




Article

Soil Organic Carbon Estimation and Transfer Framework in Agricultural Areas Based on Spatiotemporal Constraint Strategy Combined with Active and Passive Remote Sensing

Jiaxin Qian¹, Jie Yang¹, Weidong Sun^{1,*} , Lingli Zhao², Lei Shi¹ , Hongtao Shi³ , Lu Liao⁴, Chaoya Dang¹ and Qi Dou¹

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China; qianjx1996@whu.edu.cn (J.Q.); yangj@whu.edu.cn (J.Y.)

² College of Remote Sensing Information Engineering, Wuhan University, Wuhan 430072, China

³ School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

⁴ Technology Service Center of Surveying and Mapping, Sichuan Bureau of Surveying, Mapping and Geoinformation, Chengdu 610081, China

* Correspondence: widensun2012@whu.edu.cn

Abstract: Mapping soil organic carbon (SOC) plays a crucial role in agricultural productivity and water management. This study discusses the potential of active and passive remote sensing for SOC estimation modeling in agricultural areas, incorporating synthetic aperture radar (SAR) data (L-band quad-polarization and C-band dual-polarization), multi-spectrum (MS) data, and brightness temperature (TB) data. The performance of five advanced machine learning regression (MLR) models for SOC modeling was assessed, focusing on spatial interpolation accuracy and cross-spatial transfer accuracy, using two field observation datasets for modeling and validation. Results indicate that the SOC estimation accuracy when using MS data alone is comparable to that of using TB data alone, and both perform slightly better than SAR data. Radar cross-polarization ratio index, microwave polarization difference index, shortwave infrared reflectance, and soil parameters (elevation and soil moisture) demonstrate high correlation with the measured SOC. Incorporating temporal features, as opposed to single-phase features, allows each regression model to reach its upper limit of SOC estimation accuracy. The spatial interpolation accuracy of each MLR algorithm is satisfactory, with the Gaussian process regression (GPR) model demonstrating optimal modeling performance. When SAR, MS, or TB data are used individually in modeling, the estimation errors (RMSE) for SOC are 0.637 g/kg, 0.492 g/kg, and 0.229 g/kg for the SMAPVEX12 sampling campaign, and 0.706 g/kg, 0.454 g/kg, and 0.474 g/kg for the SMAPVEX16-MB sampling campaign, respectively. After incorporating soil moisture and topographic factors, the above RMSEs for SOC are further reduced by 57.8%, 35.6%, and 3.5% for the SMAPVEX12, and by 18.4%, 8.8%, and 3.4% for the SMAPVEX16-MB, respectively. However, cross-spatial transfer accuracy of the regression models remains limited (RMSE = 0.866–1.043 g/kg and 0.995–1.679 g/kg for different data sources). To address this, this study reduces uncertainties in SOC cross-spatial transfer by introducing terrain factors sensitive to SOC (RMSE = 0.457–0.516 g/kg and 0.799–1.198 g/kg for different data sources). The proposed SOC estimation and transfer framework, based on active and passive remote sensing data, provides guidance for high-resolution regional-scale SOC mapping and applications.

Keywords: soil organic carbon estimation and transfer; synthetic aperture radar; brightness temperature; multi-spectrum; machine learning regression model



Academic Editors: Tong Li, Songchao Chen, Anquan Xia, Francesco Fava and Yash Dang

Received: 18 December 2024

Revised: 10 January 2025

Accepted: 14 January 2025

Published: 19 January 2025

Citation: Qian, J.; Yang, J.; Sun, W.; Zhao, L.; Shi, L.; Shi, H.; Liao, L.; Dang, C.; Dou, Q. Soil Organic Carbon Estimation and Transfer Framework in Agricultural Areas Based on Spatiotemporal Constraint Strategy Combined with Active and Passive Remote Sensing. *Remote Sens.* **2025**, *17*, 333. <https://doi.org/10.3390/rs17020333>

Copyright: © 2025 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 carbon (SOC), as one of the largest carbon reservoirs in terrestrial ecosystems, plays a critical role in soil and ecosystem functions [1,2]. By absorbing and storing atmospheric carbon dioxide, it helps mitigate climate change [3]. SOC is a major component of soil organic matter (SOM), improving soil pore structure, enhancing water retention, and increasing nutrient availability, thereby promoting plant growth [4,5]. SOC provides energy for soil microorganisms, stimulating microbial activity and biodiversity, which are essential for nutrient cycling [6]. Moreover, SOC helps stabilize soil structure, reduces erosion, and increases resistance to soil degradation [7,8]. Overall, SOC is not only vital for agricultural productivity but also plays a key role in global carbon cycling and climate regulation [3,9].

In situ field sampling and laboratory analysis of SOC often achieve high measurement accuracy, but these methods are limited to local sampling points and cannot adequately capture large-scale distribution trends [10]. Interpolation algorithms are generally used for mapping at regional scales, but they introduce considerable uncertainty [11,12]. In recent years, significant progress has been made in estimating SOC using remote sensing techniques [1,13–15]. However, challenges related to high spatial resolution and high-precision estimation still remain. For instance, the spatial resolution of currently available SOC gridded products is limited to 250 m or 1000 m [13], although resolutions of up to 90 m can be achieved in specific regions [14]. Moreover, gridded products often lack timeliness, which is particularly critical for agricultural production and management [15].

SOC remote sensing estimation typically integrates multi-source data such as multi-spectrum (MS), synthetic aperture radar (SAR), and LiDAR to obtain more comprehensive multidimensional information [16–19]. For example, optical data, especially hyperspectral data, can capture soil spectral characteristics, SAR data are more sensitive to soil structure and moisture, and LiDAR is more responsive to soil texture [20]. Traditional models, such as linear regression and principal component analysis (PCA), have gradually been replaced by more advanced machine learning regression (MLR) algorithms, leading to improved SOC estimation accuracy [21–23]. With the advancement of remote sensing platforms and the availability of high spatiotemporal resolution data, such as Sentinel and Landsat series, researchers are now better equipped to monitor the spatiotemporal dynamics of SOC at regional and even global scales [24–26].

Although hyperspectral remote sensing can provide detailed soil spectral characteristics, thereby improving the estimation accuracy and robustness of SOC, the limited availability of hyperspectral platforms currently restricts its application potential [17,27]. It is necessary to explore the potential of other multi-source remote sensing data, such as MS, active microwave, and passive microwave data (Brightness temperature, TB), for SOC modeling, as these data are more readily accessible and can comprehensively reflect SOC across different information dimensions. Moreover, the cross-spatial transferability of SOC estimation models is crucial for their application in SOC mapping and monitoring. Enhancing the cross-spatial generalization capability of SOC estimation models to ensure they are applicable to different soil types and regions has become an important research trend [28].

This study mainly discusses the potential of several MLR algorithms, combined with multi-source data, for estimating SOC in agricultural areas with complex surface heterogeneity. It develops an SOC estimation and transfer framework that combines active and passive remote sensing data with spatiotemporal constraints and evaluates its accuracy and uncertainty using two field-measured datasets. The study addresses the following key issues:

- (1) What is the significance of SAR (L- and C-band) features, MS features, TB features, and soil parameters (elevation, slope, soil moisture, and soil roughness) for estimating SOC?
- (2) Can incorporating multi-source data enable the regression algorithms to achieve optimal SOC estimation accuracy? Can temporal features improve the estimation accuracy of SOC? Which regression algorithm is most suitable for SOC modeling?
- (3) How to improve the regional-scale spatial transferability of the SOC regression model?

2. Study Areas and Data

2.1. Study Areas

The study area is located in Manitoba, Canada (Figure 1a,b), characterized by a typical temperate continental climate, with hot, sunny summers and long, cold winters. The region is flat and open, serving as one of the primary agricultural areas, where crops such as canola, wheat, oats, soybeans, and corn are predominantly grown (Figure 1f). In situ field data for this study were obtained from the SMAPVEX12 and SMAPVEX16-MB datasets (Soil Moisture Active Passive Validation Experiments 2012 and 2016-Manitoba) [29,30]. The area's land surface is covered with diverse vegetation, features highly heterogeneous soil textures, and exhibits significant variations in surface soil moisture (SSM), making it an ideal environment for thoroughly testing SOC estimation models.

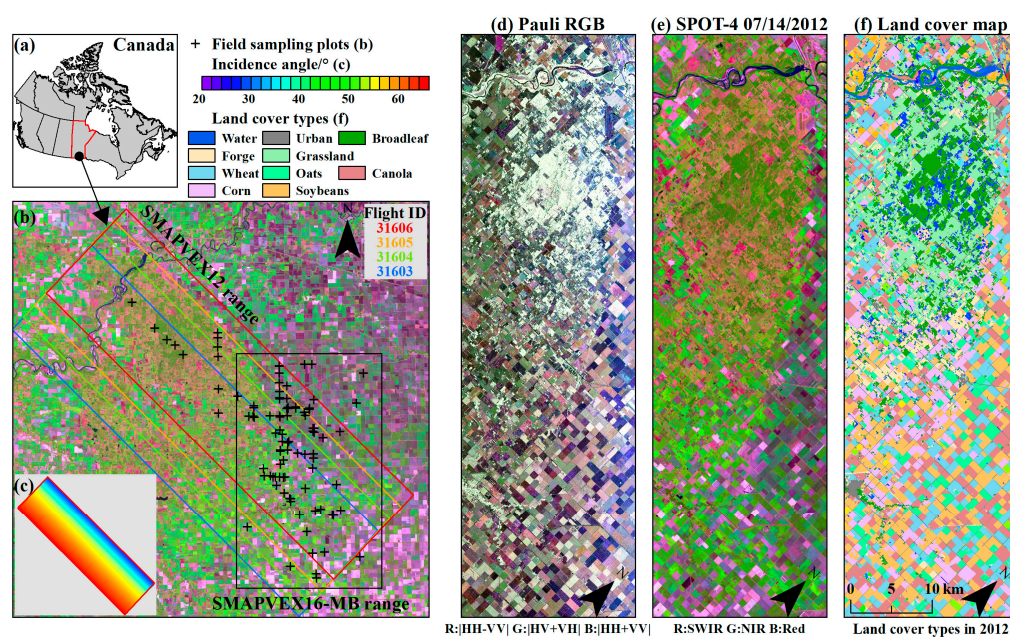


Figure 1. (a) Geographical location of the study area, (b) observation ranges of the SMAPVEX12 (color rectangular box) and SMAPVEX16-MB datasets (black rectangular box), (c) radar incidence angle of UAVSAR image, (d) Pauli decomposition image, (e) SPOT-4 image, (f) land cover map in 2012.

2.2. In Situ Sampling Data

2.2.1. SMAPVEX12 Dataset

In June and July 2012, the SMAPVEX12 sampling campaign included measurements of SSM, soil texture types (STT), soil roughness (SR), and various indicators of vegetation growth. Each sampling plot contained 2 surface SOM sampling points (0–5 cm deep), with a total of 55 plots sampled in early June [4]. A total of 110 SOM samples were collected during the SMAPVEX12 sampling campaign. STT data captured the proportions of sand, silt, and clay (f_{sand} , f_{silt} , and f_{clay}), while SR data encompassed the root mean square height

(RMSH) and correlation length (CL). Since soil parameters (SOM, STT, and SR) remained relatively stable throughout the observation period, these measurements were taken only once [4,31]. Soil samples were collected manually by digging, with each sample having a diameter and depth of 4.7 cm and 4.6 cm. Soil samples were then dried in an oven at 105 °C for 24 h to measure SSM. Additionally, the particle size was measured using the pipette method. The dried soil samples were further ground and passed through a 0.5 mm sieve, then heated at 375 °C for 16 h to obtain SOM [4,32]. SR data were obtained by digital processing after shooting with a pinboard and digital camera. All sampled data were quality controlled [4]. The sampled crops included canola, cereals (wheat and oats), soybeans, corn, and pasture. Additional details are available on the SMAPVEX12 website (<https://smapvex12.espaceweb.usherbrooke.ca>, accessed on 14 July 2014).

2.2.2. SMAPVEX16-MB Dataset

In June and July 2016, the SMAPVEX16-MB sampling campaign was carried out across 50 farmland plots. The sampling method used was similar to that of the SMAPVEX12 dataset. Each plot contained 5–6 surface SOM sampling points, covering a total of 50 plots sampled in early June [29]. A total of 252 SOM samples were collected during the SMAPVEX16-MB sampling campaign. All sampled data were quality controlled. Vegetation growth conditions were also recorded, including crops such as canola, wheat, oats, soybeans, and corn. Further details can be found on the SMAPVEX16-MB website (<https://smapvex16-mb.espaceweb.usherbrooke.ca>, accessed on 10 July 2021).

Whether in SMAPVEX12 or SMAPVEX16-MB, there are significant differences in soil's physical and geometric parameters between different plots (Figure 2). Since both sampling campaigns collected SOM, a coefficient conversion was performed to obtain SOC, as follows [33]: $SOC = SOM/1.724$.

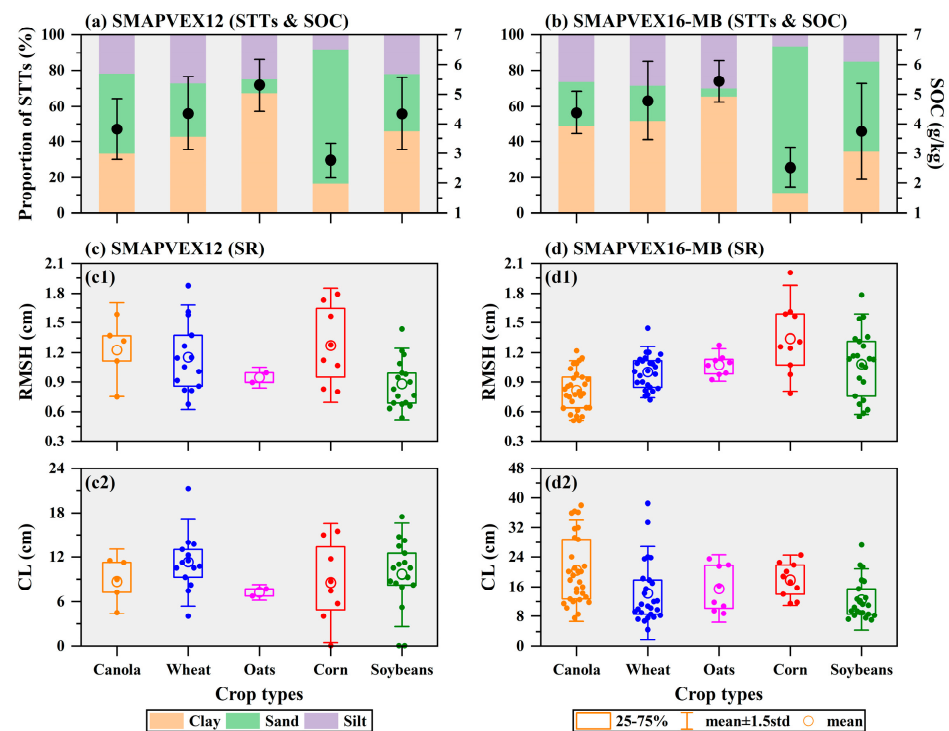


Figure 2. Measured soil parameter information: (a) proportion of measured soil texture types and SOC of different crop plots during SMAPVEX12 sampling campaign; (b) is similar to (a), SMAPVEX16-MB; (c) measured soil roughness of different crop plots during SMAPVEX12 sampling campaign; (d) is similar to (c), SMAPVEX16-MB.

2.3. Remote Sensing Data

2.3.1. Microwave Remote Sensing Data

During June to July 2012, a total of 13 phases L-band quad-polarization (quad-pol) SAR images were acquired using the uninhabited aerial vehicle synthetic aperture radar (UAVSAR) platform (Figure 1d), with ground-based synchronous observational sampling conducted for 12 phases. The L-band quad-pol SAR data consisted of horizontal–horizontal (HH), horizontal–vertical (HV), vertical–horizontal (VH), and vertical–vertical (VV) polarization modes (Table 1).

Table 1. Remote sensing features used in the regression models.

Sampling Campaign	Feature Types	Feature Descriptions
SAMPVEX12	L-band SAR features	$\sigma_{HH}, \sigma_{HV \text{ or } VH}, \sigma_{VV};$ $CPR_H^{L\text{-band}} = \sigma_{HV} / \sigma_{HH}, CPR_V^{L\text{-band}} = \sigma_{VH} / \sigma_{VV}, CoPR_{H-V}^{L\text{-band}} = \sigma_{HH} / \sigma_{VV},$ RIA [34,35]
	L-band TB features	$TB_H, TB_V, MPDI = \frac{TB_V - TB_H}{TB_V + TB_H}$ [36]
	Optical features (SPOT4/5)	$\rho_{Green}, \rho_{Red}, \rho_{NIR}, \rho_{SWIR};$ $NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$ [37], $NDMI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$ [38], $NIR_v = NDVI \times NIR$ [39]
SAMPVEX16-MB	C-band SAR features	$\sigma_{VH}, \sigma_{VV}, CPR_V^{C\text{-band}} = \sigma_{VH} / \sigma_{VV}, RIA$
	L-band TB features	$TB_H, TB_V, MPDI$
	Optical features (Sentinel-2A)	$\rho_{Blue}, \rho_{Green}, \rho_{Red}, \rho_{Red \text{ edge}1}, \rho_{Red \text{ edge}2}, \rho_{Red \text{ edge}3}, \rho_{NIR}, \rho_{SWIR1}, \rho_{SWIR2};$ $NDVI, NDMI, NIR_v, NDVI_{Red \text{ edge}} = \frac{\rho_{NIR} - \rho_{Red \text{ edge}}}{\rho_{NIR} + \rho_{Red \text{ edge}}},$ $NDSI = \frac{\rho_{SWIR1} - \rho_{SWIR2}}{\rho_{SWIR1} + \rho_{SWIR2}}$ [40–42]

The products used were the multi-look ground project complex data, derived from the original single look complex (SLC) data with a spatial resolution of 1.66 m \times 0.8 m, resulting in a final spatial resolution of 4 m \times 6 m. The radar incidence angles (RIA) range from 20.0° to 65.0° (Figure 1c). The original backscattering coefficient data were normalized to 40° using histogram matching method [43]. The backscattering coefficient images were processed by mean filtering (5 \times 5 window) and were resampled to 20–30 m spatial resolution using the nearest neighbor method to match the optical remote sensing data (Figure 3). Due to the low flight altitude of the UAVSAR platform, RIAs of different plots differ greatly, which will bring great uncertainty to the SOC modeling.

Sentinel-1A Interferometric Wide Swath (IW) single look complex (SLC) products during April to October 2016 (non-frozen stage) were chosen as the C-band dual-polarization (dual-pol) SAR data sources for analyzing backscattering coefficients in the study area. The preprocessing method is visually depicted in Figure 3. This study acquired dual-pol backscattering coefficients, which encompass both VH and VV polarization (Table 1). The RIAs were extracted from header files. To reduce the impact of RIA on the SOC models, the dual-pol backscattering coefficient images were processed using second-order cosine normalization and uniformly adjusted to 40° [44,45].

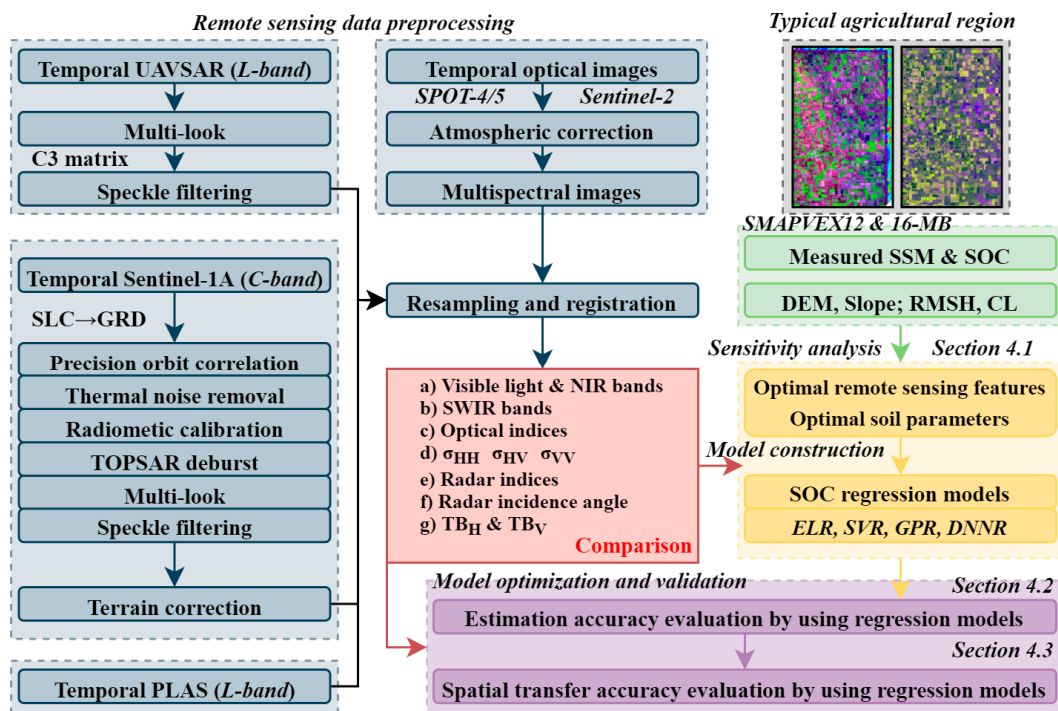


Figure 3. Technical process.

2.3.2. Optical Remote Sensing Data

The SPOT-4 and SPOT-5 Level-1A data were used as optical remote sensing data sources in 2012. After radiometric calibration, atmospheric correction, orthorectification, and geometric registration, four-band land surface reflectance (LSR) images with a spatial resolution of 10–20 m were obtained (Figure 1e), including two visible light bands (VIS; ρ_{Green} and ρ_{Red}), one near-infrared band (ρ_{NIR}), and one short-wave infrared band (ρ_{SWIR}). The Sentinel-2A Level-2A data were selected as optical data sources in 2016. After preprocessing, the LSR images were obtained with 10–20 m spatial resolution, including three VIS bands (ρ_{Blue} , ρ_{Green} , ρ_{Red}), three red edge bands ($\rho_{red\ edge1}$, $\rho_{red\ edge2}$, $\rho_{red\ edge3}$), two NIR bands (ρ_{NIR} and ρ_{nNIR}), and two SWIR bands (ρ_{SWIR1} and ρ_{SWIR2}).

2.3.3. Brightness Temperature Data and Preprocessing

The TB data were derived from the passive and active L-band system (PLAS; 1.413 GHz), which was mounted on a fixed-wing aircraft. The imaging incident angle was set to 40°. The land surface vertical polarization TB (TB_V) and horizontal polarization TB (TB_H) in the study areas were mapped [29,30]. The spatial resolution of TB_V and TB_H was about 650 m (2012) and 450 m (2016), corresponding to the scale of a sampling plot. The microwave polarization difference index (MPDI) was derived from TB_V and TB_H (Table 1) [36]. Considering the joint modeling of multi-source remote sensing data, this study omitted some of the sampling plots that were not in the TB imaging ranges. The imaging information of the above data are shown in Tables S1 and S2 in the Supplementary Materials.

3. Methods

The technical process of this study is as follows (Figure 3): (1) The collected L- and C-band SAR data, TB remote sensing data, optical remote sensing data, and field observation data were preprocessed; (2) A correlation analysis was performed on temporal remote sensing features, the measured soil parameters (SSM and SR), terrain factors (DEM and slope), and the measured SOC; (3) SOC estimation models were constructed based

on regression algorithms; (4) The SOC estimation and transfer accuracy of different MLR models were assessed under multiple strategy constraints.

3.1. Machine Learning Regression Algorithms

3.1.1. Ensemble Learning Regression Algorithms

Ensemble learning regression (ELR) leverages multiple regression models to enhance predictive accuracy and robustness. By aggregating predictions, ELR balances the strengths and weaknesses of individual models, reducing both bias and variance while improving overall performance. ELR includes techniques such as bagging (bootstrap aggregating) and boosting. Bagging generates subsets to train independent models, with their predictions averaged for the final output. Boosting, on the other hand, iteratively trains weak learners to correct errors, with the final prediction being a weighted average of the learners' results. In this study, extreme random tree regression (ETR) and eXtreme Gradient Boosting regression (XGBR) were employed as the bagging and boosting algorithms, respectively, both of which have demonstrated great potential in the SOC modeling [46,47].

3.1.2. Support Vector Regression Algorithm

Support vector regression (SVR) is a regression method based on statistical learning theory [48,49]. Its core idea is to find an optimal regression hyperplane such that as many training data points as possible are close to this plane, while penalizing points that fall outside a certain error margin (commonly known as the ϵ -insensitive zone). This approach balances model complexity with prediction error. The SVR model focuses only on points far from the boundary of the optimal regression hyperplane (i.e., support vectors), while points within the ϵ zone do not affect the model. This mechanism simplifies the model and reduces the risk of overfitting. SVR model can handle nonlinear regression problems by using kernel functions (such as linear, radial basis function, and Gaussian kernels) to map data into a high-dimensional feature space, performing well in environments with small sample sizes and high-dimensional data [50].

3.1.3. Gaussian Process Regression Algorithm

Gaussian process regression (GPR) is a non-parametric algorithm used to model the underlying relationships in data. It relies on the principles of Gaussian processes, which enable it to model uncertainty in functions and make predictions for new data points based on previously observed data [51–53]. The GPR model utilizes flexible kernel functions that can adapt to various types of functional relationships between different data sources. This makes GPR a powerful regression method, capable of not only capturing complex relationships within the data but also providing insights into predictive uncertainties [53–55].

3.1.4. Neural Network Regression Algorithm

Deep neural network (DNN) is a feedforward neural network with multiple hidden layers [56]. DNN performs multiple layers of nonlinear transformations to automatically extract complex features from the input data. During training, DNN updates its weights gradually through the backpropagation algorithm to minimize prediction errors. Neurons in each layer map inputs to nonlinear outputs using activation functions, such as ReLU or Sigmoid, extracting higher-order features layer by layer. Deep neural network regression (DNNR) is also used in SOC modeling applications [57].

3.2. Modeling Features and Strategies

In the MLR models, the input features (X_1, X_2, X_3, \dots) include different remote sensing data (SAR, TB, or MS features), soil parameters, and terrain factors, and the output feature (Y) is the measured SOC. SAR features include different polarization backscattering

coefficients and corresponding radar indices [34,35], TB features include dual-pol TB and MPDI, while MS features include reflectance in various bands and corresponding spectral indices [37–39]. For the SMAPVEX12 campaign, the remote sensing features used for modeling include L-band quad-pol SAR features, L-band TB features, and four-band MS features (Table 1). For the SMAPVEX16-MB campaign, the remote sensing features used for modeling include C-band dual-pol SAR features, L-band TB features, and ten-band MS features (Table 1). Additionally, the impact of introducing the measured SSM, the measured SR, and terrain factors (DEM and slope) derived from the Shuttle Radar Topography Mission (elevation and slope) on the SOC estimation model was also considered.

3.3. Model Construction and Optimization

The measured SOC data are randomly divided into two parts (Part A and Part B) according to different plots and sampling points. First, Part A (half of all samples) is used as the training set and Part B (the other half of all samples) as the test set. Then, Part B is used as the training set and Part A as the test set. The estimation results of the test sets from the two modeling processes were used for the final accuracy evaluation. The probability distribution curves of SOC between the training set and the test set are largely consistent (Figure 4a,b). This modeling partitioning strategy is primarily used to evaluate the estimation modeling performance of each regression algorithm for SOC. This estimation accuracy can be regarded as the spatial interpolation accuracy of the SOC regression model. Furthermore, for evaluating cross-spatial transfer accuracy, the measured SOC data are randomly divided into two parts according to different sampling plots. Half of the sampling plots are used as the training set, and the remaining sampling plots are used as the test set. Due to the differences between sampling plots, the probability distribution curves of SOC between the training set and the test set show significant variation.

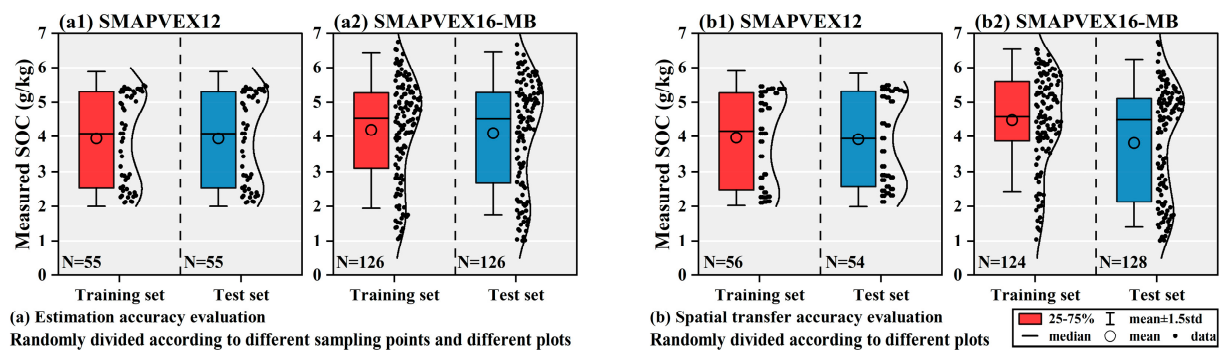


Figure 4. The distribution of the training sets and test sets of the measured SOC in modeling: (a) partition strategy of spatial interpolation accuracy, (a1) SMAPVEX12, (a2) SMAPVEX16-MB, (b) partition strategy of cross-spatial transfer accuracy, (b1) SMAPVEX12, (b2) SMAPVEX16-MB.

The Bayesian optimization (BO) algorithm was used to perform hyperparameter (HP) optimization for each basic regression model, with the acquisition function evaluated based on expected improvement per second. This approach fully exploits the potential of data-driven models for SOC modeling [58]. It is important to note that HP optimization was conducted using a 10-fold cross-validation method within the training set, without involving any data from the test set. The optimization strategies for different regression models are described below (Table S3): For the ETR model, the optimizable HPs include maximum (max) features, max depth, minimum (min) samples leaf (split), min weight fraction leaf, and min impurity decrease [46]; for the XGBR model, the optimizable HPs include subsample, learning rate, Gamma value (complexity control), max depth, min child weight, and regularization parameters (Alpha and Lambda) [47]; for the SVR model, the

optimizable HPs include Kernel function, Box constraint, Kernel scale, and Epsilon [50]; for the GPR model, the optimizable HPs include Sigma value, basis functions, kernel functions, and kernel scale [59]; for the DNNR model, the optimizable HPs include the number of neurons, maximum epochs, learning rates, and batch sizes, Lambda (L2 regularization), and activation functions [60].

3.4. Model Validation

The following indices were used to evaluate the estimation accuracy of the SOC models, including Pearson's correlation coefficient (R), mean absolute error (MAE), and root mean square error ($RMSE$). The expressions are as follows:

$$R = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \quad (1)$$

$$MAE = \frac{1}{n} \times \sum_{i=1}^n |X_{obs,i} - X_{model,i}| \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}} \quad (3)$$

where $Cov(X, Y)$ is the covariance of X and Y . $Var(X)$ and $Var(Y)$ are the variance of X and Y , respectively. n is the total number of verification data. $X_{obs,i}$ and $X_{model,i}$ are i th estimated SOC and the measured SOC, respectively.

4. Results and Discussions

4.1. Sensitivity Analysis of Multi-Source Remote Sensing Features and the Measured SOC

As shown in Figure 5a, for the SMAPVEX12 campaign, the correlation (R) between the measured SOC and L-band quad-pol SAR features in June is significantly better than in July. This is mainly because the crops in June were in the early growth stage with lower vegetation cover, allowing SAR data to penetrate better and capture the soil scattering signal [61]. However, the overall correlation is relatively low ($R = -0.373$ – 0.364), and most of the SAR features fail to pass the significance test across most observation phases ($p > 0.05$). Among the SAR features, the cross-polarization ratio (CPR_H^{L-band}) is relatively more sensitive to SOC. The correlation between the measured SOC and L-band TB features is higher ($R = -0.752$ – 0.818), with MPDI being the most effective. In addition, there are good correlations between the measured SOC and MS data ($R = -0.679$ – 0.364). The correlation of single-band reflectance features is significantly better than that of derived spectral indices, and the SWIR bands show higher correlations. Moreover, there are strong correlations between the measured SOC and temporal SSM and DEM ($R = 0.752$, -0.744), while the correlations with slope and SR (RMSH and CL) are relatively low ($p > 0.05$).

For the SMAPVEX16-MB campaign, similar conclusions are drawn (Figure 5b). The correlation between the measured SOC and C-band dual-pol SAR features is relatively low ($R = -0.353$ – 0.490), with the cross-polarization ratio (CPR_V^{C-band}) performing better than the single-polarization backscattering coefficient. Compared to L-band SAR data, the C-band SAR data have weaker penetration but better represented surface soil scattering signals. TB features still show high correlations ($R = -0.512$ – 0.656), with MPDI remaining the best indicator. The correlation between the measured SOC and MS features during the non-vegetated period ($NDVI_{mean \pm std} = 0.093 \pm 0.046$) is significant ($R = -0.800$ – -0.328 ; $p < 0.001$), with higher correlations in the two SWIR bands. Although the correlations are lower during the vegetation growth period, the two SWIR bands still show relatively high correlations ($p < 0.001$). Compared to SMAPVEX16-MB, the SMAPVEX12 sampling campaign experienced a complete “wet-to-dry” event, resulting in a greater temporal

range of SSM variation. The measured SOC and temporal SSM at all observation phases exhibit high correlations ($R = 0.660\text{--}0.875$), as does the correlation with DEM ($R = -0.772$), while the correlations with slope and soil roughness (RMSH and CL) still remain low ($p > 0.05$). These results suggest that multi-source remote sensing features hold potential for SOC modeling.

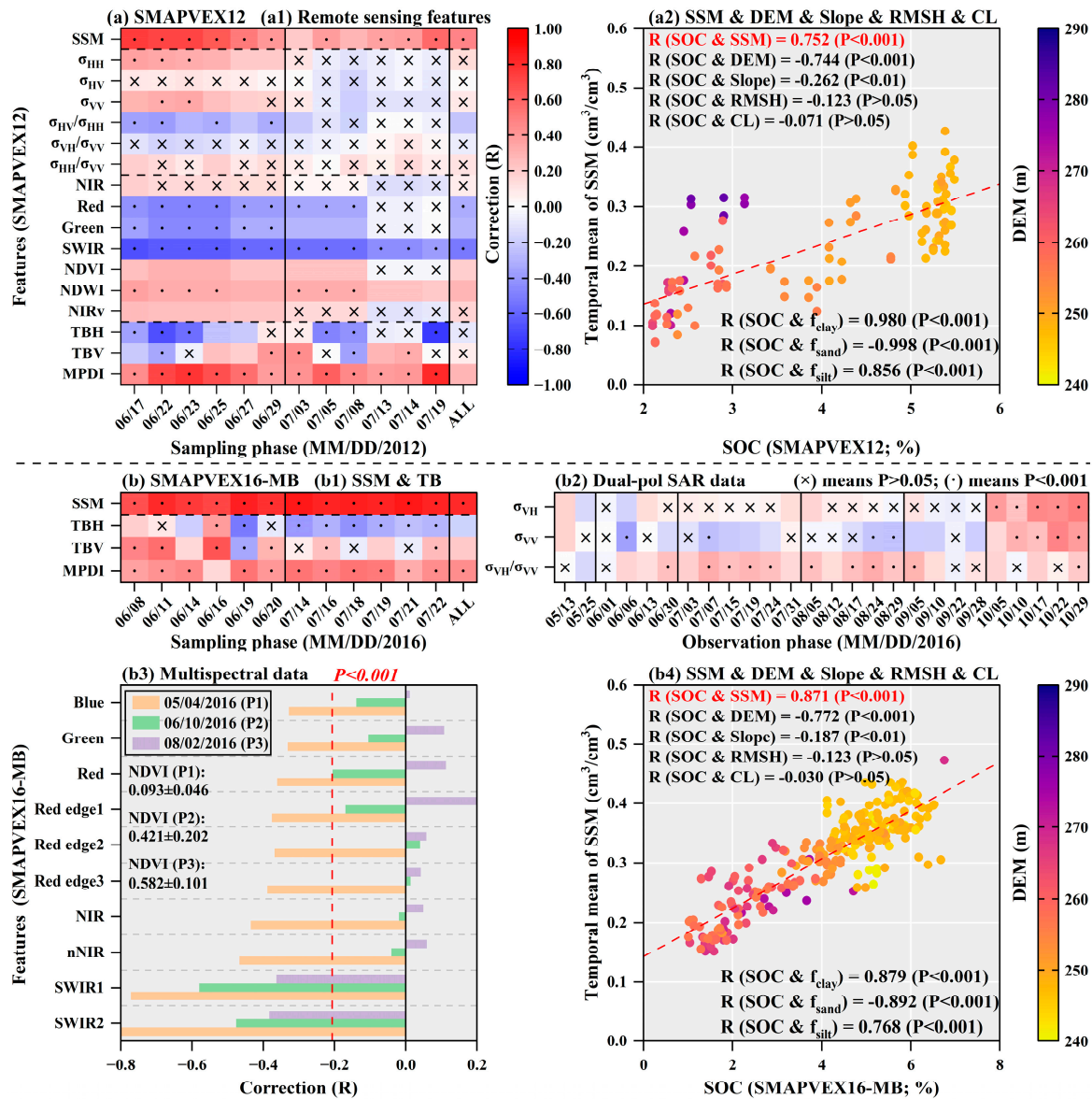


Figure 5. Temporal correlation between different remote sensing features, soil parameters, and the measured SOC: (a) SMAPVEX12 sampling campaign, (a1) the relationship between the remote sensing features (L-band quad-pol SAR data, L-band TB data, and MS data) and the measured SOC, (a2) the relationship between the soil parameters (SSM, DEM, slope, RMSH, and CL) and the measured SOC, (b) SMAPVEX16-MB sampling campaign, (b1) the relationship between the measured SSM, L-band TB data, and the measured SOC, (b2) the relationship between the C-band dual-polar SAR data (Sentinel-1A) and the measured SOC, (b3) the relationship between the MS data (Sentinel-2A) and the measured SOC, (b4) is similar to (a2).

4.2. Spatial Interpolation Accuracy Evaluation of SOC by Using Different Regression Algorithms

It is important to examine the upper and lower bounds and robustness of SOC estimation accuracy when using SAR data, TB data, MS data, soil parameters, and their combined modeling. This is crucial in practical applications, where simultaneous ac-

cess to these data types may not always be possible. As shown in Figure 6a1, for the same regression algorithm (SMAPVEX12 sampling campaign), modeling with only L-band quad-pol SAR data (Group1) yields the lowest SOC estimation accuracy, followed by modeling with only MS data (Group2), whereas using only L-band TB data (Group3) achieves the highest accuracy. Combined modeling of SAR and MS data (Group4) significantly improves estimation accuracy over single data source models but remains lower than Group3. Among the algorithms, the GPR model performs best overall (Group4; $R = 0.924$, $MAE = 0.336$ g/kg, $RMSE = 0.494$ g/kg). For Group3, the metrics are $R = 0.982$, $MAE = 0.178$ g/kg, and $RMSE = 0.229$ g/kg. An evident correlation appears, that is, samples with higher SOC generally correspond to higher SSM, and the opposite also holds true. When the soil parameters (elevation and time-series SSM) are incorporated (Figure 6a2), SOC estimation accuracy markedly improves across all feature groups ($R > 0.900$), although performance differences between feature groups remain similar, with GPR still performing best. Figure 6a3 shows Group4's results, without overestimations or underestimations ($R = 0.959$, $MAE = 0.267$ g/kg, $RMSE = 0.365$ g/kg). However, for some regression algorithms, Group4's accuracy may be slightly lower than Group1 and Group2, indicating that a single remote sensing data source can suffice for SOC modeling when the soil parameters are included.

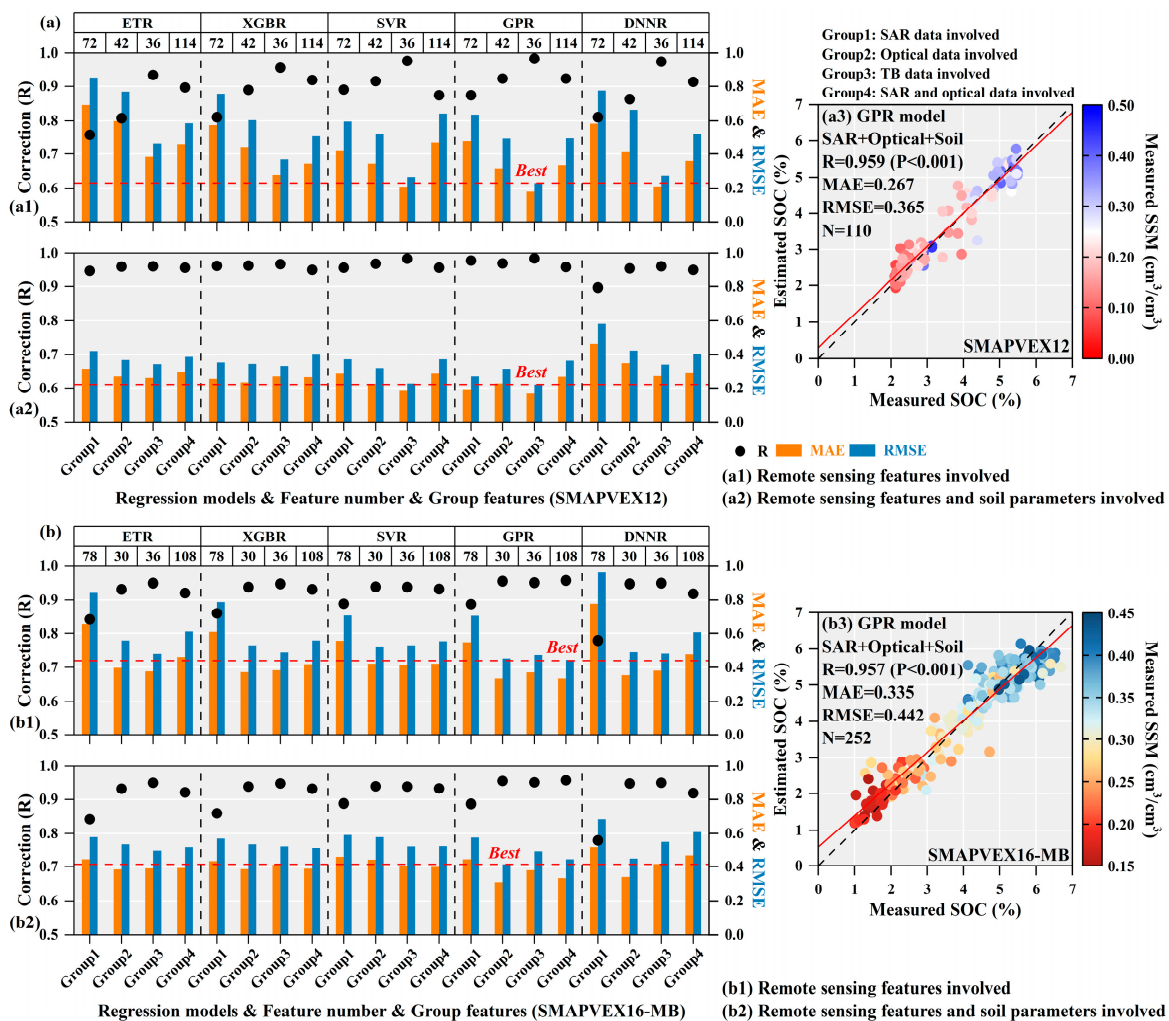


Figure 6. Estimation accuracy of SOC by using different MLR algorithms under different feature groups: (a) SMAPVEX12 sampling campaign; (a1) remote sensing features involved; (a2) remote sensing features and soil parameters involved; (a3) SOC estimation results; (b) is similar to (a), SMAPVEX16-MB sampling campaign.

As illustrated in Figure 6b, compared to the SMAPVEX12 (SOC range: 2.0–5.5 g/kg), the SOC range in the SMAPVEX16-MB sampling campaign is broader (SOC range: 1.0–6.5 g/kg), making SOC modeling more challenging for the latter. Overall, using only C-band dual-pol SAR data for modeling (Group1) still results in the lowest SOC estimation accuracy for a given regression algorithm, while models using only MS data or TB data achieve similar accuracy. Compared to four-band reflectance SPOT4/5 data, ten-band reflectance Sentinel-2 data appears more suitable for SOC modeling. Similarly, when the elevation and time-series SSM factors are introduced, Group1's accuracy improves significantly, with Group2 and Group3 showing smaller improvements (Figure 6b1,b2). Figure 6b3 shows that the GPR model performs best in modeling ($R = 0.957$, $MAE = 0.335$ g/kg, $RMSE = 0.442$ g/kg), and the scatter plots continue to reveal the close relationship between the measured SOC and the measured SSM.

Table S3 shows the hyperparameter values for each regression algorithm that achieved the best SOC estimation accuracy. It is worth noting that while excessive temporal features can lead to a “curse of dimensionality” and may reduce SOC estimation accuracy, this study finds that dimensionality reduction methods like PCA model do not enhance SOC estimation accuracy of regression models. On the contrary, introducing temporal features can yield an upper bound for SOC estimation accuracy. For the SMAPVEX12, modeling with temporal SAR features (12 phases; Table S1) reduces estimation errors by 5.1–20.4% compared to single phase SAR features, as it does for optical and TB features (4.2–10.8% and 3.3–8.6%). The same pattern holds for the SMAPVEX16-MB (26 phases; Table S2), where errors decrease by 7.5–22.9%, 5.5–13.3%, and 4.2–9.8% for each feature groups, respectively. Additionally, this study finds that including spectral indices derived from MS reflectance (Table 1) does not effectively improve SOC estimation accuracy and may even have a slight negative impact. Compared to using only reflectance features, when additional spectral indices are incorporated into the modeling, the estimation errors (RMSE) of the various regression algorithms increased by approximately 3.7–4.3%. Hence, original reflectance features are sufficient for meeting SOC regression modeling requirements.

4.3. Spatial Transfer Accuracy Evaluation of SOC by Using Different Regression Algorithms

Section 4.2 analyzes the interpolation performance of various regression algorithms for SOC modeling. However, it is also essential to investigate the cross-spatial transferability of SOC models, which is particularly crucial for SOC mapping. This analysis is conducted by using certain plots as the training set for SOC modeling, while the remaining plots serve as the test set. As shown in Figure 7a, for the SMAPVEX12 sampling campaign with the GPR model, the SOC accuracy in the validation set is acceptable when L-band quad-pol SAR data are used for modeling alone ($R = 0.895$, $MAE = 0.376$ g/kg, $RMSE = 0.576$ g/kg). However, the transfer accuracy is considerably lower ($R = 0.607$, $MAE = 0.806$ g/kg, $RMSE = 1.043$ g/kg). The same trend is observed when MS data are used, where validation accuracy ($RMSE = 0.523$ g/kg) is significantly better than transfer accuracy ($RMSE = 0.866$ g/kg). Although the inclusion of the terrain factor (DEM) markedly improves validation accuracy, there remains a considerable gap between validation and transfer accuracy ($RMSE = 0.221$ g/kg \rightarrow 0.516 g/kg; $RMSE = 0.199$ g/kg \rightarrow 0.457 g/kg; Figure 7a1–a4). Figure 7a5,a6 show the SOC validation accuracy (10-fold cross-validation accuracy) and test set accuracy (spatial transfer accuracy) of different regression algorithms with different remote sensing data for modeling. It is evident that, without the terrain factor, validation accuracy substantially exceeds transfer accuracy. The inclusion of the terrain factor significantly narrows this accuracy gap. Similar conclusions are reached for the SMAPVEX16-MB sampling campaign (Figure 7b). When C-band dual-pol SAR data are used alone, the SOC validation accuracy ($RMSE = 0.616$ g/kg) outperforms the lower trans-

fer accuracy (RMSE = 1.420 g/kg). The trend is similar with MS data (RMSE = 0.423g/kg and 0.995 g/kg). Introducing the terrain factor substantially improves both validation and transfer accuracy (RMSE = 0.496 g/kg → 1.198 g/kg; RMSE = 0.438 g/kg → 0.799 g/kg; Figure 7b1–b4), although transfer results shows notable overestimation at low values and underestimation at high values, which is consistent across different regression algorithms (Figure 7b5,b6). Figure 7 also illustrates the potential of TB data in modeling the spatial transfer of SOC. However, its relatively low spatial resolution blurs the heterogeneity of SOC within the plots, leading to the illusion of low RMSEs between the estimated and observed values (with relatively low correlation).

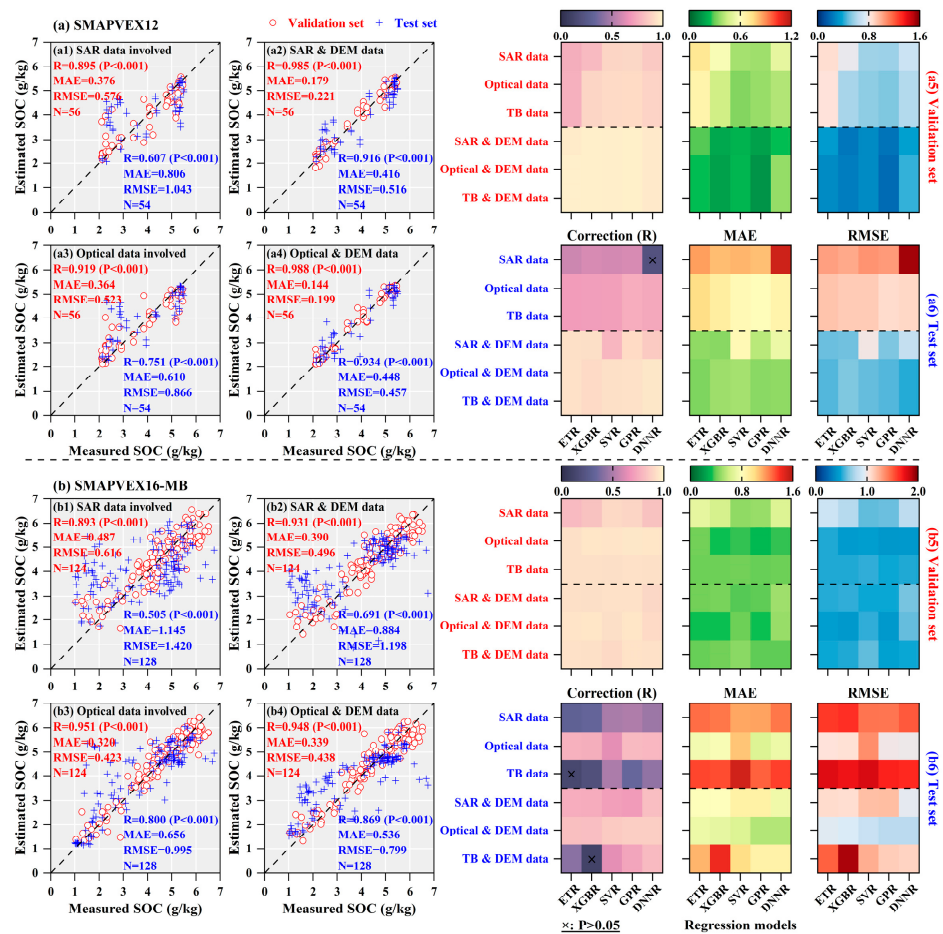


Figure 7. Spatial transfer accuracy of SOC by using different MLR algorithms under different feature groups: (a) SMAPVEX12 sampling campaign, (a1) L-band quad-pol SAR features involved, (a2) quad-pol SAR features and DEM data involved, (a3) optical features involved, (a4) optical features and DEM data involved, (a5,a6) spatial transfer accuracy (R, MAE, and RMSE) of SOC with or without DEM data participation, (b) is similar to (a), SMAPVEX16-MB sampling campaign. C-band dual-pol SAR features.

This indicates that regression models trained on limited measured data struggle to meet the practical accuracy requirements for large-scale SOC mapping. When there is a substantial difference between the actual SOC distribution at the regional scale and the SOC at sampling points, the cross-spatial transfer accuracy of the regression models can decrease significantly, with transfer results potentially becoming unreliable. In other words, the SOC mapping performance of regression models depends on the reasonable distribution of sampling points [13,14]. The study area features relatively flat terrain with an elevation change in less than 30 m, characteristic of the Red River Valley [4]. Sandy soils in the west

are slightly elevated by a moraine, separating them from the heavy clay soils in the east, which accounts for most of the elevation difference. Loam soils predominate in the south near Carman, with high SOC content (2–5%), strong water-holding capacity, and fertility. This also explains the close relationship observed between elevation and the measured SOC. Therefore, the improvement in SOC spatial transfer accuracy by including terrain factors, as observed in this study, may be regionally specific and limited. The inclusion of the SSM factor might also improve SOC spatial transfer accuracy. However, without measured and auxiliary information, obtaining high-resolution (spatial and temporal) SSM raster products remains challenging and uncertain [62,63].

4.4. Comparison with Other Studies

Currently, global-scale SOC modeling and mapping applications typically use low spatial resolution Moderate Resolution Imaging Spectroradiometer (MODIS) data (250–1000 m) [13,15,27]. For regional fine-scale studies, researchers often rely on single optical data sources (such as Landsat or Sentinel-2) for SOC estimation. For example, Zhang et al. explored the potential of various MLR models to estimate SOC in the JiangHan Plain using Landsat-8 data. The results showed that temporal NDVI data-based joint modeling more accurately estimated SOC than single-phase data (RMSE = 3.718 g/kg), but the overall correlation was relatively low ($R = 0.625$). However, due to minimal topographical variation in the study area, the research concluded that the inclusion of terrain factors had little effect on SOC modeling [64]. Castaldi et al. assessed the potential and uncertainty of SOC estimation in Germany's agricultural areas using Sentinel-2 data. This study also highlighted the irreplaceable importance of the SWIR bands in SOC estimation modeling. It further concluded that a 20 m spatial resolution is sufficient to capture SOC heterogeneity between different plots in agricultural areas [65]. Zhou et al. and Wang et al. also demonstrated that combining SAR (C- and L-band) and optical data leads to more accurate SOC estimation in agricultural regions compared to using a single data source [19,66].

However, few studies have explored the potential and value of multi-temporal, multi-source data (including passive and active microwave remote sensing, optical remote sensing, and soil parameters) in SOC estimation modeling for agricultural areas, and even fewer studies have focused on the spatial transferability performance of SOC regression models. This study finds that the SOC modeling performance using TB features obtained from airborne platforms is good; however, its relatively low spatial resolution (450–650 m) limits the fine-scale monitoring of SOC in highly heterogeneous agricultural areas. The even lower spatial resolution (<10 km) of TB data from satellite platforms further constrains its potential for SOC modeling [67]. Therefore, time-series open-access SAR and MS satellite data with relatively high spatial resolution (10–30 m) are ideal data sources and low-cost means for SOC modeling in agricultural areas. In addition, this study addresses the issue of poor spatial transfer accuracy of SOC regression algorithms, which makes it difficult to meet the requirements for large-scale mapping. This issue was improved by introducing terrain factors that are highly sensitive to SOC. However, this approach may have regional limitations. Future research still needs to explore methods for effectively improving SOC spatial transferability and mapping capabilities without relying on specific factors.

5. Conclusions

This study developed an SOC estimation and transfer framework for agricultural areas based on MLR algorithms. It explores the SOC modeling performance and potential of active and passive remote sensing data, including L-band quad-pol SAR data, C-band dual-pol SAR data, L-band TB data, and MS data. Validation and assessment were conducted using two field-observed datasets, leading to the following conclusions:

- (1) Sensitivity analysis shows a strong relationship between the measured SOC and temporal SSM in both dry and rainy seasons, indicating its potential to reflect spatial variations in regional SSM. Additionally, cross-polarization ratio (SAR feature), MPDI index (TB feature), and shortwave infrared reflectance (MS feature) demonstrate high temporal correlation with the measured SOC. Thus, using active microwave, passive microwave, or optical data alone holds potential for SOC modeling.
- (2) The SOC estimation accuracy achieved with MS data alone is comparable to that obtained with TB data alone, both performing slightly better than SAR data. Introducing temporal features can bring optimal SOC estimation accuracy across all regression algorithms. The spatial interpolation of each regression algorithm is satisfactory, with the GPR algorithm achieving the best SOC modeling performance (RMSE = 0.365 g/kg and 0.442 g/kg for SMAPVEX12 and SMAPVEX16-MB sampling campaigns).
- (3) The cross-spatial transfer accuracy of MLR algorithms remains limited (RMSE = 0.866–1.043 g/kg and 0.995–1.679 g/kg for different data sources). To reduce uncertainties in cross-spatial SOC transfer, this study incorporates terrain factors sensitive to regional-scale SOC (RMSE = 0.457–0.516 g/kg and 0.799–1.198 g/kg for different data sources). The SOC estimation and transfer framework proposed in this study provides valuable guidance for high-resolution, regional-scale SOC mapping and applications, with substantial application potential for open-access Sentinel and NISAR satellites.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs17020333/s1>. The manuscript includes Supplementary Material, which mainly contains information on the remote sensing images used (Tables S1 and S2) and the hyperparameter settings for each regression model (Table S3).

Author Contributions: Conceptualization, J.Q.; Methodology, J.Q.; Software, J.Q., C.D. and Q.D.; Investigation, J.Q.; Data curation, J.Q., C.D. and Q.D.; Writing—original draft, J.Q.; Writing—review & editing, J.Q., J.Y., W.S., L.Z., L.S. and H.S.; Visualization, J.Q.; Funding acquisition, J.Y., W.S., L.Z., L.S. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant number 42071295, 42171442, 62471337, U2033216, U22A2010, the Natural Science Foundation of Hubei Province grant number 2022CFB193, and Sichuan Science and Technology Program grant number 2023YFG0123.

Data Availability Statement: The data used in this study are all openly accessible, including SMAPVEX12 Dataset (<https://smapvex12.espaceweb.usherbrooke.ca>, accessed on 10 January 2025), SMAPVEX16-MB Dataset (<https://smapvex16-mb.espaceweb.usherbrooke.ca>, accessed on 10 January 2025), Sentinel-1A, Sentinel-2A (<https://dataspace.copernicus.eu/>, accessed on 10 January 2025), SPOT-4, SPOT-5 (<https://cnes.fr/en>, accessed on 10 January 2025), SRTM, UAVSAR, and PLAS data (<https://vertex.daac.asf.alaska.edu>, accessed on 10 January 2025).

Acknowledgments: The authors are grateful to the European Space Agency for providing Sentinel-1A and Sentinel-2A data, the French space agency for providing SPOT-4/5 data, the National Aeronautics and Space Administration for providing SMAPVEX12 and SMAPVEX16-MB datasets, UAVSAR, PALS, and SRTM data. Sincere gratitude to the scientists who contributed to the aforementioned dataset. We sincerely thank the reviewers for their comments, which have greatly helped to improve this study.

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

References

1. Angelopoulou, T.; Tziolas, N.; Balafoutis, A.; Zalidis, G.; Bochtis, D. Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sens.* **2019**, *11*, 676. [[CrossRef](#)]
2. Chappell, A.; Baldock, J.; Sanderman, J. The global significance of omitting soil erosion from soil organic carbon cycling schemes. *Nat. Clim. Change* **2016**, *6*, 187–191. [[CrossRef](#)]
3. Stockmann, U.; Adams, M.A.; Crawford, J.W.; Field, D.J.; Henakaarchchi, N.; Jenkins, M.; Minasny, B.; McBratney, A.B.; De Courcelles, V.d.R.; Singh, K. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agric. Ecosyst. Environ.* **2013**, *164*, 80–99. [[CrossRef](#)]
4. Manns, H.R.; Berg, A.A. Importance of soil organic carbon on surface soil water content variability among agricultural fields. *J. Hydrol.* **2014**, *516*, 297–303. [[CrossRef](#)]
5. Terrer, C.; Phillips, R.P.; Hungate, B.A.; Rosende, J.; Pett-Ridge, J.; Craig, M.E.; van Groenigen, K.J.; Keenan, T.F.; Sulman, B.N.; Stocker, B.D. A trade-off between plant and soil carbon storage under elevated CO₂. *Nature* **2021**, *591*, 599–603. [[CrossRef](#)]
6. Baldock, J.; Sanderman, J.; Macdonald, L.; Puccini, A.; Hawke, B.; Szarvas, S.; McGowan, J. Quantifying the allocation of soil organic carbon to biologically significant fractions. *Soil Res.* **2013**, *51*, 561–576. [[CrossRef](#)]
7. Rawls, W.; Nemes, A.; Pachepsky, Y. Effect of soil organic carbon on soil hydraulic properties. *Dev. Soil Sci.* **2004**, *30*, 95–114.
8. Rawls, W.; Pachepsky, Y.A.; Ritchie, J.; Sobecki, T.; Bloodworth, H. Effect of soil organic carbon on soil water retention. *Geoderma* **2003**, *116*, 61–76. [[CrossRef](#)]
9. Smith, P.; Fang, C.; Dawson, J.J.; Moncrieff, J.B. Impact of global warming on soil organic carbon. *Adv. Agron.* **2008**, *97*, 1–43.
10. Johns, T.J.; Angove, M.J.; Wilkens, S. Measuring soil organic carbon: Which technique and where to from here? *Soil Res.* **2015**, *53*, 717–736.
11. Bhunia, G.S.; Shit, P.K.; Maiti, R. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *J. Saudi Soc. Agric. Sci.* **2018**, *17*, 114–126. [[CrossRef](#)]
12. Danesh, M.; Taghipour, F.; Emadi, S.M.; Ghajar Sepanlou, M. The interpolation methods and neural network to estimate the spatial variability of soil organic matter affected by land use type. *Geocarto Int.* **2022**, *37*, 11306–11315. [[CrossRef](#)]
13. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)] [[PubMed](#)]
14. Liu, F.; Zhang, G.-L.; Song, X.; Li, D.; Zhao, Y.; Yang, J.; Wu, H.; Yang, F. High-resolution and three-dimensional mapping of soil texture of China. *Geoderma* **2020**, *361*, 114061. [[CrossRef](#)]
15. Poggio, L.; De Sousa, L.M.; Batjes, N.H.; Heuvelink, G.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil* **2021**, *7*, 217–240. [[CrossRef](#)]
16. Zhou, T.; Geng, Y.; Chen, J.; Liu, M.; Haase, D.; Lausch, A. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. *Ecol. Indic.* **2020**, *114*, 106288. [[CrossRef](#)]
17. Guo, L.; Sun, X.; Fu, P.; Shi, T.; Dang, L.; Chen, Y.; Linderman, M.; Zhang, G.; Zhang, Y.; Jiang, Q. Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. *Geoderma* **2021**, *398*, 115118. [[CrossRef](#)]
18. Nguyen, T.T.; Pham, T.D.; Nguyen, C.T.; Delfos, J.; Archibald, R.; Dang, K.B.; Hoang, N.B.; Guo, W.; Ngo, H.H. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Sci. Total Environ.* **2022**, *804*, 150187. [[CrossRef](#)]
19. Wang, X.; Zhang, Y.; Atkinson, P.M.; Yao, H. Predicting soil organic carbon content in Spain by combining Landsat TM and ALOS PALSAR images. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102182. [[CrossRef](#)]
20. Lacoste, M.; Minasny, B.; McBratney, A.; Michot, D.; Viaud, V.; Walter, C. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma* **2014**, *213*, 296–311. [[CrossRef](#)]
21. Heuvelink, G.B.; Angelini, M.E.; Poggio, L.; Bai, Z.; Batjes, N.H.; van den Bosch, R.; Bossio, D.; Estella, S.; Lehmann, J.; Olmedo, G.F. Machine learning in space and time for modelling soil organic carbon change. *Eur. J. Soil Sci.* **2021**, *72*, 1607–1623. [[CrossRef](#)]
22. Liu, F.; Wu, H.; Zhao, Y.; Li, D.; Yang, J.-L.; Song, X.; Shi, Z.; Zhu, A.X.; Zhang, G.-L. Mapping high resolution National Soil Information Grids of China. *Sci. Bull.* **2022**, *67*, 328–340. [[CrossRef](#)] [[PubMed](#)]
23. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sens.* **2020**, *12*, 2234. [[CrossRef](#)]
24. Yang, R.-M.; Huang, L.-M.; Zhang, X.; Zhu, C.-M.; Xu, L. Mapping the distribution, trends, and drivers of soil organic carbon in China from 1982 to 2019. *Geoderma* **2023**, *429*, 116232. [[CrossRef](#)]
25. Lamichhane, S.; Kumar, L.; Wilson, B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* **2019**, *352*, 395–413. [[CrossRef](#)]
26. Nachtergaele, F.; van Velthuisen, H.; Verelst, L.; Wiberg, D.; Henry, M.; Chiozza, F.; Yigini, Y.; Aksoy, E.; Batjes, N.; Boateng, E. *Harmonized World Soil Database Version 2.0*; FAO: Rome, Italy, 2023.

27. Wang, S.; Guan, K.; Zhang, C.; Lee, D.; Margenot, A.J.; Ge, Y.; Peng, J.; Zhou, W.; Zhou, Q.; Huang, Y. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: Assessing potential of airborne and spaceborne optical soil sensing. *Remote Sens. Environ.* **2022**, *271*, 112914. [[CrossRef](#)]
28. Bursač, P.; Kovačević, M.; Bajat, B. Instance-based transfer learning for soil organic carbon estimation. *Front. Environ. Sci.* **2022**, *10*, 1003918. [[CrossRef](#)]
29. McNairn, H.; Jackson, T.J.; Powers, J.; Bélair, S.; Berg, A.; Bullock, P.; Colliander, A.; Cosh, M.H.; Kim, S.-B.; Magagi, R. SMAPVEX16 Database Report. *Agric. Agri-Food Can.* **2017**.
30. McNairn, H.; Jackson, T.J.; Wiseman, G.; Belair, S.; Berg, A.; Bullock, P.; Colliander, A.; Cosh, M.H.; Kim, S.-B.; Magagi, R. The soil moisture active passive validation experiment 2012 (SMAPVEX12): Prelaunch calibration and validation of the SMAP soil moisture algorithms. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2784–2801. [[CrossRef](#)]
31. Manns, H.; Maxwell, C.; Emery, R. The effect of ground cover or initial organic carbon on soil fungi, aggregation, moisture and organic carbon in one season with oat (*Avena sativa*) plots. *Soil Tillage Res.* **2007**, *96*, 83–94. [[CrossRef](#)]
32. Wang, Q.; Li, Y.; Wang, Y. Optimizing the weight loss-on-ignition methodology to quantify organic and carbonate carbon of sediments from diverse sources. *Environ. Monit. Assess.* **2011**, *174*, 241–257. [[CrossRef](#)] [[PubMed](#)]
33. Baldock, J.; Nelson, P. *Soil Organic Matter*; CRC Press: Boca Raton, FL, USA, 2000.
34. Mladenova, I.E.; Jackson, T.J.; Bindlish, R.; Hensley, S. Incidence angle normalization of radar backscatter data. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 1791–1804. [[CrossRef](#)]
35. Ulaby, F.; Moore, R.; Fung, A. *Microwave Remote Sensing: Active and Passive. Volume 2-Radar Remote Sensing and Surface Scattering and Emission Theory*; NASA: Washington, DC, USA, 1982.
36. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J. The soil moisture active passive (SMAP) mission. *Proc. IEEE* **2010**, *98*, 704–716. [[CrossRef](#)]
37. Owe, M.; de Jeu, R.; Walker, J. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1643–1654. [[CrossRef](#)]
38. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
39. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
40. Basak, D.; Pal, S.; Patranabis, D.C. Support vector regression. *Neural Inf. Process. Lett. Rev.* **2007**, *11*, 203–224.
41. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In Proceedings of the 10th International Conference on Neural Information Processing Systems, Denver, CO, USA, 3–5 December 1996; Volume 9.
42. Awad, M.; Khanna, R.; Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Apress: Berkeley, CA, USA, 2015; pp. 67–80.
43. Camps-Valls, G.; Verrelst, J.; Munoz-Mari, J.; Laparra, V.; Mateo-Jimenez, F.; Gomez-Dans, J. A survey on Gaussian processes for earth-observation data analysis: A comprehensive investigation. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 58–78. [[CrossRef](#)]
44. Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [[CrossRef](#)]
45. Williams, C.; Rasmussen, C. Gaussian processes for regression. In Proceedings of the 9th International Conference on Neural Information Processing Systems, Denver, CO, USA, 27 November–2 December 1995; Volume 8.
46. Xue, Z.; Zhang, Y.; Zhang, L.; Li, H. Ensemble learning embedded with Gaussian process regression for soil moisture estimation: A case study of the continental us. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [[CrossRef](#)]
47. Qian, J.; Jie, Y.; Weidong, S.; Lingli, Z.; Lei, S.; Chaoya, D. Evaluation and improvement of temporal robustness and transfer performance of surface soil moisture estimated by machine learning regression algorithms. *Comput. Electron. Agric.* **2024**, *217*, 108518.
48. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A survey of deep neural network architectures and their applications. *Neurocomputing* **2017**, *234*, 11–26. [[CrossRef](#)]
49. Liu, Q.; He, L.; Guo, L.; Wang, M.; Deng, D.; Lv, P.; Wang, R.; Jia, Z.; Hu, Z.; Wu, G. Digital mapping of soil organic carbon density using newly developed bare soil spectral indices and deep neural network. *Catena* **2022**, *219*, 106603. [[CrossRef](#)]
50. Ji, K.; Wu, Y. Scattering mechanism extraction by a modified cloude-pottier decomposition for dual polarization SAR. *Remote Sens.* **2015**, *7*, 7447–7470. [[CrossRef](#)]
51. Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.-F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [[CrossRef](#)]
52. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA Special Publications: Washington, DC, USA, 1974; Volume 351, p. 309.
53. Horler, D.; Ahern, F. Forestry information content of Thematic Mapper data. *Int. J. Remote Sens.* **1986**, *7*, 405–428. [[CrossRef](#)]
54. Badgley, G.; Field, C.B.; Berry, J.A. Canopy near-infrared reflectance and terrestrial photosynthesis. *Sci. Adv.* **2017**, *3*, e1602244. [[CrossRef](#)]

55. 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](#)]
56. Gitelson, A.; Merzlyak, M.N. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [[CrossRef](#)]
57. Liu, Y.; Qian, J.; Yue, H. Comprehensive evaluation of Sentinel-2 red edge and shortwave-infrared bands to estimate soil moisture. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7448–7465. [[CrossRef](#)]
58. Wu, J.; Chen, X.-Y.; Zhang, H.; Xiong, L.-D.; Lei, H.; Deng, S.-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.
59. Rasmussen, C.E. Gaussian processes in machine learning. In *Summer School on Machine Learning*; Springer: New York, NY, USA, 2003; pp. 63–71.
60. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
61. Shi, H.; Zhao, L.; Yang, J.; Lopez-Sanchez, J.M.; Zhao, J.; Sun, W.; Shi, L.; Li, P. Soil moisture retrieval over agricultural fields from L-band multi-incidence and multitemporal PolSAR observations using polarimetric decomposition techniques. *Remote Sens. Environ.* **2021**, *261*, 112485. [[CrossRef](#)]
62. Zhu, L.; Si, R.; Shen, X.; Walker, J.P. An advanced change detection method for time-series soil moisture retrieval from Sentinel-1. *Remote Sens. Environ.* **2022**, *279*, 113137. [[CrossRef](#)]
63. Paloscia, S.; Pettinato, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Reppucci, A. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.* **2013**, *134*, 234–248. [[CrossRef](#)]
64. Zhang, Y.; Guo, L.; Chen, Y.; Shi, T.; Luo, M.; Ju, Q.; Zhang, H.; Wang, S. Prediction of Soil Organic Carbon based on Landsat 8 Monthly NDVI Data for the Jiangnan Plain in Hubei Province, China. *Remote Sens.* **2019**, *11*, 1683. [[CrossRef](#)]
65. Castaldi, F.; Hueni, A.; Chabrillat, S.; Ward, K.; Buttafuoco, G.; Bomans, B.; Vreys, K.; Brell, M.; van Wesemael, B. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 267–282. [[CrossRef](#)]
66. Zhou, Y.; Zhao, X.; Guo, X.; Li, Y. Mapping of soil organic carbon using machine learning models: Combination of optical and radar remote sensing data. *Soil Sci. Soc. Am. J.* **2022**, *86*, 293–310. [[CrossRef](#)]
67. Long, D.G.; Brodzik, M.J.; Hardman, M.A. Enhanced-resolution SMAP brightness temperature image products. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4151–4163. [[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.