

Article

Collaborative Utilization of Sentinel-1/2 and DEM Data for Mapping the Soil Organic Carbon in Forested Areas Based on the Random Forest

Zeqiang Wang ^{1,2,3}, Dongyou Zhang ^{1,2,*}, Xibo Xu ^{3,*}, Tingyu Lu ^{1,2} and Guanghui Yang ⁴

¹ College of Geographical Sciences, Harbin Normal University, Harbin 150025, China; wangzeqiang@stu.hrbnu.edu.cn (Z.W.); hrblutingyu@163.com (T.L.)

² Heilongjiang Province Key Laboratory of Geographical Environment Monitoring and Spatial Information Service in Cold Regions, Harbin Normal University, Harbin 150025, China

³ Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China

⁴ Jilin Emergency Warning Information Dissemination Center, Changchun 130062, China; yghui0517@163.com

* Correspondence: zhangdy@hrbnu.edu.cn (D.Z.); xuxibo@stu.sdnu.edu.cn (X.X.); Tel.: +86-133-0481-2601 (D.Z.)

Abstract: Optical remote sensing data are widely used for constructing soil organic carbon (SOC) mapping models. However, it is challenging to map SOC in forested areas because atmospheric water vapor affects the results derived from optical remote sensing data. To address this issue, we utilized Sentinel-1, Sentinel-2, and digital elevation model (DEM) data to obtain a comprehensive feature set (including S1-based textural indices, S2-based spectral indices, and DEM-derived indices) to map the SOC content in forested areas. The features set were the predictor variables, and the measured SOC content was the dependent variable. The random forest algorithm was used to establish the SOC model. The ratio of performance to inter-quartile range (RPIQ) was 2.92 when the S2-based spectral indices were used as predictor variables. When the comprehensive feature set was utilized as the model input, the model achieved an RPIQ of 4.13 ($R^2 = 0.91$, root mean square error (RMSE) = 9.18), representing a 41.44% improvement in model accuracy. The average SOC content in the Greater Khingan Mountains was 43.75 g kg⁻¹. The northern and southwestern parts had higher SOC contents (>54.93 g kg⁻¹), while the southeastern and northwestern parts had lower contents (<39.83 g kg⁻¹). This discrepancy was primarily attributed to agricultural activities. The results indicate that using a comprehensive feature set and the random forest algorithm is a reliable approach for estimating the spatial distribution of the SOC content in forested areas and is suitable for forest ecology and carbon management studies.

Keywords: soil organic carbon; Sentinel-1; Sentinel-2; spatial distribution; random forest; forested area



Citation: Wang, Z.; Zhang, D.; Xu, X.; Lu, T.; Yang, G. Collaborative Utilization of Sentinel-1/2 and DEM Data for Mapping the Soil Organic Carbon in Forested Areas Based on the Random Forest. *Forests* **2024**, *15*, 218. <https://doi.org/10.3390/f15010218>

Received: 12 December 2023

Revised: 16 January 2024

Accepted: 18 January 2024

Published: 22 January 2024



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

1. Introduction

The soil is the largest carbon reservoir on the Earth's surface and is vital in maintaining the carbon balance and mitigating climate change [1,2]. Soils in forested areas contain large amounts of organic carbon, with high rates of carbon absorption and conversion. The accumulation of organic carbon in soils can reduce the carbon dioxide concentration in the atmosphere, mitigating global warming, especially in forested areas [3–5]. Therefore, understanding the content and spatial distribution of soil organic carbon (SOC) in forest areas is crucial for regulating the global carbon equilibrium and mitigating climate change [6,7].

Three methods have been commonly used to map SOC in forest areas. The first method involves establishing field sampling points and conducting chemical analysis in the laboratory, combined with a geostatistical analysis to estimate SOC distribution [8,9]. This method is simple and accurate but labor-intensive, and it is difficult to update the data. The second method is the soil-landscape model proposed by [10]. It predicts the

SOC content using environmental factors related to the SOC as predictive variables and the measured SOC content as the dependent variable to establish a distribution prediction model for SOC. This method is typically specific to a region and is not widely applicable to areas outside the study area [11]. The third method uses spectral information from optical images (such as Sentinel-2) to obtain vegetation information related to the SOC to predict the SOC content [12–15]. This approach enables SOC mapping using remote sensing in forested areas because spectral information on the soil cannot be obtained directly due to vegetation cover in forested areas.

However, the atmospheric water vapor content is relatively high over forest areas due to tree transpiration and temperature differences, affecting the spectral reflectance data and resulting in significant uncertainty in SOC mapping relying solely on optical remote sensing data [16,17]. Therefore, it is challenging to estimate the SOC content in forested areas using only spectral features, and the model accuracy and stability are insufficient.

Therefore, this study utilizes Sentinel-1 (S1), Sentinel-2 (S2), and digital elevation model (DEM) data to obtain a comprehensive feature set for predicting the SOC content in forest areas. The features set are the predictor variables, and the measured SOC content is the dependent variable. The random forest (RF) algorithm is used to map SOC in forested areas at the regional scale.

The Sentinel-1 mission comprises a constellation of two polar-orbiting satellites equipped with active microwave sensors. The C-band synthetic aperture radar (SAR) sensor onboard the satellites has a spatial resolution of 5–20 m and captures backscatter information from multiple polarization channels on the Earth's surface. SAR is highly sensitive to surface roughness and is not affected by cloud cover or water vapor; thus, it has been used as complementary data to optical imagery to establish SOC mapping models [18,19]. Additionally, the relief significantly affects soil formation, resulting in spatial heterogeneity in soil properties at the regional scale [20]. The topographic relief influences SOC accumulation, therefore, topographic variables are indicators of the SOC content.

Random Forest (RF) models are effective for determining the mapping relationship between the SOC content and a comprehensive feature set. Zhang et al. [15] used remote sensing data and climatic and topographical variables as covariates with the RF algorithm to estimate the SOC content of coastal wetland soil. Nabiollahi et al. [21] used the RF model combined with Landsat 8 images to predict SOC contents in forests, wetlands, and cultivated land. The RF model exhibited high robustness. Xu et al. [22] employed an optimized RF model alongside the remote sensing indices to map the SOC content across the Weibei Plain in China. The aforementioned researchers utilized remote sensing data and the RF model to estimate SOC, obtaining satisfactory accuracy. The RF algorithm is a machine learning method that uses several parallel weak estimators and voting on the results of all decision trees. It is not affected by variable collinearity, has high generalization ability, and is applicable to describe the nonlinear relationships between the features and the soil properties [23–25].

The objectives of this study are to: (1) construct a comprehensive feature set to predict the SOC content in forested areas using S1, S2, and DEM data; and (2) establish an RF-based SOC mapping model for forested areas.

2. Materials and Methods

2.1. Study Area

The study area (Greater Khingan Mountains) is located northwest of Heilongjiang Province and northeast of the Inner Mongolia Autonomous Region (Figure 1). It covers an area of about 84,800 km² (50°07'02" N~53°33'42" N and 121°10'53" E~127°01'21" E). The area has a cold temperate monsoon climate. The average temperature is −22.7 °C in winter and 16.5 °C in summer. The annual temperature difference is large, and the annual average rainfall is about 460 mm. The terrain is high in the southwest and low in the northeast. The average altitude is 573 m. The dominant soil types are Podzol and Phaeozems [26].

Forest area (e.g., coniferous forests and broad-leaved forests) account for 79.83% of the Greater Khingan Mountains. Vegetation has a positive effect on the accumulation of SOC. Therefore, this region is an important area for forestry products in China.

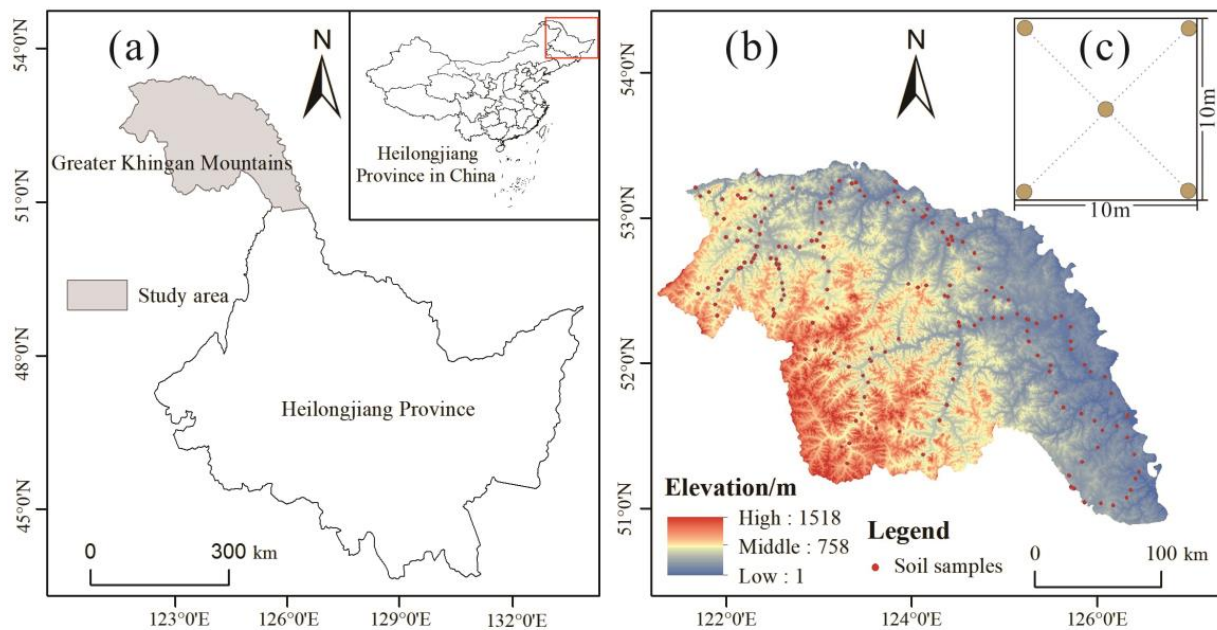


Figure 1. Location of sampling sites in the study area: (a) map of the study area; (b) location of soil samples; (c) five-point sampling method.

2.2. Sample Collection and Chemical Analysis

The soil type, topography, and transportation conditions in the Greater Khingan Mountains were considered to determine the location of 177 soil sampling sites (Figure 1b) using ArcGIS 10.2 software (ESRI Inc., Redlands, CA, USA). Field sampling was conducted in July 2018 to minimize the impact of ice and snow cover in the satellite images. Surface soil samples (0–20 cm) were collected using the five-point sampling method (Figure 1c) and a soil drill [27]. The soil samples at each sampling point were mixed (500–600 g), packed into an airtight plastic bag, and sent to the laboratory for subsequent processing.

The latitude and longitude coordinates of the sample locations were recorded using a handheld GPS, and photos of the surrounding environment were obtained with a camera. The samples were air-dried (25 °C), and weeds, gravel, and other debris were removed. They were ground and passed through a 2 mm nylon mesh. The soil samples were weighed (0.1 g) and placed in a porcelain boat. Then, 2 mol/L hydrochloric acid was added to remove inorganic carbon, followed by oven-drying at 70 °C. The samples were placed in a Multi N/C 3100 analyzer (Jena, Germany) to determine the SOC content [28].

2.3. Deriving a Comprehensive Feature Set for Characterizing SOC in Forested Areas

2.3.1. S1, S2, and DEM Data Acquisition and Preprocessing

The European Space Agency (ESA) launched S1, an Earth observation satellite. It has a C-band SAR sensor capable of penetrating clouds and fog, enabling all-weather and all-day monitoring, and providing early warning data. Six Interferometric Wide-Swath Ground Range Detected images acquired on 1 October 2018 and 17 October 2018 were downloaded from the ESA Open Access Center <https://scihub.copernicus.eu/> (accessed on 10 January 2023). The radar tool of the Sentinel Application Platform (SNAP) software (version, 2020) was used to convert the C-band SAR raw intensity signal data into polarization backscattering coefficients of the vertical transmit-vertical receive (VV) and the vertical transmit-horizontal receive (VH) channels. Preprocessing included thermal and border noise removal, orbit correction, radiometric calibration, speckle filtering, terrain correction, and conversion from the decibel (dB) scale to the linear scale [29,30].

S2 consists of S2A and S2B, which were launched by the ESA in 2015 and 2017, respectively. The purpose is for environmental monitoring and disaster relief. The S2 images have 13 bands, and the satellite has a 10-day revisit period. The red-edge bands (Band 5, Band 6, and Band 7) are highly sensitive to vegetation properties [31]. More details on S2 have been described in [32]. Seventeen scenes of S2 Level-1C images were downloaded from the ESA Center. The images were acquired on 1 October 2018 and 17 October 2018. Radiometric calibration and atmospheric correction of the Level-1C data were executed in SNAP software to obtain ground reflectance (Level-1A) data. The images were resampled to a 10 m resolution. Band 1 is the coastal band, and bands 9 and 10 are aerosol and atmospheric bands, respectively. These were not used in the SOC modeling [33].

DEM contains ground elevation data, which provide information on SOC in forested areas. The 30 m DEM data were downloaded from the Geospatial Data Cloud website <http://www.gscloud.cn/> (accessed on 15 January 2023). They were resampled to 10 m resolution and co-registered with the S1 and S2 data.

2.3.2. S2-Based Spectral Indices

A spectral index is a spectral band combination containing useful information in remote sensing images. They reduce the complexity of data processing. Therefore, in addition to the 10 bands of the S2 imagery, 16 spectral indices closely related to the SOC content were selected from the literature to predict the SOC content in forest areas: clay mineral ratio (CMR), normalized difference built-up index (NDBI), normalized difference vegetation index (NDVI), forest wetland index (FWI), normalized difference tillage index (NDTI), enhanced vegetation index (EVI), red edge position index (REPI), soil-adjusted vegetation index (SAVI), green normalized difference vegetation index (GNDVI), inverted red edge chlorophyll index (IRECI), infrared percentage vegetation index (IPVI), ratio vegetation index (RVI), difference vegetation index (DVI), modified normalized difference water index (MNDWI), red green ratio index (RGRI), and green vegetation index (GVI). Table 1 lists the index equations.

Table 1. S2-based spectral indices and their equations.

ID	Equations	Reference
1	Single bands	---
2	CMR = SWIR1/SWIR2	[34]
3	NDBI = (SWIR1 – NIR)/(SWIR1 + NIR)	[35]
4	NDVI = (NIR – Red)/(NIR + Red)	[36]
5	FWI = (NIR – Red)/SWIR2	[37]
6	NDTI = (SWIR1 – SWIR2)/(SWIR1 + SWIR2)	[38]
7	EVI = 2.5(NIR – Red)/(NIR + 6Red – 7.5Blue + 1)	[39]
8	REPI = 700 + 35(((Red + Red–edge 3)/2) – Red–edge 1)/(Red–edge 2 – Red–edge 1)	[40]
9	SAVI = 1.5((NIR – Red)/(NIR + Red + 0.5))	[41]
10	GNDVI = (NIR – Green)/(NIR + Green)	[42]
11	IRECI = (Red–edge 3 – Red)/(Red–edge 1 + Red–edge 2)	[43]
12	IPVI = NIR/(NIR + Red)	[44]
13	RVI = NIR/Red	[45]
14	DVI = NIR – Red	[46]
15	MNDWI = (Green – SWIR1)/(Green + SWIR2)	[47]
16	RGRI = Red/Blue	[48]
17	GVI = NIR/Green	[49]

2.3.3. S1-Based Textural Indices

S1-based textural indices were derived from the S1 radar image. Texture information represents the macrostructure and microstructure of the forest area. Two polarimetric backscattering coefficients of the S1 image were used to extract four textural indices, including the difference between the polarization coefficients of VV and VH (DI), the ratio of the polarization coefficients of VV and VH (RI), the normalized difference index of the

polarization coefficients between VV and VH (NDI), and the average of the sum of the polarization coefficients of VV and VH (AVE). Additionally, texture features based on the gray-level co-occurrence matrix were extracted [50], including the mean (MEA), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (SM), and correlation (COR). Table 2 lists the equations for the textural indices.

Table 2. S1-based textural indices and their equations.

ID	Equations	Reference
1	VH = Vertical transmit-horizontal receive polarization	---
2	VV = Vertical transmit-vertical receive polarization	---
3	DI = VH − VV	[51]
4	RI = VH/VV	[51]
5	NDI = (VH − VV)/(VH + VV)	[51]
6	AVE = (VH + VV)/2	[51]
7	$MEA = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j, d, \theta)$	[52]
8	$VAR = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (i - \mu)^2 f(i, j, d, \theta)$	[52]
9	$HON = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \frac{f(i, j, d, \theta)}{1 + (i - j)^2}$	[52]
10	$CON = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (i - j)^2 f(i, j, d, \theta)$	[52]
11	$DIS = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i - j f(i, j, d, \theta)$	[52]
12	$ENT = - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j) \lg f(i, j, d, \theta)$	[52]
13	$SM = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j, d, \theta)^2$	[52]
14	$COR = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \frac{(i - \mu)(j - \mu) f(i, j, d, \theta)^2}{\sigma^2}$	[52]

Note: $f(i, j)$ represents the frequency of simultaneous occurrence of pixel pairs with gray levels of i and j ($i, j = 0, 1, 2, 3, \dots, N$); d is the relative distance represented by the number of pixels; θ denotes the four directions: 0° , 45° , 90° , and 135° .

2.3.4. DEM-Derived Indices

DEM-derived indices are critical surface parameters that reflect the distribution of soil material and energy [53]. Therefore, 15 DEM-derived indices were extracted to characterize SOC in forested areas: analytical hillshading (AH), aspect (A), closed depressions (CD), convergence index (CI), plan curvature (PC), relative slope position (RSP), topographic wetness index (TWI), channel network base level (CNBL), relative slope position (RSP), length-slope factor (LSF), profile curvature (PFC), slope (S), channel network distance (CND), total catchment area (TCA), and valley depth (VD). The DEM-derived indices were calculated using SAGA software (Automated Geoscientific Analyses, Marburg, Germany, version 7.9) [54].

2.3.5. Predictor Covariate Selection

The Spearman correlation coefficient-variance inflation factor (SCC-VIF) was used to select the best predictive variables. It considers the multicollinearity between parameters and the degree of information redundancy. Using predictive variables closely related to the SOC content can improve the training efficiency, robustness, and performance of the mapping model. The SCC-VIF is calculated as follows:

$$SCC = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

$$VIF = \frac{1}{1 - R^2} \quad (2)$$

where n is the total number of soil samples, d_i is the difference in the SOC content between two samples, and $\sum d_i^2$ is the squared sum of the differences between the two samples. A $VIF < 3$ indicates no multicollinearity between the predictors, $VIF > 3$ indicates potential multicollinearity, $VIF > 5$ indicates a low probability of multicollinearity, and $VIF > 10$ indicates a high probability of multicollinearity [55].

2.4. Mapping Model and Evaluation

The RF algorithm was developed by Breiman in 2001 [56,57]. It is a technique for ensemble learning that consists of multiple decision trees. Variables are selected randomly using the bootstrap method. The results of multiple decision trees are aggregated by voting [58–60]. Therefore, the RF model has high generalization ability and is not prone to overfitting. The number of decision trees ($n_estimators$) affects model accuracy. The estimation error of the model is determined based on the out-of-bag (OOB) sample, and optimal $n_estimators$ are selected. We used the following parameter settings: $n_estimators = 800$; $oob_score = True$; $random_state = 0$. The other settings were set to their default values [61]. The RF model was implemented in the Python 3.8 platform with the scikit-learn package.

The 177 soil samples were randomly split into calibration ($n = 156$) and validation ($n = 21$) datasets with a ratio of 7:3, which were utilized for tenfold cross validation and independent validation of the model, respectively. The quantitative evaluation indices of the model's prediction performance included the coefficient of determination (R^2), root mean square error (RMSE), and the ratio of performance to inter-quartile range (RPIQ) (Equations (3)–(5)). The R^2 value represents the proportion of the variance in the dependent variable that is explained by the independent variable. The RMSE represents the difference between the predicted and measured values. Ref. [62] have shown that when $RPIQ \geq 4.05$, the model has more accurate prediction ability. When $3.37 \leq RPIQ < 4.05$, the estimation result is good. When $2.70 \leq RPIQ < 3.37$, the model can provide approximate estimation. When $2.02 \leq RPIQ < 2.70$, it shows that the prediction result has a large gap with the real value and needs further improvement [63]. Larger R^2 and RPIQ values and smaller RMSE values indicate higher estimation precision and resilience of the model [64].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$RPIQ = \frac{IQ}{RMSE} \quad (5)$$

In the above equation, n ($i = 1, 2, 3, \dots, n$) is the number of soil samples, y_i and $\hat{y}_i$ are the measured and predicted SOC contents value of sample i , respectively, $\bar{y}_i$ is the average SOC content, IQ is the quartile distance, i.e., the difference between the third quartile (75% of the number of soil samples in the validation set ranked from small to large) and the first quartile (25% of the number of soil samples in the validation set ranked from small to large).

2.5. Flowchart

The flowchart for SOC mapping in forest areas utilizing S1, S2, and DEM data and the RF algorithm is shown in Figure 2. The procedure begins with soil sample collection and acquisition of the S1, S2, and DEM data, followed by the construction of the comprehensive feature set (S1-based textural indices, S2-based spectral indices, and DEM-derived indices).

The SCC-VIF method is then employed to determine the optimal parameters that characterize the SOC content in forest areas. These are the independent variables in the model, and the measured SOC content is the dependent variable. An RF-based SOC mapping model is established, and the prediction accuracy is evaluated.

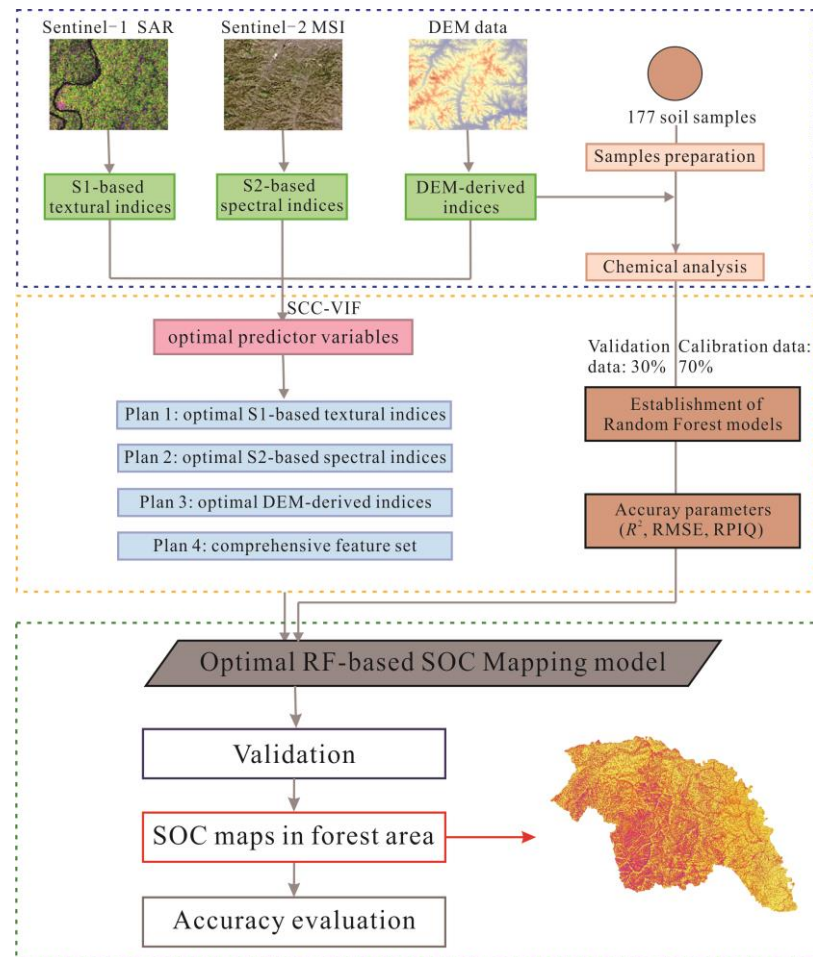


Figure 2. Flowchart for mapping the soil organic carbon (SOC) content of forested area using Sentinel-1/2 and DEM data and the random forest algorithm.

3. Results

3.1. Descriptive Statistics of the SOC Content

Figure 3 displays the descriptive statistics of the SOC content for the total, calibration, and validation sample sets. The SOC content of the total sample ranged from 11.59 to 161.01 g kg⁻¹, with an average of 43.75 g kg⁻¹. The coefficient of variation (CV) is typically used to assess the level of dispersion in a dataset. The CVs of the total, calibration, and validation sample sets exceeded 0.36, indicating a high degree of variation [65]. A soil sample set with high CV values indicates high spatial heterogeneity. The kurtosis and skewness describe the shape of the frequency distribution of the datasets. Their values are 0 and 3, respectively, when a dataset is normally distributed. The skewness values of total, calibration, and validation sample sets exceeded 0, indicating positive skewness, suggesting that the SOC content in forested areas was unevenly distributed and influenced by external disturbances. Therefore, it is necessary to perform SOC modeling in the Greater Khingan Mountains.

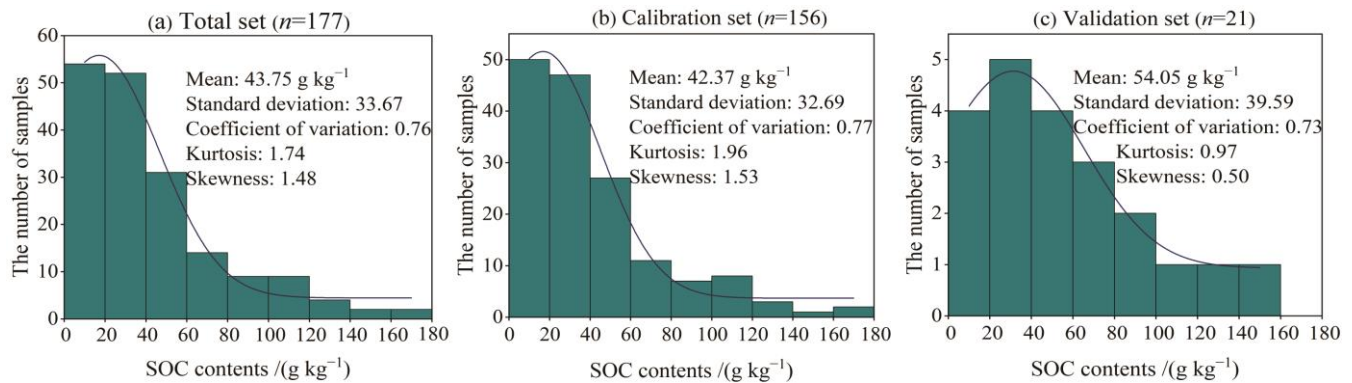


Figure 3. Descriptive statistics of the SOC contents in the (a) total; (b) calibration; and (c) validation sample sets.

3.2. Feature Selection for Predicting the SOC Content in Forested Areas

3.2.1. Optimal S2-Based Spectral Indices

The SCC-VIF results of the S2-based spectral indices and SOC are presented in Figure 4. Based on prior knowledge and multiple calculations, an SCC value of 0.8 was used as a threshold. S2-based spectral indices with SCC values greater than 0.8 were selected for the VIF test, and the remaining spectral indices were not included in the subsequent calculation. The selected S2-based spectral indices were B6, B7, B8a, B11, MNDWI, NDTI, CMR, NDVI, and IPVI. The VIF values of these nine selected S2-based spectral indices are shown in Figure 4b. Based on the multicollinearity test, S2-based spectral indices with VIFs greater than 5 were removed. Therefore, the selected S2-based spectral indices were B6, B8a, and CMR. Overall, the chosen S2 spectral indices consisted of B6, B8a, CMR, DVI, EVI, RVI, RGRI, REPI, and FWI.

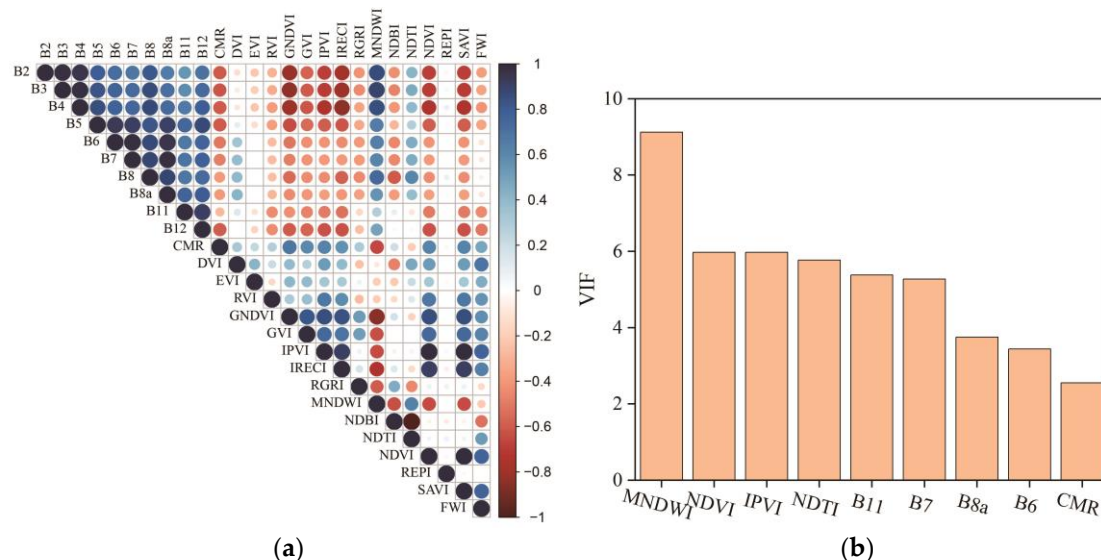


Figure 4. The variance inflation factor (VIF) value of the S2-based spectral indices screened by Spearman correlation coefficient (SCC). (a) Heat map of the SCC; and (b) VIF values of the S2-based spectral indices selected by the SCC.

3.2.2. Optimal S1-Based Textural Indices

The SCC-VIF results of S1-based textural indices and SOC are displayed in Figure 5. Based on prior knowledge and multiple operations, an SCC threshold of 0.8 was utilized. S1-based textural indices with SCC values exceeding 0.8 were selected, and the remaining S1-based textural indices were excluded from further calculations. The selected S1-based textural indices consisted of VH_E, VH_C, VH_H, AVE, VV_H, VV_SM, VV_E, VV_D, and

VV_V. Figure 5b illustrates the VIF values of the nine selected S1-based texture indices. Based on the multicollinearity test, the S1-based textural indices with VIF values exceeding 5 were removed, and VH_E and AVE were selected. Overall, the chosen S1-based textural indices comprised VH_E, AVE, DI, RI, NDI, VH_Cor, and VV_Cor.

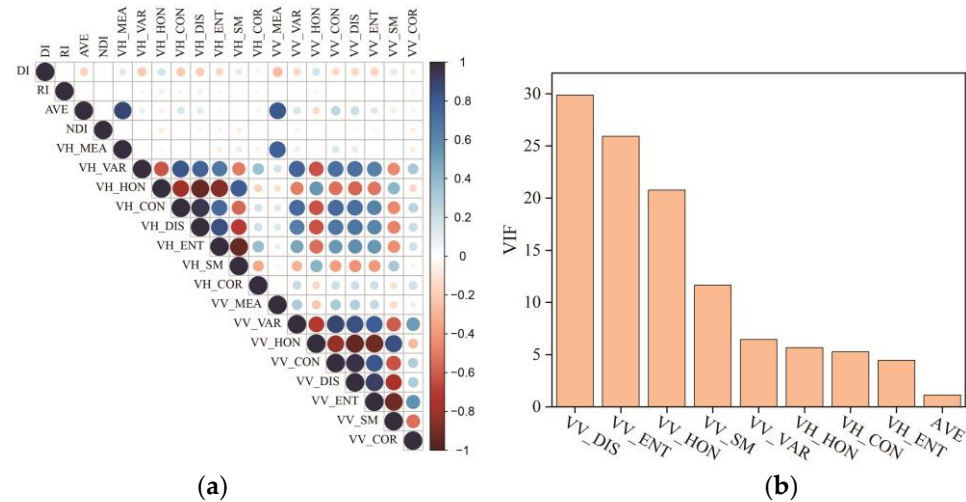


Figure 5. The variance inflation factor (VIF) value of the S1-based textural indices screened by Spearman correlation coefficient (SCC). (a) Heat map of the SCC; and (b) VIF values of the S2-based textural indices selected by the SCC.

3.2.3. Optimal DEM-Derived Indices

The SCC-VIF results of the DEM-derived indices and SOC are presented in Figure 6. Based on prior knowledge and multiple calculations, an SCC threshold of 0.6 was utilized. The remaining DEM-derived indices exhibited low collinearity and did not require VIF testing. The selected DEM-derived indices were RSP, TCA, PFC, CNBL, and LSF. Figure 6b displays the VIF values of the five selected DEM-derived indices. Based on the multicollinearity test, DEM-derived indices with VIF values exceeding 5 were removed, and RSP, TCA, PFC, CNBL, and LSF were selected. Overall, the chosen DEM-derived indices comprised RSP, TCA, PFC, CNBL, LSF, AH, A, CD, CI, PC, and TWI.

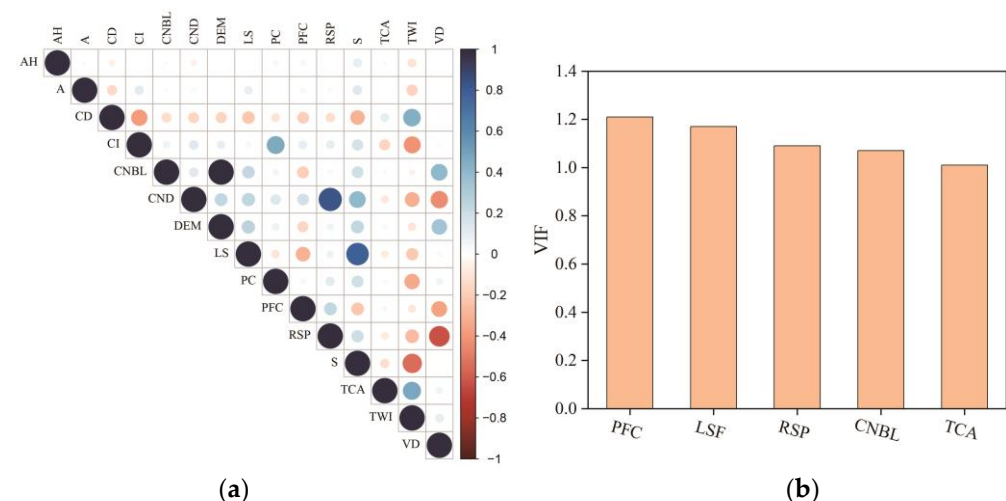


Figure 6. The variance inflation factor (VIF) value of the DEM-derived indices screened by Spearman correlation coefficient (SCC). (a) Heat map of the SCC; and (b) VIF values of the DEM-derived indices selected by the SCC.

3.3. Establishment of SOC Model with the Selected Indices as Predictor Covariates

Selecting SOC-sensitive indices as predictor covariates helps establish a precise SOC mapping model for forested areas. We used 9 S2-based spectral indices (B6, B8a, CMR, DVI, EVI, RVI, RGRI, REPI, and FWI), 7 S1-based textural indices (VH_E, AVE, DI, RI, NDI, VH_Cor, and VV_Cor), and 13 DEM-derived indices (RSP, TCA, PFC, CNBL, LSF, AH, A, CD, CI, PC, and TWI) as independent variables. The measured SOC content was the dependent variable, and the RF model was used as the SOC mapping model.

The SOC estimation model contained predictor covariates and their combinations, including S1-based textural indices, S2-based spectral indices, and DEM-derived indices and their combinations. Figure 7 displays the estimation results on the validation set. Among the models with single indices as inputs, the model based on the DEM-derived indices had the best performance, with an RPIQ value of 3.56 (Figure 7c). The SOC estimation model based on the S1-based textural indices and S2-based spectral indices had RPIQ values of 2.71 and 2.92, respectively, (Figure 7a,b), indicating that models with single indices provided good estimations ($RPIQ > 2.70$). Figure 7d shows the results of the model with the comprehensive feature set. Most of the data in the scatter plots were closely aligned with the 1:1 line, suggesting this model achieved the best estimation accuracy ($R^2 = 0.91$, $RMSE = 9.78$, $RPIQ = 4.13$). The R^2 and RPIQ were 9.76% (0.82 versus 0.91) and 16.01% (3.56 versus 4.13) higher, respectively, and the RMSE was 13.80% (10.65 versus 9.18) lower than those of the model with only DEM-derived indices. The model with the comprehensive feature set had higher accuracy for predicting the SOC content than the model with single indices. The optimal performance was obtained from the model with a combination of S1-based textural indices + S2-based spectral indices + DEM-derived indices as the inputs.

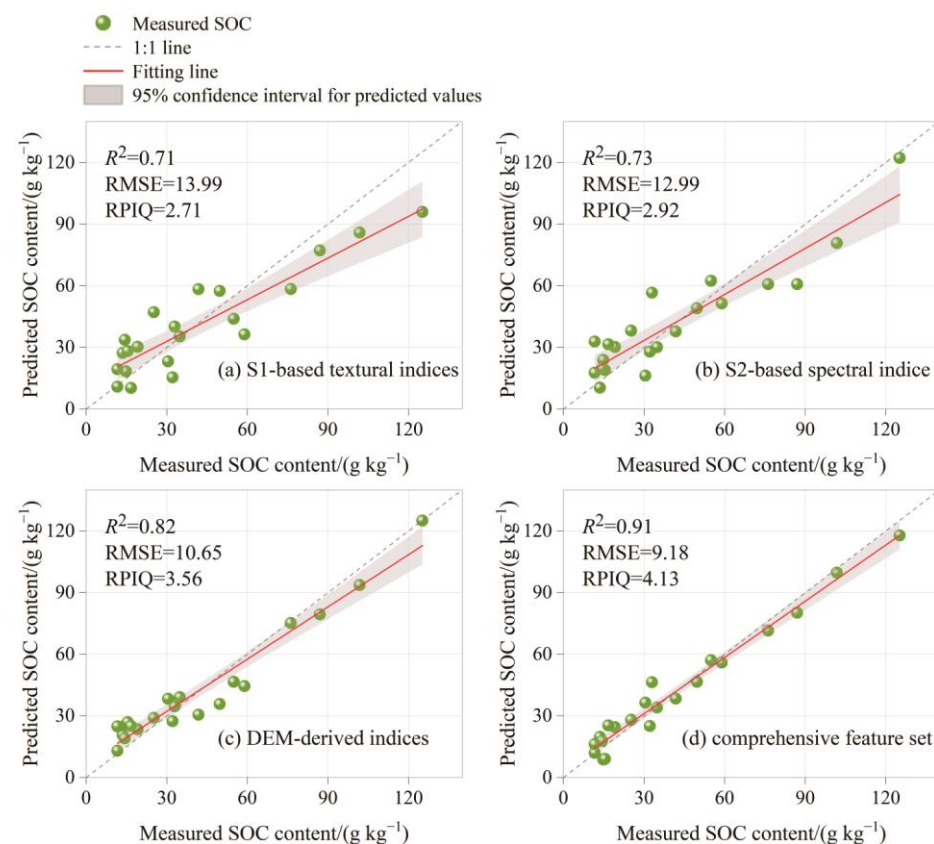


Figure 7. Scatter plots of measured versus predicted SOC contents of the models constructed with the (a) S1-based textural indices; (b) S2-based spectral indices; (c) DEM-derived indices; and (d) comprehensive feature set.

3.4. SOC Distribution Map

We applied the optimal model to the full research region to assess model's generalization ability. The SOC content in the Greater Khingan Mountains is shown in Figure 8.

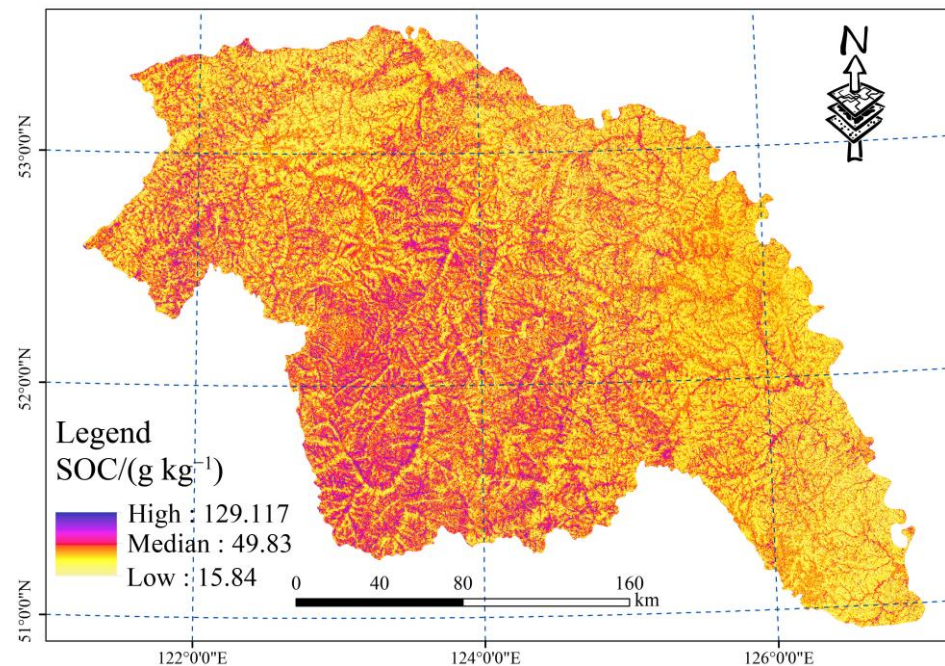


Figure 8. Spatial distribution of soil organic carbon (SOC) content in the Greater Khingan Mountains.

The SOC distribution has a strip pattern, with higher values in the north and southwest. The northern and southwestern regions have Phaeozems with a high humus content due to dense vegetation cover and favorable low temperatures for humus fixation and accumulation. Conversely, the southeast and northwest regions have lower SOC contents ($<39.83 \text{ g kg}^{-1}$). The southeast is dominated by cultivated land and grassland significantly affected by human activities, such as reclamation, resulting in relatively low SOC contents. The SOC content varies significantly, with the lowest value at 15.84 g kg^{-1} , the highest at 129.12 g kg^{-1} , and an average of 49.83 g kg^{-1} . This significant variation confirms the results of the descriptive statistics of the SOC content. The majority of SOC contents in the Greater Khingan Mountains are in the range of $34.94\text{--}54.93 \text{ g kg}^{-1}$, accounting for 60.05% of the total. SOC values below 15.84 g kg^{-1} and above 69.15 g kg^{-1} represent 20.71% and 3.76%, respectively. However, it is important to note that the model's range for the predicted SOC content ($15.84 \text{ g kg}^{-1}\text{--}129.12 \text{ g kg}^{-1}$) is narrower than the field-measured range ($11\text{--}161 \text{ g kg}^{-1}$), indicating some limitations in estimating high SOC contents due to the complex geographical environment of the Greater Khingan Mountains.

4. Discussion

In the SOC mapping models, the S2-based spectral indices were the most important, followed by the DEM-derived and S1-based textural indices. The S2-based spectral indices, S1-based textural indices, and DEM-derived indices accounted for 60.44%, 10.92%, and 28.65% of the total, respectively (Figure 9).

Vegetation has rich spectral information and provides indirect features (i.e., S2-based spectral indices) for estimating the SOC content in forested areas [66]. The MSI sensor onboard Sentinel-2 is highly sensitive to vegetation properties, particularly the newly added red-edge band [67]. Among the selected spectral parameters, B8a and B6 were the most important single bands; they are often considered indicators of plant growth. Additionally, REPI (VI = 8.78%), RGRI (VI = 6.14), CMR (VI = 5.04), and FWI (VI = 4.36) were the most important indices in mapping model construction. REPI has a high level of sensitivity towards the red edge area of plant leaves and can indicate chlorophyll content [68]. The

RGRI has been primarily used to assess the development of tree leaves and can also serve as a parameter for estimating the SOC content in forested areas [48]. The CMR is related to soil minerals and clay content and has been selected as a predictive covariate in SOC spectral estimation models [69–71]. The FWI can distinguish healthy and unhealthy forest areas and has been used for the indirect evaluation of the SOC content in forested areas [21,72,73].

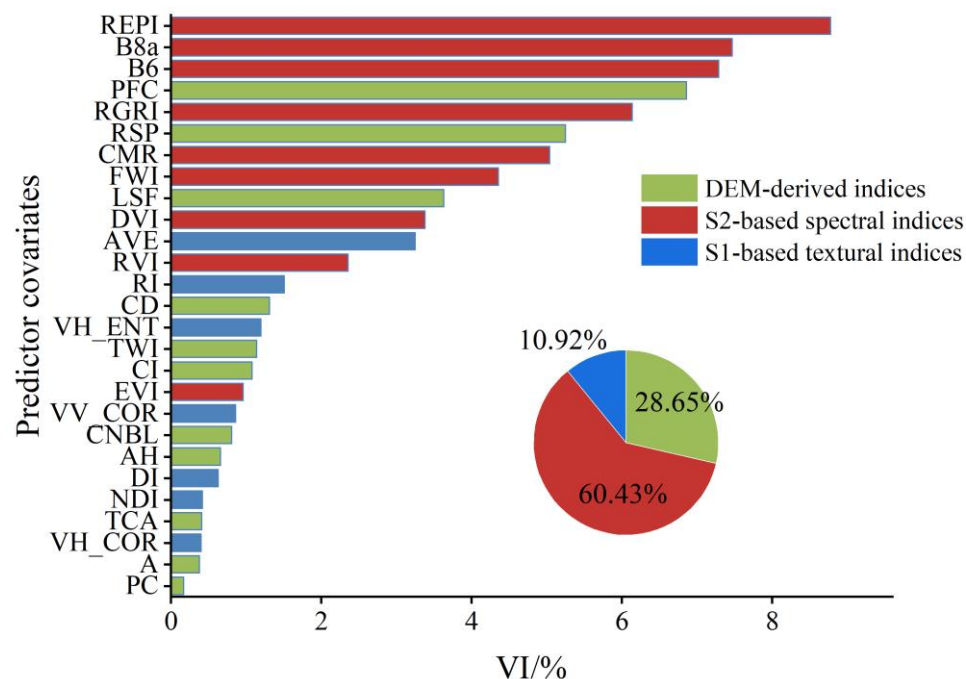


Figure 9. Variable importance (VI) of the predictor covariates in the SOC mapping model of forested areas.

DEM and indices derived from them are critical parameters for characterizing topographic relief and have a vital role in the material flow during the development of SOC [23,74,75]. Therefore, these indices can reflect the distribution status of SOC. In this study, DEM-derived indices were critical in the SOC mapping model. The PFC had the highest VI value (6.86%), followed by RSP (5.25%) and LSF (3.63%). The PFC is used to describe local variations in the terrain curvature, which can affect the velocity and direction of water flow and material. This index can be used to identify areas with abrupt changes in slope or aspect, explaining the spatial distribution of the SOC concentration [76]. The RSP is used to identify the slope shape and is a crucial parameter for characterizing the spatial distribution of the SOC content [20]. The LSF describes the influence of the slope on soil erosion and is also a crucial parameter in determining the SOC content [77]. The Greater Khingan Mountains have hilly terrain and uneven ground, affecting the allocation of matter and energy inside the soil. Therefore, DEM-derived indices are key parameters in SOC mapping models.

S1-based textural indices reflect the structure and grayscale conditions of the land surface; thus, they are an important indicator of the SOC content in forested areas. The AVE, RI, and VH_E had the highest VI values (3.25%, 1.51%, and 2.25%, respectively). The inclusion of S1-based textural indices as predictor covariates significantly improves the accuracy of SOC mapping models. Similarly, [78–81] and [30] also found that using S1-based textural indices as predictor covariates in SOC mapping models improved the mapping accuracy, particularly in areas with high cloud cover or dense vegetation.

The RF algorithm has significant advantages for describing the nonlinear relationship between the SOC content and predictive features. SOC in forested areas has a complex formation mechanism due to long-term interactions between forest litter and root exudates. Therefore, the SOC content is typically correlated with vegetation and topography properties and texture features. In addition, partial least squares, artificial neural networks,

and support vector machine algorithms have been employed as SOC mapping models for forested regions, but their performance is lower than that of the RF model. The RF model outperforms these models regarding sample requirements, multicollinearity, and efficiency.

The difference in the image acquisition time (October) and the soil sample collection time (July) likely had a negligible effect because the SOC content changes slowly over time [82]. At the second level, due to the effect of clouds, the signal-to-noise ratio of the images acquired in July was very low due to clouds. Therefore, we selected images from the nearest period (October). Although the vegetation and texture may have changed slightly, the conditions still reflect the SOC content. The model's R^2 was 0.91, indicating relatively high accuracy. Future studies will aim for a closer match between the sampling time and the image acquisition time.

In summary, the combination of S2-based spectral indices, S1-based textural indices, and DEM-derived indices (i.e., a comprehensive feature set) provided a better understanding of the content and spatial distribution of SOC in forest regions than the use of single index types. Additionally, the RF algorithm provided high accuracy for SOC content mapping in forested areas.

5. Conclusions

This paper proposed a robust approach for mapping SOC in forested areas using Sentinel-1/2 and DEM data. S1-based textural indices, S2-based spectral indices, and DEM-derived indices were used as predictive variables, and the measured SOC content was the dependent variable in the SOC mapping model. The RPIQ of the models with the different index types were 2.71, 2.92, and 3.56, respectively. The model with the combined indices had the highest performance (RPIQ = 4.13). Its RPIQ values were 52.40%, 41.44%, and 16.01% higher than that of the models with the single index types. The SOC mapping results revealed a relatively high SOC content in the Greater Khingan Mountains, with an average value of 43.75 g kg^{-1} . The northern and southwestern parts had higher SOC contents ($>54.93 \text{ g kg}^{-1}$), and the southeastern and northwestern parts had lower contents ($<39.83 \text{ g kg}^{-1}$), which was primarily attributed to agricultural activities. The research findings indicate that the combination of S1-based textural indices, S2-based spectral indices, and DEM-derived indices and the RF algorithm provides a reliable technique for predicting the distribution of the SOC content in forested areas.

Author Contributions: Writing and review were conducted by Z.W. and D.Z.; the methodology section was written by Z.W. and X.X.; the software implementation was conducted by G.Y. and T.L.; funding was obtained by D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Nation Nature Science Foundation of China (Project No. 41671064), the Natural Science Foundation of Heilongjiang Province (Project No. LH2021D012), and the Key Laboratory of Environmental Change and Natural Disaster of the Ministry of Education, Beijing Normal University (Project No. 2023-KF-12).

Data Availability Statement: Data for this study can be obtained by contacting the authors.

Acknowledgments: We thank the reviewers for their helpful revisions.

Conflicts of Interest: The authors have declared no conflicts of interest.

References

1. Li, X.; Ding, J.; Liu, J.; Ge, X.; Zhang, J. Digital Mapping of Soil Organic Carbon Using Sentinel Series Data: A Case Study of the Ebinur Lake Watershed in Xinjiang. *Remote Sens.* **2021**, *13*, 769. [CrossRef]
2. Sommer, R.; Bossio, D. Dynamics and climate change mitigation potential of soil organic carbon sequestration. *J. Environ. Manag.* **2014**, *144*, 83–87. [CrossRef]
3. Li, H.; Wu, Y.; Liu, S.; Xiao, J.; Zhao, W.; Chen, J.; Alexandrov, G.; Cao, Y. Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Glob. Chang. Biol.* **2022**, *28*, 3394–3410. [CrossRef]
4. Stockmann, U.; Padarian, J.; McBratney, A.; Minasny, B.; de Brogniez, D.; Montanarella, L.; Hong, S.Y.; Rawlins, B.G.; Field, D.J. Global soil organic carbon assessment. *Glob. Food Secur.* **2015**, *6*, 9–16. [CrossRef]

5. Fantappiè, M.; L'Abate, G.; Costantini, E. The influence of climate change on the soil organic carbon content in Italy from 1961 to 2008. *Geomorphology* **2011**, *135*, 343–352. [\[CrossRef\]](#)
6. Magnussen, S.; Köhl, M.; Olschofsky, K. Error propagation in stock-difference and gain-loss estimates of a forest biomass carbon balance. *Eur. J. For. Res.* **2014**, *133*, 1137–1155. [\[CrossRef\]](#)
7. Lorenz, K.; Lal, R. Biochar application to soil for climate change mitigation by soil organic carbon sequestration. *J. Plant Nutr. Soil Sci.* **2014**, *177*, 651–670. [\[CrossRef\]](#)
8. Mallik, S.; Bhowmik, T.; Mishra, U.; Paul, N. Mapping and prediction of soil organic carbon by an advanced geostatistical technique using remote sensing and terrain data. *Geocarto Int.* **2022**, *37*, 2198–2214. [\[CrossRef\]](#)
9. Marchetti, A.; Piccini, C.; Francaviglia, R.; Mabit, L. Spatial distribution of soil organic matter using geostatistics: A key indicator to assess soil degradation status in central Italy. *Pedosphere* **2012**, *22*, 230–242. [\[CrossRef\]](#)
10. McBratney, A.B.; Santos, M.M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [\[CrossRef\]](#)
11. Cambule, A.; Rossiter, D.; Stoorvogel, J.; Smaling, E. Soil organic carbon stocks in the Limpopo National Park, Mozambique: Amount, spatial distribution and uncertainty. *Geoderma* **2014**, *213*, 46–56. [\[CrossRef\]](#)
12. Castaldi, F.; Palombo, A.; Santini, F.; Pascucci, S.; Pignatti, S.; Casa, R. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sens. Environ.* **2016**, *179*, 54–65. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. Wang, C.; Feng, M.-C.; Yang, W.-D.; Ding, G.-W.; Wang, H.-Q.; Li, Z.-H.; Sun, H.; Shi, C.-C. Use of spectral character to evaluate soil organic matter. *Soil Sci. Soc. Am. J.* **2016**, *80*, 1078–1088. [\[CrossRef\]](#)
15. Zhang, Y.; Kou, C.; Liu, M.; Man, W.; Li, F.; Lu, C.; Song, J.; Song, T.; Zhang, Q.; Li, X. Estimation of Coastal Wetland Soil Organic Carbon Content in Western Bohai Bay Using Remote Sensing, Climate, and Topographic Data. *Remote Sens.* **2023**, *15*, 4241. [\[CrossRef\]](#)
16. Zhang, J.; Lin, X. Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. *Int. J. Image Data Fusion* **2017**, *8*, 1–31. [\[CrossRef\]](#)
17. Prudente, V.H.R.; Martins, V.S.; Vieira, D.C.; Rodrigues de França e Silva, N.; Adami, M.; Sanches, I.D.A. Limitations of cloud cover for optical remote sensing of agricultural areas across South America. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100414. [\[CrossRef\]](#)
18. Yang, R.-M.; Guo, W.-W. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. *Environ. Monit. Assess.* **2019**, *191*, 462. [\[CrossRef\]](#)
19. Orynbaiqyzy, A.; Plank, S.; Vetrita, Y.; Martinis, S.; Santoso, I.; Ismanto, R.D.; Chusnayah, F.; Tjahjaningsih, A.; Genzano, N.; Marchese, F. Joint use of Sentinel-2 and Sentinel-1 data for rapid mapping of volcanic eruption deposits in Southeast Asia. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103166. [\[CrossRef\]](#)
20. Guo, Z.; Adhikari, K.; Chellasamy, M.; Greve, M.B.; Owens, P.R.; Greve, M.H. Selection of terrain attributes and its scale dependency on soil organic carbon prediction. *Geoderma* **2019**, *340*, 303–312. [\[CrossRef\]](#)
21. Nabiollahi, K.; Eskandari, S.; Taghizadeh-Mehrjardi, R.; Kerry, R.; Triantafyllis, J. Assessing soil organic carbon stocks under land-use change scenarios using random forest models. *Carbon Manag.* **2019**, *10*, 63–77. [\[CrossRef\]](#)
22. Xu, X.; Chen, Y.; Dai, X.; Lei, T.; Wang, S.; Li, K. An improved Vis-NIR estimation model of soil organic matter through the artificial samples enhanced calibration set. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4626–4637. [\[CrossRef\]](#)
23. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* **2019**, *338*, 445–452. [\[CrossRef\]](#)
24. Matinfar, H.R.; Maghsodi, Z.; Mousavi, S.R.; Rahmani, A. Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale. *Catena* **2021**, *202*, 105258. [\[CrossRef\]](#)
25. Azizi, K.; Ayoubi, S.; Nabiollahi, K.; Garosi, Y.; Gislum, R. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. *J. Geochem. Explor.* **2022**, *233*, 106921. [\[CrossRef\]](#)
26. Nachtergaele, F.O.; Spaargaren, O.; Deckers, J.A.; Ahrens, B. New developments in soil classification: World reference base for soil resources. *Geoderma* **2000**, *96*, 345–357. [\[CrossRef\]](#)
27. Pradhan, B.; Awasthi, K.; Bajracharya, R. Soil organic carbon stocks under different forest types in Pokhara Khola sub-watershed: A case study from Dhading district of Nepal. *WIT Trans. Ecol. Environ.* **2012**, *157*, 535–546.
28. Dong, X.; Liu, C.; Wu, X.; Man, H.; Wu, X.; Ma, D.; Li, M.; Zang, S. Linking soil organic carbon mineralization with soil variables and bacterial communities in a permafrost-affected tussock wetland during laboratory incubation. *Catena* **2023**, *221*, 106783. [\[CrossRef\]](#)
29. Pham, T.D.; Yokoya, N.; Nguyen, T.T.T.; Le, N.N.; Ha, N.T.; Xia, J.; Takeuchi, W.; Pham, T.D. Improvement of mangrove soil carbon stocks estimation in North Vietnam using Sentinel-2 data and machine learning approach. *GISci. Remote Sens.* **2021**, *58*, 68–87. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#) [\[PubMed\]](#)

31. Toming, K.; Kutser, T.; Laas, A.; Sepp, M.; Paavel, B.; Nõges, T. First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sens.* **2016**, *8*, 640. [\[CrossRef\]](#)
32. Wang, J.; Ding, J.; Yu, D.; Ma, X.; Zhang, Z.; Ge, X.; Teng, D.; Li, X.; Liang, J.; Lizaga, I. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma* **2019**, *353*, 172–187. [\[CrossRef\]](#)
33. Sibanda, M.; Mutanga, O.; Dube, T.; Vundla, T.S.; Mafongoya, P.L. Estimating LAI and mapping canopy storage capacity for hydrological applications in wattle infested ecosystems using Sentinel-2 MSI derived red edge bands. *GISci. Remote Sens.* **2019**, *56*, 68–86. [\[CrossRef\]](#)
34. Sionneau, T.; Bout-Roumazeilles, V.; Biscaye, P.; van Vliet-Lanoë, B.; Bory, A. Clay mineral distributions in and around the Mississippi River watershed and Northern Gulf of Mexico: Sources and transport patterns. *Quat. Sci. Rev.* **2008**, *27*, 1740–1751. [\[CrossRef\]](#)
35. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [\[CrossRef\]](#)
36. Rouse, J.W., Jr.; Haas, R.H.; Deering, D.; Schell, J.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*; NASA/GSFC Type III Final Rep.: Greenbelt, MD, USA, 1974; p. 371.
37. Seilheimer, T.; Chow-Fraser, P. Development and use of the Wetland Fish Index to assess the quality of coastal wetlands in the Laurentian Great Lakes. *Can. J. Fish. Aquat. Sci.* **2006**, *63*, 354–366. [\[CrossRef\]](#)
38. Fouad, A. Soil Salinity Detection Using Satellite Remote Sensing. Master's Thesis, ITC Faculty Geo-Information Science and Earth Observation, Enschede, The Netherlands, 2003.
39. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [\[CrossRef\]](#)
40. Curran, P.J.; Dungan, J.L.; Gholz, H.L. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiol.* **1990**, *7*, 33–48. [\[CrossRef\]](#)
41. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [\[CrossRef\]](#)
42. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [\[CrossRef\]](#)
43. Frampton, W.J.; Dash, J.; Watmough, G.; Milton, E.J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 83–92. [\[CrossRef\]](#)
44. Crippen, R.E. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [\[CrossRef\]](#)
45. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [\[CrossRef\]](#)
46. Jiang, Z.; Huete, A.R.; Chen, J.; Chen, Y.; Li, J.; Yan, G.; Zhang, X. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sens. Environ.* **2006**, *101*, 366–378. [\[CrossRef\]](#)
47. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [\[CrossRef\]](#)
48. Verrelst, J.; Schaepman, M.E.; Koetz, B.; Kneubühler, M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sens. Environ.* **2008**, *112*, 2341–2353. [\[CrossRef\]](#)
49. Jiang, L.; Kogan, F.N.; Guo, W.; Tarpley, J.D.; Mitchell, K.E.; Ek, M.B.; Tian, Y.; Zheng, W.; Zou, C.Z.; Ramsay, B.H. Real-time weekly global green vegetation fraction derived from advanced very high resolution radiometer-based NOAA operational global vegetation index (GVI) system. *J. Geophys. Res. Atmos.* **2010**, *115*, D11. [\[CrossRef\]](#)
50. Huang, C.; Zhang, C.; Li, H. Assessment of the Impact of Rubber Plantation Expansion on Regional Carbon Storage Based on Time Series Remote Sensing and the InVEST Model. *Remote Sens.* **2022**, *14*, 6234. [\[CrossRef\]](#)
51. Chen, Q.; Yang, H.; Li, L.; Liu, X. A novel statistical texture feature for SAR building damage assessment in different polarization modes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *13*, 154–165. [\[CrossRef\]](#)
52. Caballero, G.R.; Platzeck, G.; Pezzola, A.; Casella, A.; Winschel, C.; Silva, S.S.; Ludueña, E.; Pasqualotto, N.; Delegido, J. Assessment of multi-date sentinel-1 polarizations and GLCM texture features capacity for onion and sunflower classification in an irrigated valley: An object level approach. *Agronomy* **2020**, *10*, 845. [\[CrossRef\]](#)
53. Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Sci. Total Environ.* **2020**, *729*, 138244. [\[CrossRef\]](#)
54. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [\[CrossRef\]](#)
55. O'Brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690. [\[CrossRef\]](#)
56. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [\[CrossRef\]](#)
57. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
58. John, K.; Abraham Isong, I.; Michael Kebonye, N.; Okon Ayito, E.; Chapman Agyeman, P.; Marcus Afu, S. Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. *Land* **2020**, *9*, 487. [\[CrossRef\]](#)

59. Chen, W.; Ran, H.; Cao, X.; Wang, J.; Teng, D.; Chen, J.; Zheng, X. Estimating PM_{2.5} with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China. *Sci. Total Environ.* **2020**, *746*, 141093. [\[CrossRef\]](#)
60. Ji, W.; Adamchuk, V.I.; Chen, S.; Su, A.S.M.; Ismail, A.; Gan, Q.; Shi, Z.; Biswas, A. Simultaneous measurement of multiple soil properties through proximal sensor data fusion: A case study. *Geoderma* **2019**, *341*, 111–128. [\[CrossRef\]](#)
61. Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*, 3. [\[CrossRef\]](#)
62. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.-M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **2010**, *29*, 1073–1081. [\[CrossRef\]](#)
63. Xu, X.; Wang, Z.; Song, X.; Zhan, W.; Yang, S. A remote sensing-based strategy for mapping potentially toxic elements of soils: Temporal-spatial-spectral covariates combined with random forest. *Environ. Res.* **2024**, *240*, 117570. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Williams, P.; Norris, K. *Near-Infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists, Inc.: St. Paul, MN, USA, 1987.
65. Wilding, L. Spatial variability: Its documentation, accommodation and implication to soil surveys. In Proceedings of the Soil Spatial Variability, Las Vegas, NV, USA, 30 November–1 December 1984; pp. 166–194.
66. Zeng, Y.; Hao, D.; Huete, A.; Dechant, B.; Berry, J.; Chen, J.M.; Joiner, J.; Frankenberg, C.; Bond-Lamberty, B.; Ryu, Y. Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nat. Rev. Earth Environ.* **2022**, *3*, 477–493. [\[CrossRef\]](#)
67. Dvorakova, K.; Heiden, U.; Pepers, K.; Staats, G.; van Os, G.; van Wesemael, B. Improving soil organic carbon predictions from a Sentinel-2 soil composite by assessing surface conditions and uncertainties. *Geoderma* **2023**, *429*, 116128. [\[CrossRef\]](#)
68. Gholizadeh, A.; Mišurec, J.; Kopačková, V.; Mielke, C.; Rogass, C. Assessment of red-edge position extraction techniques: A case study for norway spruce forests using hymap and simulated sentinel-2 data. *Forests* **2016**, *7*, 226. [\[CrossRef\]](#)
69. Schweizer, S.A.; Mueller, C.W.; Höschen, C.; Ivanov, P.; Kögel-Knabner, I. The role of clay content and mineral surface area for soil organic carbon storage in an arable topequence. *Biogeochemistry* **2021**, *156*, 401–420. [\[CrossRef\]](#)
70. Xue, B.; Huang, L.; Li, X.; Lu, J.; Gao, R.; Kamran, M.; Fahad, S. Effect of clay mineralogy and soil organic carbon in aggregates under straw incorporation. *Agronomy* **2022**, *12*, 534. [\[CrossRef\]](#)
71. Das, A.; Purakayastha, T.J.; Ahmed, N.; Das, R.; Biswas, S.; Shivay, Y.S.; Sehgal, V.K.; Rani, K.; Trivedi, A.; Tigga, P. Influence of Clay Mineralogy on Soil Organic Carbon Stabilization under Tropical Climate, India. *J. Soil Sci. Plant Nutr.* **2023**, *23*, 1003–1018. [\[CrossRef\]](#)
72. Song, J.; Gao, J.; Zhang, Y.; Li, F.; Man, W.; Liu, M.; Wang, J.; Li, M.; Zheng, H.; Yang, X. Estimation of soil organic carbon content in coastal wetlands with measured VIS-NIR spectroscopy using optimized support vector machines and random forests. *Remote Sens.* **2022**, *14*, 4372. [\[CrossRef\]](#)
73. Uhan, B.; Windham-Myers, L.; Bliss, N.; Nahlik, A.M.; Sundquist, E.T.; Stagg, C.L. Improved wetland soil organic carbon stocks of the conterminous US through data harmonization. *Front. Soil Sci.* **2021**, *1*, 16. [\[CrossRef\]](#)
74. Tziachris, P.; Aschonitis, V.; Chatzistathis, T.; Papadopoulou, M. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *Catena* **2019**, *174*, 206–216. [\[CrossRef\]](#)
75. 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. [\[CrossRef\]](#)
76. Kennelly, P.J. Terrain maps displaying hill-shading with curvature. *Geomorphology* **2008**, *102*, 567–577. [\[CrossRef\]](#)
77. Jakšić, S.; Ninkov, J.; Milić, S.; Vasin, J.; Živanov, M.; Jakšić, D.; Komlen, V. Influence of slope gradient and aspect on soil organic carbon content in the region of Niš, Serbia. *Sustainability* **2021**, *13*, 8332. [\[CrossRef\]](#)
78. Duarte, E.; Zagal, E.; Barrera, J.A.; Dube, F.; Casco, F.; Hernández, A.J. Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic. *Eur. J. Remote Sens.* **2022**, *55*, 213–231. [\[CrossRef\]](#)
79. Azizi, K.; Garosi, Y.; Ayoubi, S.; Tajik, S. Integration of Sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. *Soil Tillage Res.* **2023**, *229*, 105681. [\[CrossRef\]](#)
80. Chi, Y.; Zhao, M.; Sun, J.; Xie, Z.; Wang, E. Mapping soil total nitrogen in an estuarine area with high landscape fragmentation using a multiple-scale approach. *Geoderma* **2019**, *339*, 70–84. [\[CrossRef\]](#)
81. Shafizadeh-Moghadam, H.; Minaei, F.; Talebi-khiyavi, H.; Xu, T.; Homaei, M. Synergetic use of multi-temporal Sentinel-1, Sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. *Catena* **2022**, *212*, 106077. [\[CrossRef\]](#)
82. Dou, X.; Wang, X.; Liu, H.; Zhang, X.; Meng, L.; Pan, Y.; Yu, Z.; Cui, Y. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China. *Geoderma* **2019**, *356*, 113896. [\[CrossRef\]](#)

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.