*remote sensing*

*Article*

# Spatial Prediction of Soil Texture at the Field Scale Using Synthetic Images and Partitioning Strategies

**Yiang Wang [1,2], Shinai Ma [1], Shuai Bao [1], Yuxin Ma [1], Yan Zhang [1], Dianyao Wang [3], Yihan Ma [1] and Huanjun Liu [1,*]**

1 State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China; wangyiang@iga.ac.cn (Y.W.)
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 College of Resources and Environmental Sciences, Northeast Agricultural University, Harbin 150030, China; zxshirley@neau.edu.cn
* Correspondence: liuhuanjun@iga.ac.cn

## Highlights

**What are the main findings?**

- We reveal how different image compositing elements—data type, time window, and compositing modes—affect the prediction accuracy of soil texture at the field scale, and demonstrate the applicability of a local partitioning regression strategy.
- This study innovatively constructs a comprehensive strategy of "moisture spectral indices + specific compositing time window + specific compositing mode + soil type partitioning" for high-precision predicting soil texture at the field scale.

**What are the implications of the main findings?**

- In the field of smart agriculture, soil property data at the field scale drive precision decision-making for agricultural inputs such as seeds and fertilizers. Currently, most soil texture prediction studies are conducted at large scales, which often fail to characterize field-scale conditions effectively. This comprehensive strategy delivers robust and reliable accuracy, shows great potential for small-scale predictions, and provides a new paradigm for field-scale mapping.
- Soil texture exhibits significant spatial variability at the field scale, and traditional remote sensing monitoring methods are limited by data intermittency, which constrains small-scale prediction research. This problem has been effectively solved through the integration of synthetic imagery and a partitioning strategy.

## Abstract

In the field of smart agriculture, soil property data at the field scale drives the precision decision-making of agricultural inputs such as seeds and chemical fertilizers. However, soil texture has significant spatial variability at the field scale, and traditional remote sensing monitoring methods have certain data intermittency, which limits small-scale prediction research. In this study, based on the Google Earth Engine platform, soil synthetic images were generated according to different time intervals using mean compositing and median compositing modes, image bands were extracted, and spectral indices were introduced; combined with the random forest algorithm, the effects of different compositing time windows, compositing modes, and compositing data types on prediction accuracy were evaluated; and three partitioning strategies based on crop growth, soil synthetic image brightness, and soil type were adopted to conduct local partitioning regression of soil texture. The results show that: (1) The use of mean compositing of multi-year

Check for updates

Academic Editor: Nicolas Baghdadi

Received: 22 December 2025
Revised: 5 January 2026
Accepted: 9 January 2026
Published: 14 January 2026
Corrected: 11 March 2026

May images from 2021 to 2024 can improve prediction accuracy. When this method is combined with the "band reflectance + spectral indices" dataset, compared with other compositing methods, the $R^2$ of clay particles, silt particles, and sand particles can be increased by 8.89%, 9.50%, and 2.48% on average. (2) Compared with using only image band data, the introduction of spectral indices can significantly improve the prediction accuracy of soil texture at the field scale, and the $R^2$ of clay particles, silt particles, and sand particles is increased by 4.58%, 3.43%, and 4.59% on average, respectively. (3) Global regression is superior to local partitioning regression; however, the local partitioning regression strategy based on soil type has good accuracy performance. Under the optimal compositing method, the average $R^2$ of soil particles of each size fraction is only 1.08% lower than that of global regression, which has great application potential. This study innovatively constructs a comprehensive strategy of "moisture spectral indices + specific compositing time window + specific compositing mode + soil type partitioning", providing a new paradigm for soil texture prediction at the field scale in Northeastern China, and lays the foundation for data-driven water and fertilizer decision-making.

**Keywords:** soil texture; synthetic imagery; partition strategy; field scale

## 1. Introduction

In the precision management system of smart agriculture, field-scale soil property data is the core foundation for realizing efficient resource utilization and production optimization [1–3]. As a key indicator of soil physical properties, soil texture characterizes the water and fertilizer retention capacity of soil, and directly affects the crop growth environment and water and fertilizer requirements [4,5]. As the basic unit of agricultural production, the internal texture heterogeneity of a field is often ignored by large-scale data, which leads to extensive agricultural production and management measures [6]. Studies have confirmed that accurately obtaining the distribution information of sand, silt and clay particles at the field scale is a key prerequisite for improving the management efficiency of smart agriculture, which can support intelligent decisions such as variable fertilization and precision irrigation, and reduce the risk of resource waste and environmental stress [7]. Traditional collection methods of soil texture are mainly based on manual sampling and fixed-point observation, which have the shortcomings of being time-consuming, labor-intensive and having insufficient spatial representativeness [8]. Moreover, they easily ignore the soil texture heterogeneity caused by human factors and differences in soil parent material, making it difficult to accurately characterize the spatial differentiation characteristics [9]. The collection method integrating spatial information technology is becoming a key direction to break through this limitation. Scholars have formed the theoretical basis of spatial mapping in this field: sand, silt and clay particles have spectral responses in the visible (400–700 nm), near-infrared (700–2500 nm) and shortwave infrared bands [10,11]; for example, sand particles will increase the reflectance, and clay particles will form absorption valleys [12], and this response is highly matched with the bands of mainstream satellite sensors such as Sentinel-2, Landsat 8, and ZY1E/F [13], which provides a basis for the application of remote sensing technology in soil texture prediction.

With the development of technologies such as deep learning and machine learning, better algorithm options are provided [14], which gives remote sensing images great advantages in soil texture prediction. However, the complexity of the soil system amplifies the micro-spatial heterogeneity at the field scale [15]. Soil moisture fluctuations caused by meteorological activities further interfere with spectral signals; meanwhile, the intermittency of remote sensing data leads to the coverage of crop canopy and residues on the soil surface, reducing the ability to obtain soil information; in addition, the mismatch between data acquisition and precipitation time causes spatio-temporal heterogeneity of spectral characteristics, increasing the difficulty of soil texture prediction. In response to this situation, scholars have carried out extensive research on synthetic images for soil property mapping based on satellite data in recent years: in terms of spatial resolution, Major et al. (2025) proposed a super-resolution synthesis method to improve the spatial resolution of Sentinel-2 images to 2.5 m while maintaining spectral accuracy, and it showed good spectral verification results in agricultural applications [16]. In terms of spectral resolution, Yuan et al. (2022) proposed a multi-resolution collaborative fusion method, which reduced spectral loss when integrating SAR, multispectral and hyperspectral images to improve the mapping accuracy in areas with complex surface features [17]. In terms of temporal resolution, Maynard et al. (2017) utilized the advantage of multi-temporal data in effectively weakening the interference of soil moisture and vegetation coverage, and used 28-year Landsat time-series data combined with Support Vector Machine to map the soil texture distribution in semi-arid areas, which significantly improved the accuracy compared with single-temporal data [18]. This series of studies shows the potential of remote sensing synthetic images in soil mapping research, which can provide finer spectral resolution and spatio-temporal resolution to characterize ground object information [19], and can overcome the influence of clouds, shadows and different agricultural management methods that single-phase images are prone to cause [20]. However, in the field of soil texture prediction, there is a lack of systematic prediction research similar to that on soil organic matter and other soil property indicators [21]. At present, in the research on field-scale soil texture prediction, there is still a lack of good strategies to deal with the complex spatial heterogeneity of cultivated land and the intermittency of remote sensing data, which highlights the importance of images from different synthesis methods for soil texture prediction.

The spatial heterogeneity of soil texture is affected by soil parent material, climate and human activities [22,23], which reduces the correlation stability between spectrum and texture, and this may cause the global regression model to be easily disturbed by local factors in mapping [24,25]. In response to this situation, Wang et al. (2025) determined the environmental variables affecting organic matter prediction in different soil type regions through principal component analysis, and established prediction models by region to improve the prediction accuracy [26]. This provides a new idea for fine mapping of soil texture at the field scale. On the other hand, the current soil texture mapping scale mainly focuses on large scales [27]. Field-scale soil texture information is faced with the situation of high-precision demand but complex and changeable characteristics that are difficult to accurately characterize [28]. If the advantages of synthetic images in capturing ground object information and the advantages of local partition regression strategy in dealing with regional complexity can be effectively integrated, it is expected to improve the accuracy of field-scale soil texture mapping.

To solve the above problems, this study first used Sentinel-2 remote sensing images to synthesize soil images according to three time intervals (April, May, and April–May) using two compositing modes of median and mean, and carried out soil texture prediction through the Random Forest algorithm, mainly to reveal the influence of soil image synthesis methods on the prediction process and the influence of time windows on the prediction accuracy. On the other hand, this study explored the improvement effect of spectral indices on prediction accuracy in the environment of complex spatial heterogeneity. Finally, based on image brightness, crop growth and soil type partitioning, the application potential of the local partition regression model in field-scale soil texture prediction was attempted.

This study reveals the influence mechanism of different image synthesis elements (data type, time window, compositing mode) on the prediction accuracy of field-scale soil texture, verifies the applicability of the local partition regression strategy, and finally constructs a high-precision field-scale soil texture prediction method.

## 2. Materials and Methods

### 2.1. Study Area

The study area is located at Friendship Farm (46°09′–46°28′N, 131°27′–132°15′E) in Shuangyashan City, Heilongjiang Province, and is an important commodity grain production base in China. The entire area comprises about 126,700 hectares of cultivated land and adopts a one-crop-per-year planting pattern, with rice, corn, and soybean as the main crops. The period from early April to late May each year is designated as the "Bare Land Period," characterized by the absence of both crop and snow cover [29]. This area was listed as one of China's first National Agricultural Modernization Demonstration Zones in 2021 and has played a significant exemplary role in the development of smart agriculture. The area lies within the hinterland of the Sanjiang Plain, formed by the confluence of the Heilong, Songhua, and Wusuli Rivers. It is bordered by the Zhangguangcai and Wanda Mountains to the southwest, features flat terraces in the central part, and connects to the Qixing River Wetland in the southeast. The altitude ranges from approximately 59 m to 300 m, gradually decreasing from southwest to northeast. The region experiences a temperate continental monsoon climate with distinct seasonal variations, and the annual accumulated temperature is approximately 2500 °C. The complex and diverse natural conditions give rise to rich soil types, primarily meadow soil, mollisols, and bog soil, with smaller distributions of luvisols and dark brown forest soil. These soils are highly representative of Northeast China.

The four experimental fields (Figure 1) in the study area are all located in dryland agricultural areas. Agricultural production practices follow unified standards, and crop growth periods are consistent across the fields. The fields are distributed across different altitudes, and through the collection of undisturbed 1 m soil profiles, they were found to exhibit distinct soil type characteristics, providing high representativeness for this area. Area 1 (A1) is a luvisol plot, covering approximately 80.3 hectares. Area 2 (A2) is a mollisols plot, covering about 43.6 hectares. Area 3 (A3) is a meadow soil plot, with an area of roughly 80.5 hectares. Area 4 (A4) is a bog soil plot, covering approximately 51.8 hectares.

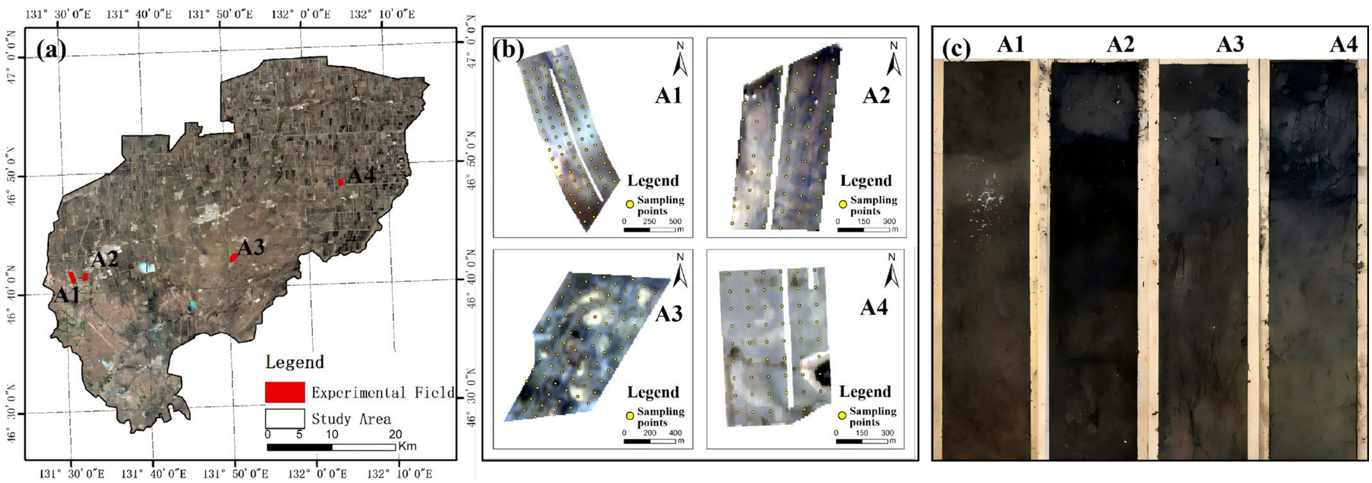


**Figure 1.** Overview of the study area Figure. (**a**) Location map of the experimental field; (**b**) Sampling points distribution map; (**c**) Soil profile samples from the entire section of the experimental field.

### 2.2. Collection and Pretreatment of the Soil Sample Data

In this study, surface soil (0–20 cm) samples were collected on a regular grid at 100 m intervals across four study areas, yielding a total of 264 samples (Figure 1). Each sample was obtained using a five-point composite sampling method. A portable GPS unit was used to record the location of each sampling point. Soil particle size was measured using a laser particle size analyzer [30], and soil texture was determined according to the International Soil Texture Classification system.

### 2.3. Image Selection and Preprocessing

In this study, Sentinel-2 MSI Level-2A data (Table 1) were obtained through the Google Earth Engine (GEE) platform. All Sentinel-2 high-resolution images covering the study area during the spring bare soil period (April and May) from 2021 to 2024 were selected. Atmospheric correction was performed using the Sen2Cor algorithm [31], and cloud masking was applied using the "QA60" band [32]. The proportion of cloud-free pixels was ≥95%, resulting in a cloud-free Sentinel-2 high-resolution image collection covering the study area during the bare soil period from 2021 to 2024. Synthetic images with different modes were generated based on the median and mean values of pixels in multi-temporal images. The composite time intervals included April across multiple years (2021–2024), May across multiple years, and April to May across multiple years. After determining the optimal synthetic image, the dataset was synthesized according to different partitioning strategies listed in Table 2.

**Table 1.** Based on Sentinel-2 images synthetic images.

| Band Type | Wavelength (nm) | Synthetic Image Band Numbering |
|---|---|---|
| Blue | 458–523 | Band1 |
| Green | 543–578 | Band2 |
| Red | 650–680 | Band3 |
| Red Edge1 | 698–713 | Band4 |
| Red Edge2 | 733–748 | Band5 |
| Red Edge3 | 773–793 | Band6 |
| NIR | 785–900 | Band7 |
| Red Edge4 | 855–875 | Band8 |
| SWIR1 | 1566–1655 | Band9 |
| SWIR2 | 2100–2280 | Band10 |

**Table 2.** Datasets based on different partitioning strategies.

| Composite Mode | Months | Partitioning Strategies | Regression Methods | Label |
|---|---|---|---|---|
| Mean | May | Non-partitioned | Global regression | Mean_May_GR |
| Mean | May | NDVI | Local partition regression | Mean_May_NPLR |
| Mean | May | Brightness | Local partition regression | Mean_May_BPLR |
| Mean | May | Soil type | Local partition regression | Mean_May_STLR |
| Median | April | Non-partitioned | Global regression | Median_April_GR |
| Median | April | NDVI | Local partition regression | Median_April_NPLR |
| Median | April | Brightness | Local partition regression | Median_April_BPLR |
| Median | April | Soil type | Local partition regression | Median_April_STLR |

### 2.4. Construction of the Spectral Indices

Existing studies have shown that soil texture is a core property affecting soil water distribution and movement. In this study, well-established water-related spectral indices [33] were selected, including the Normalized Difference Moisture Index (NDMI), which exhibits high sensitivity to surface soil moisture and can effectively support soil texture mapping [28]; the Normalized Difference Snow Index (NDSI), which utilizes surface reflectance characteristics during the bare soil period—Yue et al. (2019) found that NDSI can effectively distinguish the spectral response differences between sandy and clayey soils; and the Normalized Difference Water Index (NDWI), which correlates indirectly with soil porosity and particle composition through its sensitivity to surface moisture [34]. Huang et al. (2025) demonstrated that NDWI correlates well with silt content, playing a key role in texture prediction [35]. Furthermore, Mgohele et al. (2024) proved that the synergistic application of multiple indices can better overcome the limitations of a single index across different land use types [36]. For this study, Sentinel-2 high-resolution imagery was used to create synthetic imagery of the study area during the spring bare soil period (April to May) from 2021 to 2024, and the corresponding spectral indices were calculated. The formulas for NDMI, NDSI, and NDWI are as follows: (1) $P_{\mathrm{NIR}}$ refers to Band 7 in the above synthetic image, $P_{\mathrm{SWIR}}$ refers to Band 9; (2) $P_{\mathrm{Green}}$ refers to Band 2, $P_{\mathrm{SWIR}}$ refers to Band 9; (3) $P_{\mathrm{Green}}$ refers to Band 2, $P_{\mathrm{NIR}}$ refers to Band 7.

$$\mathrm{NDMI} = \frac{P_{NIR} - P_{SWIR}}{P_{NIR} + P_{SWIR}} \tag{1}$$

$$\mathrm{NDSI} = \frac{P_{Green} - P_{SWIR}}{P_{Green} + P_{SWIR}} \tag{2}$$

$$\mathrm{NDWI} = \frac{P_{Green} - P_{NIR}}{P_{Green} + P_{NIR}} \tag{3}$$

### *2.5. Partitioning Algorithms*

#### 2.5.1. NDVI Partitioning

Sandy soil warms more rapidly in spring, leading to earlier crop emergence and resulting in higher NDVI values at the seedling stage (June) [37]. Furthermore, NDVI values in this period do not exhibit oversaturation, allowing for effective observation of crop differences induced by soil properties [38]. Based on this imaging rationale, this study generated mean synthetic images for the June 1st to 25th seedling stage from 2021 to 2024. Using the mean seedling-stage NDVI of crops across the study area as the basis for partitioning, we applied the "Cascaded Simple K-means" algorithm available in Google Earth Engine (GEE). This algorithm evaluates K-means clustering results quantitatively based on the Calinski-Harabasz Index [39], which determined the optimal number of zones to be three. The variability within the sub-areas is less than that between the sub-areas. And the sample points within each zone are reasonably distributed (Table 3).

#### 2.5.2. Image Brightness Partitioning

The Image Brightness Index correlates with soil texture by capturing the reflectance differences in soil particles. Sandy soils, with high reflectivity, correspond to high index values, whereas clayey soils correspond to low values [40]. This index effectively characterizes the spatial distribution of soil texture and reduces prediction errors. Using the mean value brightness of image from April to May from 2021 to 2024 [41] as the basis for partitioning, we applied the "Cascaded Simple K-means" algorithm available in Google Earth Engine (GEE). This algorithm evaluates K-means clustering results quantitatively based on the Calinski-Harabasz Index [39], which determined the optimal number of zones to be three. The variability within the sub-areas is less than that between the sub-areas. And the sample points within each zone are reasonably distributed (Table 3).

#### 2.5.3. Soil Type Partitioning

There are significant differences in soil texture among different soil types [42]. Based on field profile surveys, this study divides the study area into four soil types (Luvisols, Mollisols, Meadow Soil, and Bog Soil), with a reasonable distribution of sample points within each region (Table 3), and completes local regression analysis.

**Table 3.** Sampling points statistics based on partitioning strategies.

| Partitioning Strategies | Set | Particles | N | Max (%) | Min (%) | Mean (%) | SD (%) | CV (%) |
|---|---|---|---|---|---|---|---|---|
| NDVI | Cluster1 | Clay | 55 | 12.52 | 4.54 | 9.35 | 1.90 | 20.27 |
| | | Silt | 55 | 62.90 | 22.82 | 51.93 | 8.97 | 17.27 |
| | | Sand | 55 | 72.63 | 25.25 | 38.72 | 10.52 | 27.17 |
| | Cluster2 | Clay | 88 | 13.72 | 2.86 | 10.10 | 2.33 | 23.08 |
| | | Silt | 88 | 59.28 | 19.99 | 50.14 | 7.69 | 15.35 |
| | | Sand | 88 | 76.40 | 27.93 | 39.76 | 9.68 | 24.34 |
| | Cluster3 | Clay | 121 | 15.84 | 2.02 | 10.96 | 2.71 | 24.77 |
| | | Silt | 121 | 59.35 | 14.68 | 49.32 | 10.14 | 20.56 |
| | | Sand | 121 | 83.30 | 24.95 | 39.72 | 12.75 | 32.10 |
| Multi-year May images are synthesized using the mean value of image brightness | Cluster1 | Clay | 107 | 15.84 | 4.55 | 11.62 | 1.96 | 16.88 |
| | | Silt | 107 | 62.90 | 44.12 | 56.00 | 2.78 | 4.96 |
| | | Sand | 107 | 51.33 | 24.95 | 32.39 | 3.80 | 11.72 |
| | Cluster2 | Clay | 97 | 13.30 | 3.87 | 10.52 | 1.72 | 16.35 |
| | | Silt | 97 | 61.05 | 20.66 | 49.16 | 6.99 | 14.21 |
| | | Sand | 97 | 75.47 | 29.28 | 40.32 | 8.50 | 21.10 |
| | Cluster3 | Clay | 60 | 10.91 | 2.02 | 7.59 | 2.52 | 33.14 |
| | | Silt | 60 | 55.68 | 14.68 | 40.33 | 11.45 | 28.38 |
| | | Sand | 60 | 83.30 | 34.63 | 52.08 | 13.86 | 26.62 |
| Multi-year May images are synthesized using the median value of image brightness | Cluster1 | Clay | 115 | 15.84 | 4.55 | 11.64 | 1.97 | 16.94 |
| | | Silt | 115 | 62.90 | 22.89 | 55.52 | 4.07 | 7.33 |
| | | Sand | 115 | 71.72 | 24.95 | 32.84 | 5.17 | 15.73 |
| | Cluster2 | Clay | 96 | 13.17 | 2.02 | 10.10 | 2.05 | 20.31 |
| | | Silt | 96 | 61.05 | 61.05 | 47.79 | 8.61 | 18.01 |
| | | Sand | 96 | 83.30 | 29.28 | 42.11 | 10.47 | 24.87 |
| | Cluster3 | Clay | 53 | 10.97 | 2.57 | 7.72 | 2.39 | 30.91 |
| | | Silt | 53 | 59.70 | 16.16 | 41.60 | 10.79 | 25.94 |
| | | Sand | 53 | 81.27 | 29.33 | 50.68 | 13.08 | 25.81 |
| Soil Type | Mollisols | Clay | 57 | 15.84 | 10.28 | 12.96 | 0.97 | 7.50 |
| | | Silt | 57 | 59.35 | 45.58 | 55.54 | 2.16 | 3.89 |
| | | Sand | 57 | 44.14 | 24.95 | 31.50 | 2.98 | 9.46 |
| | Luvisols | Clay | 77 | 13.53 | 2.02 | 8.88 | 2.27 | 25.59 |
| | | Silt | 77 | 56.07 | 14.68 | 45.21 | 8.97 | 19.83 |
| | | Sand | 77 | 83.30 | 32.44 | 45.91 | 11.12 | 24.21 |
| | Meadow Soil | Clay | 77 | 13.30 | 2.57 | 10.27 | 2.54 | 24.72 |
| | | Silt | 77 | 57.68 | 16.16 | 47.02 | 10.50 | 22.33 |
| | | Sand | 77 | 81.27 | 29.77 | 42.71 | 12.96 | 30.34 |
| | Bog Soil | Clay | 53 | 12.88 | 4.55 | 9.79 | 1.51 | 15.47 |
| | | Silt | 53 | 62.90 | 44.12 | 56.44 | 3.59 | 6.37 |
| | | Sand | 53 | 51.33 | 24.69 | 33.77 | 4.77 | 14.14 |
| Whole dataset | | Clay | 264 | 15.84 | 2.02 | 10.31 | 2.43 | 23.60 |
| | | Silt | 264 | 62.90 | 14.68 | 50.68 | 8.97 | 17.69 |
| | | Sand | 264 | 83.30 | 24.69 | 39.01 | 10.99 | 28.16 |

### *2.6. Regression Strategies*

Soil texture regression was performed using Python 3.7 combined with the random forest algorithm. This is a nonlinear ensemble tree-based model, which performs regression of input variables against soil texture through decision rules and then determines the results of the ensemble estimator through voting, thus effectively handling high-dimensional input variables [43]. In this study, the out-of-bag prediction error was minimized through parameters such as the number of trees (Ntree), the number of sampling variables (Mtry), and the minimum number of terminal nodes (Nodesize), thereby improving the accuracy of the model.

Regarding the parameter settings of the regression model, this study employed the Optuna hyperparameter optimization framework to optimize the performance of the random forest model. Optuna helps automatically fine-tune the hyperparameters of machine learning models to improve model performance. Researchers can easily define the hyperparameter search space and explore it through efficient algorithms to determine the optimal parameter combination [44]. Optuna can not only automatically adjust the hyperparameters of the model but also effectively balance model complexity and generalization ability, thereby improving prediction accuracy and efficiency and achieving better mapping results [45].

### *2.7. Accuracy Verification*

To evaluate the accuracy of different datasets in model construction, this study employed proportionate stratified sampling across different regions based on various partitioning strategies. After reserving 20% of the data as validation data, the 10-fold cross-validation method was used to evaluate the accuracy of the datasets. In this method, the datasets are divided into 10 parts, 9 of which are used as training datasets and 1 as a validation dataset. The model evaluation is performed through 10 iterations, and the final accuracy is the mean of the prediction accuracies from these 10 iterations [46]. Meanwhile, this study adopted two indicators to evaluate model performance: Root Mean Squared Error (RMSE) and the Coefficient of Determination ($R^2$). Generally, a good prediction result corresponds to a higher $R^2$ value and a lower RMSE value. These indices are established in the following equation, where n denotes the sample size, $y_i$ denotes the observed value at the $i$-th sampling point, $\hat{y}_i$ denotes the predicted value at the $i$-th sampling point.

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

### *2.8. Flow Chart*

The flowchart of this study is shown in Figure 2.

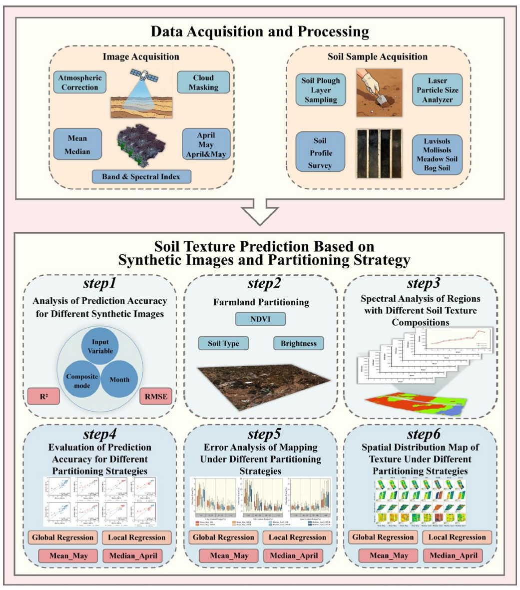


**Figure 2.** Flow chart of this study.

## 3. Results

### 3.1. Spectral Characteristics of Regions with Different Soil Texture Compositions

The data sets of "band + index" were constructed from multi-year May mean synthetic images and multi-year April median synthetic images, respectively; three partitioning strategies were applied to the synthetic images: based on multi-year June NDVI values (divided into three categories), multi-year average soil brightness index (divided into three categories), and soil types (divided into four categories). The spectral reflectance curves of the synthetic images were extracted according to these three strategies (Figure 3). The trend of spectral curves under different partitioning strategies remained consistent with the non-partitioned case, showing a distinct upward trend from Band 1 (approximately 458–523 nm) to Band 9 (approximately 1566–1655 nm), and a rapid downward trend from Band 9 to Band 10 (approximately 2100–2280 nm). However, the average spectral values of each cluster differed significantly across partitioning strategies. Compared with soil type partitioning and NDVI partitioning, the differences among the average spectral curves extracted through brightness partitioning were more pronounced, primarily due to the marked variations in silt and sand particles across regions under this strategy. The average spectral curves of Bog Soil and Mollisols clusters obtained by soil type partitioning were relatively similar, while those of Luvisols and Meadow Soil were similar only within the 860 nm to 2200 nm range; nevertheless, the overall reflectance of the former two was lower than that of the latter two, largely due to their lower sand particle content (Table 3) and similar average composition [11]. Overall, across different partitioning strategies, the differences in average spectral values among clusters were greater in the Mean May data set than in the Median April data set.

### 3.2. Mapping Accuracy of Soil Texture Using Global Regression with Different Synthetic Images

In this study, spectral reflectance and spectral indices extracted from images under two composite modes across different time windows were used to form respective datasets for soil texture prediction. It can be seen from Table 4 that the prediction accuracy for sand and clay is significantly better than that for silt, and the results for all particle sizes show that the "band + index" dataset is superior to the band-only dataset. The prediction accuracies of the mean and median composite modes vary greatly across different time periods. Relatively high prediction accuracy was yielded by median synthesis using multi-year April data: (Clay: $R^2$ = 0.7182, RMSE = 1.3852; Silt: $R^2$ = 0.7139, RMSE = 4.6035; Sand: $R^2$ = 0.7235, RMSE = 5.2609); For the mean composite mode, the highest soil texture prediction accuracy is achieved by using multi-year May data (Clay: $R^2$ = 0.7462, RMSE = 1.3364; Silt: $R^2$ = 0.7020, RMSE = 4.6989; Sand: $R^2$ = 0.7186, RMSE = 5.3073). Overall, the mean composite mode demonstrated higher prediction accuracy.

**Table 4.** Comparison of the prediction accuracy of particle content by different compositing methods.

| Dataset | Composite Mode | Month(s) | Clay | | Silt | | Sand | |
|---|---|---|---|---|---|---|---|---|
| | | | $R^2$ | RMSE (%) | $R^2$ | RMSE (%) | $R^2$ | RMSE (%) |
| Band | Mean | April | 0.6707 | 1.4973 | 0.5842 | 5.5503 | 0.6356 | 6.0391 |
| | | May | 0.7268 | 1.3637 | 0.6720 | 4.9296 | 0.6868 | 5.5990 |
| | | April and May | 0.6401 | 1.5653 | 0.6631 | 4.9960 | 0.6511 | 5.9094 |
| | Median | April | 0.6832 | 1.4687 | 0.6971 | 4.7368 | 0.6968 | 5.5094 |
| | | May | 0.6208 | 1.6068 | 0.5815 | 5.5681 | 0.6869 | 5.5978 |
| | | April and May | 0.6482 | 1.5477 | 0.5800 | 5.5783 | 0.6819 | 5.6422 |
| Band and Spectral Index | Mean | April | 0.6882 | 1.4570 | 0.5846 | 5.5472 | 0.7164 | 5.3283 |
| | | May | 0.7462 | 1.3364 | 0.7020 | 4.6989 | 0.7186 | 5.3073 |
| | | April and May | 0.6664 | 1.5071 | 0.6850 | 4.8305 | 0.6609 | 5.8259 |
| | Median | April | 0.7182 | 1.3852 | 0.7139 | 4.6035 | 0.7235 | 5.2609 |
| | | May | 0.6673 | 1.5052 | 0.6172 | 5.3256 | 0.7105 | 5.3834 |
| | | April and May | 0.6862 | 1.4616 | 0.6045 | 5.4128 | 0.6946 | 5.5289 |

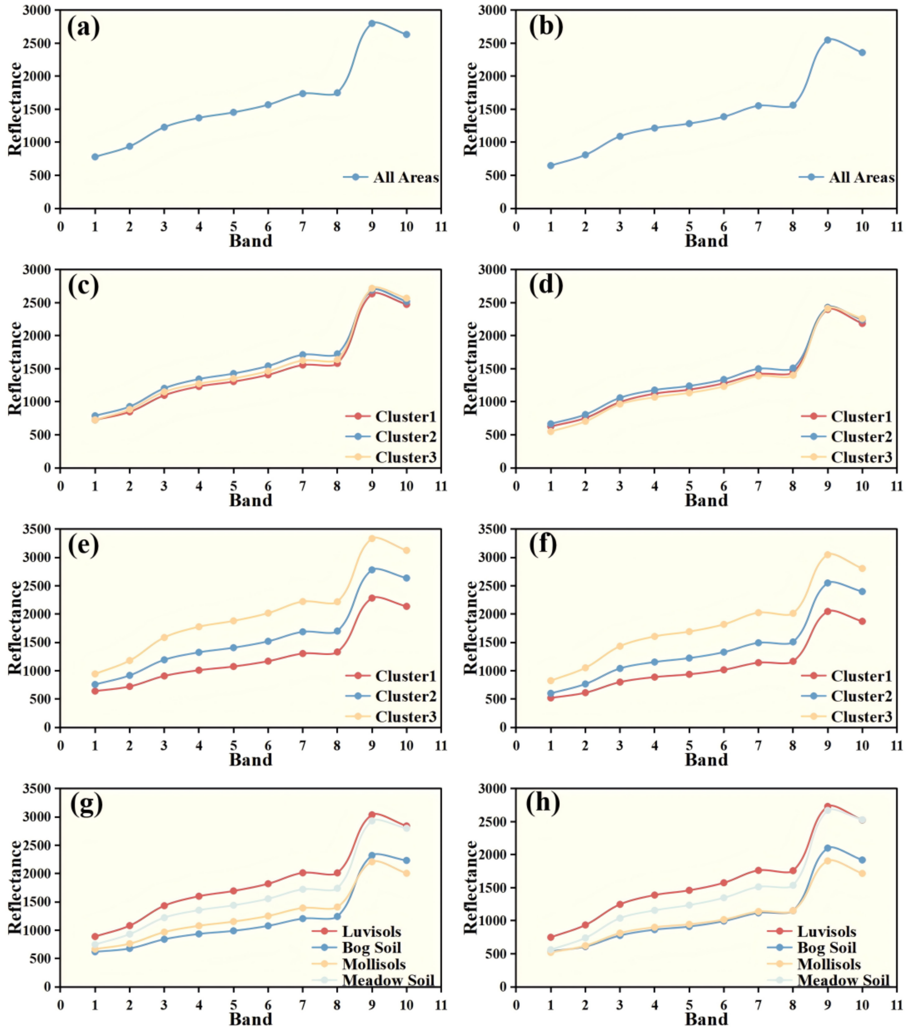


**Figure 3.** The variation curves of the average soil texture values in different regions under different Partitioning strategies with spectral bands. (**a**) Multi-year May images are synthesized using the mean value; (**b**) Multi-year April images are synthesized using the median value; (**c**) Multi-year May images are synthesized using the mean value of NDVI partitions; (**d**) Multi-year April images are synthesized using the median value of NDVI partitions; (**e**) Multi-year May images are synthesized using the mean value of brightness partitions; (**f**) Multi-year April images are synthesized using the median value of brightness partitions; (**g**) Multi-year May images are synthesized using the mean value of soil type partitions; (**h**) Multi-year April images are synthesized using the median value of soil type partitions.

### *3.3. Soil Texture Mapping Accuracy with Different Partition Strategies*

Table 5 and Figures 4–6 show the prediction accuracy of different particle size fractions. As shown in the figures: compared with the multi-year May mean synthetic image, the median April synthetic image delivers more stable accuracy for brightness partitioning (with an average $R^2$ increase of 20.49%) and for NDVI partitioning (with an average $R^2$ increase of 7.6%), whereas the opposite trend is observed for soil type partitioning. For the multi-year May mean synthetic image, the prediction accuracy of regression based on local soil type partitioning is the highest (Clay: $R^2$ = 0.7206, RMSE = 1.1848; Silt: $R^2$ = 0.7079, RMSE = 4.3250; Sand: $R^2$ = 0.7151, RMSE = 5.3670;), which is relatively close to the accuracy of global regression but still significantly better than that of regression based on brightness partitioning and NDVI partitioning (the average $R^2$ of different particle size fractions is increased by 24.3% and 19.9%, respectively). For the median April synthetic image, the $R^2$ of regression based on local brightness partitioning is the highest (Clay: $R^2$ = 0.6997, RMSE = 1.8158; Silt: $R^2$ = 0.6746, RMSE = 7.1439; Sand: $R^2$ = 0.7036, RMSE = 8.6460), which is only slightly lower than the accuracy of global regression while being significantly better than that of regression based on soil type partitioning and local NDVI partitioning (the average $R^2$ of different particle size fractions is increased by 4.8% and 8.0%, respectively); however, in terms of the RMSE index, the soil partitioning strategy performs better (Clay: $R^2$ = 0.6924, RMSE = 1.2918; Silt: $R^2$ = 0.6241, RMSE = 4.8007; Sand: $R^2$ = 0.6666, RMSE = 5.8824). By comparing the regression accuracy of each soil particle fraction, it can be found that when the brightness and NDVI partitioning regression strategies are adopted, the regression accuracy of silt is generally better than that of clay and sand, whereas the opposite is true for the soil type partitioning strategy.

**Table 5.** Comparison of prediction accuracy of particulate matter content under different partitioning strategies.

| Set | Clay | | Silt | | Sand | |
|---|---|---|---|---|---|---|
| | $R^2$ | RMSE (%) | $R^2$ | RMSE (%) | $R^2$ | RMSE (%) |
| Mean_May_GR | 0.7462 | 1.3364 | 0.7020 | 4.6989 | 0.7186 | 5.3073 |
| Mean_May_NPLR | 0.5656 | 1.9006 | 0.6256 | 7.7980 | 0.5965 | 8.7996 |
| Mean_May_BPLR | 0.5463 | 1.9265 | 0.5994 | 6.3129 | 0.5788 | 8.1512 |
| Mean_May_STLR | 0.7206 | 1.1848 | 0.7079 | 4.3250 | 0.7151 | 5.3670 |
| Median_April_GR | 0.7182 | 1.3852 | 0.7139 | 4.6035 | 0.7235 | 5.2609 |
| Median_April_NPLR | 0.6354 | 1.9172 | 0.6409 | 7.1664 | 0.6472 | 8.8503 |
| Median_April_BPLR | 0.6997 | 1.8158 | 0.6746 | 7.1439 | 0.7036 | 8.6460 |
| Median_April_STLR | 0.6924 | 1.2918 | 0.6241 | 4.8007 | 0.6666 | 5.8824 |

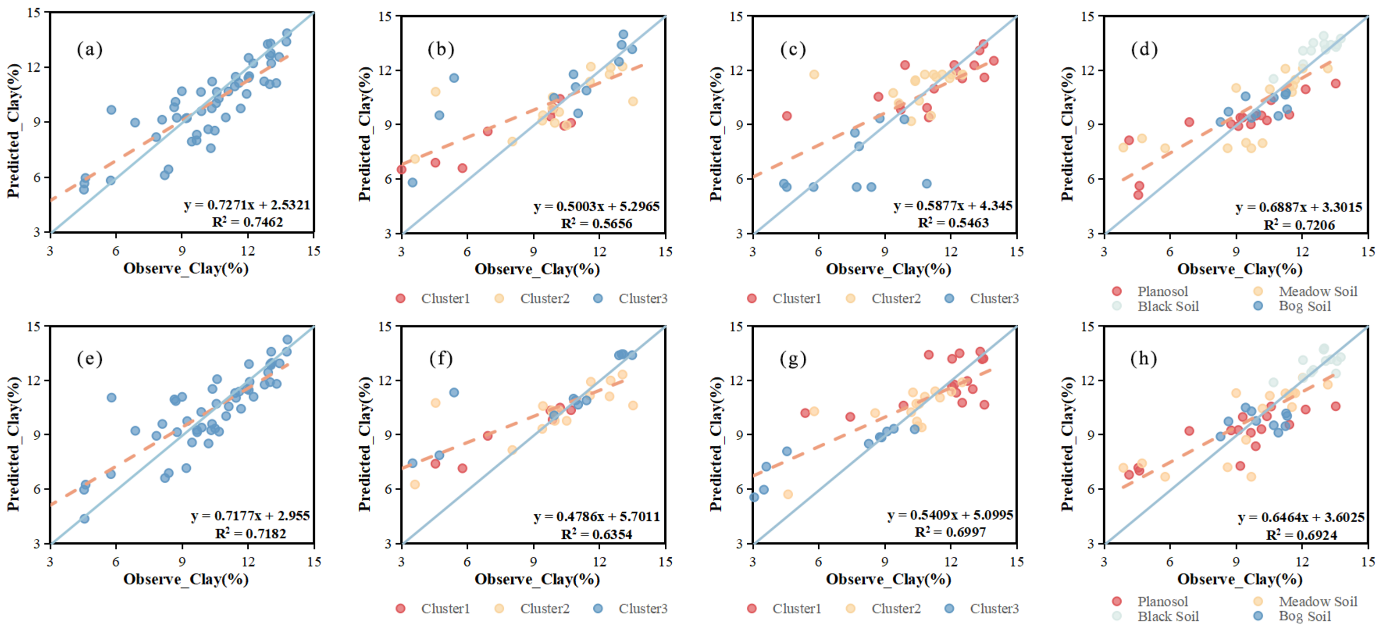


**Figure 4.** Comparison of the accuracy of global regression and local partition regression in predicting clay particles from different synthetic images. (**a**) Mean_May_GR; (**b**) Mean_May_NPLR; (**c**) Mean_May_BPLR; (**d**) Mean_May_STLR; (**e**) Median_April_GR; (**f**) Median_April_NPLR; (**g**) Median_April_BPLR; (**h**) Median_April_STLR. In the figure, the blue line is implemented as a reference line, and the orange dashed line is the regression line.

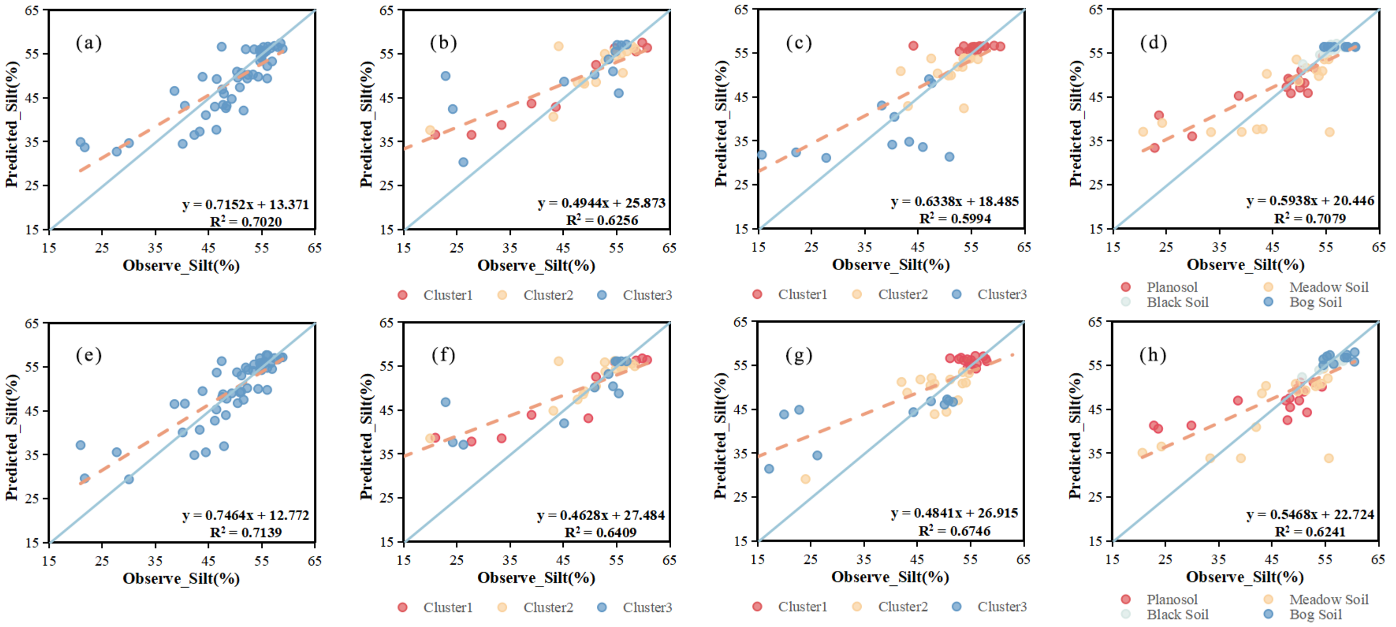


**Figure 5.** Comparison of the accuracy of global regression and local partition regression in predicting silt particles from different synthetic images. (**a**) Mean_May_GR; (**b**) Mean_May_NPLR; (**c**) Mean _May_BPLR; (**d**) Mean_May_STLR; (**e**) Median_April_GR; (**f**) Median_April_NPLR; (**g**) Median_April_BPLR; (**h**) Median_April_STLR. In the figure, the blue line is implemented as a reference line, and the orange dashed line is the regression line.

### *3.4. Error Analysis of Soil Texture Mapping Using Different Strategies*

Figure 7 shows the absolute errors in the global regression and local partition regression of soil texture using two types of data within the optimal time window. The distribution trends of absolute errors are consistent between the two data types under different strategies. The prediction error for clay gradually increases as its content decreases, with a significant rise when the clay content is below 6%. In silt prediction, the error increases gradually with decreasing content, showing a marked increase when the content falls below 45%. For sand, the error increases gradually with higher content, rising notably when the content exceeds 40%. Overall, however, the absolute error of the Mean May composite data is smaller than that of the Median April composite data. For both data types, global regression reduces the absolute error when clay < 6%, silt < 45%, and sand > 40%. The local partition regression strategy based on NDVI partitioning reduces the absolute error in the moderate ranges of clay, silt, and sand (9% < Clay < 12%, 45% < Silt < 50%, 35% < Sand < 40%). The local partition regression strategy based on brightness partitioning reduces the absolute error in the moderate ranges of silt and sand (Silt > 50%, 30% < Sand < 35%). The local partition regression strategy based on soil partitioning reduces the absolute error when clay > 6%, silt < 55%, and sand > 35%.

### *3.5. Spatial Distribution of Soil Texture Mapping Using Different Strategies*

In this study, soil texture mapping in the study area was performed using global regression and local regression, with Sentinel-2 imagery and the optimal time windows for mean and median compositing modes (Figures 8–10). Overall, clay and silt particle contents are relatively high in mollisols and bog soil areas, whereas sand particle content is relatively high in luvisols and bog soil areas. For clay particle prediction, the spatial variation in Mean_May synthetic imagery is smoother than in Median_May synthetic imagery; similarly, global regression produces smoother results than local partitioning regression. For silt particles, Mean_May synthetic imagery again shows smoother variation than Median_May. While global regression yields smoother results overall, local regression based on brightness partitioning performs better in luvisol and meadow soil areas compared to other partitioning strategies. In mollisols and bog soil areas, local regression using soil type partitioning produces smoother results than other zonal approaches. The predictions for sand particles follow a similar pattern to those for silt particles.

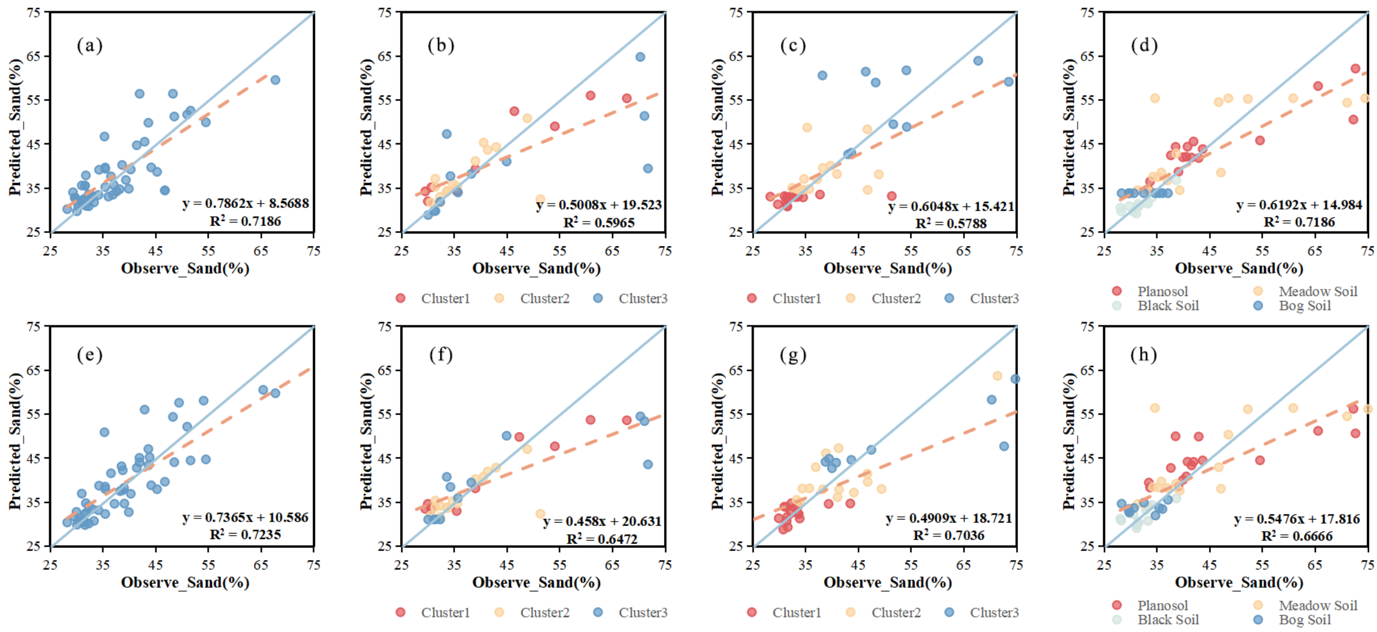


**Figure 6.** Comparison of the accuracy of global regression and local partition regression in predicting sand particles from different synthetic images. (**a**) Mean_May_GR; (**b**) Mean_May_NPLR; (**c**) Mean _May_BPLR; (**d**) Mean_May_STLR; (**e**) Median_April_GR; (**f**) Median_April_NPLR; (**g**) Median_April_BPLR; (**h**) Median_April_STLR. In the figure, the blue line is implemented as a reference line, and the orange dashed line is the regression line.

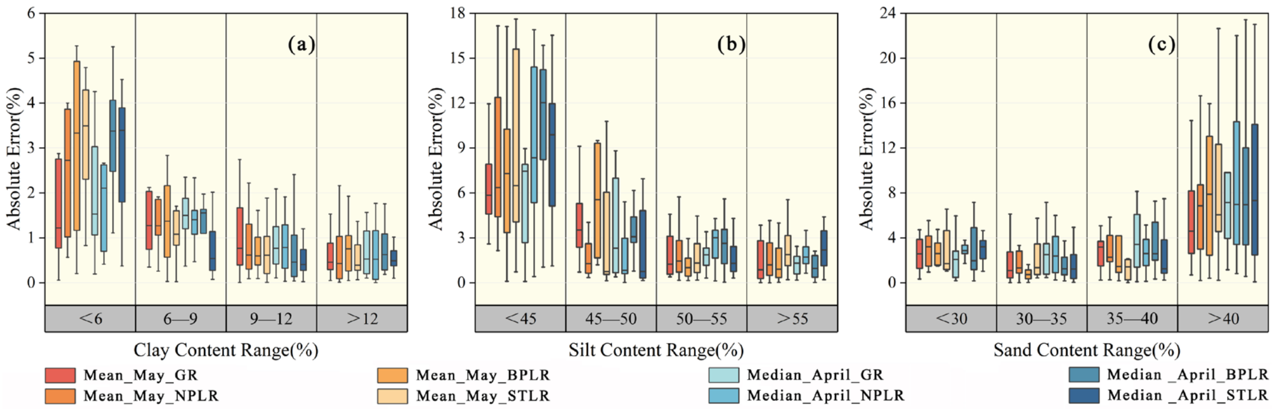


**Figure 7.** Absolute error analysis using global regression and local partition regression for predicting soil texture of different synthetic images. (**a**) Clay; (**b**) Silt; (**c**) Sand.

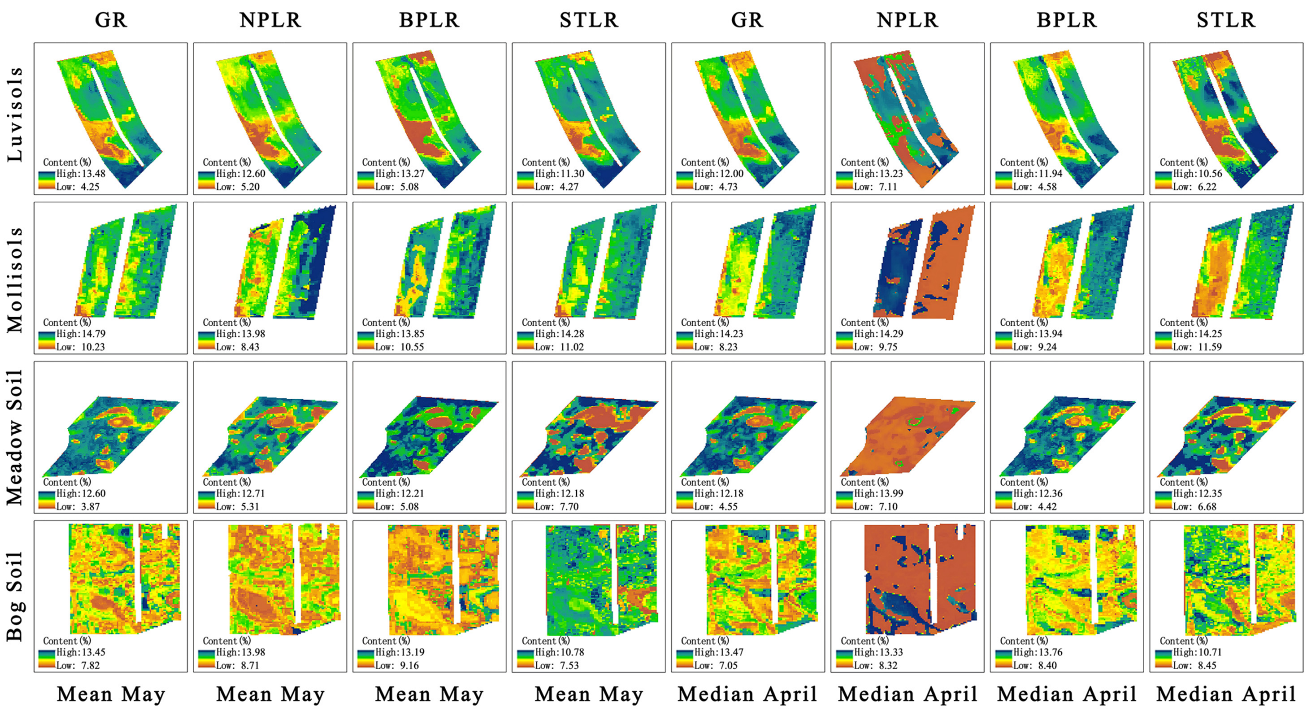


**Figure 8.** Comparison of clay mapping results of global and local partition regression models in the optimal time Windows of different images.

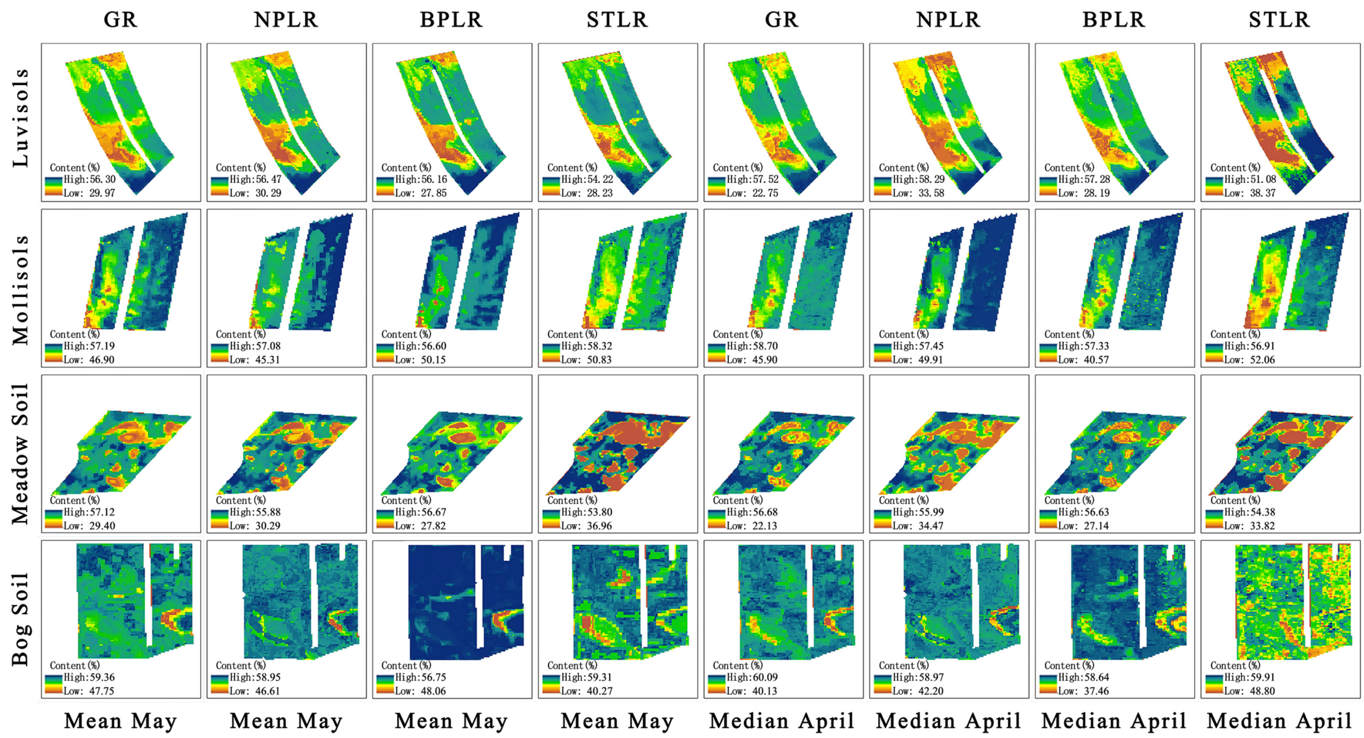


**Figure 9.** Comparison of silt mapping results of global and local partition regression models in the optimal time Windows of different images.

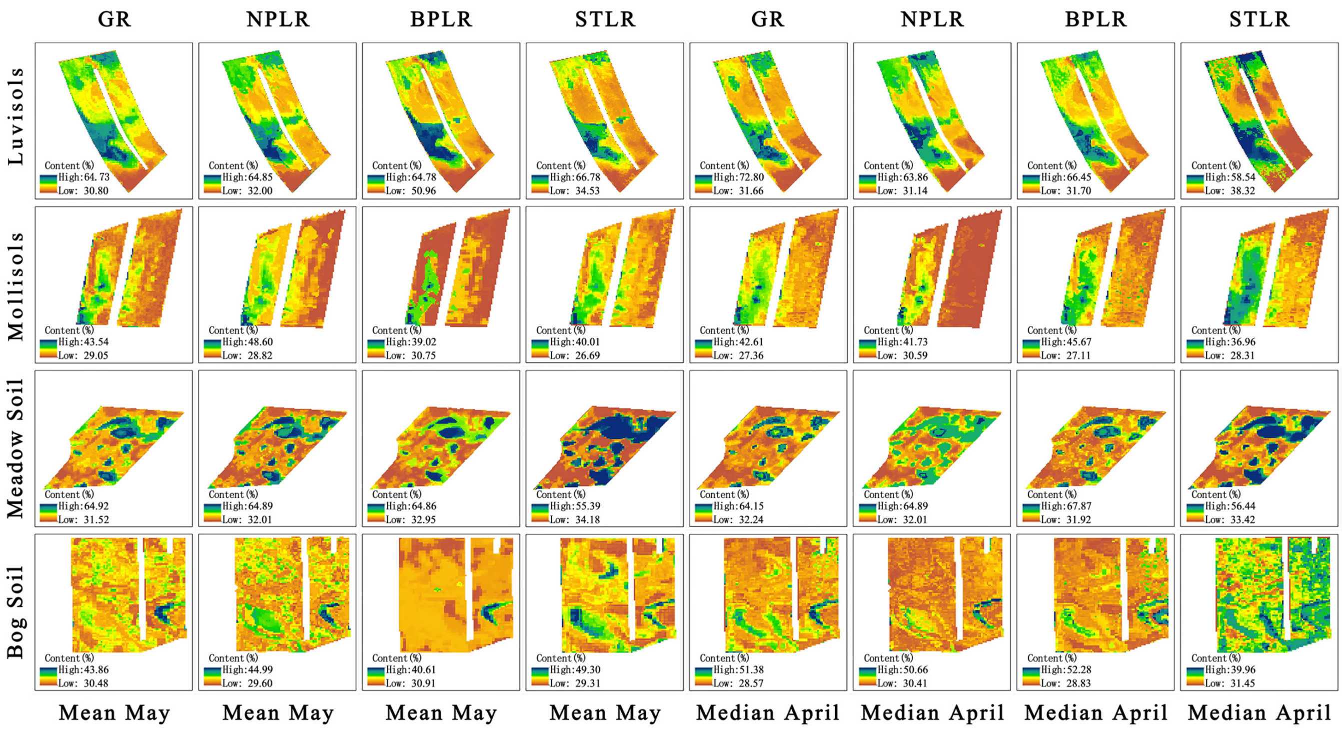


**Figure 10.** Comparison of sand mapping results of global and local partition regression models in the optimal time Windows of different images.

## 4. Discussion

In this study, the data types, temporal composition, and composition modes of synthetic images were investigated at the field scale. The capability of different composite elements to improve the prediction accuracy of soil texture was clarified, and the effectiveness of localized partition-based prediction at the field scale was explored using the optimal synthetic image. An integrated strategy of "Moisture Spectral Index + Specific Composite Time Window + Specific Composite Mode + Soil Type Partitioning" was innovatively developed for inverting soil texture at the field scale in Northeast China or other geographical areas with a distinct bare soil period in spring where crops are grown annually.

### *4.1. Influence of the Spectral Index on the Soil Texture Prediction Accuracy*

This study evaluated the effects of different data types on the accuracy of image synthesis for soil texture prediction. The introduction of spectral indices was found to improve prediction accuracy compared to using band reflectance alone. This finding confirms, at the field scale, that original spectral reflectance forms the foundation for such predictions, as it contains fundamental information about the interaction between soil particles and electromagnetic waves. For instance, features like the clay mineral absorption peak near 2200 nm and the high near-infrared reflectance characteristic of sand particles are directly captured by the original spectrum [47]. Simultaneously, the results demonstrate at this scale that incorporating moisture-related spectral indices can indirectly enhance the characterization of soil properties and improve the predictive performance of synthetic images. This aligns with conclusions from studies at different scales: such indices can both quantify and correct for moisture's effect on spectra [48] and suppress its interference [49,50]. Furthermore, the observed improvement in accuracy with spectral indices suggests underlying issues of spectral redundancy and sensitivity to soil moisture interference at the field scale. This presents a contrast to large-scale studies. In extensive spatial analyses, uneven soil moisture distribution due to meteorological variation can significantly lower reflectance in bands like 1450 nm and 1900 nm in local areas [51], thereby masking soil texture information. At the small, field scale, while soil differences lead to varying field capacities, the prediction results for both contexts reveal the limited predictive power of the original spectrum alone.

Summarizing related research helps explain the accuracy differences among data types: the original spectrum prioritizes information integrity but does not discriminate between signal and noise; "general" spectral indices emphasize feature enhancement and dimensionality reduction; moisture-related indices specifically address moisture interference by separating moisture and texture features, thereby resolving the problem of overlapping spectral signals [52]. In conclusion, moisture-related spectral indices play a crucial role in predicting soil texture at the field scale within the freeze–thaw and wet-dry transition regions of Northeast China.

### *4.2. Selection of the Best Time Window for Soil Texture Prediction*

It was found from Table 5 that the highest prediction accuracy was achieved by synthesizing the mean values of multi-year May images. For this study, data from the meteorological station (131°49′2″E, 46°45′26″N) in Youyi Farm were collected. This high accuracy is mainly attributed to the fact that soil physical properties are affected by freeze–thaw action in April [53], which influences soil spectral values. In May, with rising temperatures and increased precipitation (Table 6), the freeze–thaw action weakens. Although precipitation is consistent within the field, differences in field water-holding capacity across areas with varying soil textures [5] make the fluctuation of spectral values more likely to highlight the distribution characteristics of soil texture [54]. Furthermore, sowing activities are usually carried out in this region in May, and soil disturbance also affects soil water content [55]. The above factors jointly lead to differences in the amount of water change in areas with different soil textures, which further accentuates texture differences and enhances the predictive advantage in May.

**Table 6.** Precipitation and temperature in April and May 2021–2024 in the study area.

| Year | Total Precipitation (mm) | | Monthly Average Temperature (°C) | |
|---|---|---|---|---|
| | April | May | April | May |
| 2021 | 15.7 | 76.4 | 6.8 | 14.9 |
| 2022 | 33.9 | 47.1 | 6.6 | 14.6 |
| 2023 | 31.0 | 38.0 | 6.2 | 15.7 |
| 2024 | 7.2 | 83.9 | 7.0 | 14.1 |

In recent studies, Zhang et al. (2025) compared single-temporal and dual-temporal synthesized images in the same area [56]. They found that the latter improved prediction accuracy, which was further enhanced by adjusting the time window of the dual phases. This indicates that synthesized images acquire more bare soil information, thereby improving the prediction accuracy of soil texture [57]. This conclusion supports the necessity of exploring synthesized image prediction at the field scale in this study and proves that changing the time window of synthesized images during the spring bare soil period affects prediction accuracy at different scales.

However, in large-scale studies, the single-phase April image achieved higher prediction accuracy compared with the May image data [58,59], which differs from the conclusion of this study. Conducted only at the field scale, this study differs from large-scale investigations. During image synthesis at this scale, the impacts of regional differences in surface straw, climatic conditions, and paddy field distribution on the spectrum are relatively small [60,61]. Additionally, Petsetidi et al. (2025) found that soil moisture is the core limiting factor for soil texture prediction, and its fluctuation significantly affects the reliability of the prediction method [62]. In the cultivated land of Northeast China, the spatial heterogeneity of spring soil moisture is relatively large due to differences in soil types or the influence of freeze–thaw action [63,64]. Therefore, selecting the time window of synthesized images according to environmental characteristics is crucial for predicting soil texture at the field scale, which is also a key issue explored in this study. In summary, changes in human activities and meteorological conditions across different periods affect soil water content, leading to variations in the prediction accuracy of soil texture. This change can be utilized to enhance prediction accuracy by reasonably selecting the time window.

### *4.3. Influence of the Image Synthesis Method on the Soil Texture Prediction Accuracy*

This study found that both the mean compositing of multi-year May images and the median compositing of multi-year April images can achieve high accuracy, but the principles by which the two modes improve accuracy are different. Median compositing selects the median value of image pixels from the same period over multiple years, which can effectively eliminate interference from extreme values. Its advantage lies in the fact that soil texture, as a stable attribute, exhibits small long-term fluctuations in spectral signals; median compositing filters out interference from instantaneous moisture anomalies caused by precipitation or drought by selecting the median, thereby retaining the core spectral information related to soil texture [65]. Mean compositing reduces noise by averaging spectral values over multiple years, but it is more sensitive to soil moisture, and its prediction effect is influenced by moisture stability. In areas with stable moisture conditions, mean compositing can smooth out moisture fluctuations; however, in areas with strong spatiotemporal heterogeneity of moisture, the mean will blur the spectral signals of moisture and texture, resulting in decreased prediction accuracy [66]. This is because mean compositing cannot distinguish between outliers and valid signals, and it incorporates the spectral characteristics of short-term high humidity or drought into the average, masking the stable spectral responses related to texture. This study focuses on soil texture prediction at the field scale, contrasting with the challenge of large spatial heterogeneity in prediction areas faced in large-scale studies mentioned above. At the field scale, using the April median compositing mode for prediction can better eliminate the influence of extreme image values such as those from freeze–thaw cycles; whereas using the May mean compositing mode for prediction can under conditions of relatively abundant and consistent precipitation, leverage the advantages of spectral indices to highlight the sensitivity of different soil texture areas to moisture, thereby achieving higher prediction accuracy.

Regarding the differences in prediction accuracy for various soil particle sizes, when using mean compositing of multi-year May images, the average $R^2$ for different particle sizes is 0.7087; when using median compositing of multi-year April images, the average $R^2$ is 0.7054. These values are similar, indicating that the overall prediction effects of different compositing methods on various particle sizes are generally consistent. But significant differences exist for each specific particle size; for example, the median compositing method exhibits high accuracy in predicting sand grain content, which may be attributed to the interference of extreme humidity conditions on the spectral signals of sand grains. On the other hand, May is the optimal time window for the mean compositing mode, and April is the optimal time window for the median compositing mode. Compared to other time windows for the two compositing modes, after introducing spectral indices, both the mean compositing of multi-year May images and the median compositing of multi-year April images significantly improved the prediction accuracy for silt particles (with $R^2$ increasing by 2.73% and 4.14%, respectively). The results indicate that the compositing mode affects the selection of the optimal time window. During the image compositing process, certain correlations and constraints exist among various compositing factors. The formulation of the optimal compositing method does not depend on a single factor but is the result of the synergistic effect of multiple factors.

### *4.4. Influence of Different Partition Strategies on Soil Texture Prediction Accuracy*

This study selected three indicators as the basis for partitioning: first, partitioning by crop growth, which is closely related to soil properties [67]. Partitioning by crop growth can indirectly identify areas with similar texture conditions; second, partitioning by image brightness, which directly reflects the moisture and particle reflection characteristics of the soil surface. This helps separate spectral differences caused by moisture interference; third, partitioning by soil type, which clusters samples with similar physical and chemical properties, reduces spectral confusion due to differences in organic matter and mineral composition among soil types, and makes texture-sensitive spectral signals more prominent within the same partition [68].

This study shows that under different partitioning indicators, the average spectral reflectance curves differ significantly within each partition, making local partition regression feasible at the field scale. However, in prediction, only the soil-type local partition regression strategy and global regression exhibit high prediction accuracy and favorable error performance—a trend opposite to traditional large-scale prediction results. In previous large-scale soil property mapping studies, local partition regression has been a key technique for improving accuracy, performing significantly better than global unified regression models. Its core advantage lies in weakening local heterogeneity and strengthening the correlation between local spectra and soil properties through partitioning [69]. This may occur because, although the prediction area at the field scale is small, texture heterogeneity within some soil types remains significant, and differences between soil types are also considerable (Table 3). On the other hand, it may be due to the fact that the degree of heterogeneity inherent in the field scale is not sufficient to offset the negative impact brought about by the increased complexity of the model introduced by partitioning. Moreover, the spectral response of soil texture is jointly influenced by soil type and moisture status [70], while crop growth, as a dynamic surface feedback, is affected not only by texture but also by factors such as temperature and nutrients [71]. This indicates that a single-indicator local partitioning strategy has certain limitations in local partition regression of soil texture at the field scale.

Figures 4–6 reveal that the Mean_May_STLR dataset shows the smallest accuracy difference in predicting soil particles at all levels (Clay: $R^2$ = 0.7206, Silt: $R^2$ = 0.7079, Sand: $R^2$ = 0.7151), with a maximum $R^2$ difference of 1.79% among different particle sizes. Under the same synthetic image, the maximum $R^2$ difference among particle sizes in the Mean_May_BPLR dataset is 9.72%, and in the Mean_May_NPLR dataset it is 10.61%. Under the soil type partitioning strategy, the maximum $R^2$ difference in the Median_April_STLR dataset is 9.86%. The error analysis in Figure 7 indicates significant errors in the extreme value region, reflecting the limitations of the model at the edge of data distribution. This may be due to the scarcity of extreme value samples or a certain upper limit in the ability of the random forest algorithm to handle extreme samples, leading to greater model errors, rather than the result of a certain partitioning strategy. Figures 8–10 demonstrate that the soil type partitioning strategy achieves good error performance for specific soil particle sizes and yields a smoother spatial distribution in mapping Mollisols and marsh soil. This suggests that although this local partitioning strategy has certain limitations in prediction accuracy, it helps reduce differences in prediction accuracy and mapping error across soil particle levels. Additionally, it is observed that the synthesis methods of images at the field scale affects the effectiveness of the partitioning strategy in improving accuracy.

In summary, although the soil type partitioning strategy has limited accuracy, as a multi-attribute indicator it can comprehensively characterize soil physical and chemical properties, reduce variability in interference factors within partitions, enhance the effectiveness of spectral indices in mitigating moisture interference, enable the model to capture the correlation between spectral reflectance and sand, silt, and clay contents at the local scale, and achieve high prediction accuracy. Moreover, applying this method at the field scale makes soil type investigation easier compared to large-scale applications.

### *4.5. Limitations and Prospects*

There remain certain limitations in this study: First, conventional spectral indices (such as the Soil-Adjusted Vegetation Index, SAVI) were not introduced for comparative analysis during the research process; so, the advantages of water-related spectral indices could not be fully highlighted; Second, soil types are complex and diverse. At present, specific experimental analysis is lacking on how to fully utilize the particular characteristics of field-scale cultivated land to guide the formulation of synthetic strategies, and the effectiveness of its generalization requires further verification; In addition, local partitioning regression, it remains an urgent issue to explore how to leverage the advantages of non-single attribute indicators of soil types, how compositing adjustment strategies (such as selecting the May mean compositing mode for mollisols and the April median compositing mode for luvisols) can enhance these advantages, and what the mechanistic differences are between this strategy and global regression.

To overcome the above limitations, future research will introduce conventional spectral indices and soil moisture data, clarify the unique advantages of water-related spectral indices through further comparative analysis, and quantify the contribution of each water spectral index to the prediction results through a feature selection algorithm. Regarding compositing strategies, field-scale environmental variable analysis will be conducted for different soil type regions to clarify image synthetic methods suitable for each soil type. This is of great significance for improving prediction accuracy and reducing the differences in prediction accuracy among soil particles at all levels. It will also help to elucidate the mechanism by which synthetic images improve prediction accuracy, thereby breaking the “black box” of their role. In terms of partitioning strategies, building on the advantages of synthetic images adapted to different soil types, research scenarios will be further expanded. Attempts will be made to improve prediction accuracy across different crop types and climate regions, thereby providing a field-scale scientific basis for precise guidance of agricultural production and effectively promoting the development of smart agriculture.

## 5. Conclusions

In this study, to address the problem that soil complexity and the temporal discontinuity of remote sensing images lead to inaccurate soil texture characterization at the field scale, Sentinel-2 image data were used to explore the effects of different input data, compositing modes, and compositing time windows on soil texture prediction. Field-scale soil texture prediction was then performed by integrating a local partitioning regression strategy. An innovative comprehensive prediction framework suitable for dry farmland in Northeast China at the field scale was developed, which integrates a water spectral index, a specific compositing time window, a specific compositing mode, and soil-type partitioning, to overcome the dual challenges of moisture interference and spatial heterogeneity. The main conclusions are as follows: (1) The highest accuracy for field-scale soil texture prediction was achieved by compositing multi-year May images using the mean compositing mode. This outcome results from the synergistic interaction between the compositing mode and the selected time window. Significant correlations and constraints exist among the factors involved in image compositing. Therefore, determining the optimal compositing strategy depends not on a single factor, but on the combined effect of multiple factors. (2) In field-scale soil texture prediction, incorporating a water spectral index significantly improved the prediction accuracy compared to using spectral reflectance alone. This improvement was consistent across different spring compositing times and different image synthetic methods. (3) Comparative analysis identified May as the optimal compositing time window. Furthermore, different compositing methods could enhance the positive effect of this specific time window on prediction accuracy. (4) This study found that for field-scale soil texture prediction, global regression overall performed better than most local partitioning regression strategies. However, local partitioning regression based on soil types performed no worse than global regression and demonstrated considerable application potential. This is because this partitioning approach can reduce the variability in interfering factors within certain zones, thereby improving the stability of local regression. This study provides new conclusions and a technical reference for fine-scale soil mapping at the field level. Future research could explore more efficient image synthetic methods and partitioning regression strategies in different soil type regions.

**Author Contributions:** Conceptualization, Y.W.; methodology, S.M., S.B. and H.L.; software, D.W., Y.Z. and Y.W.; validation, Y.M. (Yihan Ma) and Y.W.; formal analysis, Y.M. (Yuxin Ma) and Y.W.; data curation, Y.Z. and Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W.; project administration, Y.Z., H.L. and S.M.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Jilin Provincial Natural Science Foundation (Distinguished Young Scholars Fund), project "Quantitative Remote Sensing of Soil" (20250101009JJ).

**Data Availability Statement:** Data are not available due to privacy restrictions; please contact the corresponding author.

**Acknowledgments:** We would like to thank Youyi Farm for providing the relevant meteorological data. The data content covers: monthly precipitation data and monthly average temperature from 2021 to 2024.

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

## References

1. Alfred, R.; Obit, J.H.; Chin, C.P.Y.; Haviluddin, H.; Lim, Y. Towards Paddy Rice Smart Farming: A Review on Big Data, Machine Learning, and Rice Production Tasks. *IEEE Access* **2021**, *9*, 50358–50380. [CrossRef]
2. Hundal, G.S.; Laux, C.M.; Buckmaster, D.; Sutton, M.J.; Langemeier, M. Exploring Barriers to the Adoption of Internet of Things-Based Precision Agriculture Practices. *Agriculture* **2023**, *13*, 163. [CrossRef]
3. McFadden, J.; Njuki, E.; Griffin, T. *Precision Agriculture in the Digital Era: Recent Adoption on US Farms*; US Department of Agriculture: Washington, DC, USA, 2023.
4. Dexter, A.R. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. *Geoderma* **2004**, *120*, 201–214. [CrossRef]
5. Harefa, D. The Influence of Soil Texture Types on Land Resilience to Drought in South Nias. *J. Sapta Agrica* **2025**, *4*, 13–30. [CrossRef]
6. Roberton, S.D. A Data Driven Approach for Diagnosis and Management of Yield Variability Attributed to Soil Constraints. Ph.D. Thesis, University of Southern Queensland, Toowoomba, Queensland, Australia, 2019.
7. Ge, Y.; Thomasson, J.A.; Sui, R. Remote sensing of soil properties in precision agriculture: A review. *Front. Earth Sci.* **2011**, *5*, 229–238. [CrossRef]
8. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]

9. Ito, R.; Yanai, J.; Nakao, A. Interactive effect of parent material and topography on spatial variability of paddy soil material characteristics in the alluvial plain. *Soil Sci. Plant Nutr.* **2023**, *69*, 99–108. [CrossRef]
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. Divya, Y.; Sanjeevi, S.; Ilamparuthi, K. A study on the hyperspectral signatures of sandy soils with varying texture and water content. *Arab. J. Geosci.* **2014**, *7*, 3537–3545. [CrossRef]
12. Wang, X.; Ding, J.; Han, L.; Tan, J.; Ge, X. Enhancing soil particle content prediction accuracy: Advanced hyperspectral analysis and machine learning models. *J. Soils Sediments* **2024**, *24*, 3443–3458. [CrossRef]
13. Song, Q.; Gao, X.; Song, Y.; Li, Q.; Chen, Z.; Li, R.; Cai, S. Estimation and mapping of soil texture content based on unmanned aerial vehicle hyperspectral imaging. *Sci. Rep.* **2023**, *13*, 14097. [CrossRef] [PubMed]
14. Omondiagbe, O.P.; Lilburne, L.; Licorish, S.A.; MacDonell, S.G. Soil texture prediction with automated deep convolutional neural networks and population-based learning. *Geoderma* **2023**, *436*, 116521. [CrossRef]
15. Taghdis, S.; Farpoor, M.H.; Mahmoodabadi, M. Pedological assessments along an arid and semi-arid transect using soil spectral behavior analysis. *Catena* **2022**, *214*, 106288. [CrossRef]
16. Major, D.; Horváth, Z.; Kröber, F.; Augustin, H.; Sudmanns, M.; Ševčík, P.; Baraldi, A.; Berg, A.; Cornel, D.; Tiede, D. A holistic approach for multi-spectral Sentinel-2 super-resolution and spectral evaluation. *Int. J. Remote Sens.* **2025**, *46*, 7437–7464. [CrossRef]
17. Yuan, Y.; Meng, X.; Sun, W.; Yang, G.; Wang, L.; Peng, J.; Wang, Y. Multi-Resolution Collaborative Fusion of SAR, Multispectral and Hyperspectral Images for Coastal Wetlands Mapping. *Remote Sens.* **2022**, *14*, 3492. [CrossRef]
18. Maynard, J.J.; Levi, M.R. Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. *Geoderma* **2017**, *285*, 94–109. [CrossRef]
19. Meng, X.; Bao, Y.; Zhang, X.; Luo, C.; Liu, H. A long-term global Mollisols SOC content prediction framework: Integrating prior knowledge, geographical partitioning, and deep learning models with spatio-temporal validation. *Remote Sens. Environ.* **2025**, *318*, 114592. [CrossRef]
20. Chen, X.; Meng, X.; Shao, F.; Sun, W. PSSTFN: A progressive spatial-temporal-spectral fusion network for remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5530312. [CrossRef]
21. Chen, X.; Yuan, F.; Ata-Ul-Karim, S.T.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. A bibliometric analysis of research on remote sensing-based monitoring of soil organic matter conducted between 2003 and 2023. *Artif. Intell. Agric.* **2025**, *15*, 26–38. [CrossRef]
22. Bajracharya, R.M.; Elliot, W.J.; Lal, R. Interrill Erodibility of Some Ohio Soils Based on Field Rainfall Simulation. *Soil Sci. Soc. Am. J.* **1992**, *56*, 267–272. [CrossRef]
23. Karlen, D.L. *Effects of Soil and Crop Management Practices on Soil Quality*; Agricultural Research Service, Department of Agriculture: Ames, IA, USA, 2003.
24. Zhou, J.G.; Dong, D.M.; Li, Y.Y. Local attribute-similarity weighting regression algorithm for interpolating soil property values. *Int. J. Agric. Biol. Eng.* **2017**, *10*, 95–103. [CrossRef]
25. Kock, A.L.; Ramphisa-Nghondzweni, P.D.; Van Zijl, G. Development of soil spectroscopy models for the Western Highveld region, South Africa: Why do we need local data? *Eur. J. Soil Sci.* **2024**, *75*, e70014. [CrossRef]
26. Wang, L.; Liu, H.; Wang, X.; Xu, X.; He, L.; Luo, C.; Li, Y.; Zhang, X.; Zang, D.; Zheng, S.; et al. Identifying Optimal Variables to Predict Soil Organic Carbon in Sandy, Saline, and Black Soil Regions: Remote Sensing, Terrain, or Climate Factors? *Remote Sens.* **2025**, *17*, 237. [CrossRef]
27. Roy, S.; Ansal, D.K.; Kumar, M. Soil Texture Prediction: Advances in Remote Sensing, Image Analysis, and Machine Learning. In Proceedings of the 2025 Fourth International Conference on Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 9–10 May 2025.
28. Wittstruck, L.; Waske, B.; Jarmer, T. Multi-Modal Vision Transformer for high-resolution soil texture prediction of German agricultural soils using remote sensing imagery. *Remote Sens. Environ.* **2025**, *331*, 114752. [CrossRef]
29. Luo, C.; Zhang, W.; Zhang, X.; Liu, H. Mapping of soil organic matter in a typical black soil area using Landsat-8 synthetic images at different time periods. *Catena* **2023**, *231*, 107336. [CrossRef]
30. Mccave, I.N.; Bryant, R.J.; Cook, H.F.; Coughanowr, C.A. Evaluation of laser-diffraction-size analyser for use with natural sediments. *J. Sediment. Res.* **1986**, *56*, 561–564. [CrossRef]
31. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Gascon, F. Sen2Cor for Sentinel-2. In Proceedings of the Image and Signal Processing for Remote Sensing XXIII, SPIE Remote Sensing, Warsaw, Poland, 11–14 September 2017; SPIE: Bellingham, WA, USA, 2017; Volume 10427, pp. 37–48.
32. Tarrio, K.; Tang, X.; Masek, J.G.; Claverie, M.; Ju, J.; Qiu, S.; Zhu, Z.; Woodcock, C.E. Comparison of cloud detection algorithms for Sentinel-2 imagery. *Sci. Remote Sens.* **2020**, *2*, 100010. [CrossRef]
33. Ma, S.; Zhou, Y.; Gowda, P.H.; Dong, J.; Zhang, G.; Kakani, V.G.; Wagle, P.; Chen, L.; Flynn, K.C.; Jiang, W. Application of the water-related spectral reflectance indices: A review. *Ecol. Indic.* **2019**, *98*, 68–79. [CrossRef]

34. Yue, J.; Tian, J.; Tian, Q.; Xu, K.; Xu, N. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 216–230. [CrossRef]
35. Huang, H.; Liu, Y.; Liu, Y.; Tong, Z.; Ren, Z.; Xie, Y. The Spatial Distribution and Driving Mechanism of Soil Organic Matter in Hilly Basin Areas Based on Genetic Algorithm Variable Combination Optimization and SHAP Interpretation. *Remote Sens.* **2025**, *17*, 1186. [CrossRef]
36. Mgohele, R.N.; Massawe, B.H.J.; Shitindi, M.J.; Sanga, H.G.; Omar, M.M. Prediction of soil texture using remote sensing data. A systematic review. *Front. Remote Sens.* **2024**, *5*, 1461537. [CrossRef]
37. Al-Mulla, Y.; Huggins, R.D.; Stöckle, C.O. Modeling the Emergence of Winter Wheat in Response to Soil Temperature, Water Potential, and Planting Depth. *Trans. ASABE* **2014**, *57*, 761–775. [CrossRef]
38. Falco, N.; Wainwright, H.M.; Dafflon, B.; Ulrich, C.; Soom, F.; Peterson, J.E.; Hubbard, S.S. Influence of soil heterogeneity on soybean plant development and crop yield evaluated using time-series of UAV and ground-based geophysical imagery. *Sci. Rep.* **2021**, *11*, 7046. [CrossRef] [PubMed]
39. Calinski, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
40. Mirzaeitalarposhti, R.; Shafizadeh-Moghadam, H.; Taghizadeh-Mehrjardi, R.; Demyan, M.S. Digital Soil Texture Mapping and Spatial Transferability of Machine Learning Models Using Sentinel-1, Sentinel-2, and Terrain-Derived Covariates. *Remote Sens.* **2022**, *14*, 5909. [CrossRef]
41. Vieira, A.S.; do Valle Junior, R.F.; Rodrigues, V.S.; da Silva Quinaia, T.L.; Mendes, R.G.; Valera, C.A.; Fernandes, L.F.S.; Pacheco, F.A.L. Estimating water erosion from the brightness index of orbital images: A framework for the prognosis of degraded pastures. *Sci. Total Environ.* **2021**, *776*, 146019. [CrossRef]
42. Lei, Z.; XiaoLin, Q.; Tao, Z.; ChangBao, M.; JianBing, L.; HuaiYu, L.; AiGuo, X.; RenLian, Z.; Ge, L. Analysis of the Characteristics and Influencing Factors of Soil Particle Composition in Ningxia. *Sci. Agric. Sin.* **2023**, *56*, 4272–4287. [CrossRef]
43. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
44. Almarzooq, H.; Bin Waheed, U. Automating hyperparameter optimization in geophysics with Optuna: A comparative study. *Geophys. Prospect.* **2024**, *72*, 1778–1788. [CrossRef]
45. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019.
46. Dietterich, T.G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput.* **1998**, *10*, 1895–1923. [CrossRef]
47. Wetterlind, J.; Simmler, M.; Castaldi, F.; Borvka, L.; Gabriel, J.L.; Gomes, L.C.; Khosravi, V.; Kvrak, C.; Koparan, M.H.; Lázaro-López, A. Influence of Soil Texture on the Estimation of Soil Organic Carbon from Sentinel-2 Temporal Mosaics at 34 European Sites. *Eur. J. Soil Sci.* **2025**, *76*, e70054. [CrossRef]
48. Diek, S.; Chabrillat, S.; Nocita, M.; Schaepman, M.E.; de Jong, R. Minimizing soil moisture variations in multi-temporal airborne imaging spectrometer data for digital soil mapping. *Geoderma* **2019**, *337*, 607–621. [CrossRef]
49. Liu, Y.; Lu, Y.; Chen, D.; Zheng, W.; Ma, Y.; Pan, X. Simultaneous estimation of multiple soil properties under moist conditions using fractional-order derivative of vis-NIR spectra and deep learning. *Geoderma* **2023**, *438*, 116653. [CrossRef]
50. Yang, J.; Li, X.; Ma, X.; Ge, X. Potential of Hyperspectral Data Combined with Optimal Band Combination Algorithm for Estimating Soil Organic Carbon Content in Lakeside Oasis. *Land Degrad. Dev.* **2024**, *35*, 3291–3305. [CrossRef]
51. McGuirk, S.L.; Cairns, I.H. Relationships between Soil Moisture and Visible–NIR Soil Reflectance: A Review Presenting New Analyses and Data to Fill the Gaps. *Geotechnics* **2024**, *4*, 78–108. [CrossRef]
52. Munnaf, M.A.; Mouazen, A.M. Spectra transfer based learning for predicting and classifying soil texture with short-ranged Vis-NIRS sensor. *Soil Tillage Res.* **2023**, *225*, 105545. [CrossRef]
53. Ma, Q.; Zhang, K.; Jabro, J.D.; Ren, L.; Liu, H. Freeze–thaw cycles effects on soil physical properties under different degraded conditions in Northeast China. *Environ. Earth Sci.* **2019**, *78*, 321. [CrossRef]
54. Seidel, M.; Vohland, M.; Greenberg, I.; Ludwig, B.; Ortner, M.; Thiele-Bruhn, S.; Hutengs, C. Soil moisture effects on predictive VNIR and MIR modeling of soil organic carbon and clay content. *Geoderma* **2022**, *427*, 116103. [CrossRef]
55. He, J.; Li, H.; Kuhn, N.J.; Wang, Q.; Zhang, X. Effect of ridge tillage, no-tillage, and conventional tillage on soil temperature, water use, and crop performance in cold and semi-arid areas in Northeast China. *Soil Res.* **2010**, *48*, 737–744. [CrossRef]
56. Zhang, J.; Ding, J.; Zhang, Z.; Wang, J.; Zeng, X.; Ge, X. Study on the inversion and spatiotemporal variation mechanism of soil salinization at multiple depths in typical oases in arid areas: A case study of Wei-Ku Oasis. *Agric. Water Manag.* **2025**, *315*, 1089765. [CrossRef]
57. Silvero, N.E.Q.; Demattê, J.A.M.; Amorim, M.T.A.; Santos, N.V.D.; Rizzo, R.; Safanelli, J.L.; Poppiel, R.R.; Mendes, W.d.S.; Bonfatti, B.R. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. *Remote Sens. Environ.* **2021**, *252*, 112117. [CrossRef]

58. Zhang, X.; Ma, Y.; Ma, S.; Qin, C.; Wang, Y.; Liu, H.; Chen, L.; Zhu, X. Mapping Soil Available Nitrogen Using Crop-Specific Growth Information and Remote Sensing. *Agriculture* **2025**, *15*, 1531. [CrossRef]
59. Gao, L.; Zhang, Y.; Zang, D.; Yang, Q.; Liu, H.; Luo, C. Using Satellites to Monitor Soil Texture in Typical Black Soil Areas and Assess Its Impact on Crop Growth. *Agriculture* **2025**, *15*, 912. [CrossRef]
60. Xiang, X.; Du, J.; Jacinthe, P.A.; Zhao, B.; Zhou, H.; Liu, H.; Song, K. Integration of tillage indices and textural features of Sentinel-2A multispectral images for maize residue cover estimation. *Soil Tillage Res.* **2022**, *221*, 105405. [CrossRef]
61. Zhang, Y.; Luo, C.; Zhang, W.; Wu, Z.; Zang, D. Mapping Soil Organic Matter in Black Soil Cropland Areas Using Remote Sensing and Environmental Covariates. *Agriculture* **2025**, *15*, 339. [CrossRef]
62. Petsetidi, P.A.; Kargas, G.; Sotirakoglou, K. Investigation of Topsoil Salinity and Soil Texture Using the EM38-MK2 and the WET-2 Sensors in Greece. *AgriEngineering* **2025**, *7*, 347. [CrossRef]
63. Han, X.Z.; Li, N. Research Progress of Black Soil in Northeast China. *Sci. Geogr. Sin.* **2018**, *38*, 1032–1041. [CrossRef]
64. Fu, Q.; Hou, R.; Li, T.; Jiang, R.; Yan, P.; Ma, Z.; Zhou, Z. Effects of soil water and heat relationship under various snow cover during freezing-thawing periods in Songnen Plain, China. *Sci. Rep.* **2018**, *8*, 1325. [CrossRef]
65. Diek, S.; Schaepman, M.E.; De Jong, R. Creating Multi-Temporal Composites of Airborne Imaging Spectroscopy Data in Support of Digital Soil Mapping. *Remote Sens.* **2016**, *8*, 906. [CrossRef]
66. Lei, F.; Crow, W.T.; Shen, H.; Su, C.H.; Holmes, T.R.H.; Parinussa, R.M.; Wang, G. Assessment of the impact of spatial heterogeneity on microwave satellite soil moisture periodic error. *Remote Sens. Environ.* **2018**, *205*, 85–99. [CrossRef]
67. Zhang, W.; Dou, W.; Gao, L.; Li, X.; Luo, C. Enhancing Soil Texture Mapping and Drought Stress Assessment Through Dual-Phase Remote Sensing in Typical Black Soil Regions. *Sustainability* **2025**, *17*, 10793. [CrossRef]
68. Summers, D.; Lewis, M.; Ostendorf, B.; Chittleborough, D. Unmixing of soil types and estimation of soil exposure with simulated hyperspectral imagery. *Int. J. Remote Sens.* **2011**, *32*, 6507–6526. [CrossRef]
69. Luo, C.; Wang, Y.; Zhang, X.; Zhang, W.; Liu, H. Spatial prediction of soil organic matter content using multiyear synthetic images and partitioning algorithms. *Catena* **2022**, *211*, 106023. [CrossRef]
70. Hutengs, C.; Seidel, M.; Oertel, F.; Ludwig, B.; Vohland, M. In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. *Geoderma* **2019**, *355*, 113900. [CrossRef]
71. Zhang, Y.; Walker, J.P.; Pauwels, V.R.N.; Sadeh, Y. Assimilation of Wheat and Soil States into the APSIM-Wheat Crop Model: A Case Study. *Remote Sens.* **2022**, *14*, 65. [CrossRef]