Soil & Tillage Research 256 (2026) 106886

Contents lists available at ScienceDirect

Soil & Tillage Research

journal homepage: www.elsevier.com/locate/still

Quantitative prediction of black soil horizon thickness at the watershed scale in Northeast China's black soil region

Weimin Ruan [a,b], Huanjun Liu [a,b,*], Yueyu Sui [a,b], Changkun Wang [b,c], Chong Luo [a,b], Xiangtian Meng [a,b], Chang Dong [a]

[a] State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

[c] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, No. 71 East Beijing Rd., Nanjing 210008, China

ARTICLE INFO

Keywords:
Black soil region
Sloping farmland
Black soil horizon thickness prediction
Optical data
Topographic factors
Vegetation index
Machine learning

ABSTRACT

The thickness of the black soil horizon in sloping farmland within China's black soil region is primarily affected by various elements, including terrain, climate, and anthropogenic activity. While conventional studies on soil thickness prediction primarily rely on topographic variables and vegetation indices, they often overlook the potential importance of spectral data. This study employs Sentinel-2 optical imagery data from May 2023, during the bare soil phase, in conjunction with topographic features and vegetation indices, to investigate the efficacy of various input variables in predicting soil thickness in sloping farms within the watersheds of the black soil region. The model was trained and validated using 157 sampling points of black soil horizon thickness (BSHT) through three machine learning techniques: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANNs), to assess the influence of various variable combinations on soil thickness prediction. The findings show that models incorporating spectral information (R^2 ranging from 0.62 to 0.71) have better explanatory power for predicting BSHT than models without (R^2 ranging from 0.68 to 0.75). While topographic factors were strong predictors, including spectral information significantly enhanced prediction accuracy. The findings indicated that RF exhibited superior prediction accuracy compared to XGBoost and ANNs among the three methodologies. This study's findings yield novel insights for accurately predicting soil thickness on sloping farmland within the black soil region and furnish scientific support for soil conservation and sustainable agricultural growth.

1. Introduction

Soil thickness, the vertical distance from the soil surface to the consolidated material, is a critical factor influencing vegetation development and land suitability (Chen et al., 2021). Soil erosion can lead to soil loss and reduce soil fertility, affecting land productivity. Soil thickness is, therefore, an important indicator of soil health. Measuring soil thickness is essential for assessing land productivity potential soil quality and promoting the sustainable use of soil resources. Research indicates that prolonged human activities, including tillage and hedgerow cultivation, can substantially modify soil thickness in agricultural fields (Fu et al., 2011). The Northeast Black Soil Region, a vital area for grain production in China, has undergone substantial alterations in the thickness of its black soil horizon due to extensive human activity. Since the initiation of extensive land reclamation in the 1950s, significant soil and water erosion has resulted in an average yearly loss of 0.3–1 cm in the black soil region. This decline has led to a significant decrease in the thickness of the black soil horizon, which has diminished from 60–70 cm in the 1950s to the present 20–30 cm (Xu et al., 2010). The thickness of the black soil horizon directly affects soil fertility and crop yield, so its reduction poses a serious threat to the region's agricultural productivity. In addition, the high spatial heterogeneity of the thickness of the black soil horizon and the high cost of obtaining high-precision distribution data pose significant challenges to accurately assessing and managing

* Corresponding author at: State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China.

E-mail addresses: ruanweimin@iga.ac.cn (W. Ruan), huanjunliu@yeah.net (H. Liu), suiyy@iga.ac.cn (Y. Sui), ckwang@issas.ac.cn (C. Wang), luochong@iga.ac.cn (C. Luo), mxt0123neau@yeah.net (X. Meng), dongchang@iga.ac.cn (C. Dong).

https://doi.org/10.1016/j.still.2025.106886
Received 17 March 2025; Received in revised form 21 August 2025; Accepted 19 September 2025
Available online 25 September 2025

this critical resource.

Since the previous century, comprehensive studies on black soil resources have been undertaken in Northeast China, encompassing essential data including black soil horizon thickness (BSHT), physicochemical characteristics, and agricultural circumstances (Pendleton et al., 1935; National Soil Census Office, 1994, pp. 276–286). However, these investigations mainly focused on the basic recording of the thickness of the black soil horizon and lacked in-depth research on its laws of variation and driving factors at different spatial scales. This lack of spatial variation analysis has led to our knowing very little about the laws of variation in BSHT under the influence of different topographic, climatic conditions, and human activities. Furthermore, invasive data collection methods require considerable effort and significant expenses, requiring extensive field sampling and laboratory processing (Frances and Lubczynski, 2011). Thus, acquiring spatial distribution data on the thickness of black soil horizons continues to be complicated.

Presently, strategies for predicting soil thickness predominantly focus on examining the processes affecting soil formation and the related landscape variables (Zhang et al., 2022). According to the concepts and data sources that inform the predictions, these methodologies can be classified into process-based models, point-sample interpolation techniques, and empirical statistical models (Kuriakose et al., 2009). Despite significant progress in their particular applications, these technologies encounter constraints when utilized across extensive areas and intricate environmental situations.

Process-based models may thoroughly elucidate the mechanisms underlying soil thickness formation. Dietrich et al. (1995) proposed a process-based model that forecasts the spatial distribution of soil thickness by resolving a mass balance equation between the bedrock soil formation rate and diffusive soil transport divergence. Heimsath et al. (1997) researched landscape morphology at the point of dynamic equilibrium between erosion and soil generation rates, indicating that soil productivity diminishes exponentially as soil thickness increases. This equilibrium idea offers a theoretical foundation for deducing soil thickness distribution in stable landscapes. D′Odorico (2000) assessed the probability distribution of soil thickness by examining the correlation between soil productivity and erosion rate. These studies mainly focus on explaining soil thickness distribution under stable landscapes, while in-depth exploration is still lacking regarding the effects of different types of geomorphic evolution models and soil transport processes over long timescales. Therefore, Roering (2008) further developed a nonlinear model that incorporates slope and soil thickness dependence, which better preserves ridge morphology in the study area and demonstrates different erosion rate distributions and soil thickness variations over long timescales. However, this research places greater emphasis on explaining the formation of landform morphology rather than directly providing a convenient method for soil thickness prediction. Liu et al. (2013) proposed a model based on the dynamic equation of soil thickness evolution. Under specific assumptions (such as no recent tectonic uplift), they derived a simplified analytical solution, thus avoiding complex numerical computations and extensive adjustment of experimental parameters. These two types of models still have certain limitations in practical applications, primarily in that they focus on equilibrium theory or single dynamic processes. For the first time, Bonfatti et al. (2018) systematically coupled the Soil Production Function with the Landscape Evolution Model, taking into account both steady-state and non-steady-state scenarios. This approach is capable of simultaneously simulating the spatial distribution and temporal variation of soil thickness, and quantitatively analyzing the mechanisms of soil thickness change under different terrain and hydrological conditions within a watershed.

The interpolation method uses soil thickness data from known measurement points to estimate the distribution of soil thickness at unsampled locations through spatial interpolation techniques. The inverse distance weighting method is simple but tends to overlook systematic influencing factors, sometimes resulting in low predictive efficiency. Topographic attributes have a significant impact on the spatial distribution of soil thickness, and these attributes can be used as auxiliary information to improve the prediction accuracy of soil thickness. Therefore, Penížek and Borůvka (2006) utilized topographic features and adopted kriging, regression kriging, and co-kriging methods to construct spatial prediction models for soil thickness. Their study used multiple topographic attributes, such as slope, aspect, and elevation, to predict soil thickness, and the experimental results indicated that regression kriging and co-kriging provided better prediction performance. Similarly, Sarkar et al. (2013) employed the regression kriging model, incorporating various topographic and environmental factors as covariates to improve the prediction of soil thickness. In the study of the spatial distribution of hard soil horizon thickness in landslide-prone areas, Yanto et al. (2022) compared several interpolation methods, including ordinary kriging, to generate regional maps of hard soil horizon thickness. Hard soil horizon thickness predicted by the ordinary kriging method showed significant spatial correlation with recorded landslide disaster points in the area. Beyond specific geological disaster applications, kriging interpolation methods have also been widely applied in broader soil resource studies. To gain a deeper understanding of the effects of human activities and natural factors on soil resource distribution, Yan et al. (2021b) fitted and compared different interpolation models (including spherical, exponential, Gaussian, and linear models) for soil thickness in the Yimeng Mountains region of China, selecting the model that best fit the regional spatial variation for kriging interpolation. With the optimized interpolation model, their study produced spatial distribution maps of soil thickness, revealing the spatial variation characteristics of soil thickness under different land use types (such as sloping farmland, forest land, and grassland).

The environmental covariate modeling method improves prediction accuracy by modeling the relationship between soil thickness and multidimensional environmental variables. This method uses terrain parameters, vegetation coverage, climate data, and other environmental factors as input variables and uses algorithms such as machine learning or regression analysis to establish a prediction model to estimate soil thickness. This method is based on the close correlation between environmental factors and soil properties during soil formation. Establishing a response relationship between environmental variables and soil thickness can predict the spatial distribution of soil thickness in larger areas. Li et al. (2020) employed machine learning techniques to model and delineate extensive soil thickness, utilizing 11 quantitative and four qualitative environmental variables. They examined the principal factors influencing soil thickness and accounted for 64 % of the variation in soil thickness. Xiao et al. (2023) tackled the challenge of precisely mapping soil thickness across extensive regions by employing an optimized geomorphological index soil thickness model (GIST) alongside a random forest regression model that integrates geomorphological parameters (GIST-RF) in the Wanzhou district of the Three Gorges Reservoir area, China. The findings indicated that the GIST model was less adept at managing the significant geographic variability of soil thickness, whereas the GIST-RF model excelled, achieving a mean absolute error of 3.52 m and a root mean square error of 4.56 m. In their study, Zhang et al. (2021) proposed an integrated model framework that utilizes the random forest approach to estimate the thickness of the black soil horizon at the watershed scale. This approach addresses the deficiency of spatial variation data on soil thickness in Northeast China's black soil region. This model can explain 61 % of the spatial variation in BSHT and is superior to traditional interpolation techniques in prediction accuracy.

Although progress has been made in predicting soil thickness in black soil regions, there are still multiple challenges at complex terrains and fine spatial scales. Firstly, existing environmental factors and traditional empirical models exhibit low prediction accuracy when addressing the highly heterogeneous and multi-factor-influenced black soil areas. While process-based models can explain the mechanisms, their practical application is limited due to difficulties in parameter acquisition (Yan et al., 2021a). Secondly, interpolation techniques assume that soil and environmental factors are evenly distributed within a region, but black soil areas exhibit strong spatial heterogeneity, which contradicts this assumption, leading to prediction inaccuracies that do not meet the fine-scale management needs of small watersheds (Feng et al., 2022). Although spectral information is crucial for soil erosion identification and spatial distribution analysis, it has not received sufficient attention in existing models. Black soil regions possess unique erosion and deposition processes that exhibit distinct spectral characteristics, yet these features remain underutilized. To improve prediction accuracy, it is necessary to develop an innovative approach that integrates multidimensional environmental variables and regional characteristics.

This study is the first to innovatively integrate spectral information, topographic features, and vegetation indices at the small watershed scale in the Northeast Black Soil Region, employing machine learning methods such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANNs) to thoroughly investigate the spatial distribution patterns of BSHT. Unlike previous research, this study takes into account the unique erosion-deposition processes of sloping farmland in this region: due to runoff erosion on sloping farmland in the black soil area, significant erosion typically occurs at the shoulder positions, resulting in a thinner black soil horizon, while the foot of the slope often features a thicker black soil horizon due to deposition. Additionally, changes in soil thickness at the summit and backslope are strongly influenced by geomorphological features and hydrological processes. These specific topography–hydrology interactions often render traditional prediction methods ineffective in explaining the spatial variability of soil thickness at the small watershed scale in the black soil region. In this study, RF, XGBoost, and ANNs are utilized to explore the synergistic effects of spectral information, topographic features, and vegetation indices to improve the accuracy of soil thickness prediction. The specific objectives of this research include: (1) evaluating the explanatory power of multidimensional environmental variables, especially bare soil imagery, for the variation in BSHT; (2) comparing the predictive accuracy of the RF, XGBoost, and ANNs models in this region; and (3) exploring the combined predictive advantages of topographic features and spectral information under the unique environmental conditions of sloping farmland in the black soil region. This study provides a novel modeling approach for assessing BSHT at the small watershed scale in the Northeast Black Soil Region and offers reliable data support for agricultural productivity assessment

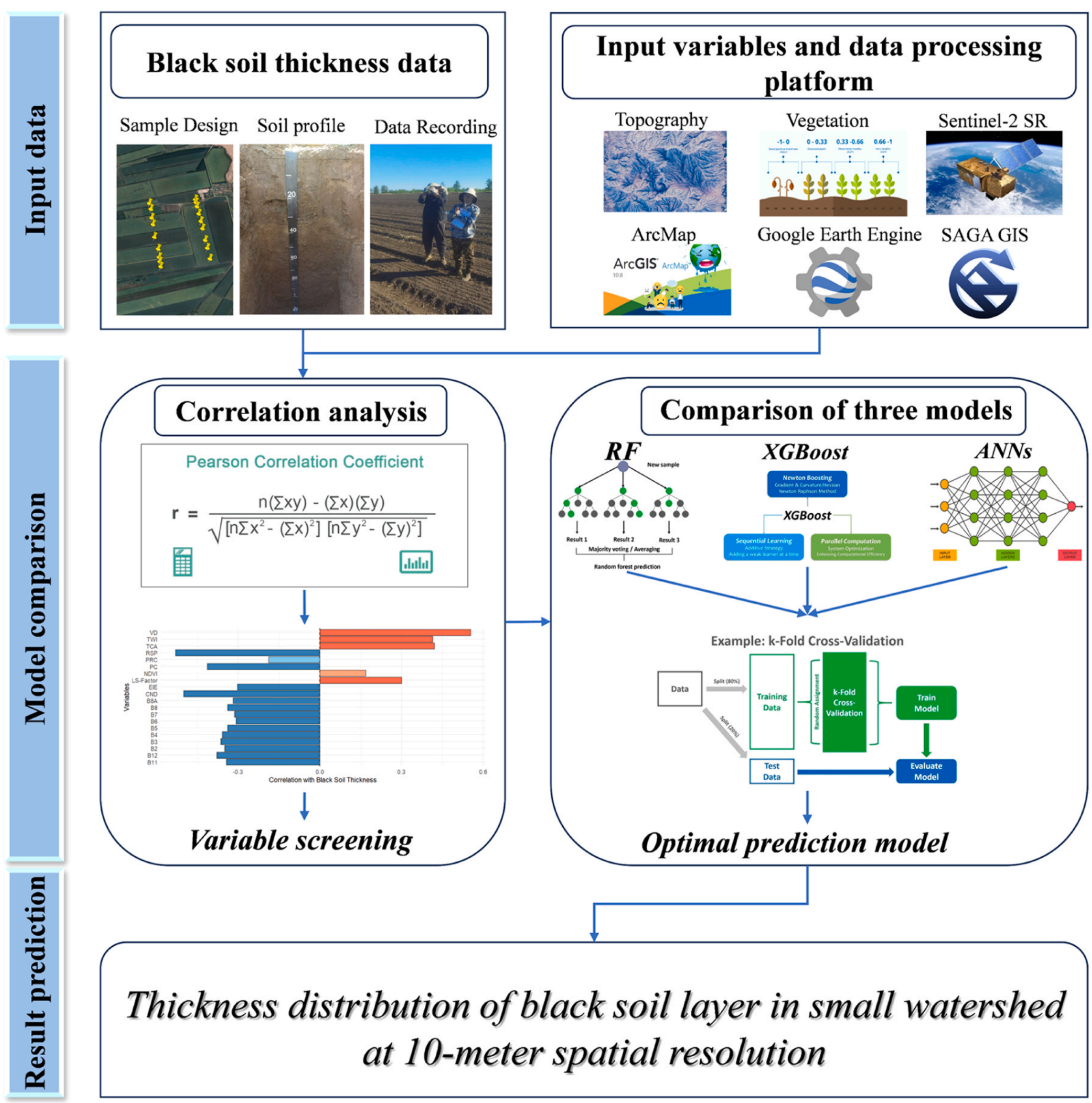

Fig. 1. Flowchart of BSHT prediction based on multi-source environmental variables.

and soil degradation prevention and control. The detailed research methodology is illustrated in Fig. 1.

2. Materials and methods

2.1. Study site and soil sampling

This research focused on three contiguous minor watersheds in Nenjiang City, China. The data for these watersheds were obtained from the HydroBASINS Level 8 watershed vector dataset (Lehner and Grill, 2013), which offers high-resolution watershed boundary details. The examined watersheds are located in the northern region of the black soil zone, including an area of 3106 square kilometers (Fig. 2a), with coordinates spanning from 49.4770°N to 48.3471°N latitude and 124.5048°E to 126.1899°E longitude, featuring diverse topographies and land use classifications. The elevation of the watersheds varies from 194 m to 488 m, with a mean elevation of 301 m (Fig. 2b). Approximately 78 % of the land in the study region is cultivable, predominantly sown with soybeans. The area experiences a temperate semi-humid continental monsoon climate, with an average annual temperature of around 0 degrees Celsius and an average annual precipitation of 534 millimeters (Zhang et al., 2023). The rainy season predominantly transpires from May to October, comprising 85.6 % of the yearly rainfall. Snowfall predominantly occurs from November to April in the subsequent year (He et al., 2023). The Chinese Soil Classification System categorizes black soil as Udic Isohumisols, which are further subdivided into four subcategories: conventional black soils, surface-gleyed black soils, meadow black soils, and chernozems (Nearing et al., 2017). In the U.S. Soil Taxonomy, black soil is categorized as Mollisols; however, in the World Reference Base for Soil Resources (WRB), it is designated Phaeozems (Liu et al., 2012).

2.2. Data sources

2.2.1. BSHT measurements

Given the significant geographic heterogeneity of the region, we established sampling stations at multiple locations in slope farmland using high-resolution Digital Elevation Model (DEM) data. Through spatial overlay analysis of existing soil type maps and Sentinel data, we selected a total of 157 representative sampling points. These points were chosen based on the Multiresolution Index of Valley Bottom Flatness and the Multiresolution Index of Ridge Top Flatness, and are distributed

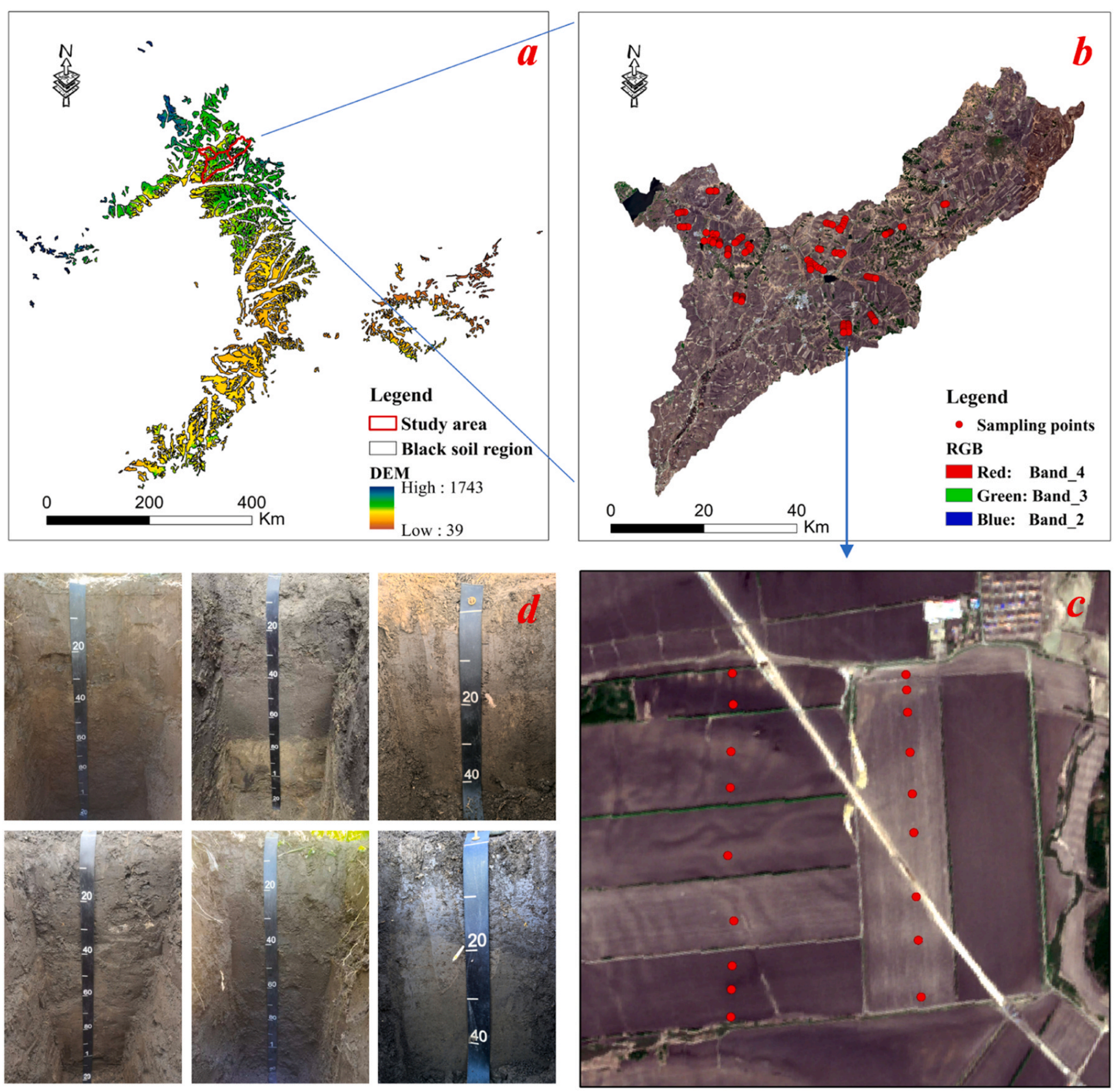

Fig. 2. (a) Distribution of black soil areas in China and the study area; (b) Distribution of sampling points in the study area; (c) Local distribution of sampling points; (d) Profile of sampling points.

across different topographic positions, including the foot, middle, and top of slopes, to elucidate the variation of BSHT with topography (Han et al., 2025). To ensure the representativeness of the samples, the regional topography was systematically classified based on different flatness indices, and multiple sampling transects were established in each type of slope area, thus comprehensively covering the topographic features of the study area. In May 2024, we performed field measurements of BSHT and documented the sampling point locations utilizing Global Positioning System technology (Fig. 2). At each sampling location, a profile was excavated to assess and document the thickness of the black soil horizon. At the same time, GPS technology was employed to capture the geographical coordinates of each site for further data processing and spatial analysis. While several areas did not fully encompass the entire slope type, most sampling points addressed the principal characteristics of the slope type. The sample data were randomly partitioned into modeling and validation sets at 2:1 to guarantee the model's dependability and validity. This study defines BSHT according to the categorization standards of the surface horizon (Pendleton et al., 1935) and the A horizon (National Soil Census Office, 1994, pp. 276–286).

2.2.2. Remote sensing data

Image reflectance is intricately linked to soil attributes, including organic matter, moisture content, and mineral composition, directly affecting its spectral reflectance characteristics. The intricacy and distinctiveness of erosion on sloping farms in black soil regions arise from the extensive presence of rivers and undulating topography, along with the prevalent agricultural method of ridge farming. These factors collectively lead to discrepancies in the thickness of black soil (Wang et al., 2023). These modifications are seen in the varying spectral properties of reflectance. Generally, the toe area has a relatively low reflectivity due to significant sedimentation, waterlogging, and high organic matter content, resulting in a thick black soil horizon. The slope shoulder area is usually the location with the most vigorous erosion due to the effects of accelerated water flow and gravity, and the soil is thin, with a high reflectivity. The intensity of erosion or deposition in the slope top and the slope back areas depends on the specific topographic and hydrodynamic conditions and may be expressed as erosion or stability. The area at the foot of the slope is located in the transition zone between the slope back and the toe of the slope. It is usually an area where deposition and erosion act together, and the soil thickness varies greatly (Fig. 3).

During the bare soil period in Northeast China, the soil surface is not covered by vegetation. By analyzing the reflectance of multiple spectral bands, spatial variations in soil physicochemical properties such as organic matter content and texture can be identified. These characteristics are related to soil development and erosion processes, providing important remote sensing data for the spatial prediction of black soil layer thickness. In this study, Sentinel-2 Level-2A imagery was processed on the Google Earth Engine platform; this dataset provides surface reflectance products that have undergone atmospheric correction and orthorectification by Sen2Cor. The specific data preprocessing steps are as follows: First, images from May 2023 with cloud coverage of less than 5 % were selected. This period is the bare soil season in Northeast China, making soil surface features easily identifiable by remote sensing. Then, the QA60 band was used to perform cloud masking on all images, and image mosaicking and temporal averaging were conducted to ensure complete coverage of the study area while removing the effects of clouds and shadows, ultimately resulting in a high-quality, cloud-free time-series image set. As for band selection, bands with little relevance to soil features—B1 (Coastal aerosol), B9 (Water vapor), and B10 (Cirrus)—were excluded. Instead, nine bands were selected: B2 (Blue), B3 (Green), B4 (Red), B5 (Vegetation Red Edge), B6 (Vegetation Red Edge), B7 (Vegetation Red Edge), B8 (Near Infrared), B8A (Narrow Near Infrared), B11 (Short Wave Infrared), and B12 (Short Wave Infrared), all of which are better suited for soil feature identification (Zhang et al., 2024b).

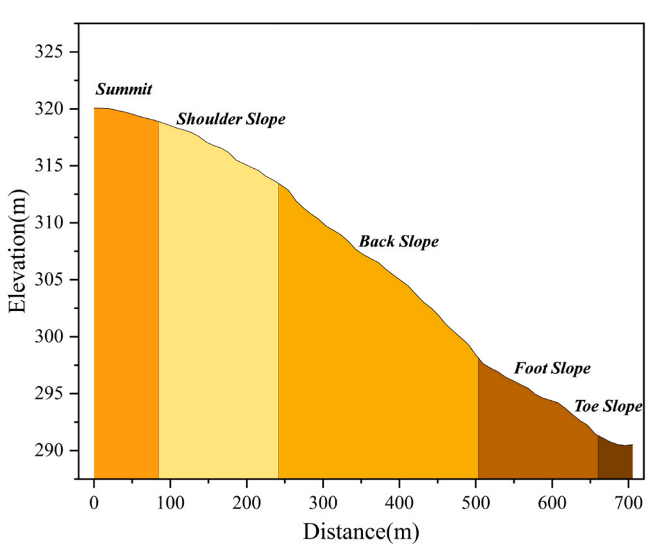

Fig. 3. Schematic diagram of slope position zoning.

2.2.3. Topographic factors

Studies demonstrate that topographic factors typically exhibit substantial spatial relationships with soil characteristics. Integrating these variables enhances the prediction of soil attributes (McKenzie and Ryan, 1999). This study employed high-resolution DEM data obtained from 91 Satellite Map Assistant, featuring a spatial resolution of 6.5 m. To ensure consistency with the spatial resolution of Sentinel-2 imagery, the DEM data were resampled to a 10-meter resolution with bilinear interpolation in ArcGIS 10.6.1. Eight topographic characteristics were derived from the resampled DEM data and utilized as input variables for simulating the thickness of the black soil horizon (Table 1). In conjunction with the DEM data, the topographic parameters were utilized to elucidate the geographical variability of BSHT. Topographic factors were extracted and calculated using SAGA software, while spatial overlay analysis was executed with additional data sources to maintain consistency in model input data. Fig. 4 illustrates the topographic elements used for the study, which encompass slope, valley depth, topographic wetness index, and relative slope position index, among others.

2.2.4. Vegetation data

The Normalized Difference Vegetation Index (NDVI) is a prevalent vegetation index utilized in remote sensing. It quantitatively represents surface vegetation's biomass, health status, and growth conditions (Huang et al., 2021). In sloped agricultural regions, the prolonged impacts of soil erosion and degradation result in a progressive reduction of

Table 1
Topographic factors used for modeling analysis of BSHT.

Topographic Factors	Software	Unit	Reference
Valley Depth(VD)	SAGA 9.5.1	Meter	Conrad et al. (2015)
Plan Curvature(PC)		Meter	
Profile Curvature(PRC)		Meter	
Total Catchment Area (TCA)		Square meters	
Topographic Wetness Index(TWI)		-	
Channel Network Distance(CND)		Meter	
LS-Factor		-	
Relative Slope Position (RSP)		-	
Elevation (ELE)	ArcGIS 10.6.1	Meter	Zevenbergen and Thorne (1987)

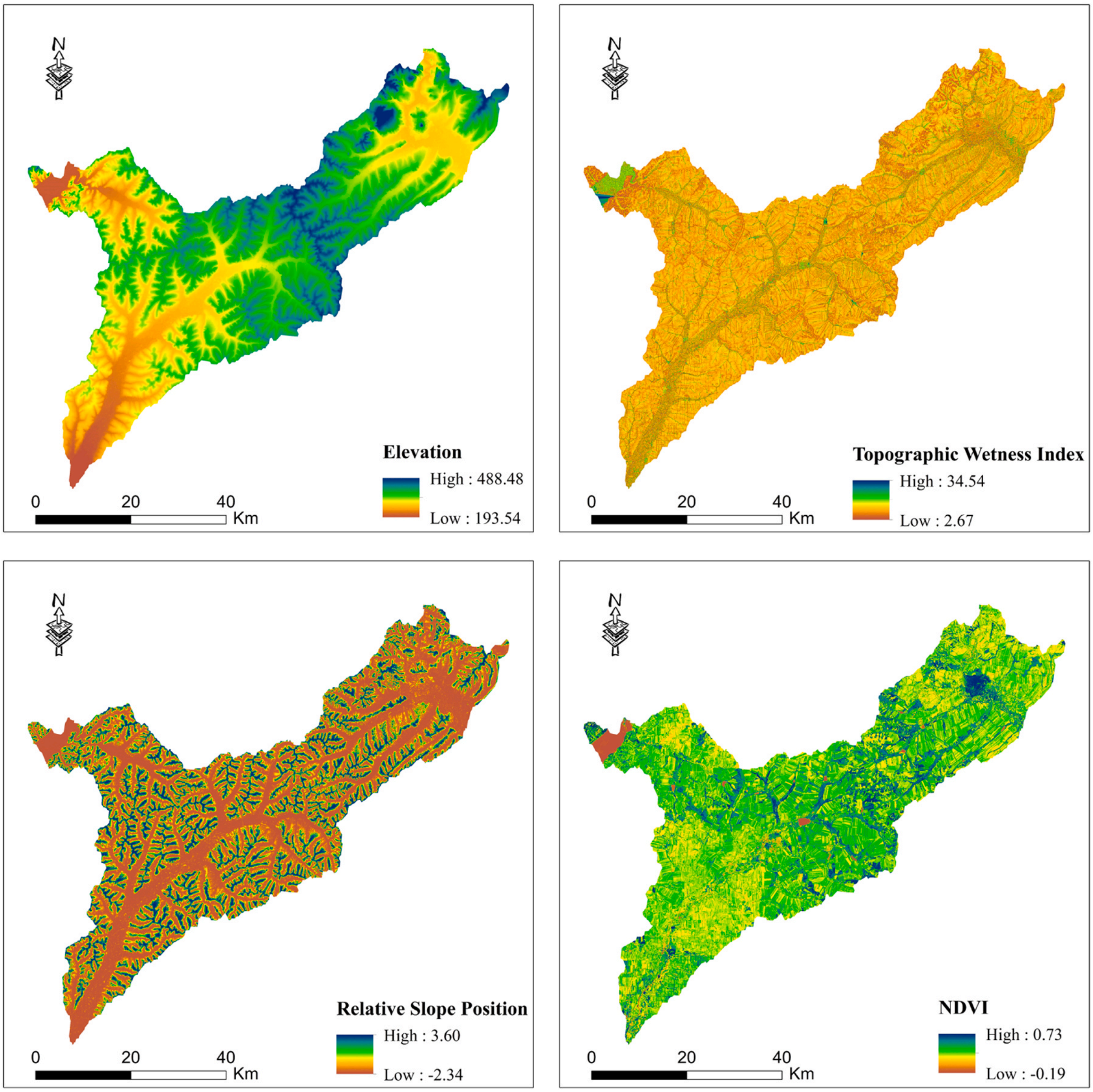

Fig. 4. Topographic factors (slope, valley depth, topographic wetness index, relative slope position) and NDVI derived from the digital elevation model and Sentinel-2 imagery in the studied area.

the black soil horizon's thickness, subsequently influencing vegetation growth conditions. Analyzing the fluctuating trends of NDVI enables the effective identification of possible soil degradation zones. The spatial distribution characteristics of NDVI can be utilized to deduce the variation patterns and trends of BSHT. The technique of calculation is as follows:

$$NDVI = \frac{NIR + RED}{NIR - RED} \quad (1)$$

This study employed the red band (B4) and near-infrared band (B8) of Sentinel-2 data to compute the monthly average Normalized Difference Vegetation Index (NDVI) for June, July, August, and September 2023. Photos exhibiting cloud cover below 5 % were picked on the GEE platform, NDVI values were computed for each pixel, and the NDVI photos for each month were averaged to provide matching monthly NDVI images. In the following investigation, we examined the relationship between NDVI values throughout various months and the thickness of the black soil horizon. The results showed that the NDVI value in June was positively correlated with the thickness of the black soil horizon. The thickness of the black soil horizon may play an important role during the critical stages of vegetation growth. Therefore, we decided to use the NDVI data in June as the input variable for the model.

2.3. Methodology

2.3.1. Random forest

RF is a commonly used regression algorithm that constructs multiple decision trees and averages their predictions to obtain the final prediction value. It is particularly suitable for dealing with complex and non-linear relationships while effectively reducing the risk of overfitting. The application of random forests is particularly suitable for soil thickness prediction because soil thickness is affected by many factors, including topography, soil parameters, and vegetation cover, which often exhibit significant spatial heterogeneity and non-linear characteristics (Xue et al., 2024). Moreover, RF does not depend on feature scaling and is appropriate for many data formats. During training, RF utilizes the bootstrap approach to randomly select about two-thirds of the samples from the training set to build each decision tree, while the unselected samples (out-of-bag samples) are employed for model validation and feature importance assessment. During tree growth, each node identifies the ideal split point from a randomly selected subset of characteristics, thereby augmenting tree variety, mitigating the risk of model overfitting, and improving generalization capability.

In this study, we performed the analysis based on the random forest model (randomForest package, version 4.7.1.1) in the R programming environment (version 4.3.2). We optimized the model performance by adjusting key parameters: the number of decision trees (ntree) was set to 500 to improve stability, the number of randomly selected features (mtry) was set to one-third of the total number of input features by default to balance model complexity and generalization ability, and the minimum node size (nodesize) was adjusted to 4 to enhance the model's adaptability to regression tasks. The remaining parameters were left at their default settings, as no effect on model performance was found during experimentation.

2.3.2. Extreme gradient boosting

XGBoost is a potent machine learning technique that uses gradient-boosted decision trees and excels, especially when dealing with high-dimensional and complex data. The basic principle is to combine a series of weak learners (i.e., decision trees) into a single robust model by continuously optimizing the loss function. In each iteration, XGBoost does not directly train a new decision tree to fit the original data but calculates the residual between the prediction result of the current model and the actual value and uses this residual as a new target for learning. XGBoost then trains a new decision tree to predict these residuals and gradually reduces the overall loss by adding the tree's predictions to the previous model. Therefore, each new tree is a correction to the prediction error of the previous tree, and by iterating continuously, the model can gradually improve the accuracy of its predictions. XGBoost, in contrast to conventional boosting techniques like Adaptive Boosting, does not modify the sample distribution; instead, it minimizes mistakes by enhancing the gradient boosting process. XGBoost incorporates regularization terms into the objective function, utilizing L1 (the sum of absolute weights) and L2 (the sum of squared weights) regularization to limit the degrees of freedom of the model parameters, contingent upon model complexity (including tree depth and the number of leaf nodes) and leaf node weights. This regularization technique efficiently mitigates overfitting and improves the model's generalization capacity (Chen and Guestrin, 2016).

This study established a maximum decision tree depth of 5 through experimental validation to enhance model resilience and efficiency, creating a simplified decision stump model to mitigate complexity and avert overfitting. The learning rate was established at 0.1 to regulate the cumulative effect of each new tree on the overall model across iterations, hence assuring model stability. The number of iterations was established at 500 to guarantee that the model could sufficiently learn the feature patterns in the data. All other parameters were configured to their default values, and no notable effect on the outcomes was detected during the studies. The research utilized R language version 4.3.2 and the XGBoost package version 1.7.8.1.

2.3.3. Artificial neural network

ANNs are computational models inspired by biological neural networks designed to model complex nonlinear relationships in data by simulating the way the human brain processes information (Yang and Wang, 2020). ANNs are composed of multiple artificial neurons (or nodes). Each neuron receives input signals, performs weighted summation operations, and outputs results through an activation function. ANNs generally include an input layer, one or more hidden layers, and an output layer, with signals transmitted between nodes via weighted connections. During training, ANNs dynamically adjust the weights of these connections using optimization algorithms such as gradient descent to minimize the value of a loss function and thus gradually improve the model's predictive power. These optimization algorithms are combined with learning rules, such as the backpropagation algorithm, to update the weights and biases during training. In addition, the structural design of the network, such as the number of layers and nodes, also directly impacts the performance of ANNs.

To prevent overfitting in the neural network model, L2 regularization was applied in this study. The regularization coefficient (decay) was set to 0.01 to limit the size of the model weights. The neural network consisted of an input layer, a single hidden layer with 4 neurons, and an output layer. Due to the randomness in weight initialization and optimization paths involved in ANNs during training, the maximum number of iterations (maxit) was set to 500 to ensure sufficient convergence of the model. For other parameters, default values were used, and experimental results validated their rationality, showing no significant impact on model performance. The study's operating environment used R language version 4.3.2, with the nnet package version 7.3.19.

2.3.4. Model evaluation and validation

In this work, all soil data were randomly partitioned into training and testing datasets to guarantee that model training and validation were representative. Environmental parameters were designated predictor variables, with the black soilhorizon's thickness as the modeling response variable. Ten-fold cross-validation was employed during model evaluation to guarantee the stability and correctness of the models utilized. To ensure the stability and reliability of the model, we used ten-fold cross-validation. The predictive performance of the model is evaluated using the following four standard metrics: the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). R^2 reflects the degree of fit of the model, while RMSE is used to evaluate the predictive accuracy. MAE quantifies the average absolute deviation between the predicted and observed values, while MAPE presents the relative size of the prediction error in percentage terms.

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

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

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

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (5)$$

Where n is the sample size, $\bar{y}$ is the average thickness of the black soil horizon, and y_i and $\hat{y}_i$ are the observed and predicted values of the BSHT at the i-th location, respectively.

3. Results

3.1. Response characteristics of spectral reflectance to different BSHTes

By extracting and averaging spectral curves from BSHT sample points at depths of 0–25 cm, 25–50 cm, 50–75 cm, and 75–100 cm, we noted that black soils of differing thicknesses exhibit similar shapes in the Sentinel-2 spectral curves while demonstrating significant variations in reflectance values (Fig. 5). The spectral reflectance increases progressively with wavelength from the visible light spectrum to shortwave infrared 1 (B2-B11), while it decreases gradually with increasing wavelength from shortwave infrared 1 to shortwave infrared 2 (B11-B12). Moreover, when the thickness of the black soil horizon increases, the overall soil reflectance typically decreases. This may correlate with the heightened organic matter and moisture levels in denser black soils,

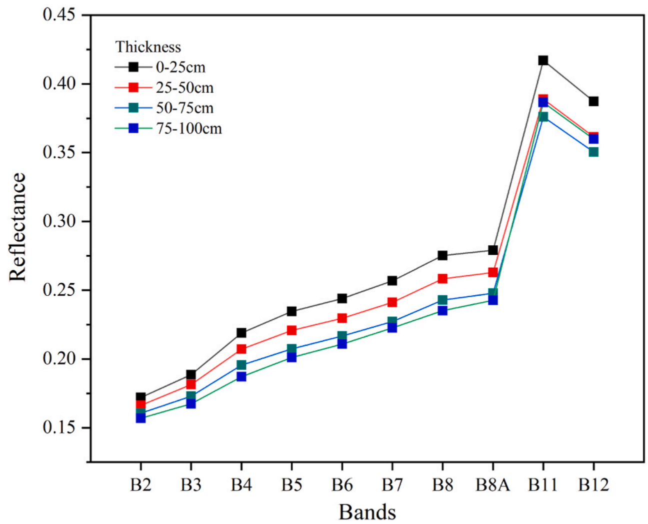

Fig. 5. Variations in spectral reflectance at different thicknesses of the black soil horizon.

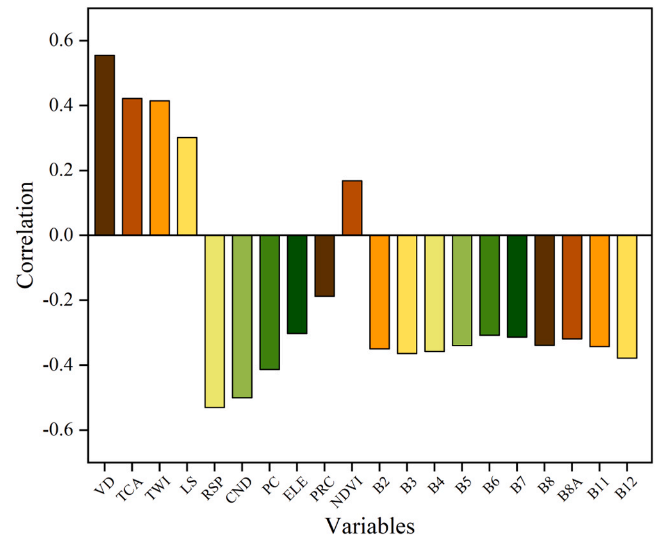

Fig. 6. Correlation of different Indicators with BSHT, where the correlation between NDVI and PC is statistically significant ($p < 0.05$), while the correlations of other variables with BSHT show more substantial statistical significance ($p < 0.01$).

enhancing light absorption and reducing reflection. In the shortwave infrared bands (B11 and B12), the reflectance of black soil at thicknesses of 25–50 cm and 75–100 cm is comparable; however, the reflectance at a thickness of 50–75 cm is the lowest, signifying superior light absorption characteristics.

3.2. Statistical analysis and correlation insights

Table 2 presents the statistical results about the thickness of the black soil horizon in the research region. The thickness of the black soil horizon varies from 19 cm to 98 cm, with a mean of 42.15 cm. The standard deviation (SD) of BSHT is 17.75 cm, and the coefficient of variation (CV) is 42.10 %, signifying substantial variability in BSHT. Pearson correlation analysis investigated the link between BSHT and variables, including NDVI, topographic parameters, and spectral reflectance (Fig. 6). The research revealed a correlation coefficient ranging from −0.53–0.56, indicating a statistically significant association between these variables and black soil horizon thickness ($P < 0.01$ or $P < 0.05$). Among the topographic factors, VD ($r = 0.56$, $P < 0.01$) showed the strongest positive correlation, indicating that an increase in the depth of the valley may significantly affect the change in the target variable, showing its high importance in the model; followed by TCA ($r = 0.42$, $P < 0.01$) and TWI ($r = 0.41$, $P < 0.01$), all of which were significantly positively correlated with the dependent variable, while RSP ($r = -0.53$, $P < 0.01$) and CND ($r = -0.50$, $P < 0.01$) showed a strong negative correlation and had a significant impact on the dependent variable. All bands (B2 to B12) showed highly significant negative correlations among the spectral reflectances, with correlation coefficients ranging from −0.22 to −0.36. Among them, B12 had the strongest negative correlation ($r = -0.36$, $P < 0.01$), indicating that changes in spectral reflectance can indirectly reflect the trend of soil thickness changes. In contrast, NDVI has a lower correlation ($r = 0.17$, $P < 0.05$), and the linear relationship between NDVI and BSHT is weak.

Table 2
Descriptive statistics of BSHT.

N	Mean(cm)	Minimum(cm)	Maximum(cm)	SD(cm)	CV(%)
157	42.15	19	98	17.75	42.10

(N: Sample size; Mean: Average value of the data; Minimum: Minimum value of the data; Maximum: Maximum value of the data; SD: Standard deviation of the data; CV: Relative variability of the data)

3.3. Model performance and accuracy

This work constructed six soil thickness prediction models utilizing three methods—RF, XGBoost, and ANNs—alongside two variable sets (a combination of topographic factors and NDVI; a combination of topographic factors, NDVI, and spectral information). The results revealed that the RF model exhibited superior predictive ability across both variable combinations, achieving an RMSE of 8.88 cm, an R^2 of 0.75, an MAE of 6.66 cm, and a MAPE of 18.09 % with the inclusion of spectral information. The XGBoost model had an RMSE of 9.41 cm, an R^2 of 0.68, an MAE of 7.57 cm, and a MAPE of 19.970 % after incorporating spectrum information, demonstrating a decrease in RMSE and MAPE relative to the model without spectral data. The ANNs model exhibited a superior R^2 (0.710) with the incorporation of spectral information, suggesting certain benefits in elucidating variable variability; however, the RMSE (12.72 cm), MAE (8.89 cm), and MAPE (19.00 %) were comparatively elevated, and its overall efficacy was inferior to that of RF and XGBoost. The results indicate that the random forest model performed best in predicting BSHT, and the introduction of multispectral band reflectance significantly improved the predictive performance of all models, indicating the importance of this variable in soil thickness prediction (Fig. 7).

3.4. Spatial distribution of black soil thickness

Three models were employed to assess the regional variability of BSHT in the study area, revealing significant spatial heterogeneity in BSHT. The RF model predicted a thickness range from 24.06 cm to 76.78 cm, with an average of 40.63 cm; the XGBoost model predicted a range from 17.01 cm to 93.82 cm, with an average thickness of 41.23 cm; and the ANNs model provided a relatively wider prediction range, from 6.92 cm to 83.73 cm, with an average thickness of 44.25 cm. The overall trend shows that BSHT gradually increases as elevation decreases within the watershed, indicating the important regulatory role of topography in the distribution of black soil. Although there are some differences in the predicted values among the models, the spatial distribution of BSHT is primarily controlled by topography.

Fig. 8 shows that all three models successfully captured the overall trend of BSHT variation within the watershed: the black soil horizon is thinner in high-altitude areas and relatively thicker in low-altitude areas. The RF model's lower root mean square error (RMSE) and mean

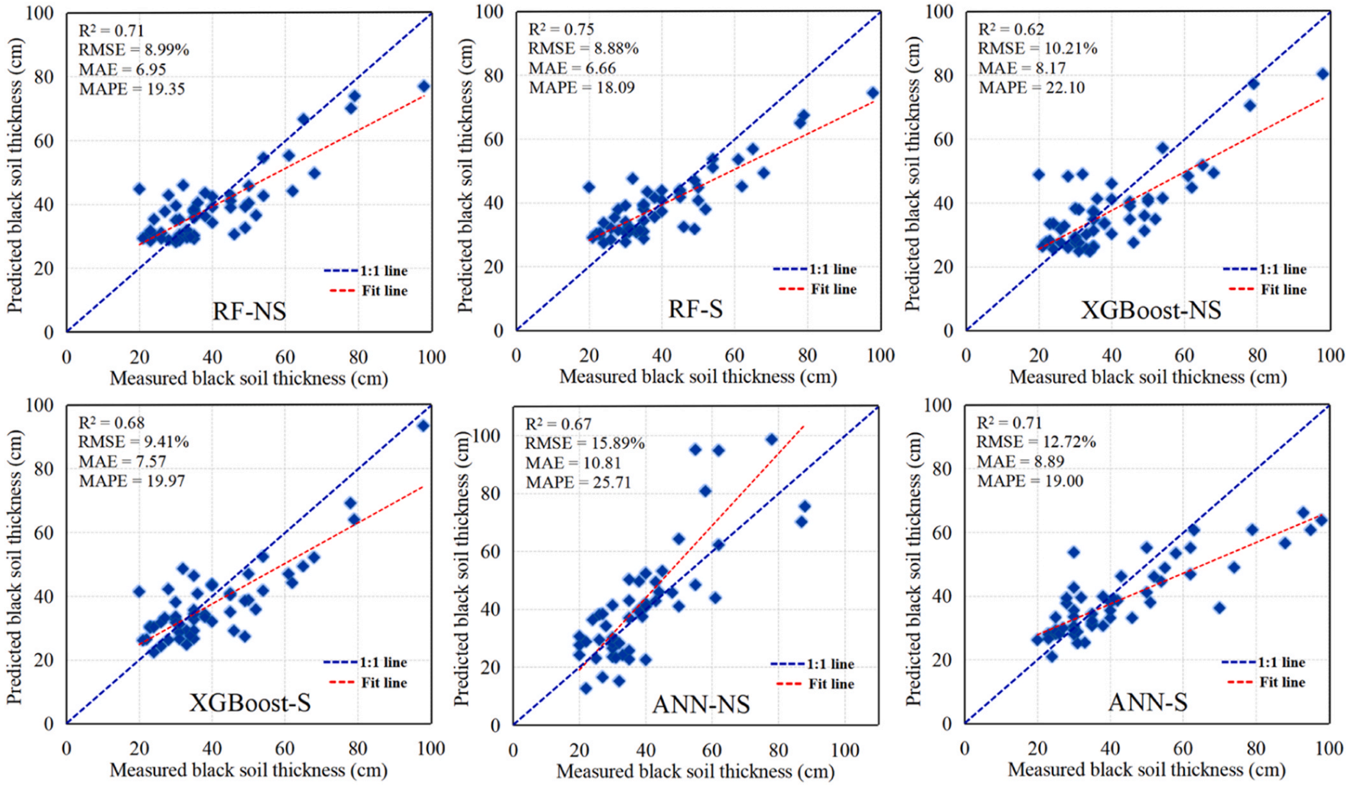

Fig. 7. Soil thickness prediction performance of Random Forest, XGBoost, and ANNs under different variable combinations (Spectral information (S) and Non-spectral information (NS)).

absolute error (MAE) indicate prediction stability and accuracy. The XGBoost model performed well in handling complex nonlinear interactions, mainly predicting a more pronounced increase in thickness in some low-altitude regions. ANNs exhibited more significant predictive variability between high and low altitudes in the transition zones, reflecting their high sensitivity to local data characteristics.

3.5. Comparison of importance of variables in different models

Through the normalization analysis of variable importance in the RF, XGBoost, and ANNs models, it can be seen that there are specific differences in the dependency on different variables among the three models (Fig. 9). In the RF model, the five most important variables are RSP, CND, TCA, VD, and TWI; in the XGBoost model, the most important variables in order are RSP, TCA, CND, VD, and TWI. In the ANNs model, the differences in variable importance are relatively small, with the most important variables including B8A, B12, LS-Factor, B8, and B2, which are primarily spectral information and topographic factors, demonstrating a balance between spectral information and topographic factors, with the contribution of spectral information being significantly higher than other models. These results reflect the ability of the ANNs model to integrate multiple data features. The combined performance of the three models suggests that topographic factors still dominate in predicting the thickness of the black soil horizon. At the same time, spectral information is an important complementary variable that helps improve the models' overall predictive performance.

4. Discussion

In soil thickness prediction studies, traditional methods usually rely on topographic factors and vegetation indices to characterize the spatial distribution of soils. Although these methods are effective to a certain extent, they tend to neglect the application of spectral information, especially in sloping cropland farmland. As these areas are susceptible to soil erosion, their spectral properties differ. Therefore, incorporating spectral information into soil thickness prediction models and evaluating the influencing factors of different modeling accuracies can provide more accurate predictions of the spatial distribution of BSHT.

4.1. Importance of input quantities and their interpretation

The results of this study are highly consistent with previous research, further confirming the important role of topographic features in soil thickness modeling. In this study, VD demonstrates a significant positive link with BSHT, while ElE exhibits a significant negative correlation. This suggests that valley terrains, owing to their more level features, have reduced erosion and are more prone to develop and sustain a thicker soil horizon. The research by Li et al. (2017) further verified this conclusion, demonstrating that elevation and valley depth are key influencing factors of active layer soil thickness change. Elevation diminishes soil formation potential by modifying local climatic conditions and exacerbating erosion processes, but valley depth considerably augments soil thickness by promoting sedimentation. Mehnatkesh et al. (2013) identified slope, wetness index, specific catchment area, and sediment transport index as significant topographic factors influencing the spatial variability of soil thickness in semi-arid hilly regions through multiple linear regression analysis. Among these, the wetness index and catchment area showed a highly significant positive correlation with soil depth, with the wetness index having the highest correlation, indicating that the topographic wetness index and catchment area are the most important variables for predicting soil thickness in this region. The conclusion of this study is partially consistent with previous research, validating the importance of TWI and TCA in predicting BSHT. However, this study also found that while TWI is a key factor, it is not the most important factor among all the variables. This mismatch may indicate the intricate mechanisms via which topographic factors influence soil properties across various geographies and soil types, alongside the impact of the multifaceted interactions of numerous factors on

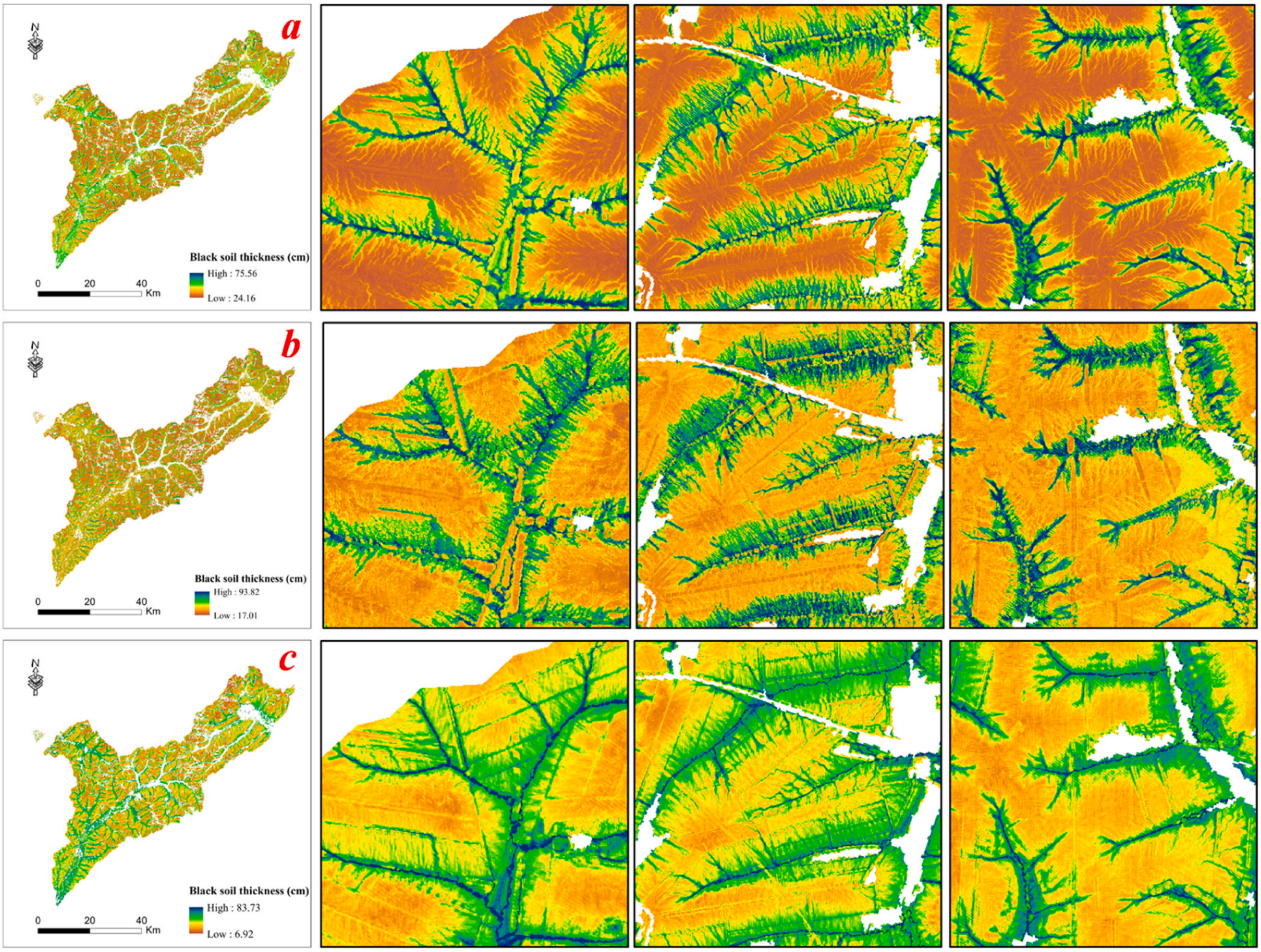

Fig. 8. Spatial distribution of BSHT inferred by RF (a), XGBoost (b), and ANNs (c) models.

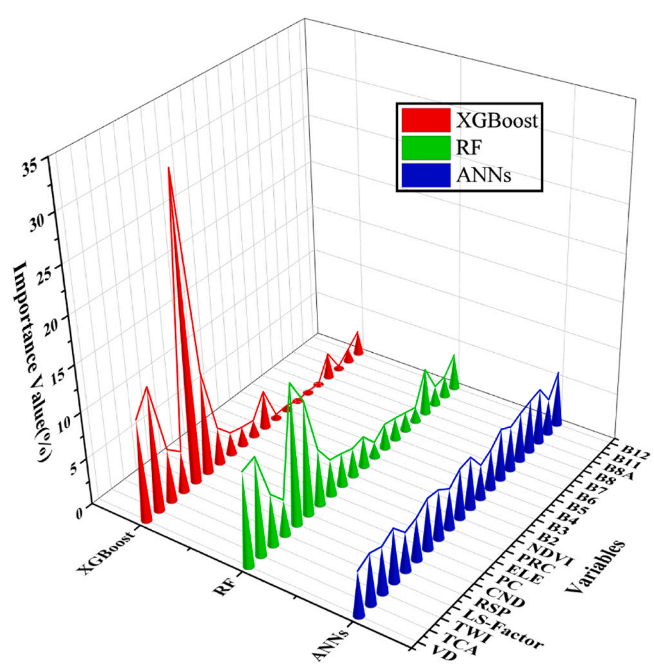

Fig. 9. Importance of variables for predicting black soil thickness using three models.

BSHT. Zhi et al. (2021) used the boosted regression tree model to identify key environmental variables affecting soil thickness. The study showed that RSP is one of the important variables influencing soil thickness. In this study, RSP can help identify the black soil layer thickness at different slope positions, as soil thickness is typically thicker in depressions and thinner in elevated areas. By using high-resolution DEM data combined with field-measured soil thickness, a more accurate prediction of the spatial distribution of soil can be made. CND represents the horizontal distance from a surface point to the nearest river network. Water flow gradually converges in areas close to the river, making these regions prone to forming depositional environments where soil thickness is generally greater. In highland or slope areas farther from the river, the dispersive effect of water flow is stronger, with erosion being dominant, often resulting in thinner soil horizons. The examined area is typical sloped agriculture; therefore, this topographic factor substantially impacts the distribution of black soil and its horizon thickness. Patton et al. (2018) proposed a simple empirical model based on high-resolution topographic data for predicting soil thickness. The study found a strong linear relationship between soil thickness and topographic curvature, with a significant linear correlation between soil thickness and curvature observed in both convergent and divergent terrain areas. In this study, PC showed a highly significant and high correlation, while PRC showed a significant but low correlation. Both influence trends on BSHT are similar to prior research findings. The LS-Factor combines slope length and gradient impacts, representing topography's function in hydrodynamic erosion processes.

Sun et al. (2024) constructed a high-resolution global set of topographic factors (LS-Factor) based on SRTM data. The study pointed out that the LS-Factor is critically important for assessing soil erosion risk, especially in steep slope areas, where high values of the LS-Factor significantly intensify soil erosion severity. In this investigation, the LS-Factor revealed a substantial positive link with BSHT, with high values largely distributed in toe slope positions and runoff path areas. Although high LS-Factor values are often linked with intense erosion, in runoff channel locations near the bottom of slopes, the lower kinetic energy of water flow leads to the deposition of eroded particles, generating thicker black soil. The greater thickness of the black soil horizon in these locations can be attributed to the lower flow velocity and depositional effects, particularly in relatively flat sections at the slope bottom where water deposits eroded material, resulting in increased soil thickness.

Topographic features can influence the formation and properties of soil, but plant growth is also jointly affected by various factors such as soil moisture and temperature. These hydrothermal conditions are often closely related to soil thickness. Generally speaking, the thicker the black soil layer, the stronger the soil's capacity to retain water and heat, which creates a favorable environment for plant growth. Therefore, as an indicator of vegetation coverage, NDVI can, to some extent, reflect soil moisture and temperature conditions as well as the thickness of the black soil horizon. This study has a substantial association between NDVI and BSHT, although the correlation is not strong, and its importance is relatively modest in the RF and XGBoost models. This contrasts with the findings of Zahedi et al. (2017), where NDVI was one of the essential variables for forecasting soil depth. NDVI considerably boosted the model's predicted accuracy in deep soils and agricultural areas through its synergistic interaction with land use and slope variables. However, in this study, the predictive value of NDVI was restricted, and its capacity to predict black soil horizon width was not as significant as other environmental variables. This implies that the applicability of NDVI differs across different study participants and environmental settings, though its combination with other factors still maintains some reference value.

4.2. Advantages of introducing spectral information

Traditional soil thickness prediction studies frequently employ topographic characteristics and vegetation indices as the primary input variables (Malone and Searle, 2020). Changes in topography affect water permeability, sediment loss, and accumulation, drainage conditions, and the distribution of water potential energy, influencing soil formation, erosion rates, and, ultimately, the spatial variation of soil depth. However, such studies often miss spectrum information, particularly the value of optical band data from remote sensing imaging in revealing soil characteristics. Spectral information can reflect key attributes of the soil surface, such as organic matter content, mineral composition, and moisture changes (Dharumarajan et al., 2024). These qualities are closely related to the regional variation of BSHT, especially in places with high erosion or deposition, and serve as essential markers for the spatial distribution of soil thickness.

This study found a significant negative correlation between black soil horizon thickness and spectral reflectance (correlation coefficient ranging from −0.22 to −0.36), with the B12 band showing the strongest negative correlation ($r = -0.36$, $P < 0.01$). As the thickness of the black soil horizon increases, the overall soil reflectance generally decreases, and this pattern is particularly evident in the shortwave infrared bands (B11-B12). The different reflectance values observed in the black soil horizon are primarily due to the following physical and chemical mechanisms: First, thicker black soil typically contains higher organic matter content, and the higher the organic matter content, the lower the soil's spectral reflectance (Francos et al., (2021)) . Second, topographic variations directly influence the spatial distribution of moisture in slope farmland. In low-lying areas, higher moisture content enhances light absorption and scattering, thereby displaying distinct characteristics in remote sensing imagery (Liu et al., 2020). Additionally, in eroded areas, soil erosion selectively transfers fine particles rich in organic carbon, leading to a reduction in the organic carbon content in the surface soil of eroded slopes (Li et al., 2019). As the black soil horizon thins and the plowing depth increases, loess-like parent material gradually mixes with the black soil, and the iron oxides in this mixture cause the soil color to turn yellow, thus altering the soil's spectral characteristics. By combining spectral information and topographic features, the model can resolve local-scale heterogeneity that topographic analysis alone cannot identify. When relying solely on topographic parameters, these variations are typically undetectable. This advantage becomes most evident in complex slope agricultural environments, thus significantly improving prediction performance.

It should be noted that topographic features and spectral information do not function independently, but exhibit a clear synergistic effect. For example, low-lying areas (identified by the VD index) often accumulate organic-rich sediments, forming thick black soil horizons that exhibit low reflectance in the shortwave infrared bands. By comparison, shoulder areas characterized by high EIE values show increased erosion susceptibility, causing thinner black soil horizons with correspondingly higher reflectance. The integration of topographic and spectral data substantially strengthens BSHT prediction performance and contributes to a better understanding of the spatial distribution characteristics of the black soil horizon.

4.3. The need for model comparison

Model comparison is necessary due to varying model sensitivities to factors like black soil horizon thickness. This ensures selection of the optimal model for our specific research context. Through model performance comparison, researchers can discover the model that best suits a particular dataset and recognize that certain models may offer considerable advantages under specific factors or environmental situations. While no universal algorithm exists for forecasting the spatial distribution of all soil qualities, some algorithms may be more suitable than others based on the data features and mapping aims (Khaledian and Miller, 2020). After doing model comparisons in this study, it was observed that the R^2 values for all three models were relatively high, each above 0.6. Among all, the RF model displayed the highest accuracy. This study shares a high degree of resemblance with previous research. De Sousa et al. (2023) found that topographic features play a significant role in identifying soil depth, whereas surface temperature and NDVI also play essential roles in forecasting soil depth. When evaluating the performance of several models, the RF model outperformed the support vector machine in soil depth prediction, with an average R^2 value of 0.77, indicating the superiority of the RF model for this assignment. Moreover, this study indicated that different models evaluate the value of variables differently, demonstrating differential sensitivities and adaptability of the models to input data elements. For instance, in RF and XGBoost, topographic-related variables (such as ELE, CND, and TCA) exhibited increased importance, while ANNs emphasized spectrum information (such as NDVI and band data). Therefore, different algorithms should be chosen to balance data types and research aims. Additionally, this study's results reveal that the model selection and the matching of data features are critical for the accuracy of soil thickness prediction. For example, RF and XGBoost are more dependable with small datasets, whereas ANNs can better capture complicated nonlinear interactions when dealing with more enormous datasets (Micheli-Tzanakou, 2011). In regional-scale soil thickness studies, a combination of methods facilitates the selection of optimal models and enhances the understanding of the effects of variable relationships on the distribution of BSHT.

4.4. Limitations and future research

This work focuses on the predictive analysis of cultivated land soil thickness but also has certain drawbacks. First, the research area is confined to cultivated land, without examining the soil thickness characteristics of other land use types (such as forest land or grassland), which may limit the applicability of the results to a wider variety of land types. Some research indicates that land use practices and erosion processes significantly affect the temporal variation of soil thickness (Zhang et al., 2024a). Secondly, this study primarily focuses on the influence of topographic factors and spectral information on soil thickness. However, it does not fully consider other factors that may significantly affect the thickness of the cultivated black soil horizon, such as soil management practices and the historical legacy effects of human activities. These components are critical for the long-term alterations and recovery capacity of soil. As Vanwalleghem et al. (2017) observed in their research, drastic changes in land use considerably exacerbate soil erosion, significantly when human activities surpass the natural recovery capacity of the soil, making changes in soil thickness more noticeable. Additionally, the study assesses soil thickness solely within a given period, neglecting temporal dynamic changes. Some studies have shown that the topsoil thickness of black soil has fallen from 60 to 70 centimeters in the 1950s to the current 20–30 centimeters, with loess parent material exposed in some regions, resulting in barren land that is unsuited for agricultural farming (Shen et al., 2019). In order to further improve the applicability and prediction accuracy of the model, future research should focus on the following aspects: first, this study only focused on the thickness of the black soil horizon in arable land, and in the future, the analysis can be extended to include different land use types, such as woodland and grassland, in order to assess the performance of the model in various environments. Second, the influencing factors are enriched by incorporating environmental factors, such as meteorological data, into the existing variables like topography and spectral data to describe the complex relationships affecting soil thickness more comprehensively. Third, the time dimension will be introduced by employing time series analysis methods to study the changes in soil thickness over time, such as the impact of long-term cultivation on soil thickness.

5. Conclusions

This study focuses on a small watershed in the black soil region of Northeast China, an area with complicated terrain and significant slope position changes, where BSHT demonstrates distinct spatial variability. By integrating topographic parameters, spectral information, and vegetation indicators and employing RF, XGBoost, and ANNs machine learning models, the study obtained high-precision predictions of soil thickness. The results reveal that multidimensional environmental variables may explain 68 %-75 % of the variability in soil thickness, including spectral information (multiband image data), considerably boosting the models' ability to explain regional changes in soil thickness. In comparing model performance, the RF model beat XGBoost and ANNs, displaying superior stability and accuracy across assessment criteria such as RMSE, R^2, MAE, and MAPE. In places with complicated topography and considerable slope position differences, the synergistic effect of spectral information and topographic parameters indicated a notable predictive advantage. This combination boosted the models' sensitivity to soil parameters (such as organic matter and moisture content), effectively enhancing forecast accuracy and reliability. Overall, this study constructed a high-precision predictive model integrating topographic factors and spectral information, revealing the spatial variability of BSHT and its main influencing factors while validating the potential of spectral information in small watershed-scale soil thickness studies. The findings give credible data support for assessing agricultural production and controlling soil degradation in the black soil region of Northeast China.

Statement

During the preparation of this work, the author utilized OpenAI GPT-4 to assist in optimizing the article's logic and providing sentence structures. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the publication's content.

CRediT authorship contribution statement

Chang Dong: Writing – review & editing. **Huanjun Liu:** Supervision, Resources, Funding acquisition, Conceptualization. **Weimin Ruan:** Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Chong Luo:** Writing – review & editing. **Xiangtian Meng:** Investigation, Data curation. **Yueyu Sui:** Resources, Data curation, Conceptualization. **Changkun Wang:** Validation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Young Scientists Innovation Funds of State Key Laboratory of Black Soils Conservation and Utilization (2023HTDGZ-QN-01) and the National Key Research and Development Program of China (2021YFD1500100). The authors sincerely thank the anonymous reviewers and members of the editorial team for their comments and contributions.

Data availability

The authors do not have permission to share data.

References

Bonfatti, Benito R., Hartemink, Alfred E., Vanwalleghem, Tom, Minasny, Budiman, Giasson, Elvio, 2018. A mechanistic model to predict soil thickness in a valley area of Rio Grande do sul. Geoderma 309, 17–31. https://doi.org/10.1016/j.geoderma.2017.08.036.

Chen, Songchao, Richer-de-Forges, Anne C., Leatitia Mulder, Vera, Martelet, Guillaume, Loiseau, Thomas, Lehmann, Sébastien, Arrouays, Dominique, 2021. Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France. CATENA 198 (March), 105062. https://doi.org/10.1016/j.catena.2020.105062.

Chen, Tianqi, Guestrin, Carlos, 2016. XGBoost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8 (7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Dharumarajan, S., Lalitha, M., Kalaiselvi, B., Kaliraj, S., Adhikari, K., Vasundhara, R., Niranjana, K.V., Rajendra Hegde, C.M., Pradeep, Hittanagi, P., 2024. Remote sensing of soils: spectral signatures and spectral Indices. Remote Sensing of Soils. Elsevier, pp. 13–23. ⟨https://www.sciencedirect.com/science/article/pii/B9780443187735000338⟩.

Dietrich, William E., Robert Reiss, Mei-Ling Hsu, David, R., Montgomery, 1995. A Process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrol. Process. 9 (3–4), 383–400. https://doi.org/10.1002/hyp.3360090311.

D'Odorico, Paolo, 2000. A possible bistable evolution of soil thickness. J. Geophys. Res. Solid Earth 105 (B11), 25927–25935. https://doi.org/10.1029/2000JB900253.

Feng, L.I.U., Yang, Fei, Zhao, Yuguo, Zhang, Ganlin, Li, Decheng, 2022. Predicting soil depth in a large and complex area using machine learning and environmental correlations. J. Integr. Agric. 21 (8), 2422–2434. https://doi.org/10.1016/S2095-3119(21)63692-4.

Frances, Alain P., Lubczynski, Maciek W., 2011. Topsoil thickness prediction at the catchment scale by integration of invasive sampling, surface geophysics, remote sensing and statistical modeling. J. Hydrol. 405 (1), 31–47. https://doi.org/10.1016/j.jhydrol.2011.05.006.

Francos, Nicolas, Ogen, Yaron, Ben-Dor, Eyal, 2021. Spectral assessment of organic matter with different composition using reflectance spectroscopy. Remote Sens. 13 (8), 1549. https://doi.org/10.3390/rs13081549.
Fu, Zhiyong, Li, Zhaoxia, Cai, Chongfa, Shi, Zhihua, Xu, Qinxue, Wang, Xiaoyan, 2011. Soil thickness effect on hydrological and erosion characteristics under sloping lands: a hydropedological perspective. Geoderma 167168 (Novemb. 41–53. https://doi.org/10.1016/j.geoderma.2011.08.013.
Han, Xiaole, Liu, Jintao, Wu, Pengfei, Yu, Zhenghong, Qiao, Xiao, Yang, Hai, 2025. Predicting the thickness of alpine meadow soil on headwater hillslopes of the Qinghai-Tibet plateau. Geoderma 456, 117271. https://doi.org/10.1016/j.geoderma.2025.117271.
He, Yangbo, Gao, Yuhao, Li, Xinyue, Chen, Junxi, Yang, Jingde, Chen, Jiazhou, Cai, Chongfa, 2023. Influence of gully erosion on hydraulic properties of black Soil-Based farmland. Catena 232, 107372. https://doi.org/10.1016/j.catena.2023.107372.
Heimsath, Arjun M., Dietrich, William E., Nishiizumi, Kunihiko, Finkel, Robert C., 1997. The soil production function and landscape equilibrium. Nature 388 (6640), 358–361. https://doi.org/10.1038/41056.
Huang, Sha, Tang, Lina, Hupy, Joseph P., Wang, Yang, Shao, Guofan, 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. For. Res. 32 (1), 1–6. https://doi.org/10.1007/s11676-020-01155-1.
Khaledian, Yones, Miller, Bradley A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model. 81, 401–418. https://doi.org/10.1016/j.apm.2019.12.016.
Kuriakose, Sekhar L., Devkota, Sanjaya, Rossiter, D.G., Jetten, V.G., 2009. Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of kerala, India. Catena 79 (1), 27–38. https://doi.org/10.1016/j.catena.2009.05.005.
Lehner, Bernhard, Grill, Günther, 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol. Process. 27 (15), 2171–2186. https://doi.org/10.1002/hyp.9740.
Li, Aidi, Tan, Xing, Wu, Wei, Liu, Hongbin, Zhu, Jie, 2017. Predicting Active-Layer soil thickness using topographic variables at a small watershed scale. Plos One 12 (9), e0183742. https://doi.org/10.1371/journal.pone.0183742.
Li, Tong, Zhang, Haicheng, Wang, Xiaoyuan, Cheng, Shulan, Fang, Huajun, Liu, Gang, Yuan, Wenping, 2019. Soil erosion affects variations of soil organic carbon and soil respiration along a slope in northeast China. Ecol. Process. 8 (1), 28. https://doi.org/10.1186/s13717-019-0184-6.
Li, Xinchuan, Luo, Juhua, Jin, Xiuliang, He, Qiaoning, Niu, Yun, 2020. Improving soil thickness estimations based on multiple environmental variables with stacking ensemble methods. Remote Sens. 12 (21), 3609. https://doi.org/10.3390/rs12213609.
Liu, Jintao, Chen, Xi, Lin, Henry, Liu, Hu, Song, Huiqing, 2013. A simple Geomorphic-based analytical model for predicting the spatial distribution of soil thickness in headwater hillslopes and catchments. Water Resour. Res. 49 (11), 7733–7746. https://doi.org/10.1002/2013WR013834.
Liu, Muxing, Wang, Qiuyue, Guo, Li, Yi, Jun, Lin, Henry, Zhu, Qing, Fan, Bihang, Zhang, Hailin, 2020. Influence of canopy and topographic position on soil moisture response to rainfall in a hilly catchment of three gorges reservoir area, China. J. Geogr. Sci. 30 (6), 949–968. https://doi.org/10.1007/s11442-020-1764-1.
Liu, Xiaobing, Lee Burras, Charles, Kravchenko, Yuri S., Duran, Artigas, Huffman, Ted, Morras, Hector, Studdert, Guillermo, Zhang, Xingyi, Cruse, Richard M., Yuan, Xiaohui, 2012. Overview of mollisols in the world: distribution, land use and management. Can. J. Soil Sci. 92 (3), 383–402. https://doi.org/10.4141/cjss2010-058.
Malone, Brendan, Searle, Ross, 2020. Improvements to The Australian national soil thickness map using an integrated data mining approach. Geoderma 377, 114579. https://doi.org/10.1016/j.geoderma.2020.114579.
McKenzie, Neil J., Ryan, Philip J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89 (1), 67–94. https://doi.org/10.1016/S0016-7061(98)00137-2.
Mehnatkesh, Abdolmohammad, Ayoubi, Shamsollah, Jalalian, Ahmad, Sahrawat, Kanwar L., 2013. Relationships between soil depth and terrain attributes in a semi arid hilly region in Western Iran. J. Mt. Sci. 10 (1), 163–172. https://doi.org/10.1007/s11629-013-2427-9.
Micheli-Tzanakou, Evangelia, 2011. Artificial neural networks: an overview. Netw. Comput. Neural Syst. 22 (1–4), 208–230. https://doi.org/10.3109/0954898X.2011.638355.
National Soil Census Office. (1994). Soil types in China. Beijing: China agricultural press (In Chinese).
Nearing, Mark A., Xie, Yun, Liu, Baoyuan, Ye, Yu, 2017. Natural and anthropogenic rates of soil erosion. Int. Soil Water Conserv. Res. 5 (2), 77–84. https://doi.org/10.1016/j.iswcr.2017.04.001.
Patton, Nicholas R., Lohse, Kathleen A., Godsey, Sarah E., Crosby, Benjamin T., Seyfried, Mark S., 2018. Predicting soil thickness on soil mantled hillslopes. Nat. Commun. 9 (1), 3329. https://doi.org/10.1038/s41467-018-05743-y.
Pendleton, Robert, Larimore, Hou, K., 1935. A reconnaissance soil survey of the harbin region. Soils Bull. 11, 42–106.
Penížek, V., Borůvka, L., 2006. Soil depth prediction supported by primary terrain attributes: a comparison of methods. Plant Soil Environ. 52 (9), 424–430. https://doi.org/10.17221/3461-PSE.
Roering, J.J., 2008. How well can hillslope evolution models "explain" topography? Simulating soil transport and production with high-resolution topographic data. Geol. Soc. Am. Bull. 120 (9-10), 1248–1262. https://doi.org/10.1130/B26283.1.
Sarkar, Shraban, Roy, Archana K., Martha, Tapas R., 2013. Soil depth estimation through Soil-Landscape modelling using regression kriging in a himalayan terrain. Int. J. Geogr. Inf. Sci. ⟨https://www.tandfonline.com/doi/abs/10.1080/13658816.2013.814780⟩ (December).
Shen, Congying, Wang, Yu, Zhao, Lanpo, Xu, Xiaohong, Yang, Xiankun, Liu, Xiaolin, 2019. Characteristics of material migration during soil erosion in sloped farmland in the black soil region of northeast China, 1940082919856835 Trop. Conserv. Sci. 12 (January). https://doi.org/10.1177/1940082919856835.
de Sousa, Gabriel, Pimenta Barbosa, Mahboobeh, Tayebi, Lucas, Rabelo Campos, Lucas T., Greschuk, Merilyn, Taynara Accorsi Amorim, Jorge, Tadeu Fim Rosas, Fellipe, Alcantara de Oliveira Mello, Songchao, Chen, Shamsollah Ayoubi, Demattê, José A. M., 2023. Improvement of spatial prediction of soil depth via earth observation. Catena 223, 106915. https://doi.org/10.1016/j.catena.2023.106915.
Sun, Yuwei, Zhang, Hongming, Yang, Qinke, Li, Rui, Liu, Baoyuan, Zhao, Xining, Shi, Haijing, Li, Hongyi, Ren, Yuhan, Fan, Xiao, 2024. A new High-Resolution global topographic factor dataset calculated based on SRTM. Sci. Data 11 (1), 101. https://doi.org/10.1038/s41597-024-02917-w.
Vanwalleghem, Tom, Gómez, J.A., Infante Amate, J., González De Molina, M., Vanderlinden, Karl, Guzmán, Gema, Laguna, Ana, Juan Vicente Giráldez, 2017. 'Impact of historical land use and soil management change on soil erosion and agricultural sustainability during the anthropocene'. Anthropocene 17, 13–29. https://doi.org/10.1016/j.ancene.2017.01.002.
Wang, Yuxian, Xu, Yingying, Yang, Huiying, Shen, Huibo, Zhao, Lei, Zhu, Baoguo, Wang, Jiangxu, Guo, Lifeng, 2023. Effect of slope shape on soil aggregate stability of slope farmland in black soil region. Front. Environ. Sci. 11 (February). https://doi.org/10.3389/fenvs.2023.1127043.
Xiao, Ting, Segoni, Samuele, Liang, Xin, Yin, Kunlong, Casagli, Nicola, 2023. Generating soil thickness maps by means of Geomorphological-Empirical approach and random forest algorithm in wanzhou county, three gorges reservoir. Geosci. Front. 14 (2), 101514. https://doi.org/10.1016/j.gsf.2022.101514.
Xu, X.Z., Xu, Y., Chen, S.C., Xu, S.G., Zhang, H.W., 2010. Soil loss and conservation in the black soil region of northeast China: a retrospective study. Environ. Sci. Policy 13 (8), 793–800. https://doi.org/10.1016/j.envsci.2010.07.004.
Xue, Zhenghai, Yi, Xiaoyu, Feng, Wenkai, Kong, Linghao, Wu, Mingtang, 2024. Prediction and mapping of soil thickness in alpine canyon regions based on whale optimization algorithm optimized random forest: a case study of baihetan reservoir area in China. Comput. Geosci. 191, 105667. https://doi.org/10.1016/j.cageo.2024.105667.
Yan, Qina, Wainwright, Haruko, Dafflon, Baptiste, Uhlemann, Sebastian, Steefel, Carl I., Falco, Nicola, Kwang, Jeffrey, Hubbard, Susan S., 2021a. A hybrid Data–Model approach to map soil thickness in mountain hillslopes. Earth Surf. Dyn. 9 (5), 1347–1361. https://doi.org/10.5194/esurf-9-1347-2021.
Yan, Tingting, Zhao, Weijun, Zhu, Qingke, Xu, Fujin, Gao, Zhikang, 2021b. Spatial distribution characteristics of the soil thickness on different land use types in the yimeng mountain area, China. Alex. Eng. J. 60 (1), 511–520. https://doi.org/10.1016/j.aej.2020.09.024.
Yang, Guangyu Robert, Wang, Xiao-Jing, 2020. Artificial neural networks for neuroscientists: a primer. Neuron 107 (6), 1048–1070. https://doi.org/10.1016/j.neuron.2020.09.005.
Yanto, Arwan, Apriyono, Purwanto, Bekti Santoso, Sumiyanto, 2022. Landslide susceptible areas identification using IDW and ordinary kriging interpolation techniques from hard soil depth at middle Western central java, Indonesia. Nat. Hazards 110 (2), 1405–1416. https://doi.org/10.1007/s11069-021-04982-5.
Zahedi, Salahudin, Shahedi, Kaka, Habibnejad Rawshan, Mahmod, Solimani, Karim, Dadkhah, Kourosh, 2017. Soil depth modelling using terrain analysis and satellite imagery: the case study of qeshlaq mountainous watershed (Kurdistan, Iran). J. Agric. Eng. 48 (3), 167–174. https://doi.org/10.4081/jae.2017.595.
Zevenbergen, Lyle W., Thorne, Colin R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landf. 12 (1), 47–56. https://doi.org/10.1002/esp.3290120107.
Zhang, Mengjie, Zhang, Wei, Zhang, Keli, Yu, Yue, Liu, Liang, 2023. Centennial scale temporal responses of soil magnetic susceptibility and spatial variation to human cultivation on hillslopes in northeast China. Soil Tillage Res. 234 (October), 105865. https://doi.org/10.1016/j.still.2023.105865.
Zhang, Shuai, Liu, Gang, Chen, Shuli, Rasmussen, Craig, Liu, Baoyuan, 2021. Assessing soil thickness in a black soil watershed in northeast China using random forest and field observations. Int. Soil Water Conserv. Res. 9 (1), 49–57. https://doi.org/10.1016/j.iswcr.2020.09.004.
Zhang, Yakun, Hartemink, Alfred E., Vanwalleghem, Tom, Bonfatti, Benito Roberto, Moen, Steven, 2024a. Climate and land use changes explain variation in the a horizon and soil thickness in the United States. Commun. Earth Environ. 5 (1), 129. https://doi.org/10.1038/s43247-024-01299-6.
Zhang, Yaohua, Xu, Xianli, Li, Zhenwei, Yi, Ruzhou, Xu, Chaohao, Luo, Wei, 2022. Modelling soil thickness using environmental attributes in karst watersheds. Catena 212, 106053. https://doi.org/10.1016/j.catena.2022.106053.
Zhang, Yu, Luo, Chong, Zhang, Yuhong, Gao, Liren, Wang, Yihao, Wu, Zexin, Zhang, Wenqi, Liu, Huanjun, 2024b. Integration of bare soil and crop growth remote sensing data to improve the accuracy of soil organic matter mapping in black soil areas. Soil Tillage Res. 244 (December), 106269. https://doi.org/10.1016/j.still.2024.106269.
Zhi, Junjun, Zhou, Zequn, Cao, Xinyue, 2021. Exploring the determinants and distribution patterns of soil mattic horizon thickness in a typical alpine environment using boosted regression trees. Ecol. Indic. 133, 108373. https://doi.org/10.1016/j.ecolind.2021.108373.