

Article

Improving the Estimation Accuracy of Soil Organic Matter Content Based on the Spectral Reflectance from Soils with Different Grain Sizes

Xayida Subi ^{1,2}, Mamattursun Eziz ^{1,2,*} and Ning Wang ²
¹ College of Geographical Science and Tourism, Xinjiang Normal University, Urumqi 830054, China; xayida104@126.com

² Xinjiang Laboratory of Arid Zone Lake Environment and Resources, Xinjiang Normal University, Urumqi 830054, China; 13659908171@163.com

* Correspondence: oasiseco@xjnu.edu.cn

Abstract: Accurate and rapid estimation of soil organic matter (SOM) content is of great significance for advancing precision agriculture. Compared with traditional chemical methods, the hyperspectral estimation is superior in rapidly estimating SOM content. Soil grain size affects soil spectral reflectance, thereby affecting the accuracy of hyperspectral estimation. However, the appropriate soil grain size for the hyperspectral analysis is nearly unknown. This study propose a best hyperspectral estimation method for determining SOM content of farmland soil in the Ibinur Lake Irrigation Area (ILIA) of the northwest arid zones of China. The original spectral reflectance of the 20-mesh (0.85 mm) and 60-mesh (0.25 mm) sieved soil were obtained, and the feature wavebands were selected using five types of spectral transformations. Then, hyperspectral estimation models were constructed based on the partial least squares regression (PLSR), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost) models. Results show that the SOM content had relatively higher correlation coefficient with spectral reflectance of the 0.85 mm sieved soil than that of the 0.25 mm sieved soil. The transformation of original spectral reflectance of soil effectively enhanced the spectral characteristics related to SOM content. Soil grain size obviously affected spectral reflectance and the accuracy of hyperspectral estimation models. The overall stability and estimation accuracy of RF model was significantly higher compared with the PLSR, SVM, and XGBoost. Finally, the RF model combined with the root mean first-order differentiation (RMSFD) of spectral reflectance of the 0.85 mm sieved soil ($R^2 = 0.82$, $RMSE = 2.37$, $RPD = 2.27$) was identified as the best method for estimating SOM content of farmland soil in the ILIA.

Keywords: soil grain size; SOM; soil spectrum; hyperspectral remote sensing; machine learning



Citation: Subi, X.; Eziz, M.; Wang, N. Improving the Estimation Accuracy of Soil Organic Matter Content Based on the Spectral Reflectance from Soils with Different Grain Sizes. *Land* **2024**, *13*, 1111. <https://doi.org/10.3390/land13071111>

Academic Editors: Yinnian Liu, Bing Zhang, Liangpei Zhang, Chenchao Xiao, Yueming Wang, Yongguang Zhang, Qingli Li, Yuwei Chen, Jianxin Jia and Mingyang Zhang

Received: 22 June 2024

Revised: 17 July 2024

Accepted: 19 July 2024

Published: 22 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soil organic matter (SOM) is an indicator for evaluating soil quality and is crucial for soil health [1]. SOM plays a pivotal role in soil fertility and agricultural productivity [2]. The significant role of SOM also extends to influencing the global carbon budget, mitigating environmental pollution, and influencing regional climate change [3,4]. Accurately assessing SOM content is essential for identifying areas requiring fertilization within sustainable soil management practices and sustainable agriculture [5]. In addition, the rapid estimation of SOM content is crucial for understanding the spatial distribution of soil fertility [6]. However, the dynamics of SOM content, influenced by various temporal and spatial factors, can result in variability and spatial heterogeneity of SOM, particularly in agricultural soils [7,8]. The differences in natural environment, soil types, and the complex heterogeneity of soils in different areas, further limit the estimation of SOM content [9]. Therefore, developing a rapid and effective monitoring technique for SOM content is challenging due to the limitations associated with the differences in regional soil environment.

Hyperspectral remote sensing technology is known for its feasibility in accurately and rapidly monitoring of soil properties. It has been used for estimating soil moisture [10], soil salt content [11], soil total nitrogen [12], heavy metals [13,14], and SOM content [15]. Hyperspectral estimation of SOM content by using soil spectral reflectance is of great significance for advancing modern precision agriculture [1]. The monitoring of soil spectral signatures and then hyperspectral estimation of SOM content supports wider environmental and agricultural endeavors [15]. However, hyperspectral estimation accuracy of SOM content depends on the quality of data processing and model construction, though its overall accuracy may be slightly lower than traditional methods. Nonetheless, hyperspectral estimation of SOM offers advantages in terms of higher temporal and spatial resolution, which are critical for effectively estimating SOM content [16].

Soil spectral reflectance is a comprehensive indicator of the spectral behavior of physical and chemical properties of soil [17]. Soil grain size leads to obvious differences in physical properties of soil, as well as characteristic changes in soil spectral reflectance [18]. It has been proven that the soil grain size significantly affects soil properties, including pore structure, fungal hyphae, as well as SOM [2]. In general, soil grain size also affects the spectral reflectance data of soil. The average spectral reflectance of soil varies with grain size across all wavelength bands, highlighting the influence of soil texture on spectral reflectance data [19]. Even for the same type of soil, different soil grain sizes affect their spectral characteristics [20,21]. In addition, the estimation accuracy of hyperspectral models based on spectral reflectance from soils with different grain sizes are different [18]. In summary, the soil grain size can affect the accuracy and stability of hyperspectral estimation models [22,23].

There is no consensus on the optimal soil grain size in SOM estimation, and the effects of soil grain size on the hyperspectral estimation accuracy of SOM content are nearly unknown, especially for farmlands in arid zones. The main objectives of this research are to (a) obtain the feature spectral wavebands for SOM contents of farmland soils in the northwest arid zones of China; (b) detect the effects of soil grain size on the spectral reflectance related to SOM; (c) clarify the effects of soil grain size on the hyperspectral estimation accuracy of SOM content; (d) identify a best hyperspectral method for rapidly estimating SOM content of farmland soil by means of the partial least squares regression (PLSR), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGBoost) models. Results of this study would offer a technical reference for selecting the optimum soil grain size during the hyperspectral estimation of SOM content.

2. Materials and Methods

2.1. Data Acquisition

This study was conducted in the Ibinur Lake Irrigation Area (ILIA) of the NW arid zones of China, with an experimental area range of 80°50′–83°00′ E and 44°20′–45°10′ N (Figure 1), which covers an area of 3300 km². Climate type of the experimental area belongs to the temperate continental dry climate, with an annual average precipitation of 105.27 mm, 80% of which is mainly concentrated from June to September. The annual average evaporation reaches 2221.3 mm, and the annual average temperature is 7.8 °C. Main soil types in the ILIA are irrigation desert soil, sandy soil, saline soil, and calcareous soil [24].

The field investigation, soil sampling, chemical analysis, and spectral measurement of this study were conducted in May of 2023. A total of 106 surface soil specimens from the top 20 cm (0–20 cm) layer were collected from farmlands (mainly cultivated land including cotton, corn, beets, and wheat) according with the soil sampling standard detailed in “NY/T 395—2000” [25]. The locations of sample sites are also shown in Figure 1.

Five sub-samples (approximately 400 g) were collected at each sample site (100 m × 100 m areas), and mixed as one typical soil sample (about 2 kg). After three days of air-drying in laboratory, the non-soil materials such as plant roots and stones in the collected samples were removed. The collected soil samples divided into two groups. One group was ground and then passed through a 60-mesh (grain size of ≤0.25 mm) sieve for determining the

SOM content, and the other group was ground and passed through a 20-mesh (grain size of ≤ 0.85 mm) and 60-mesh sieve, respectively, for measuring the soil spectral reflectance.

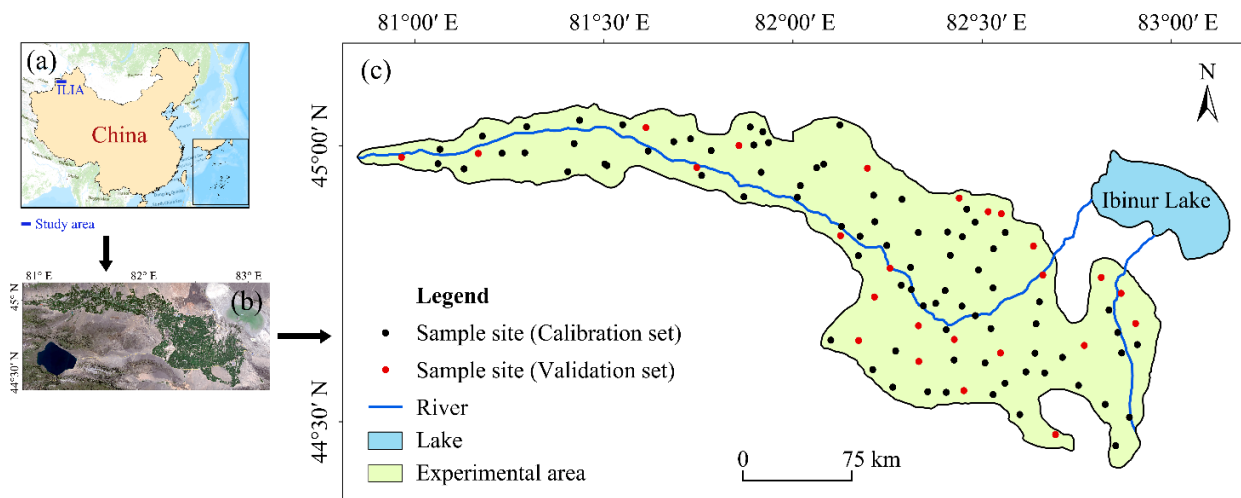


Figure 1. Map of the study area. (a) Location of ILIR; (b) Satellite image of ILIR; (c) Sample sites.

The SOM content was determined according with the National Standard of China detailed in NY/T 1121.6—2006 [26]. Soil spectral reflectance extraction is accomplished by a FieldSpec®3 portable object spectrometer (Analytical Spectral Devices, Boulder, CO, USA) with spectral resolution of 1 nm. The spectrometer was switched for 30 min before the spectral extraction, and it was corrected using a black and white board. The spectral reflectance data of 350–1750 nm (including visible light band (350–1000 nm) and near infrared band (900–1700 nm)) from two different grain size of soils were obtained. The changes in soil moisture may affect the predictive performance of the models, especially in scenarios where it is necessary to capture long-term trends or seasonal variations. The uncertainty of soil moisture may lead to an increase in the uncertainty of model predictions [10]. Therefore, when developing and using these models, the influence of soil moisture dynamics needs to be considered and the water absorption band should be removed to improve the accuracy and robustness of the models. The spectral data within 350–399 nm and 1301–1430 nm were excluded to reduce abnormal soil spectrum [27].

Each soil sample was scanned 10 times and 10 spectral reflectance curves were obtained, then the average value of them was taken as the final spectral reflectance data. Consequently, the Savitzky–Golay (S–G) algorithm was used to smooth the final spectral reflectance data to improve the signal-to-noise ratio [28]. Finally, the spectral reflectance curves of the 0.85 mm and 0.25 mm sieved soils after the above spectral pretreatment were obtained.

2.2. Spectral Feature Extraction

The original soil spectral reflectance data of the two different grain size soils were mathematically transformed into the first-order differentiation (FD), logarithmic FD (LTFD), root mean FD (RMSFD), reciprocal logarithmic FD (ATFD), and logarithmic reciprocal FD (RLFD) to enhance the spectral information related to SOM content, as well as to reduce unpredictable interference of environmental background [9]. To select the feature wavebands, the correlation analysis was performed between SOM content and original and transformed spectral reflectance data of soils with two different grain sizes. The Pearson's correlation coefficient (r) between SOM content and spectral reflectance were calculated, and the significance of correlation analysis was tested at the $p < 0.01$ level (two-tailed), whereas the threshold for the r was set at ± 0.248 [29]. Then, wavebands with absolute correlation coefficients more than 0.248 were selected as the feature wavebands, and used for following model construction. The correlation coefficient (r) can be determined according to special-

ized literature Zhong et al. The r value ranges from -1 to $+1$, and based on the absolute value of magnitude, the strength of the correlation is classified as: maximum correlation ($1.0 \geq r \geq 0.80$), strong correlation ($0.60 \leq r < 0.80$), moderate correlation ($0.40 \leq r < 0.60$), weak correlation ($0.20 \leq r < 0.40$), and weakest correlation ($0 \leq r < 0.20$) [30].

2.3. Model Construction

The samples of spectral extraction were divided into a calibration set (81 samples) and a validation set (25 samples) according with the Kennard Stone algorithm [1]. The calibration set was used to construct and train models, whereas the validation set was used to test and evaluate the model performance. Among various algorithms, the partial least squares regression (PLSR), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGBoost) models were adopted to establish the hyperspectral estimation models of SOM. PLSR is a statistical method designed to model the relationship between a set of independent variables (x) and a dependent variable (y), especially useful in scenarios of multicollinearity among the independent variables or when the number of predictors exceeds the number of observations [13]. SVM is a supervised learning algorithm for regression, it identifies a hyperplane that optimally represents the relationship between input and output variables while permitting a certain degree of deviation or error. SVM is effective for non-linear relationships and is applied in diverse research fields [31]. RF is an ensemble learning algorithm for regression analysis that enhances prediction accuracy and stability by combining multiple decision trees [32]. XGBoost is an advanced machine learning algorithm frequently used for regression and classification tasks, which is known for its efficiency for non-linear relationships [33]. The data of different soil types may have different distribution patterns and characteristics, which requires the model to have a strong generalization ability to adapt to unseen soil types. These models need to ensure their generalization ability through appropriate parameter adjustment and model validation. The PLSR, the SVM and the RF were constructed for predicting two different sieves SOM content in this study. Based on Python, the “random-state” of three models was set as 69. Due to the randomness of the RF model, the number of parameters (“n-estimators” and another “random-state”) will disturb the predictive performance of the model. Under the consideration of model performance, model running time, sample number and other factors, the number of parameters (“n-estimators” and another “random-state”) of the RF model was set in the range from 1 to 99.

2.4. Model Evaluation Indices

To compare the accuracy and reliability of the constructed hyperspectral estimation models, the R^2 (coefficient of determination), $RMSE$ (root mean square error), and RPD (residual predictive deviation) of validation set were used. These three indices were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_m - y_e)^2}{\sum_{i=1}^n (y_{ave} - y_e)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_m - y_e)^2}{n}} \quad (2)$$

$$RPD = S.D / RMSE \quad (3)$$

where y_m and y_e represent the ground-measured and hyperspectral estimated values of SOM content of sample i , respectively. The y_{ave} represents the average value of the ground-measured SOM contents, and n is total number of the collected soil samples.

In general, a robust hyperspectral estimation model has higher R^2 and RPD values but a lower $RMSE$ [30,34]. R^2 is used to assess the stability and estimation accuracy in reflectance spectroscopy studies [35]. The stability and prediction accuracy of R^2 is classified into five categories: $R^2 > 0.90$ indicates an “excellent prediction”, whereas $0.82 \leq R^2 < 0.90$ indicates a “good prediction”, $0.66 \leq R^2 < 0.82$ indicates an “approximate quantitative

prediction”, $0.50 \leq R^2 < 0.66$ indicates a “poor prediction”, and $R^2 < 0.50$ denotes an “unsuccessful prediction” [35,36]. The *RMSE* is used to evaluate the estimation quality of the model. The lower *RMSE* indicates the higher estimation quality of the model.

The *RPD* is defined as the ratio of the standard deviation (S.D) of the ground-measured data to the *RMSE* of the cross-validation. It is used to evaluate the estimation ability of the hyperspectral model [37]. The estimation ability of *RPD* is divided into five categories: $1.40 < RPD$ indicates a “poor model and/or estimation”, whereas $1.40 \leq RPD < 1.80$ indicates a “fair model and/or estimation”, $1.80 \leq RPD < 2.00$ indicates a “good model and/or estimation”, $2.00 \leq RPD < 2.50$ indicates a “very good quantitative model and/or estimation”, $RPD > 2.50$ indicates an “excellent model and/or estimation” [37,38].

3. Results

3.1. Descriptive Analysis of SOM

Table 1 details the basic statistical outcomes for the calibration, validation, and total sets of SOM content of farmland soil in the ILIA. The pH, salt content, and EC values of the collected soil samples were also given. The S.D and CV (coefficient of variation) values of SOM content used to quantify data variability. It can be seen that the SOM content of the total revealed a range from 6.04 to 31.60 g/kg, with an average value of 17.18 g/kg. The average pH value of the collected samples was 8.85, indicating an alkaline soil condition, whereas the average salt content was noted as 0.40 g/kg, and the average EC was measured at 1432.30 us/cm.

Table 1. Descriptive statistics of the soil properties.

Sample Type	<i>n</i>	SOM				pH	Salt (g/kg)	EC (us/cm)
		Range (g/kg)	Average (g/kg)	S.D (g/kg)	CV (%)			
Calibration set	81	6.04–31.60	16.94	4.99	29.48	8.84	0.39	1368.42
Validation set	25	10.40–28.10	17.96	5.38	29.94	8.88	0.45	1639.28
Total set	106	6.04–31.60	17.18	5.22	30.39	8.85	0.40	1432.30

Specifically, the average SOM content (16.94 g/kg for the calibration set vs. 17.96 g/kg for the validation set), S.D (4.99 g/kg for the calibration set vs. 5.38 g/kg for the validation set), and CV (29.48% for the calibration set vs. 29.94% for the validation set) were remarkably consistent between the calibration and validation sets. This similarity proves that the division of dataset in this study was appropriate, which is very applicable for the subsequent model construction [9].

3.2. Spectral Reflectance of Soils with Different Grain Sizes

Figure 2 illustrates the S-G smoothed soil spectral reflectance curves for the 0.85 mm (Figure 2a) and 0.25 mm (Figure 2b) sieved soil samples, respectively.

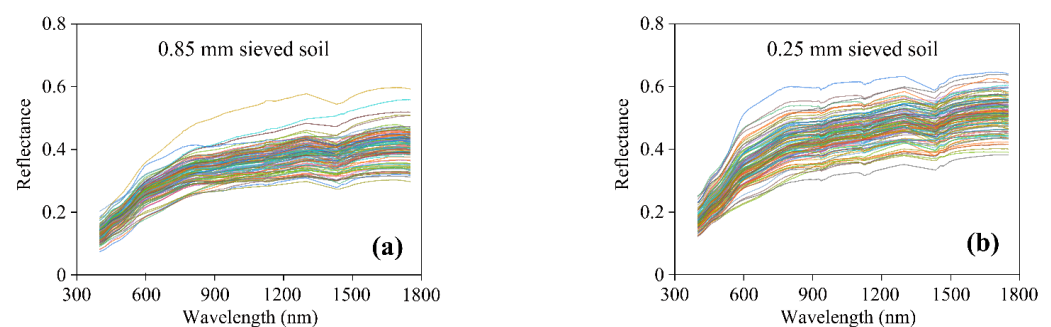


Figure 2. The original soil spectral reflectance curves processed through S-G smoothing. (Each line represents the spectrum of the soil sample ($n = 106$)). (a) 0.85 mm sieved soil; (b) 0.25 mm sieved soil.

As shown in Figure 2, the range of the original spectral reflectance of the 0.85 mm and 0.25 mm sieved soil samples were 0.074–0.598 and 0.122–0.647, respectively. The average reflectance value of the 0.85 mm and 0.25 mm sieved soil samples were 0.334 and 0.423, respectively. As shown here, the spectral reflectance curves of these two sieved soil exhibited a rapid increase and upward trend at the 400–800 nm wavelength range, while showing relatively stable at the 800–1300 nm and 1430–1750 nm wavelength range. The steeper slope observed in the 400–600 nm range may be attributed to the presence of iron in the soil. However, the trend of the spectral reflectance curves decreased at the 1300–1430 nm wavelength range. Besides, the spectral reflectance curves of these two sieved soil starts to change significantly at around 600 nm, and the spectral curves of investigated soil samples exhibited consistency in shape, trend, and the positions of main peaks and valleys. Generally, the spectrum of the 0.25 mm sieved soil was slightly higher than that of the 0.85 mm sieved soil (Figure 2). This result indicating the obvious effects of soil grain sizes on the soil spectral reflectance.

3.3. Correlations between Soil Spectral Reflectance and SOM Content

The correlation between the soil spectral reflectance (including the original and mathematically transformed spectral reflectance) and SOM content of the collected soil samples was analyzed (Figure 3). It can be seen that the SOM content exhibited relatively weak association with the original spectral reflectance (R) of both the 0.25 mm sieved soil ($r = -0.227$, at the weak correlation level) and the 0.85 mm sieved soil ($r = -0.415$, at the moderate correlation level) (Figure 3a). As for the original spectral reflectance, the 0.85 mm sieved soil exhibited better correlation with SOM content compared with the 0.25 mm sieved soil. It is evident that the original spectral reflectance curve of the 0.25 mm sieved soil does not meet the correlation test threshold of ± 0.248 . The original spectral reflectance of soil, especially the 0.25 mm sieved soil, had poor performance in the correlation between SOM content of farmland soil in the ILIA.

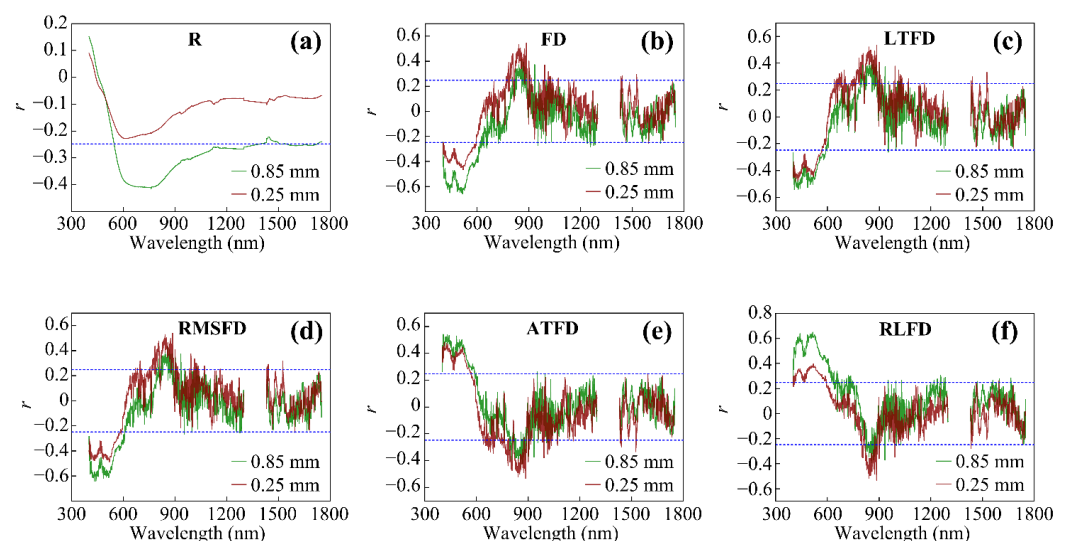


Figure 3. Correlation between the soil spectral reflectance and SOM content. (a) R ; (b) FD ; (c) $LTFD$; (d) $RMSFD$; (e) $ATFD$; (f) $RLFD$.

The correlations between SOM content and the five types of transformed spectral reflectance data including FD (Figure 3b), $LTFD$ (Figure 3c), $RMSFD$ (Figure 3d), $ATFD$ (Figure 3e), and $RLFD$ (Figure 3f) were significantly improved. Specifically, as for the 0.25 mm sieved soil, the absolute values of the maximum correlation coefficients between SOM content and FD , $LTFD$, $RMSFD$, $ATFD$, and $RLFD$ transformed soil spectral reflectance data were 0.544 (at 885 nm), 0.533 (at 885 nm), 0.541 (at 885 nm), 0.533 (at 885 nm), and 0.532 (at 885 nm), respectively, at a moderate correlation level ($0.40 \leq r < 0.60$). Meanwhile,

as for the 0.85 mm sieved soil, the maximum absolute correlation coefficients between SOM content and the FD, RMSFD, and RLFD transformed spectral data were 0.658 (at 515 nm), 0.641 (at 441 nm), and 0.651 (at 515 nm), respectively, at a strong correlation level ($0.60 \leq r < 0.80$), whereas the maximum absolute correlation coefficients between SOM content LTFD and ATFD transformed spectral data were 0.543 (at 407 nm) and 0.543 (at 407 nm), respectively, at a moderate correlation level.

Among the five types of mathematically transformed spectral reflectance data, the FD, RMSFD, and RLFD fall into the strong correlation level ($0.60 \leq r < 0.80$), whereas LTFD and ATFD fall into the moderate correlation ($0.40 \leq r < 0.60$). The results indicate that the FD, RMSFD, and RLFD transformation can effectively minimize environmental interference or eliminate baseline drift during spectral data collection, thereby enhancing spectral features of soil and facilitating the identification of effective wavebands. It can be concluded that the mathematical transformation of the original soil spectral reflectance can effectively enhance the correlation between the SOM content and soil spectral reflectance, which is consistent with results of related study [39]. Thus, applying appropriate mathematical transformations to the original spectrum constitutes an effective strategy for enhancing accuracy of hyperspectral estimation model. Notably, the FD, RMSFD, and RLFD transformation of the original soil spectral reflectance of the 0.85 mm sieved soil exhibited the more significant ($r > 0.6$) impact on the spectral characteristics of the soil.

In addition, the feature wavebands primarily located within the 407–885 nm range of visible light, and achieving a maximum correlation coefficient of -0.658 , at the strong correlation level. In the near-infrared waveband, there is no notable correlation between SOM content and spectral reflectance. The near-infrared band is typically highly sensitive to SOM content. Therefore, SOM displays relatively lower reflectivity in the near-infrared waveband due to its absorption of most near-infrared light [40]. Consequently, spectral reflectance in the near-infrared band exhibited a negative correlation with SOM content.

3.4. Model Construction and Evaluation

Based on the correlation coefficient between the soil spectral reflectance data and SOM content, wavebands with the absolute correlation coefficient value more than 0.248 were taken as the feature wavebands. Then, taking the selected feature wavebands as the independent variables (x), whereas taking the SOM content as the dependent variables (y), the PLSR, SVM, RF, and XGBoost algorithms were employed to construct hyperspectral estimation models of SOM content of farmland soil in the ILIA. Three evaluation indices including the R^2 , RMSE, and RPD for the constructed hyperspectral estimation models were obtained to compare the performance of constructed models (Table 2).

Table 2. Model evaluation indices of the hyperspectral estimation models.

Models	Indices	The 0.85 mm Sieved Soil					The 0.25 mm Sieved Soil				
		FD	LTFD	RMSFD	ATFD	RLFD	FD	LTFD	RMSFD	ATFD	RLFD
PLSR	R^2	0.56	0.61	0.59	0.54	0.62	0.41	0.44	0.40	0.44	0.48
	RMSE	3.55	3.51	3.48	3.74	3.40	4.11	4.08	4.16	4.08	3.90
	RPD	1.52	1.53	1.55	1.44	1.58	1.31	1.32	1.29	1.32	1.38
SVM	R^2	0.70	0.64	0.74	0.63	0.69	0.44	0.73	0.62	0.73	0.38
	RMSE	4.17	4.58	4.29	4.54	4.22	4.93	4.96	4.83	4.95	4.91
	RPD	1.29	1.17	1.25	1.19	1.27	1.09	1.08	1.11	1.09	1.10
RF	R^2	0.74	0.64	0.82	0.59	0.75	0.58	0.55	0.60	0.44	0.72
	RMSE	2.97	3.62	2.37	3.41	3.00	3.53	3.71	3.62	4.07	3.05
	RPD	1.81	1.49	2.27	1.58	1.79	1.52	1.45	1.49	1.32	1.76
XGBoost	R^2	0.39	0.28	0.37	0.13	0.37	0.10	0.10	0.03	0.06	0.21
	RMSE	3.76	3.52	3.79	3.81	4.16	4.54	4.53	5.21	3.78	4.45
	RPD	1.43	1.53	1.42	1.41	1.29	1.19	1.19	1.03	1.42	1.21

3.4.1. PLSR Model

Table 2 showed that, as for the 0.85 mm sieved soil, the R^2 values of the constructed PLSR model combined with FD, LTFD, RMSFD, ATFD, and RLFD were 0.56, 0.61, 0.59, 0.54, and 0.62, respectively, at the “poor prediction” level based on the classification criteria of the R^2 . Meanwhile, as for the 0.25 mm sieved soil, the R^2 values of the PLSR model combined with FD, LTFD, RMSFD, ATFD, and RLFD were 0.41, 0.44, 0.40, 0.44, and 0.48, respectively, indicating an “unsuccessful prediction” category. According to the R^2 values of the constructed PLSR models for two types of soil grain sizes, the stability and estimation accuracy of PLSR model by using the 0.85 mm sieved soil were relatively higher than that of the 0.25 mm sieved soil.

The ranges of $RMSE$ values of the constructed PLSR model across the five types of spectral transformations were 3.40–3.74 for the 0.85 mm sieved soil, whereas 3.90–4.16 for the 0.25 mm sieved soil. The $RMSE$ of the 0.85 mm sieved soil was lower than that of the 0.25 mm sieved soil. It indicates that the estimation quality of PLSR model by using the 0.85 mm grain size soil was higher than the 0.25 mm grain size soil. Moreover, the ranges of RPD of PLSR model across the five types of spectral transformations were 1.44–1.58 for the 0.85 mm sieved soil, whereas 1.29–1.38 for the 0.25 mm sieved soil. Based on the classification criteria of RPD , the estimation ability of the constructed PLSR model for the 0.85 mm sieved soil fall into a “fair model and/or estimation”, whereas PLSR model for the 0.25 mm sieved soil belonged to the “poor model and/or estimation” category. The above analysis indicates that the stability, estimation accuracy, estimation quality, and estimation ability of the constructed PLSR model were poor based on three model evaluation indices. However, the spectral reflectance data of the 0.85 mm sieved soil were better for constructing PLSR model compared with the 0.25 mm sieved soil. Therefore, the RLFD transformed spectral reflectance of the 0.85 mm sieved soil is superior when constructing hyperspectral estimation model of SOM content by using the PLSR model.

The scatter plot of SOM content for the ground-measured and predicted by the selected PLSR method (with the highest R^2 and RPD , and lowest $RMSE$) was exhibited in Figure 4. In Figure 4, the reasons for choosing the linear equation are based on the simplicity of the model, statistical foundations, performance metrics, data characteristics and predictive accuracy. Despite the differences between the predicted and actual values, previous related research indicates that the linear model is still regarded as a practical and effective predictive tool [11,39]. Results of the 0.85 mm and 0.25 mm sieved soils were compared. It can be seen that RLFD transformed spectral reflectance of the 0.85 mm sieved soil had a relatively higher performance. Therefore, the 0.85 mm-RLFD-PLSR ($R^2 = 0.62$, $RMSE = 3.40$, $RPD = 1.58$) can be identified as a better PLSR method for estimating SOM content of farmland soil in the ILIA. However, based on the evaluation indices of PLSR models, the overall performance of all the constructed PLSR models were not reliable.

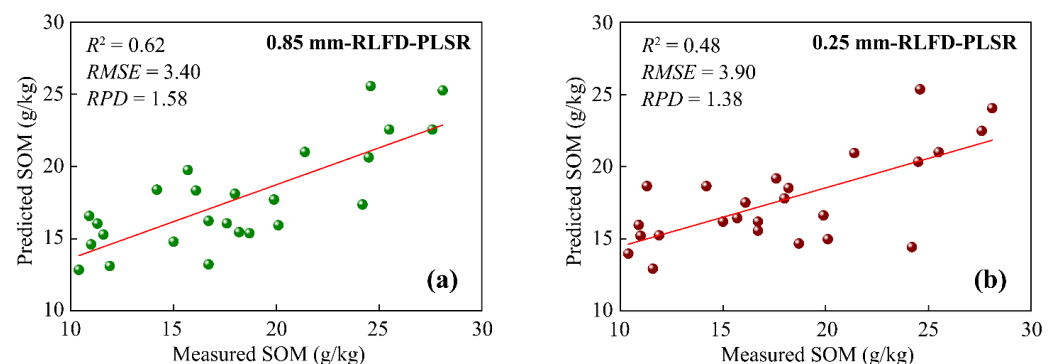


Figure 4. Comparison of the measured and predicted values of SOM content by PLSR modeling. (a) 0.85 mm-RLFD-PLSR; (b) 0.25 mm-RLFD-PLSR.

3.4.2. SVM Model

As shown in Table 2, as for the 0.85 mm sieved soil, the R^2 values of the constructed SVM model combined with FD, LTFD, RMSFD, ATFD, and RLFD were 0.70, 0.64, 0.74, 0.63, and 0.69, respectively, at the “approximate quantitative prediction” category for FD, RMSFD, and RLFD, and the “poor prediction” for other two spectral transformations. Meanwhile, as for the 0.25 mm sieved soil, the R^2 values of the SVM model combined with FD, LTFD, RMSFD, ATFD, and RLFD were 0.44, 0.73, 0.62, 0.73, and 0.38, respectively, with the “approximate quantitative prediction” for LTFD and ATFD, the “poor prediction” for RMSFD, and the “unsuccessful prediction” for other two spectral transformations. The ranges of RMSE values of the constructed SVM model across the five types of spectral transformations were 4.17–4.58 for the 0.85 mm sieved soil, whereas 4.83–4.96 for the 0.25 mm sieved soil (Table 2). The difference in the R^2 and RMSE values for these two different soil grain sizes were relatively small.

Moreover, the ranges of RPD values of SVM model across the five types of spectral transformations for both the 0.85 mm and 0.25 mm sieved soil were less than 1.40. Based on the classification criteria of RPD, the estimation ability of the constructed SVM model indicated a “poor model and/or estimation” level. Relatively speaking, spectral reflectance data of the 0.85 mm sieved soil were better for constructing SVM model compared with the 0.25 mm sieved soil.

The scatter plot of SOM content for the ground-measured and predicted by the selected SVM method (with the highest R^2 and RPD, and lowest RMSE) was exhibited in Figure 5. Results of the 0.85 mm and 0.25 mm sieved soils were compared. It is clear that the RMSFD transformed spectral reflectance of the 0.85 mm sieved soil had a relatively higher performance in estimation of SOM content. However, the 0.85 mm-RMSFD-SVM ($R^2 = 0.74$, $RMSE = 4.29$, $RPD = 1.25$) can be identified as a better SVM method for estimating SOM content of farmland soil in the ILIA. Overall, based on the model evaluation indices, the overall performance of all the constructed SVM models were also not reliable.

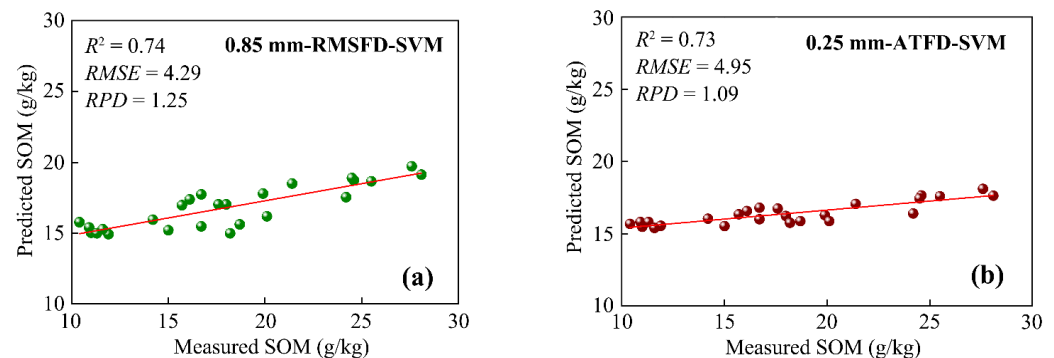


Figure 5. Comparison of the measured and predicted values of SOM content by SVM modeling. (a) 0.85 mm-RMSFD-SVM; (b) 0.25 mm-ATFD-SVM.

3.4.3. RF Model

As for the 0.85 mm sieved soil, the R^2 values of the constructed RF model combined with FD, LTFD, RMSFD, ATFD, and RLFD were 0.74, 0.64, 0.82, 0.59, and 0.75, respectively, with a “good prediction” for RMSFD, an “approximate quantitative prediction” for FD and RLFD, and a “poor prediction” for other two spectral transformations (Table 2). Meanwhile, as for the 0.25 mm sieved soil, the R^2 values of the RF model combined with FD, LTFD, RMSFD, ATFD, and RLFD were 0.58, 0.55, 0.60, 0.44, and 0.72, respectively, with an “approximate quantitative prediction” for RLFD, an “unsuccessful prediction” for ATFD, and a “poor prediction” for other three spectral transformations. According to the R^2 values of RF models for these two different grain size of soils, the stability and estimation accuracy of RF model by using the 0.85 mm sieved soil were obviously higher than that of the 0.25 mm sieved soil.

The ranges of *RMSE* values of the constructed RF model across the five types of spectral transformations were 2.37–3.00 for the 0.85 mm sieved soil, whereas the 3.05–4.07 for the 0.25 mm sieved soil. The *RMSE* of the 0.85 mm sieved soil was obviously lower than that of the 0.25 mm sieved soil. It proves that the estimation quality of RF model by using the 0.85 mm grain size soil was higher than the 0.25 mm grain size soil.

It should be noted that the *RPD* values of the RF model combined with FD, LTFD, RMSFD, ATFD, and RLFD of the 0.85 mm sieved soil were 1.81, 1.49, 2.27, 1.58, and 1.79, respectively (Table 2). Based on the classification criteria of *RPD*, the estimation ability of RF model indicated a “very good quantitative model and/or estimation” for RMSFD, a “good model and/or estimation” for FD, and a “fair model and/or estimation” for other three spectral transformations. The *RPD* values of the RF model combined with FD, LTFD, RMSFD, ATFD, and RLFD of the 0.25 mm sieved soil were 1.52, 1.45, 1.49, 1.32, and 1.76, respectively, with a “fair model and/or estimation” for FD, LTFD, RMSFD, and RLFD transformations, and a “poor model and/or estimation” for ATFD transformation.

Based on three model evaluation indices, the stability, estimation accuracy, estimation quality, and estimation ability of RF model were superior compared with PLSR and SVM. As analyzed here, spectral reflectance data of the 0.85 mm sieved soil were relatively better for constructing RF model compared with the 0.25 mm sieved soil. Therefore, the RMSFD transformed spectral reflectance of the 0.85 mm sieved soil is superior when constructing hyperspectral estimation model of SOM content by using the RF model.

The scatter plot of SOM content for the ground-measured and predicted by the selected RF method (with the highest R^2 and *RPD*, and lowest *RMSE*) was exhibited in Figure 6. Results of the 0.85 mm and 0.25 mm sieved soils were compared. It can be seen that the RMSFD transformed spectral reflectance of the 0.85 mm sieved soil had a higher performance in estimation of SOM content. Overall, the 0.85 mm-RMSFD-RF ($R^2 = 0.82$, *RMSE* = 2.37, *RPD* = 2.27) can be identified as the best RF method for estimating SOM content of farmland soil in the ILIA. However, the RF model is very applicable for estimating SOM content.

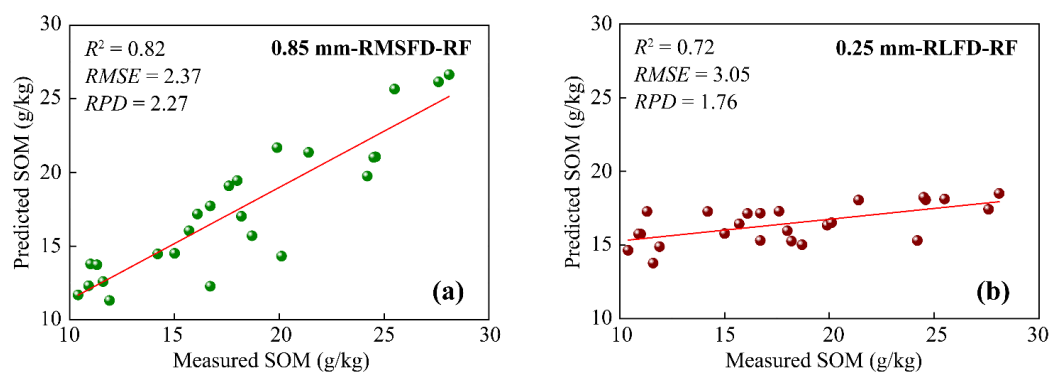


Figure 6. Comparison of the measured and predicted values of SOM content by RF modeling. (a) 0.85 mm-RMSFD-RF; (b) 0.25 mm-RLFD-RF.

3.4.4. XGBoost Model

As given in Table 2, the R^2 values of the XGBoost model combined with five types of spectral transformation for both the 0.85 mm and 0.25 mm sieved soils were less than 0.39, indicating an “unsuccessful prediction”. The R^2 values for the 0.85 mm grain size soil were higher than that the 0.25 mm grain size soil. Besides, the ranges of *RMSE* values of XGBoost model across the five types of spectral transformations were 3.52–4.16 for the 0.85 mm sieved soil, while 3.78–5.21 for the 0.25 mm sieved soil.

The *RMSE* of the 0.85 mm sieved soil was lower than that of the 0.25 mm sieved soil. However, the estimation quality of XGBoost model by using the 0.85 mm grain size soil was relatively better than the 0.25 mm grain size soil. The ranges of *RPD* values of the XGBoost model across the five types of spectral transformations were smaller than that

of PLSR, SVM, and RF, with a “poor (or fair) model and/or estimation”. It indicates that the stability, estimation accuracy, estimation quality, and estimation ability of the XGBoost model were very poor.

The scatter plot of SOM content for the ground-measured and predicted by the selected XGBoost method (with the highest R^2 and RPD , and lowest $RMSE$) was exhibited in Figure 7. Results of the 0.85 mm and 0.25 mm sieved soils were also compared. Figure 7 illustrated that the FD transformed spectral reflectance of the 0.85 mm sieved soil had a relatively higher performance in estimation of SOM content. Therefore, the 0.85 mm-FD-XGBoost ($R^2 = 0.39$, $RMSE = 3.76$, $RPD = 1.43$) can be selected as a better XGBoost method for estimating SOM content of farmland soil in the ILIA. Overall, based on the model evaluation indices, the overall performance of all the constructed XGBoost models were very poor and not reliable.

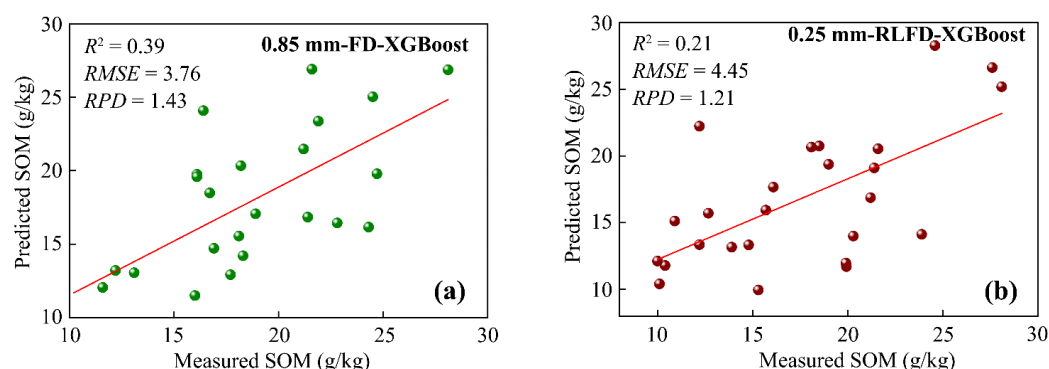


Figure 7. Comparison of the measured and predicted values of SOM content by XGBoost modeling. (a) 0.85 mm-FD-XGBoost; (b) 0.25 mm-RLFD- XGBoost.

4. Discussion

In this work, the overall performance of the constructed hyperspectral estimation models can be ranked as: RF > SVM > PLSR > XGBoost. It should be noted that the RF model had a significantly higher R^2 and RPD value and relatively lower $RMSE$ values compared with PLSR, SVM, and XGBoost models. Therefore, the RF was selected the best model for predicting SOM content of farmlands in the ILIA. Results of this study are inconsistent with the research findings of some previous studies. For example, Zheng et al. reported that the PLSR had the best estimation accuracy of SOM content of coastal soil [41]. Wei et al. constructed a hyperspectral inversion model for SOM content of farmland soils and suggested that the AdaBoost algorithm had the best accuracy compared with the Ridge Regression (RR), Kernel RR (KRR), and Bayesian RR (BRR) [17]. Zhang et al. also constructed a SOM estimation model, and their results showed that the estimation accuracy of SVM surpassed than that of the back propagation neural network (BPNN) [42]. Recently, Li et al. suggested that the CNN (convolutional neural network) had high accuracy in predicting SOM content [5]. Bai et al. indicated that the PLSR model based on outer-product analysis (OPA) achieved the best estimation accuracy of SOM content [1].

It is worth noting that the RF model had better accuracy than the PLSR model for SOM content in the Ogan-Kuqa River Oasis of NW arid zones of China [43]. This result is consistent with our research findings. Similarly, the best accuracy for hyperspectral estimation of heavy metals in farmland soils was obtained by using the RF Model [44]. However, due to the effects of the regional geographical environment and physicochemical features of various soil types in different areas, the optimal hyperspectral model for estimating SOM content varies considerably [9].

From the perspective of inversion accuracy, most preprocessed spectra have higher modeling accuracy than the original spectra. This is because in the process of obtaining spectral information, external interference can introduce noise, which hinders the accurate reflection of the spectral characteristics of features. However, spectral preprocessing

techniques can reduce spectral noise and highlight spectral feature information (Figure 2). The original spectral reflectance of the 0.25 mm sieved soil was slightly higher than that of the 0.85 mm sieved soil. This result indicates that the smaller the soil grain size, the higher the soil spectral reflectance. The reason is that the void among the smaller soil grains are smaller than the bigger soil grains, which enhance the spectral reflectance of soil [2]. Moreover, the lower spectral reflectance of soils with the 0.85 mm grain size may be attributed to light scattering and changes in optical path length [20]. Soil with a sieved of 0.25 mm has a smaller porosity, which reduces the absorption and scattering of light, thereby increasing the reflectance. On the other hand, soil with a sieved of 0.85 mm has a larger porosity, and the path of light within the soil is longer, which increases the absorption and scattering of light, leading to a decrease in reflectance. Therefore, when analyzing the spectral reflectance of soil, it is very important to consider the influence of soil grain sizes. By studying the spectral characteristics of soils with different grain sizes, scientific bases can be provided for soil classification, quality assessment, and land resource management.

At present, it is difficult to fully extract the effective feature wavebands by using not-sieved soil samples, which limits the estimation ability of hyperspectral models. However, by selecting feature wavebands obtained from soils with appropriate grain size, deeper feature wavebands extraction can be achieved and the constructed hyperspectral estimation model has better generalization, which is consistent with our research findings [20]. Based on the R^2 , RPD , and $RMSE$ values of the constructed hyperspectral estimation models by using soils with different grain sizes, a significantly higher R^2 , RPD values and lower $RMSE$ values were observed for the 0.25 mm sieved soil. The correlations between soil spectral reflectance and SOM content were effectively improved by using the spectrum of the 0.85 mm sieved soil samples. Then the use of the feature wavebands of the 0.85 mm sieved soil significantly improved the stability and prediction ability of the constructed hyperspectral estimation model in this study. It is verified that the soil grain size affects the stability, estimation accuracy, estimation quality, and estimation ability of hyperspectral estimation of SOM content, and the 0.85 mm sieved soil is more suitable for spectral measurement and following model construction. In the case where higher grain size results in a higher R^2 value, the relationship may exist with the physical and chemical properties of the soil as follows: (1) Porosity structure: Larger grain sizes may lead to changes in the size and distribution of soil pores. A more uniform or suitable pore structure may make the related physical processes more regular, thereby improving the degree of fit of the model to the data, i.e., a higher R^2 value. (2) Particle arrangement: When grain sizes are larger, the arrangement of particles may be more orderly, which affects the soil's permeability, water retention, and other physical properties. This can make the relationship between these properties and other factors clearer, thus increasing the R^2 value. (3) Nutrient adsorption and release: Larger grains may affect the soil's ability to adsorb and release nutrients. More stable or regular nutrient dynamics may allow the model to better explain the data, leading to an increase in the R^2 value. However, to accurately determine the relationship between grain size and R -squared, as well as the specific correlation with the physical and chemical properties of the soil, further experimental research and detailed data analysis are required [20,22,23,45].

Finally, RF model based on RMSFD transformed spectral reflectance of the 0.85 mm sieved soil (0.85 mm-RMSFD-RF) can realize the effective fusion of spectral features, which can make up for the limitation of single data features, and further improve the stability and estimation ability of the constructor model. Therefore the 0.85 mm-RMSFD-RF method ($R^2 = 0.82$, $RMSE = 2.37$, $RPD = 2.27$) is the best hyperspectral estimation method of SOM content of farmland soil in the ILIA.

Based on the measured and predicted SOM content, the actual distribution and the estimated distribution patterns of SOM content based on the selected PLSR, SVM, RF, and XGBoost methods were mapped using the Ordinary Kriging (OK) interpolation method and geostatistical analysis method (Figure 8).

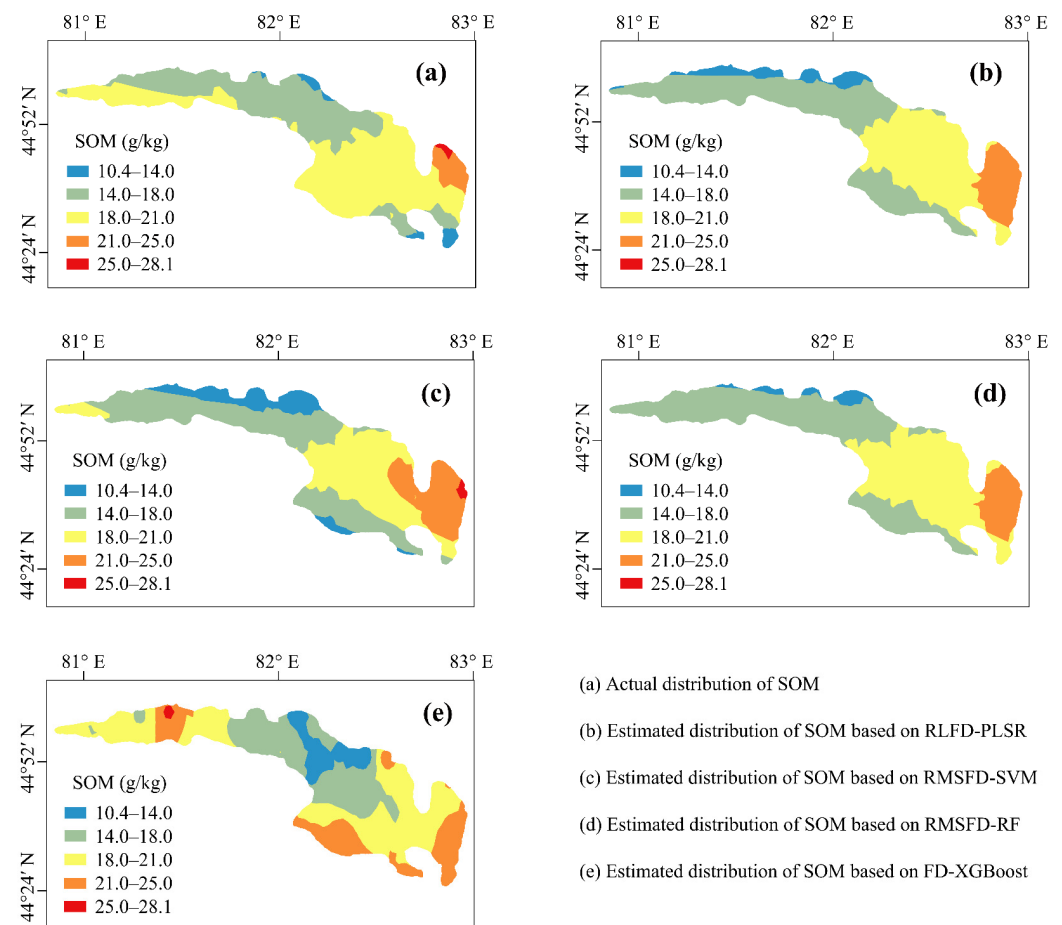


Figure 8. Spatial distribution map of SOM content based on the measured and predicted values.

It can be observed that the spatial distribution patterns of the estimated SOM content via the 0.85 mm-RMSFD-RF method (Figure 8d) was most similar with the actual distribution of SOM content (Figure 8a), with higher SOM content in the eastern and northern parts, and lower SOM content in the central parts of the ILIA. However, the spatial distribution patterns of SOM content estimated by the 0.85 mm-RLFD-PLSR (Figure 8b), the 0.85 mm-RMSFD-SVM (Figure 8c), and the 0.85 mm-FD-XGBoost (Figure 8e) methods significantly varied from the actual distribution. This result further proves that the 0.85 mm-RMSFD-RF is the best hyperspectral estimation method for SOM content of farmland soil in the ILIA. However, this work is a regional study, and the applicability of hyperspectral estimation models varies across different geographic regions due to differences in soil types and physical and chemical properties of soil [46]. Therefore, future studies are needed to explore whether the overall estimation accuracy and stability of hyperspectral estimation models using the spectrum of the 0.85 mm sieved soil was also the optimal method to other regions.

5. Conclusions

This study investigated the effects of soil grain size on the accuracy of hyperspectral estimation of SOM content of farmland soil in arid zones. The following conclusions were drawn:

- (1) The smaller the soil grain size, the higher the spectral reflectance. In the original spectral reflectance curve, the wavebands sensitive to SOM were primarily found within the 350–550 nm range ($r > 0.5$, $p < 0.01$), exhibiting a negative correlation with SOM content.
- (2) The spectral reflectance of the 0.85 mm sieved soil demonstrated relatively higher correlation coefficients with SOM content than the 0.25 mm sieved soil.

- (3) The mathematical transformation of original spectral reflectance of soil can effectively enhance the spectral characteristics related to the SOM content, and soil grain size obviously effect the accuracy of hyperspectral estimation model of SOM content.
- (4) The overall estimation accuracy and stability of the constructed hyperspectral estimation models in this study can be ranked as: RF > SVM > PLSR > XGBoost. The RF model had a significantly higher R^2 and RPD value and relatively lower RMSE values compared with the PLSR, SVM, and XGBoost models. The 0.85 mm-RMSFD-RF method ($R^2 = 0.82$, $RMSE = 2.37$, $RPD = 2.27$) was selected as the best model for estimating SOM content of farmland soil in the ILIA.

Findings of this work offer a technical reference for the hyperspectral estimation of the SOM content of farmland soil in arid zones. However, further investigation should be considered in future studies.

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

Funding: This research is funded by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2023D01E08) and the National Natural Science Foundation of China (U2003301).

Data Availability Statement: Data will be available upon request to the corresponding author.

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

References

1. Bai, Y.; Yang, W.; Wang, Z.; Cao, Y.; Li, M. Improving the estimation accuracy of soil organic matter based on the fusion of near-infrared and Raman spectroscopy using the outer-product analysis. *Comput. Electron. Agric.* **2024**, *219*, 108760. [\[CrossRef\]](#)
2. Chen, Y.; Wang, J.; Liu, G.; Yang, Y.; Liu, Z.; Deng, H. Hyperspectral estimation model of forest soil organic matter in northwest Yunnan Province, China. *Forests* **2019**, *10*, 217. [\[CrossRef\]](#)
3. He, Y.; Yang, M.; Huang, R.; Wang, Y.; Ali, W. Soil organic matter and clay zeta potential influence aggregation of a clayey red soil (Ultisol) under long-term fertilization. *Sci. Rep.* **2021**, *11*, 20498. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Zhao, L.; Fang, Q.; Hong, H.; Algeo, T.J.; Lu, A.; Yin, K.; Wang, C.; Liu, C.; Chen, L.; Xie, S. Pedogenic-weathering evolution and soil discrimination by sensor fusion combined with machine-learning-based spectral modeling. *Geoderma* **2022**, *409*, 115648. [\[CrossRef\]](#)
5. Li, H.; Ju, W.L.; Song, Y.M.; Cao, Y.Y.; Yang, W.; Li, M.Z. Soil organic matter content prediction based on two-branch convolutional neural network combining image and spectral features. *Comput. Electron. Agric.* **2024**, *217*, 108561. [\[CrossRef\]](#)
6. Hong, Y.S.; Chen, S.C.; Zhang, Y.; Chen, Y.Y.; Yu, L.; Liu, Y.F.; Liu, Y.L.; Cheng, H.; Liu, Y. Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine. *Sci. Total Environ.* **2018**, *644*, 1232–1243. [\[CrossRef\]](#)
7. Six, J.; Paustian, K. Aggregate-associated soil organic matter as an ecosystem property and a measurement tool. *Soil Biol. Biochem.* **2014**, *68*, 4–9. [\[CrossRef\]](#)
8. Keesstra, S.; Pereira, P.; Novara, A.; Brevik, E.C.; Azorin-Molina, C.; Parras-Alcántara, L.; Jordan, A.; Cerda, A. Effects of soil management techniques on soil water erosion in apricot orchards. *Sci. Total Environ.* **2016**, *357*, 551–552. [\[CrossRef\]](#)
9. Xayida, S.; Mamattursun, E.; Zhong, Q.; Li, X.G. Estimating the chromium concentration of farmland soils in an arid zone from hyperspectral reflectance by using partial least squares regression methods. *Ecol. Indic.* **2024**, *161*, 111987.
10. Jiang, X.Q.; Luo, S.J.; Ye, Q.; Li, X.C.; Jiao, W.H. Hyperspectral estimates of soil moisture content incorporating harmonic indicators and machine learning. *Agriculture* **2022**, *12*, 1188. [\[CrossRef\]](#)
11. Jiang, X.F.; Duan, H.C.; Liao, J.; Guo, P.L.; Huang, C.H.; Xue, X.A. Estimation of soil salinization by machine learning algorithms in different arid regions of northwest China. *Remote Sens.* **2022**, *14*, 347. [\[CrossRef\]](#)
12. Lin, L.X.; Gao, L.P.; Xue, F.C.; Wang, X.Y.; Zhang, S.R. Hyperspectral analysis of total nitrogen in soil using a synchronized decoloring fuzzy measured value method. *Soil Till. Res.* **2020**, *202*, 104658. [\[CrossRef\]](#)
13. Wang, Y.; Zhang, X.; Sun, W.; Wang, J.; Ding, S.; Liu, S. Effects of hyperspectral data with different spectral resolutions on the estimation of soil heavy metal content: From ground-based and airborne data to satellite-simulated data. *Sci. Total Environ.* **2022**, *838*, 156129. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ye, M.; Zhu, L.; Li, X.; Ke, Y.; Huang, Y.; Chen, B.; Yu, H.; Li, H.; Feng, H. Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. *Sci. Total Environ.* **2023**, *858*, 159798. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Shabtai, I.A.; Wilhelm, R.C.; Schweizer, S.A.; Höschen, C.; Buckley, D.H.; Lehmann, J. Calcium promotes persistent soil organic matter by altering microbial transformation of plant litter. *Nature Commun.* **2023**, *14*, 6609. [[CrossRef](#)] [[PubMed](#)]
16. Khosravi, V.; Ardejani, F.D.; Yousefi, S.; Aryafar, A. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods. *Geoderma* **2018**, *318*, 29–41. [[CrossRef](#)]
17. Wei, L.; Yuan, Z.; Wang, Z.; Zhao, L.; Zhang, Y.; Lu, X.; Cao, L. Hyperspectral inversion of soil organic matter content based on a combined spectral index model. *Sensors* **2020**, *20*, 2777. [[CrossRef](#)] [[PubMed](#)]
18. Xu, S.X.; Wang, M.Y.; Shi, X.Z.; Yu, Q.B.; Zhang, Z.Q. Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. *Sci. Total Environ.* **2021**, *754*, 142135. [[CrossRef](#)] [[PubMed](#)]
19. Ma, C.; Shen, G.; Wang, Z.; Wang, Z. Analysis of spectral characteristics for different soil particle sizes. *Chin. J. Soil Sci.* **2015**, *46*, 292–298. (In Chinese)
20. Sadeghi, M.; Babaeian, E.; Tuller, M.; Jones, S.B. Particle size effects on soil reflectance explained by an analytical radiative transfer model. *Remote Sens. Environ.* **2018**, *210*, 375–386. [[CrossRef](#)]
21. An, X.; Li, M.; Zheng, L.; Hong, S. Eliminating the interference of soil moisture and particle size on predicting soil total nitrogen content using a NIRS-based portable detector. *Comput. Electron. Agric.* **2015**, *112*, 47–53. [[CrossRef](#)]
22. Bao, Y.; He, Y.; Fang, H.; Annia, G.P. Spectral characterization and N content prediction of soil with different particle size and moisture content. *Spectro. Spec. Anal.* **2007**, *27*, 62.
23. Si, H.; Yao, Y.; Wang, D.; Liu, Y. Influence of soil particle size on the estimate of soil organic matter by hyperspectral spectroscopy. *Chin. Agric. Sci. Bull.* **2015**, *31*, 173–178. (In Chinese)
24. Muyassar, M.; Mamattursun, E.; Wang, L.L.; Xayida, S.; Wang, N.; Hu, Y.L. Pollution and ecological risk assessment of metal elements in groundwater in the Ibinur Lake Basin of NW China. *Water* **2023**, *15*, 4071. [[CrossRef](#)]
25. NY/T 395—2000; Procedural Regulations Regarding the Environment Quality Monitoring of Soil. Standards Press of China: Beijing, China, 2000. (In Chinese)
26. NY/T 1121.6—2006; Soil Testing—Part 6: Method for Determination of Soil Organic Matter. Standards Press of China: Beijing, China, 2006. (In Chinese)
27. Wei, L.; Zhang, Y.; Lu, Q.; Yuan, Z.; Li, H.; Huang, Q. Estimating the spatial distribution of soil total arsenic in the suspected contaminated area using UAV-Borne hyperspectral imagery and deep learning. *Ecol. Indic.* **2021**, *133*, 108384. [[CrossRef](#)]
28. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [[CrossRef](#)]
29. Zhong, Q.; Eziz, M.; Sawut, R.; Ainiwaer, M.; Li, H.; Wang, L. Application of a hyperspectral remote sensing model for the inversion of nickel content in urban soil. *Sustainability* **2023**, *15*, 13948. [[CrossRef](#)]
30. Cao, X.; Zhang, J.; Meng, H.; Lai, Y.; Xu, M. Remote sensing inversion of water quality parameters in the Yellow River Delta. *Ecol. Indic.* **2023**, *155*, 110914. [[CrossRef](#)]
31. Pal, M.; Foody, G.M. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2297–2307. [[CrossRef](#)]
32. Elfatih, M.A.; Onesimo, M.; Elhadi, A.; Riyad, I. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS J. Photo. Remote Sens.* **2014**, *88*, 48–59.
33. Jia, Y.; Jin, S.G.; Savi, P.; Gao, Y.; Tang, J.; Chen, Y.X.; Li, W.M. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation. *Remote Sens.* **2019**, *11*, 1655. [[CrossRef](#)]
34. Liu, W.; Li, M.; Zhang, M.; Long, S.; Guo, Z.; Wang, H.; Li, W.; Wang, D.; Hu, Y.; Wei, Y.; et al. Hyperspectral inversion of mercury in reed leaves under different levels of soil mercury contamination. *Environ. Sci. Pollut. Res. Inter.* **2020**, *27*, 22935–22945. [[CrossRef](#)] [[PubMed](#)]
35. Sun, Y.S.; Chen, S.S.; Dai, X.M.; Li, D.; Jiang, H.; Jia, K. Coupled retrieval of heavy metal nickel concentration in agricultural soil from spaceborne hyperspectral imagery. *J. Hazard. Mater.* **2023**, *446*, 130722. [[CrossRef](#)] [[PubMed](#)]
36. Vohland, M.; Besold, J.; Hill, J.; Fründ, H.C. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma* **2011**, *166*, 198–205. [[CrossRef](#)]
37. 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](#)]
38. Bian, Z.J.; Sun, L.N.; Tian, K.; Liu, B.L.; Huang, B.; Wu, L.H. Estimation of multi-media metal(loid)s around abandoned mineral processing plants using hyperspectral technology and extreme learning machine. *Environ. Sci. Pollut. Res.* **2023**, *30*, 19495–19512. [[CrossRef](#)] [[PubMed](#)]
39. Dai, X.; Liu, S.; Xiang, T.; Fu, T.; Feng, H.; Xiao, L.; Wang, Z.; Yao, Y.; Zhao, R.; Yang, X. Hyperspectral imagery reveals large spatial variations of heavy metal content in agricultural soil: A case study of remote-sensing inversion based on Orbita hyperspectral satellites (OHS) imagery. *J. Clean. Product.* **2022**, *380*, 134878. [[CrossRef](#)]
40. Fang, S.; Yang, M.; Zhao, X.; Guo, X. Spectral characteristics and quantitative estimation of SOM in red soil typical of Ji'an County, Jiangxi Province. *Acta Pedo. Sin.* **2014**, *51*, 1003–1010. (In Chinese)
41. Zheng, G.H.; Ryu, D.R.; Jiao, C.X.; Hong, C.Q. Estimation of organic matter content in coastal soil using reflectance spectroscopy. *Pedosphere* **2016**, *26*, 130–136. [[CrossRef](#)]
42. Zhang, S.; Lu, X.; Nie, G.G.; Li, Y.R.; Shao, Y.T.; Tian, Y.Q.; Fan, L.Q.; Zhang, Y.J. Estimation of soil organic matter in coastal wetlands by SVM and BP based on hyperspectral remote sensing. *Spectro. Spec. Anal.* **2020**, *40*, 556–561.

43. Zhou, Q.; Ding, J.L.; Ge, X.Y.; Li, K.; Zhang, Z.O.; Gu, Y.S. Estimation of soil organic matter in the Ogan-Kuqa River Oasis, Northwest China, based on visible and near-infrared spectroscopy and machine learning. *J. Arid Land* **2023**, *15*, 19–204. [[CrossRef](#)]
44. Tan, K.; Wang, H.; Chen, L.; Du, Q.; Du, P.; Pan, C. Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *J. Hazard. Mater.* **2020**, *382*, 120987. [[CrossRef](#)] [[PubMed](#)]
45. Wu, H.Q.; Fan, Y.M.; He, J.; Jin, G.L.; Xie, Y.; Chai, D.P.; He, L. Response of soil hyperspectral characteristics of different particle sizes to soil. *Acta Agrestia Sin.* **2014**, *22*, 266.
46. Yang, P.; Hu, J.; Hu, B.; Luo, D.; Peng, J. Estimating soil organic matter content in desert areas using in situ hyperspectral data and feature variable selection algorithms in southern Xinjiang, China. *Remote Sens.* **2022**, *14*, 5221. [[CrossRef](#)]

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