	1.08	0.01
$(R^2)''$	—				
$(R^3)''$	—				

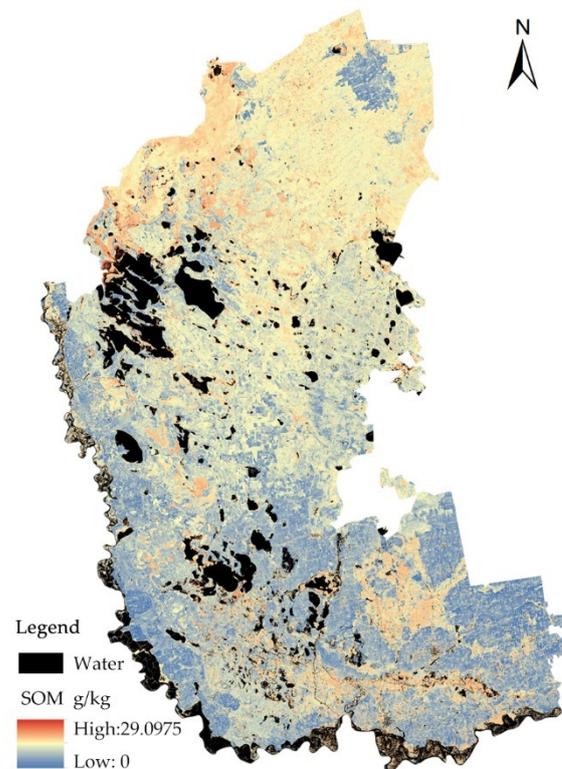
Notes: \* represents the confidence level at 0.05.

The PLSR model under the reciprocal logarithm first-order differential spectrum transformation showed the best correlation with SOM among the three methods. The information above indicates that  $1/R$  transformation performed satisfactorily in data representing spectral features and quantitative inversion models. The results of multi spectral model conducted by PLS show that the prediction capability of PLSR modeling is high using different spectral pretreatment methods.

### 3.6. Spatial Pattern Analysis of Soil Organic Matter Content

The accuracy of the above single-band, MLSR and PLSR inversion models was investigated. The models created by MLSR and PLSR both had satisfactory prediction performance. The accuracy using PLSR is higher than that of MLSR, with excellent prediction. According to the validation sample detection model, the accuracy and stability of the MLSR model are relatively stable. This shows that MLSR regression is more stable and meets the application needs in predicting the SOM content in Daqing. Reciprocal logarithm first-order differential by MLSR regression model is the optimal model. The SOM content-inversion model based on the Sentinel-2A image spectral index was selected

to invert and map the SOM content in the study area to obtain the SOM spatial distribution in Daqing (Figure 6).



**Figure 6.** Inversion result of model.

#### 4. Discussion

The inversion results are in line with the actual spatial distribution of SOM. The SOM content in the study area was generally low and uneven, with large spatial differences. The content was gradually distributed from northeast to southwest, and the SOM content in the northeast was generally higher than that in the southwest. According to the soil field survey, surface runoff accumulates in low-lying areas to form intermittent and permanent puddles due to concentrated precipitation. Poor drainage accelerates the process of salt accumulation, and the complex micro-topography causes uneven evaporation of soil water. The salt above the micro-topography is aggravated by strong evaporation and rainfall and irrigation water containing a certain amount of salt flow from high places to places in areas with poor soil permeability. Soil salinization occurs in low-lying areas after fraction evaporation. In addition, the Daqing area is located in a seasonally frozen soil area, and the freezing and thawing of the soil promotes the accumulation of salt. These aspects combined can lead to serious soil salinization. The high salt content of the soil is not conducive to the survival of vegetation, and the small amount of vegetation means that the content of humus is low and is not conducive to the survival of general decomposers [45]. Higher salinity in soil masks the spectral signature of SOM, resulting in a low inversion value of SOM content in the southwest of the study area [46,47].

The above results showed that Sentinel-2A remote-sensing images had a good correlation with SOM content in visible and near-infrared bands. The effect of multi-band modeling is better than that of single-band modeling. It is found that there is a sensitive band in the correlation between the first-order differential data and SOM content. The original remote-sensing data had the highest correlation in Band 6 ( $r = -0.730$ ). In single-band modeling, the correlation between square root transformation and SOM is best, but the modeling is not as good as the original spectrum. The results of differential processing in the PLSR model are not satisfactory, probably because the differentially processed

image eliminates some of the information of the original image. The correlation between reflectance and SOM is improved after basic transformation and differential processing in MLSR models. The single-band model only uses a very small amount of information, while for remote sensing, the data-rich multi-spectral band can only express extremely limited SOM information in a single band, which can easily cause the loss of some key information. The MLSR and PLSR models achieve convincing results, and the MLSR model has better predictive ability (based on fitting accuracy). The PLSR model, based on second-order differential processing, is ideal and relevant in practice. Reciprocal transformation will improve model prediction in MLSR and PLSR regression. Based on the reciprocal logarithm first-order derivative MLSR regression model, the inversion of SOM in the study area was carried out. The inversion result is in accordance with the actual situation, which is suitable for spectral inversion of soil organic matter content in a certain geographical area.

Compared to direct contact, remote sensing has advantages in the estimation of SOM content, such as predicting soil fertility without direct contact with the object of study, forecasting crop yields from visible and near-infrared bands, accessing the information on the surface of the earth more efficiently and affordably, and updating soil databases in many fields. The previous study found that for most places, the forecast accuracy based on high-resolution satellite was satisfactory for SOM content prediction, and mapped the soil organic matter more precisely than the airborne sensors [32]. The univariate model only considers a single variable to participate in the modeling [48], and the current research is mainly aimed at soil and crops. In this study, the single-band model built using the original spectra worked best—the mathematical transformed form reduced the model prediction—indicating the applicability of Sentinel-2A for predictions. The accuracy of SOM estimation using the MLSR model established by simple mathematical transformation and derivative transformation is higher than that of univariate model. Differential transformation is more beneficial to extract sensitive features in the soil spectrum [49]. The multivariable model integrates the features of multiple sensitive bands, alleviates the “multicollinearity” to a large extent, and improves the applicability and stability of the model. In previous studies, there have been inconsistencies between PLSR models and MLSR predictions [22,23,50,51]. The MLSR model possesses better prediction results than the PLSR model in this study, and high soil salinity in the study area may be a significant factor. In future work, we intend to develop the method further, for example, by expanding the number of soil samples, diversifying the soil types, and taking into account soil moisture and microorganisms. The multispectral examination of SOM by Sentinel-2A has received little attention and has not been thoroughly investigated. Future spectral modeling of the SOM could be integrated with different spectrum indices (e.g., salinity indices) to screen high-precision spectral parameters. Furthermore, indoor light sources can be used to generate reflectance spectra and provide a complete set of measurement data to eliminate the weather impact of spectral acquisition.

## 5. Conclusions

To maximize the correlation between spectral metrics and soil organic matter content, MLSR and PLSR modeling were applied to establish the SOM content model based on Sentinel-2A remote-sensing images. The effective and predictive capacities of different models, which combined basic transformation with differential processing, were validated. Sentinel-2A remote-sensing images had a good correlation with SOM content in visible and near-infrared bands. MLSR and PLSR models of SOM in the study area were established based on different processing and measured values, respectively. The correlation between SOM content and spectral data was improved by multi-spectral modeling after differential processing. However, the correlation between SOM content and reflectance was reduced after first-order differential, indicating that spectral information was partially lost due to differential treatment, and the relationship between spectral data and SOM was not ideal. Multi-band modeling made superior predictions compared to single-band. SOM content could be well-estimated using MLSR models. The MLSR model is more accurate

and stable than PLSR, verified by the calibration and validation samples. The accuracy of the modeling results is high and can meet research requirements. These findings give a theoretical foundation and technological support for utilizing spectroscopy to estimate soil organic matter concentration, and indicates this method can substitute traditional experimental methods for measuring organic matter, thus enabling a larger scale of long-term monitoring of changes in soil organic matter content. In this study, Sentinel-2A images made it possible to retrieve surface soil organic matter with a high spatial and temporal resolution. For soil ecosystem observations, these prediction models will need to be assessed, optimized, and used more broadly.

**Author Contributions:** Conceptualization, C.D.; methodology, C.D. and S.T.; software, S.T.; validation, T.N.; writing—original draft, S.T.; writing—review and editing, T.N. and S.T.; funding acquisition, C.D. and T.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the basic scientific research business expenses of provincial colleges and universities in Heilongjiang Province of China (grant number 2018-KYYWF-1570).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data in this study is available on request from the corresponding author. As this data is the result of a project related to the remote sensing interpretation of soil organic matter, it is not allowed to be made public unless specifically requested.

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

## References

1. Mcbratney, A.; Field, D.J.; Koch, A.J.G. The dimensions of soil security. *Geoderma* **2014**, *213*, 203–213. [[CrossRef](#)]
2. Kerdsueb, P.; Teartisup, P. The Use of Geoinformatics for Estimating Soil Organic Matter in Central Plain of Thailand. *Int. J. Environ. Sci. Dev.* **2014**, *5*, 282–285. [[CrossRef](#)]
3. Seleiman, M.F.; Al-Suhaibani, N.; Ali, N.; Akmal, M.; Alotaibi, M.; Refay, Y.; Dindaroglu, T.; Abdul-Wajid, H.H.; Battaglia, M.L. Drought Stress Impacts on Plants and Different Approaches to Alleviate Its Adverse Effects. *Plants* **2021**, *10*, 259. [[CrossRef](#)] [[PubMed](#)]
4. Lorenz, K.; Lal, R.; Preston, C.M.; Nierop, K.J.G. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules. *Geoderma* **2007**, *142*, 1–10. [[CrossRef](#)]
5. Scharlemann, J.P.; Tanner, E.V.; Hiederer, R.; Kapos, V.J.C.M. Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Manag.* **2014**, *5*, 81–91. [[CrossRef](#)]
6. Gaius, R.S.; Billings, W.D.; Chapin, F.S. Global Change and the Carbon Balance of Arctic Ecosystems: Carbon/Nutrient Interactions should as as large constraints in global terrestrial carbon cycle. *BioScience* **1992**, *42*, 433–441.
7. Dotto, A.C.; Dalmolin, R.S.D.; Caten, T.; Geoderma, G.J. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. *Geoderma* **2018**, *314*, 262–274. [[CrossRef](#)]
8. Bao, Y.; Meng, X.; Ustin, S.; Wang, X.; Tang, H.J.C. Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies. *Catena* **2020**, *195*, 104703. [[CrossRef](#)]
9. Mingliang, L.I.; Xican, L.I.; Zhang, S. Grey Relation Estimating Pattern of Soil Water Content Based on Hyper-Spectral Data. *J. Geomat. Sci. Technol.* **2016**, *28*, 27–39.
10. 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](#)]
11. Ji, W.J.; Li, X.; Li, C.X.; Zhou, Y.; Shi, Z.J.S.; Analysis, S. Using Different Data Mining Algorithms to Predict Soil Organic Matter Based on Visible-Near Infrared Spectroscopy. *Spectrosc. Spectr. Anal.* **2012**, *32*, 2393.
12. Hummel, J. Soil moisture and organic matter prediction of b-horizon soils using an nirsoil sensor. *Am. J. Manag. Care* **2002**, *8*, 963–974.
13. Rinnan, R. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. *Soil Biol. Biochem.* **2007**, *39*, 1664–1673. [[CrossRef](#)]
14. Kasim, N.; Sawut, R.; Shi, Q.; Balati, M.; Kuerban, M.; Julaiti, S. Estimation of Soil Organic Matter Content Based on Optimized Spectral Index. *Trans. Chin. Soc. Agric. Mach.* **2018**, *49*, 155–163.
15. Hansen, P.M.; Schjoerring, J.K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* **2003**, *86*, 542–553. [[CrossRef](#)]

16. Priori, S.; Fantappiè, M.; Lorenzetti, R.; Pellegrini, S.; Costantini, E.A. Field Scale Mapping of Soil Carbon Stock with Limited Sampling by Coupling Gamma Ray and Vis NIR Spectroscopy. *Soil Sci. Soc. Am. J.* **2016**, *80*, 954–964. [[CrossRef](#)]
17. Smith, K.L.; Steven, M.D.; Colls, J.J. Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks. *Remote Sens. Environ.* **2004**, *92*, 207–217. [[CrossRef](#)]
18. Mahesh, S.; Jayas, D.S.; Paliwal, J.; White, N. Comparison of Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) Methods for Protein and Hardness Predictions using the Near-Infrared (NIR) Hyperspectral Images of Bulk Samples of Canadian Wheat. *Food Bioprocess Technol.* **2015**, *8*, 31–40. [[CrossRef](#)]
19. Fernandes, M.; Coelho, A.P.; Fernandes, C.; Silva, M.; Marta, C. Estimation of soil organic matter content by modeling with artificial neural networks. *Geoderma* **2019**, *350*, 46–51. [[CrossRef](#)]
20. Dunn, B.W.; Batten, G.D.; Beecher, H.G.; Ciavarella, S. The potential of near-infrared reflectance spectroscopy for soil analysis—A case study from the Riverine Plain of south-eastern Australia. *Aust. J. Exp. Agric.* **2002**, *42*, 607–614. [[CrossRef](#)]
21. Kooistra, L.; Wanders, J.; Epema, G.F.; Leuven, R.; Buydens, L. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Anal. Chim. Acta* **2003**, *484*, 189–200. [[CrossRef](#)]
22. Yu, X.; Liu, Q.; Wang, Y.; Liu, X.; Liu, X. Evaluation of MLR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula. *Catena* **2016**, *137*, 340–349. [[CrossRef](#)]
23. Sullivan, D.G.; Shaw, J.N.; Rickman, D. IKONOS Imagery to Estimate Surface Soil Property Variability in Two Alabama Physiographies. *Soil Sci. Soc. Am. J.* **2005**, *69*, 1789–1798. [[CrossRef](#)]
24. Goldshleger, N.; Chudnovsky, A.; Ben-Binyamin, R. Predicting salinity in tomato using soil reflectance spectra. *Int. J. Remote Sens.* **2013**, *34*, 6079–6093. [[CrossRef](#)]
25. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
26. Rocha Neto, O.C.D.; Teixeira, A.D.S.; de Oliveira Leão, R.A.; Moreira, L.C.J.; Galvão, L. Hyperspectral Remote Sensing for Detecting Soil Salinization Using ProSpecTIR-VS Aerial Imagery and Sensor Simulation. *Remote Sens.* **2017**, *9*, 42. [[CrossRef](#)]
27. Xu, S.; Shi, X.; Wang, M.; Zhao, Y. Determination of rice root density at the field level using visible and near-infrared reflectance spectroscopy. *Geoderma* **2016**, *267*, 174–184. [[CrossRef](#)]
28. Martens, H.; Næs, T. Multivariate Calibration. In *Chemometrics, Mathematics and Statistics in Chemistry*; Kowalski, B.R., Ed.; Springer: Dordrecht, The Netherlands, 1984; pp. 147–156.
29. Vajsová, B.; Fasbender, D.; Wirthardt, C.; Lemajic, S.; Devos, W.J.R.S. Assessing Spatial Limits of Sentinel-2 Data on Arable Crops in the Context of Checks by Monitoring. *Remote Sens.* **2020**, *12*, 2195. [[CrossRef](#)]
30. Sadeghi, M.; Babaeian, E.; Tuller, M.; Jones, S.B. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sens. Environ.* **2017**, *198*, 52–68. [[CrossRef](#)]
31. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [[CrossRef](#)]
32. Gao, Q.; Zribi, M.; Escorihuela, M.J.; Baghdadi, N. Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. *Sensors* **2017**, *17*, 1966. [[CrossRef](#)] [[PubMed](#)]
33. Steinberg, A.; Chabrilat, S.; Stevens, A.; Segl, K.; Foerster, S. Remote sensing prediction of common surface soil properties based on vis-nir airborne and simulated enmap imaging spectroscopy data: Prediction accuracy and influence of spatial resolution. *Remote Sens.* **2016**, *8*, 613. [[CrossRef](#)]
34. Nawar, S.; Buddenbaum, H.; Hill, J.; Kozak, J.; Mouazen, A.M. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res.* **2016**, *155*, 510–522. [[CrossRef](#)]
35. Nelson, D.W.; Sommers, L.E. A rapid and accurate method for estimating organic carbon in soil. *Proc. Indiana Acad. Sci.* **1975**, *84*, 456–462.
36. Dawson, T.P.; Curran, P.J. Technical note A new technique for interpolating the reflectance red edge position. *Int. J. Remote Sens.* **1998**, *19*, 2133–2139. [[CrossRef](#)]
37. Chong, I.G.; Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112. [[CrossRef](#)]
38. Hong, Y. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. *Geoderma* **2018**, *337*, 758–769. [[CrossRef](#)]
39. Wijewardane, N.K.; Ge, Y.; Morgan, C. Prediction of soil organic and inorganic carbon at different moisture contents with dry ground VNIR: A comparative study of different approaches. *Eur. J. Soil Sci.* **2016**, *67*, 605–615. [[CrossRef](#)]
40. Al-Abbas, A.H.; Swain, P.H.; Baumgardner, M.F. Relating Organic Matter and Clay Content to the Multispectral Radiance of Soils. *Soil Sci.* **1972**, *114*, 65–82. [[CrossRef](#)]
41. Zhao, H.; Liu, P.; Qiao, B.; Wu, K. The Spatial Distribution and Prediction of Soil Heavy Metals Based on Measured Samples and Multi-Spectral Images in Tai Lake of China. *Land* **2021**, *10*, 1227. [[CrossRef](#)]
42. Wei, L.; Yuan, Z.; Zhong, Y.; Yang, L.; Zhang, Y. An Improved Gradient Boosting Regression Tree Estimation Model for Soil Heavy Metal (Arsenic) Pollution Monitoring Using Hyperspectral Remote Sensing. *Appl. Sci.* **2019**, *9*, 1943. [[CrossRef](#)]
43. Zhou, H.; Deng, Z.; Xia, Y.; Fu, M.; Neurocomputing, F.J. A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing* **2016**, *216*, 208–215. [[CrossRef](#)]

44. Liu, J.; Dong, Z.; Xia, J.; Wang, H.; Meng, T.; Zhang, R.; Han, J.; Wang, N.; Xie, J. Estimation of soil organic matter content based on CARS algorithm coupled with random forest. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *258*, 119823. [[CrossRef](#)] [[PubMed](#)]
45. Zinck, G. Remote sensing of soil salinity: Potentials and constraints. *Remote Sens. Environ.* **2003**, *85*, 1–20.
46. Garcia, E.L.A. Detecting Soil Salinity in Alfalfa Fields using Spatial Modeling and Remote Sensing. *Soil Sci. Soc. Am. J.* **2008**, *72*, 201–211.
47. Kasim, N.; Tiyyip, T.; Abliz, A.; Nurmemet, I.; Sawut, R.; Maihemuti, B. Mapping and Modeling of Soil Salinity Using WorldView-2 Data and EM38-KM2 in an Arid Region of the Keriya River, China. *Photogramm. Eng. Remote Sens.* **2018**, *84*, 43–52. [[CrossRef](#)]
48. Liesenfeld, R.; Richard, J.-F. Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *J. Empir. Financ.* **2003**, *10*, 505–531. [[CrossRef](#)]
49. Xie, S.; Ding, F.; Chen, S.; Wang, X.; Li, Y.; Ma, K. Prediction of soil organic matter content based on characteristic band selection method. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *273*, 120949. [[CrossRef](#)]
50. Escribano, P.; Palacios-Orueta, A.; Oyonarte, C.; Chabrillat, S. Spectral properties and sources of variability of ecosystem components in a Mediterranean semiarid environment. *J. Arid Environ.* **2010**, *74*, 1041–1051. [[CrossRef](#)]
51. Fontán, J.; Calvache, S.; López-Bellido, R. Soil carbon measurement in clods and sieved samples in a Mediterranean Vertisol by Visible and Near-Infrared Reflectance Spectroscopy. *Geoderma* **2010**, *156*, 93–98. [[CrossRef](#)]