

Estimating the Spatial Distribution of Soil Organic Carbon and Soil Total Nitrogen Stocks by Remote Sensing Methods: A Case Study in a Forest-dominated Alpine Watershed, Western China

chao wang (✉ wangch-03@163.com)

Lanzhou City University <https://orcid.org/0000-0002-3924-7321>

Chuanyan Zhao

Lanzhou University

Kaiming Li

Lanzhou City University

Shouzhong Peng

Northwest A&F University: Northwest Agriculture and Forestry University

Ying Wang

Key laboratory of arid climatic reducing disaster of Gansu province

Research Article

Keywords: soil organic carbon, soil total nitrogen, multiple linear regression model, remote sensing technology, Qilian mountains

Posted Date: July 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-691637/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Estimating the spatial distribution of soil organic carbon and soil total nitrogen stocks by remote sensing methods: a case study in a forest-dominated alpine watershed, western China

Chao WANG^{1*}, Chuanyan Zhao², Kaiming Li¹, Shouzhang Peng³, Ying Wang⁴

¹ School of Geography and Environmental Engineering, Lanzhou City University, Lanzhou 730070, China;

² School of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China;

³ State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest A&F University, Yangling 712100, China;

⁴ Key Laboratory of Arid Climatic Change and Reducing Disaster of Gansu Province, Key Open Laboratory of Arid Change and Disaster Reduction of CMA, Institute of Arid Meteorology, China Meteorological Administration, Lanzhou 730020, China

Abstract: Soil organic carbon and soil total nitrogen stocks are important indicators for evaluating soil health and stability. Accurately predicting the spatial distribution of soil organic carbon and total nitrogen stocks is an important basis for mitigating global warming, ensuring regional food security, and maintaining the sustainable development of ecologically fragile areas. On the basis of field sampling data and remote sensing technology, this study divided the topsoil (0–30 cm) into three soil layers of 0–10 cm, 10–20 cm, and 20–30 cm to carry out soil organic carbon and soil total nitrogen stocks estimation experiments in the Qilian Mountains in western China. A multiple linear regression model and a stepwise multiple linear regression model were used to estimate soil organic carbon and soil total nitrogen stocks. A total of 119 topsoil samples and nine remotely sensed environmental variables were collected and used for model development and validation. The results indicated that these two linear regression models showed good performance. The modified soil-adjusted vegetation index (MSAVI), perpendicular vegetation index (PVI), aspect, elevation, and solar radiation were the key environmental variables affecting soil organic carbon and total nitrogen stocks. In topsoil, remote sensing technology could be used to predict the soil properties in layers; however, as the soil depth increased, the performance of the linear regression models gradually decreased.

Keywords: soil organic carbon; soil total nitrogen; multiple linear regression model; remote sensing technology; Qilian mountains

1 Introduction

Soil organic carbon and soil total nitrogen stocks are important indicators for evaluating soil health and stability (Cardoso et al., 2013). They play important roles in the functions of terrestrial ecosystems (Turner and Chapin, 2005), and they are important factors for global and regional food security, ecological restoration, and mitigation of global climate warming (Cardoso et al., 2013; Deb et al., 2015; Oldfield et al., 2019). Affected by soil-forming factors, the soil organic carbon and soil total nitrogen stocks of terrestrial ecosystems show spatial heterogeneity. This spatial heterogeneity affects vegetation patterns (Li et al., 2017; Portell et al., 2018) and the sustainability of regional terrestrial ecosystem functions (Turner and Chapin, 2005). Therefore, accurately predicting the spatial distribution of soil organic carbon and soil total nitrogen stocks is very important for understanding the carbon and nitrogen cycles and making regional ecosystem management decisions.

The spatial distribution of soil organic carbon and soil total nitrogen stocks is affected by soil-forming factors such as topography, vegetation, soil temperature, and soil parent material (Wang et al., 2012). However, in different regions, the factors affecting soil organic carbon and total nitrogen stocks are not consistent (Ruiz Sinoga et al., 2012; Mirchooli et al., 2020). The traditional methods of predicting soil organic carbon and total nitrogen stocks involve using a certain sampling method to investigate a large number of soil samples in the region, before using statistical analysis to predict the soil organic carbon and soil total nitrogen stocks in the entire region (Scull et al., 2003; Brevik et al., 2016). This method is time-consuming and expensive. Therefore, in recent years, low-cost and high-efficiency soil digital mapping based on remote sensing technology has received much attention (Mulder et al., 2011; Angelopoulou et al., 2019). There have been many prediction models that combine environmental variables calculated from remote sensing images with field sampling data to estimate soil organic carbon and soil total nitrogen stocks on continuous land surfaces (Gomez et al., 2008; Forkuor et al., 2017; Kalambukattu et al., 2018).

These models are based on the theory of soil genetics. At the end of the 19th century, Dakuchaev proposed that soil formation was controlled by climate, biological factors, soil-forming parent material, topography, and time (Florinsky, 2012; Johnson and Schaetzl, 2015). Following the further development of the theory, Jenny (Jenny, 1994) proposed a

*Corresponding author: Chao Wang (E-Mail: wangch-03@163.com)

general formula for soil development and tried to quantitatively understand the relative importance of soil-forming factors for soil development in a certain area. According to Jenny's formula, various climate functions and topographic functions have been developed to explain the relationship between the spatial distribution of soil properties and environmental factors such as climate (Wang et al., 2012; Luo et al., 2017), topography (Wang et al., 2012) and vegetation (MAEStre and Cortina, 2002; Ruiz Sinoga et al., 2012). These studies laid the foundation for digital soil mapping based on remote sensing technology. Estimation of soil organic carbon and soil total nitrogen stocks based on remote sensing technology is usually carried out to establish a quantitative relationship between field sampling data and easily available environmental variables, before using this quantitative relationship to predict the continuous spatial distribution of soil organic carbon and soil total nitrogen (Mulder et al., 2011; Brevik et al., 2016; Kalambukattu et al., 2018). A variety of soil organic carbon and total nitrogen stocks prediction models have been developed, such as linear models (Wang et al., 2020), Kriging models (Mondal et al., 2017; Bangroo et al., 2020), geographically weighted regression models (Wang et al., 2020), random forest models (Yang et al., 2016), and some machine learning models (Keskin et al., 2019; Lin et al., 2020).

With different research areas or different models, researchers often choose different remote sensing images and environmental variables. According to the published literatures, digital soil mapping based on hyperspectral remote sensing data shows the best prospects (Angelopoulou et al., 2019). However, in some areas with diverse vegetation types and complex topography, the spatial heterogeneity of soil is very large. For example, in alpine mountainous areas, due to the obvious vertical zoning of vegetation and large height differences, soil organic carbon and soil total nitrogen stocks show great differences in a very small range. In this case, the spatial resolution of hyperspectral remote sensing images (spatial resolution generally greater than 10 m) sometimes cannot meet the mapping requirements. However, some multispectral data with better spatial resolution (spatial resolution is usually at the submeter to meter level), such as Quickbird, WorldView, IKONOS, and Geoeye-1, are more advantageous in predicting soil organic carbon and total nitrogen stocks.

This paper selected the Qilian Mountains on the eastern edge of the Qinghai–Tibet Plateau in western China to carry out experiments. On the basis of field sampling data and environmental variables derived from remote sensing images, estimation of soil organic carbon and soil total nitrogen stocks in topsoil (0–30 cm) was carried out. This research focused on the following questions: (1) Can the linear regression model combining high-spatial-resolution multispectral remote sensing data and field sampling data estimate the spatial heterogeneity of soil properties in alpine mountainous areas? (2) What are the key environmental variables for estimating soil organic carbon and soil total nitrogen stocks in alpine mountainous areas? (3) Can the combination of remote sensing technology and field sampling data estimate soil organic carbon and soil total nitrogen stocks in layers?

2 Materials and Methods

2.1 Study Area

This study was conducted in a small watershed named Tianlaochi, which is in the Qilian Mountains in western China (Figure 1). It covers 12.8 km². The watershed has an altitude ranging from 2600 to 4450 m above sea level (asl). The Tianlaochi watershed has a cold and humid alpine climate with a mean annual precipitation of over 500 mm and a mean annual temperature of −0.6 to 2.0 °C. The land covers in the study area includes steppes, bush woods (made up of *Potentilla fruticosa* L., *Salix gilashanica* C., *Caragaha jubata*, and *Spiraea alpina* Pall), sabina forests (*S.przewalskii*), spruce forests (*P. crassifolia*), alpine meadows, and bare rocks. Most of the watershed is covered with forests and bush woods (about 65%). The main soil types are gray-cinnamon soils and meadow soils. The soil texture is mainly silty sand, and the soil parent material includes pelitic rock, conglomerate, and purple-red sand shale.

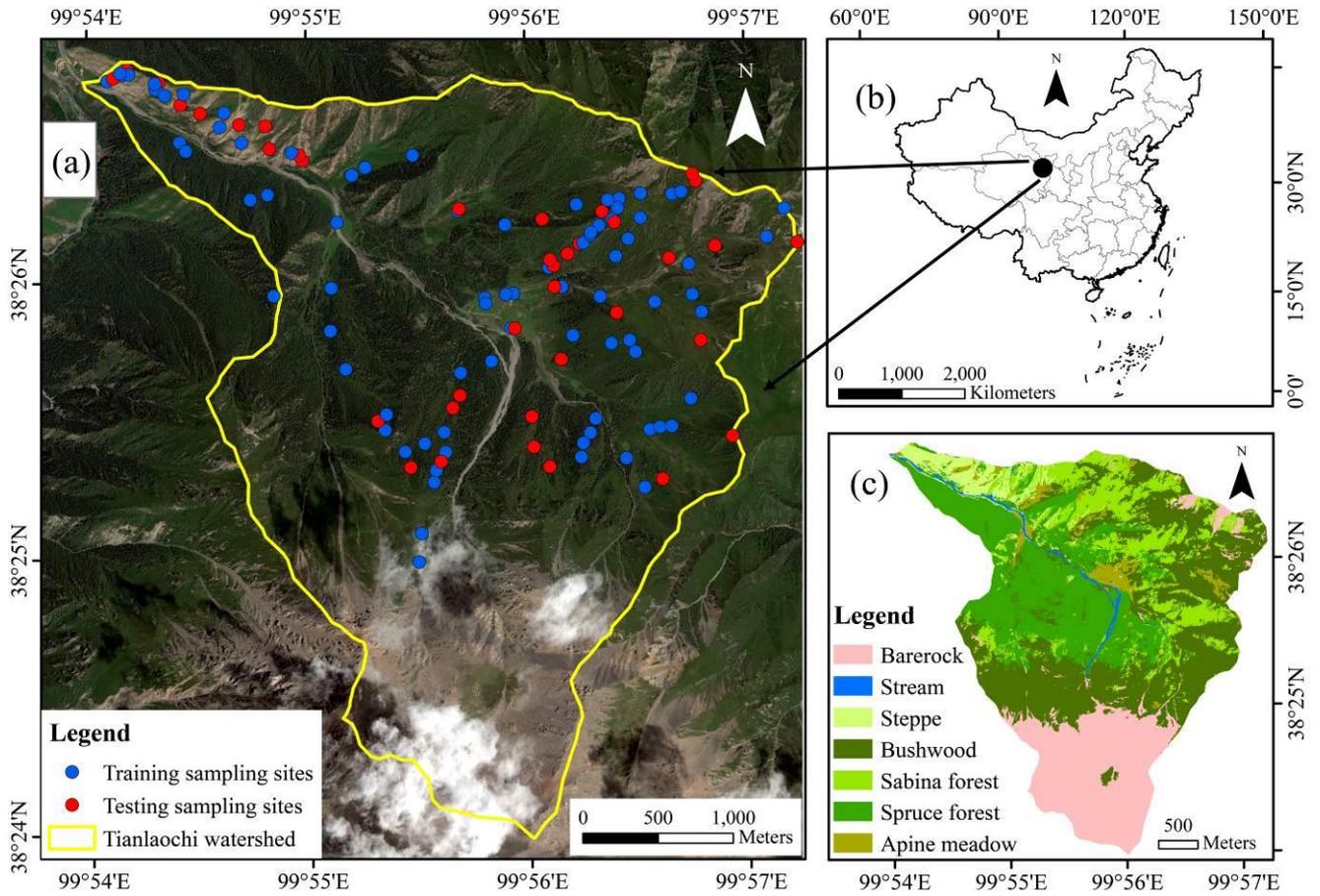


Fig. 1 Map of study area, where (a) is the spatial distribution of sampling sites; (b) is the location of study area; and (c) is the classification of the land covers.

2.2 Soil Sampling and Soil Analysis

A total of 119 sites with different vegetation types, slopes, aspects, and elevations were selected to carry out soil sampling at depths of 0–10 cm, 10–20 cm, and 20–30 cm in 2012. The sampling sites were located using a global positioning system (GPS) receiver. Soil samples were collected used from the soil layer after litter removal. They were kept separately in aluminum boxes, and two duplicate samples were collected from each site (Wang et al., 2013). The soil samples were air-dried at room temperature for 1 week and sieved through a 2 mm sieve to remove stones, roots, and coarse debris (Zheng et al., 2014). Soil passing the sieve was used to measure the soil organic carbon (SOC) and soil total nitrogen (STN) content. Soil organic carbon stock was measured using the Walkley and Black method, and soil total nitrogen stock was measured using the Kjeldahl method (Walkley and Black, 1934; De Vos et al., 2007; Sparks et al., 2020). Since almost all of the study area is bare rock above 3600 m asl, sampling and soil mapping in this study were carried out below 3600 m asl.

2.3 Predictor Variable Acquisition

The environmental variables selected in this study were all obtained using remote sensing technology. The normalized difference vegetation index (NDVI), modified soil-adjusted vegetation index (MSAVI), perpendicular vegetation index (PVI), and ratio vegetation index (RVI) were extracted from a Geoeye-1 image (2 m × 2 m) acquired on 4 August 2012. Topographic factors such as elevation, slope, aspect, topographic wetness index (TWI), and solar radiation (SR) were calculated from the digital elevation model (DEM) produced by light detection and ranging technology acquired on 12 August 2012. The data were stored in a raster format with a spatial resolution of 2 m × 2 m.

The NDVI is a classical indication of plant health used to monitor the changes in vegetation and is significantly correlated with vegetation type, health, biomass accumulation, and the leaf area index (Tucker, 1979; Bangroo et al., 2020), which is given as

$$NDVI = \frac{Band_{nir} - Band_{red}}{Band_{nir} + Band_{red}} \quad (1)$$

where $Band_{nir}$ and $Band_{red}$ stand for the spectral reflectance in the near-infrared and red regions, respectively (the same below). The NDVI values range from -1 to 1 (Bangroo et al., 2020).

The MSAVI is a modified version of the soil-adjusted vegetation index (SAVI), which uses the L function to increase the vegetation dynamic response and reduce the soil background variations (Qi et al., 1994). The MSAVI is given as

$$MSAVI = \frac{(\text{Band}_{nir} - \text{Band}_{red})(1 + L)}{(\text{Band}_{nir} + \text{Band}_{red} + L)} \quad (2)$$

where L stand for the soil brightness correction factor that can be calculated from the Band_{nir} , Band_{red} , and the slope of the soil line (Qi et al., 1994). The MSAVI values range from -1 to 1 .

The PVI is defined as the orthogonal (perpendicular) distance between a point representing the plant canopy and the soil line (Richardson and Wiegand, 1977). The PVI is better than the NDVI for leaf area index prediction and estimation of the green cover percentage of vegetation (Elvidge and Chen, 1995; Panda et al., 2010). The PVI is given as

$$PVI = \frac{1}{\sqrt{1 + a^2}} (\text{Band}_{nir} - a \times \text{Band}_{red} + b) \quad (3)$$

where a and b denote the slope and intercept of the soil line. The PVI values range from -1 to 1 .

The RVI is defined as the ratio of spectral reflectance in the red band to that in the near-infrared band, which is based on the principle that leaves absorb relatively more red than infrared light (Jordan, 1969; Xue and Su, 2017). The RVI is widely used for green biomass estimations and monitoring, especially in high-density vegetation coverage (Jordan, 1969; Xue and Su, 2017). The RVI values are greater than zero.

The TWI (topographic wetness index) is an important aspect of the hydrologic system model and provides indirect information on land cover and vegetation potential (Grabs et al., 2009; Bangroo et al., 2020). The TWI is defined as

$$TWI = \frac{\ln \alpha}{\tan \beta} \quad (4)$$

where α represents the catchment area per unit width orthogonal to the direction of the flow direction, and β refers to the slope (Bangroo et al., 2020). When the slope is 0, the equation is meaningless and cannot be calculated. In order to ensure normal operation of the equation, the minimum slope of the study area was mandated to be 2% (approximately 0.0057 degrees).

The TWI, slope, aspect, SR (solar radiation), and elevation were derived from the DEM using ArcGIS 10.7 software. All factors have demonstrated strong correlation with the soil organic carbon and soil total nitrogen (Prueger et al., 2005; Shank et al., 2011; Yang et al., 2016; Bangroo et al., 2020).

2.4 Prediction Models

2.4.1 Multiple Linear Regression Model

The multiple linear regression (MLR) model is usually used to study the relationship between a dependent variable and multiple independent variables. If the relationship between the two can be described in a linear form, a multiple linear model can be established for analysis. The MLR model assumes that there is a linear relationship between the predicted variable (dependent variables were SOC and STN in this article) and the predictor variables (independent variables were vegetation-related variables and terrain-related variables in this article) (Xie et al., 2021). The MLR model is given as

$$Y = \alpha + \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (5)$$

where Y is the predicted variable, x_i represents the predictor variables, α is the intercept, β_i is the regression coefficient, and ε is the regression residual.

2.4.2 Stepwise Multiple Linear Regression Model

The stepwise multiple linear regression (SMLR) model essentially establishes the optimal multiple linear regression equation by screening the independent variables. The basic idea of the stepwise multiple regression model is to introduce new variables into the regression model one by one. When a new variable is introduced, it is considered whether to eliminate the existing variables until no more new variables are introduced. After introducing an explanatory variable, an F -test is performed on the regression equation, and a t -test is performed on all the selected explanatory variables one by one. Therefore, in this case, the final model only included the important environmental variables that affected the spatial distribution of soil attributes (Wang et al., 2020). The equation of the SMLR model is similar to that of the MLR model.

For the fitted linear equation, the F -test is used to test its significance, and the t -test is used to test the significance of its independent variables. The descriptive statistics were implemented in IBM SPSS statistics 25 for Windows, whereas the MLR model and SMLR model were constructed in R software. All figures were produced in Sigmaplot 14.0 and ArcGIS 10.7 for Windows.

2.5 Model Validation

For the model validation, 38 out of 119 soil samples were randomly extracted from the input data to test the model's predictive accuracy using the data from the remaining 81 soil samples. The mean absolute error (MAE), root-mean-square error ($RMSE$), and coefficient of determination (R^2) were used to evaluate the model efficiency in estimating SOC and STN values. The three evaluation indices are given below (Wang et al., 2020):

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - M_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - M_i)^2} \quad (7)$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{M}_i)^2}{\sum_{i=1}^n (M_i - \bar{M}_i)^2} \quad (8)$$

where P_i and M_i are the estimated and measured values at site i , respectively, \bar{M}_i represents the mean measured values, and n is the number of samples.

3 Results

3.1 Descriptive Statistics

3.1.1 Soil Organic Carbon and Soil Total Nitrogen Stock Sampling Data Statistics

Table 1 lists the summary statistics of soil organic carbon (SOC) and soil total nitrogen (STN) stocks at different soil depths (0–10 cm, 10–20 cm, and 20–30 cm) at the sample sites. According to measurements of soil samples at the 119 sampling sites, the average SOC stocks at the depths of 0–10 cm, 10–20 cm, and 20–30 cm were 105.35 g/kg, 81.92 g/kg, and 63.25 g/kg, respectively. Under the generalized skew distribution with skew coefficients of 0.13, 0.39, and 0.56, the distribution of measured SOC stocks was immaterially skewed (skewness < 1); therefore, it can be assumed to follow a normal distribution. In addition, it can be seen from Table 1 that the training dataset and test dataset that were randomly selected from the overall sample showed good consistency with the statistical characteristics of the overall sample.

Table 1 Summary statistics of soil organic carbon (SOC) stocks (g/kg) and soil total nitrogen (STN) stocks (g/kg) at different soil depths at the sample sites.

Soil Properties	Training Dataset ($n = 81$)							Test Dataset ($n = 38$)							
	Max.	Min.	Mean	SD	CV	Skewness	Kurtosis	Max.	Min.	Mean	SD	CV	Skewness	Kurtosis	
SOC	0–10 cm	223.57	11.78	105.52	46.87	0.44	0.27	-0.23	204.34	10.55	104.99	56.39	0.54	-0.03	-0.94
	10–20 cm	191.38	7.78	83.27	42.82	0.51	0.44	-0.22	183.31	7.00	79.05	45.65	0.58	0.32	-0.33
	20–30 cm	170.96	1.09	64.87	39.06	0.60	0.61	-0.02	154.94	3.94	59.79	43.18	0.72	0.56	-0.58
STN	0–10 cm	8.24	0.97	4.15	1.84	0.44	0.37	-0.68	9.36	0.55	3.95	1.94	0.49	0.62	0.62
	10–20 cm	9.14	0.71	3.39	1.77	0.52	0.71	0.25	7.45	0.53	3.24	1.70	0.53	0.60	0.16
	20–30 cm	9.32	0.31	2.56	1.63	0.64	1.19	2.55	4.40	0.17	2.14	1.10	0.52	-0.02	-0.74

Note: Max., maximum; Min., minimum; SD, standard deviation; CV, coefficient of variation.

From the statistical results, the average STN stocks at the depths of 0–10 cm, 10–20 cm, and 20–30 cm were 4.09 g/kg, 3.34 g/kg, and 2.42 g/kg, respectively. The training dataset and the test dataset also showed good statistical characteristic similarity with the overall sample. The generalized skew distribution presented skew coefficients of 0.44 and 0.67 at the soil depths of 0–10 cm and 10–20 cm, respectively. However, the skew coefficient was 1.16 at the soil depth of 20–30 cm. In other words, the STN stocks in this soil layer did not present a normal distribution.

3.1.2 Correlations of Environmental Variables at Sample Sites

Table 2 lists the summary statistics of environmental variables at the sample sites, and Figure 2 shows the spatial distributions of environmental variables in the study area. According to the skewness coefficient results, except for NDVI, the other environmental variables showed normal distribution characteristics. It can be seen from Table 2 and Figure 2 that the different vegetation indices and terrain-related factors reflect the different characteristics of regional vegetation and terrain. Table 3 shows the correlation coefficients among the averaged SOC stocks, the averaged STN stocks, and the environmental variables. The SOC stocks were positively correlated with NDVI, RVI, and elevation and negatively correlated with solar radiation (SR) and aspect. The STN stocks were positively correlated with MSAVI, NDVI, RVI, and elevation and negatively correlated with SR, aspect, and slope. Moreover, there was multivariate collinearity among the environmental variables.

Table 2 Summary statistics of environmental variables calculated from Geoeye-1 and light detection and ranging (LiDAR) data at the sample sites.

Environment	Training Dataset ($n = 81$)							Test Dataset ($n = 38$)						
-------------	-------------------------------	--	--	--	--	--	--	---------------------------	--	--	--	--	--	--

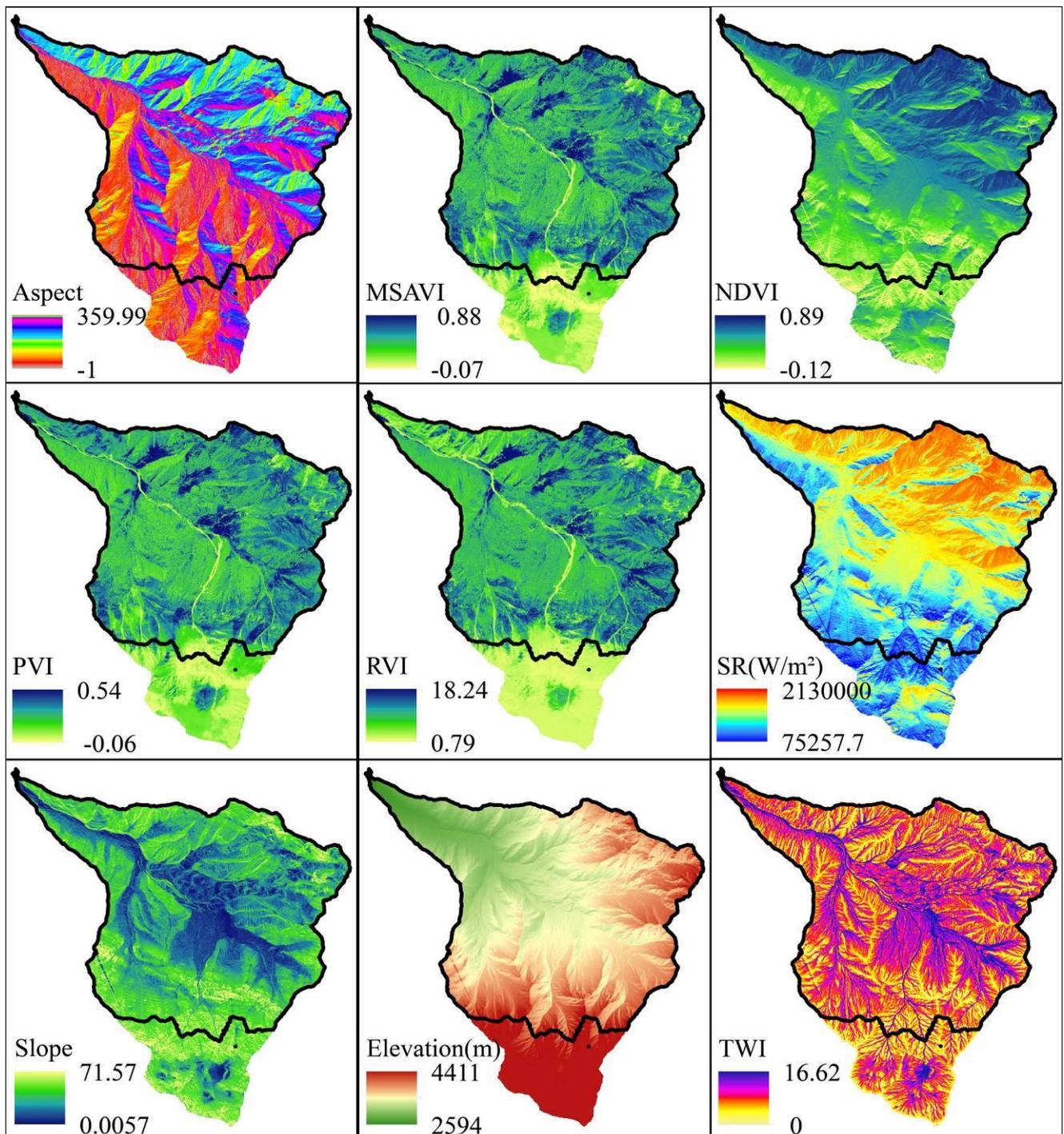
Variables	Max.	Min.	Mean	SD	CV	Skewness	Kurtosis	Max.	Min.	Mean	SD	CV	Skewness	Kurtosis
MSAVI	0.78	0.03	0.41	0.13	0.31	0.21	0.59	0.70	0.21	0.40	0.12	0.30	0.74	-0.23
NDVI	0.84	0.06	0.64	0.12	0.18	-2.09	7.56	0.78	0.38	0.63	0.10	0.16	-0.92	0.78
PVI	0.38	0.00	0.16	0.06	0.38	0.80	2.06	0.34	0.07	0.16	0.06	0.38	1.05	0.94
RVI	11.84	1.13	5.20	1.70	0.33	0.65	2.15	8.32	2.25	4.83	1.47	0.30	0.40	0.01
SR	1951110	884379	1554,530	244717	0.16	-0.86	0.22	1948700	909202	1596233	258557	0.16	-0.95	0.08
Aspect	349.04	18.19	218.10	79.99	0.37	-0.97	0.38	343.92	15.01	197.85	76.81	0.39	-0.50	0.48
Slope	46.38	6.57	25.92	9.94	0.38	-0.05	-1.00	55.21	4.75	27.93	10.91	0.39	0.13	-0.31
Elevation	3549.48	2642.35	3095.95	216.60	0.07	-0.31	-0.72	3507.32	2650.18	3099.19	232.46	0.08	-0.16	-0.79
TWI	3.95	0.32	1.65	0.72	0.44	0.71	0.37	3.49	0.20	1.62	0.76	0.47	0.27	0.04

Note: MSAVI, modified soil-adjusted vegetation index; NDVI, normalized difference vegetation index; PVI, perpendicular vegetation index; RVI, ratio of vegetation index; SR, solar radiation (W/m^2); TWI, topographic wetness index; Max., maximum; Min., minimum; SD, standard deviation; CV, coefficient of variation.

Table 3 Pearson's correlation coefficients among the averaged SOC and averaged STN stocks (g/kg) at the soil depths of 0–10 cm, 10–20 cm, and 20–30 cm and environmental variables for all samples (119 samples).

	SOC	STN	MSAVI	NDVI	PVI	RVI	SR	Aspect	Slope	Elevation
STN	0.76 **									
MSAVI	-0.05	0.23 *								
NDVI	0.35 **	0.43 **	0.73 **							
PVI	-0.14	0.15	0.98 **	0.62 **						
RVI	0.29 **	0.43 **	0.85 **	0.90 **	0.78 **					
SR	-0.55 **	-0.26 **	0.43 **	0.08	0.44 **	0.15				
Aspect	-0.32 **	-0.27 **	0.18	0.17	0.15	0.17	0.19 *			
Slope	-0.13	-0.19 *	-0.29 **	-0.30 **	-0.27 **	-0.39 **	-0.12	-0.14		
Elevation	0.56 **	0.56 **	0.15	0.43 **	0.04	0.35 **	-0.03	0.01	-0.06	
TWI	0.05	0.09	0.21 *	0.17	0.21 *	0.22 *	0.08	0.11	-0.55 **	-0.11

Note: MSAVI, modified soil-adjusted vegetation index; NDVI, normalized difference vegetation index; PVI, perpendicular vegetation index; RVI, ratio of vegetation index; SR, solar radiation (W/m^2); TWI, topographic wetness index; **, $p < 0.01$; *, $p < 0.05$.



Legend
 The area below 3600m asl
 0 0.5 1 2 3 Kilometers
 N

Fig.2 The spatial distributions of environmental variables in the Tianlaochi watershed.

3.2 Model Evaluation and Validation

3.2.1 Model Evaluation and Significance Testing of Parameters

Table 4 lists the results of significance testing for estimated parameters of the fitted SOC stock regression equations using the MLR model and SMLR model. The fitting equations of the two models in different soil layers all passed the F -test ($p < 0.001$). For the MLR model, in addition to MSAVI, NDVI, and slope, the remaining environmental variables passed the t -test ($p < 0.05$) when estimating the topsoil, and with the increase in soil depth, PVI, RVI, and TWI gradually became insignificant. For the SMLR model, the insignificant environmental variables (MSAVI, NDVI, and slope) in the MLR model were eliminated step by step. Similarly, as the soil depths increased, the significance level of PVI and TWI decreased.

Table 4 The coefficients of fitted equations and significance testing of the multiple linear regression model (MLR) and stepwise multiple linear regression (SMLR) model for soil organic carbon stocks.

Variables	Coefficient of the Fitted Equation by MLR			Coefficient of the Fitted Equation by SMLR		
	0–10 cm	10–20 cm	20–30 cm	0–10 cm	10–20 cm	20–30 cm
Constant	-23.87	-42.01	-71.70	-46.65	-74.32	-75.33
MSAVI	298.62	289.83	83.80			
NDVI	-49.49	-63.77	-38.59			
PVI	-861.30 *	-722.65*	-338.75	-367.07 ***	-244.35 *	-191.76*
RVI	17.97 **	12.51 *	10.95*	19.04 ***	12.80 ***	9.37 *
SR	-7.66×10^{-5} ***	-7.34×10^{-5} ***	-5.89×10^{-5} ***	-7.08×10^{-5} ***	-6.68×10^{-5} ***	-5.75×10^{-5} ***
Aspect	-0.16 ***	-0.16 ***	-0.12 **	-0.15 ***	-0.16 ***	-0.12 **
Slope	0.04	-0.10	0.17			
Elevation	0.07 ***	0.08 ***	0.07 ***	0.08 ***	0.08 ***	0.07 ***
TWI	11.92 *	7.88	11.12 *	11.91 *	8.99 *	10.16 *
<i>F</i> -test value	16.93 ***	14.53 ***	8.78 ***	25.46 ***	21.80 ***	13.56 ***

Note: The *F*-test was used for the regression equations, and the *t*-test was used for the regression coefficients; MSAVI, modified soil-adjusted vegetation index; NDVI, normalized difference vegetation index; PVI, perpendicular vegetation index; RVI, ratio of vegetation index; SR, solar radiation (W/m²); TWI, topographic wetness index; ***, $p < 0.001$, **, $p < 0.01$; *, $p < 0.05$, •, $p < 0.1$.

Table 5 lists the results of significance testing for estimated parameters of the fitted STN stock regression equations using the MLR model and SMLR model. The fitting equations of the two models in different soil layers also passed the *F*-test ($p < 0.001$). It should be noted that the significance test results for environmental variables in the STN estimation model were different from those in the SOC model. For the MLR model, MSAVI, PVI, SR, aspect, and elevation passed the *t*-test ($p < 0.05$) when estimating the topsoil, and PVI became insignificant in the 20–30 cm soil layer. For the SMLR model, NDVI, RVI, and slope were eliminated, RVI, slope, and TWI were eliminated, and RVI and slope were eliminated in the 0–10 cm, 10–20 cm, and 20–30 cm soil layers, respectively. In other words, in different soil layers, different environmental variables in the SMLR model were eliminated step by step.

Table 5 The coefficient of fitted equations and significance testing of the multiple linear regression (MLR) model and stepwise multiple linear regression (SMLR) model for soil total nitrogen stocks.

Variables	Coefficient of the Fitted Equation by MLR model			Coefficient of the Fitted Equation by SMLR model		
	0–10 cm	10–20 cm	20–30 cm	0–10 cm	10–20 cm	20–30 cm
Constant	-4.18	-5.85	-6.20	-6.35	-6.80	-6.13
MSAVI	23.68 *	25.12 *	18.27 •	19.60 **	26.19 **	18.87 •
NDVI	-4.18	-6.15 •	-4.87		-4.23	-4.03
PVI	-39.87 *	-38.86 *	-25.40	-30.52 *	-37.11 *	-24.84
RVI	0.29	0.24	0.13			
SR	-2.2×10^{-6} **	-1.71×10^{-6} *	-1.7×10^{-6} *	-2.14×10^{-6} **	-1.8×10^{-6} *	-1.85×10^{-6} *
Aspect	-8.86×10^{-3} ***	-7.65×10^{-3} ***	-6.76×10^{-3} ***	-8.89×10^{-3} ***	-7.54×10^{-3} ***	-6.79×10^{-3} ***
Slope	-1.65×10^{-2}	-1.26×10^{-2}	6.92×10^{-3}			
Elevation	3.74×10^{-3} ***	4.06×10^{-3} ***	3.62×10^{-3} ***	3.84×10^{-3} ***	4.09×10^{-3} ***	3.66×10^{-3} ***
TWI	0.34	0.06	0.42	0.51 *		0.38 •
<i>F</i> -test value	11.70 ***	9.31 ***	6.75 ***	17.01 ***	13.65 ***	8.83 ***

Note: The *F*-test was used for the regression equations, and the *t*-test was used for the regression coefficients; MSAVI, modified soil-adjusted vegetation index; NDVI, normalized difference vegetation index; PVI, perpendicular vegetation index; RVI, ratio of vegetation index; SR, solar radiation (W/m²); TWI, topographic wetness index; ***, $p < 0.001$, **, $p < 0.01$; *, $p < 0.05$, •, $p < 0.1$.

3.2.2 Model for Validation

The results of MLR and SMLR models in the estimations of SOC stocks, as well as their performance assessment, are presented in Fig. 3 and Fig. 4, including R^2 , *RMSE*, and *MAE*. The MLR model estimated the SOC stocks in the 0–10 cm soil layer with $R^2 = 0.68$, *RMSE* = 26.26 g/kg, and *MAE* = 21.79 g/kg for the training dataset, $R^2 = 0.81$, *RMSE* = 25.09 g/kg, and *MAE* = 18.81 g/kg for the testing dataset, and $R^2 = 0.73$, *RMSE* = 25.09 g/kg, and *MAE* = 18.81 g/kg for the total dataset (Table 6 and Fig. 3a). However, with the increase in soil depth, the estimation ability of the MLR model for SOC stocks decreased gradually. The MLR model estimated the SOC stocks in the 20–30 cm soil layer with $R^2 = 0.53$, *RMSE* = 26.71 g/kg, and *MAE* = 20.22 g/kg for the training dataset, $R^2 = 0.74$, *RMSE* = 21.81 g/kg, and *MAE* = 16.14 g/kg for the testing dataset, and $R^2 = 0.60$, *RMSE* = 25.40 g/kg, and *MAE* = 18.92 g/kg for the total dataset (Table 6 and Fig. 3b). The results suggest that the SMLR model and MLR model had a similar

estimation ability for the training dataset, testing dataset, and total dataset (Fig. 3).

Table 6 The coefficient of determination (R^2) of the multiple linear regression (MLR) model and stepwise multiple linear regression (SMLR) model for soil organic carbon stocks.

Dataset	Coefficient of determination of MLR model			Coefficient of determination of SMLR model		
	0–10 cm	10–20 cm	20–30 cm	0–10 cm	10–20 cm	20–30 cm
Total Dataset	0.73	0.71	0.60	0.73	0.70	0.60
Train Dataset	0.68	0.65	0.53	0.67	0.64	0.52
Test Dataset	0.81	0.82	0.74	0.83	0.82	0.73

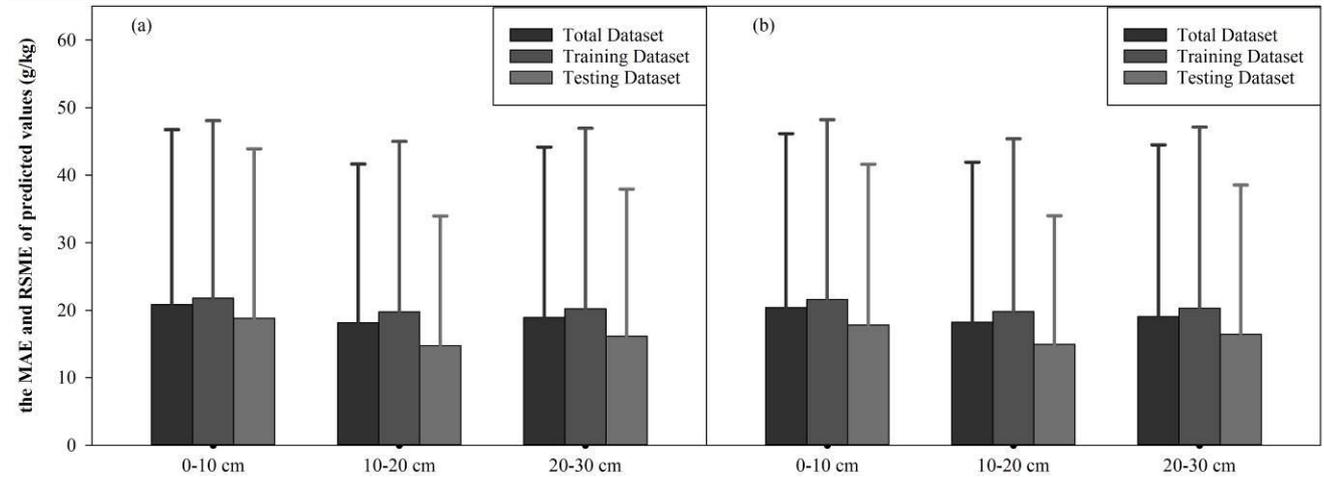


Fig. 3 Performance of the MLR model (a) and SMLR model (b) for predicting SOC stocks in the 0–10 cm, 10–20 cm, and 20–30 cm soil layers.

The results of MLR and SMLR models in the estimations of STN stocks, as well as their performance assessment, are presented in Table 7 and Fig. 4. Simulation results showed that the ability of these two models to estimate STN stocks was relatively poorer than that to estimate SOC stocks. The MLR model estimated the STN stocks in the 0–10 cm soil layer with $R^2 = 0.60$, $RMSE = 1.16$ g/kg, and $MAE = 0.94$ g/kg for the training dataset, $R^2 = 0.52$, $RMSE = 1.38$ g/kg, and $MAE = 1.05$ g/kg for the testing dataset, and $R^2 = 0.57$, $RMSE = 1.23$ g/kg, and $MAE = 0.97$ g/kg for the total dataset (Table 7 and Fig. 4a). Similar to the simulation results of SOC stocks, the estimation ability of the MLR model for STN stocks decreased as the soil depth increased. The MLR model estimated the STN stocks in the 20–30 cm soil layer with $R^2 = 0.46$, $RMSE = 1.19$ g/kg, and $MAE = 0.86$ g/kg for the training dataset, $R^2 = 0.39$, $RMSE = 1.14$ g/kg, and $MAE = 0.93$ g/kg for the testing dataset, and $R^2 = 0.41$, $RMSE = 1.17$ g/kg, and $MAE = 0.88$ g/kg for the total dataset (Table 7 and Fig. 4b). The results suggest that the SMLR model and MLR model had a similar estimation ability for the training dataset, testing dataset, and total dataset (Fig. 4).

Table 7 The coefficient of determination (R^2) of the multiple linear regression (MLR) model and stepwise multiple linear regression (SMLR) model for soil total nitrogen stocks.

Dataset	Coefficient of determination of MLR model			Coefficient of determination of SMLR model		
	0–10 cm	10–20 cm	20–30 cm	0–10 cm	10–20 cm	20–30 cm
Total Dataset	0.57	0.50	0.41	0.54	0.50	0.40
Train Dataset	0.60	0.54	0.46	0.58	0.53	0.46
Test Dataset	0.52	0.44	0.39	0.48	0.39	0.38

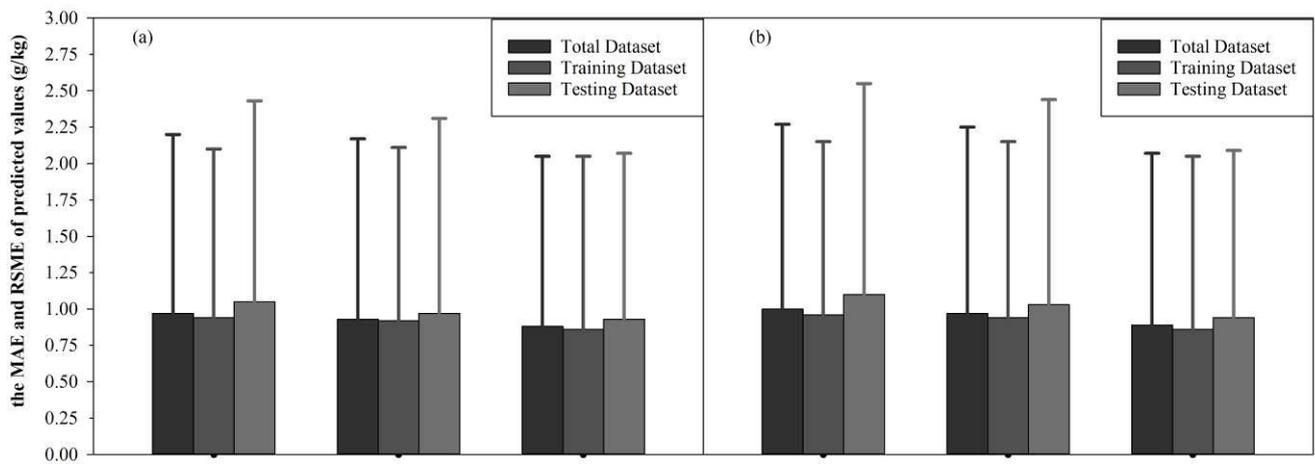
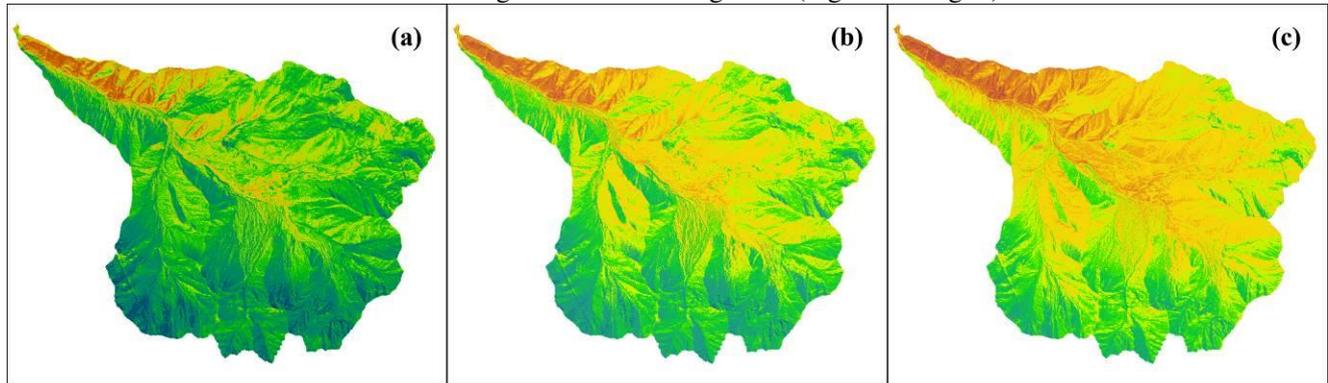


Fig. 4 Performance of the MLR model (a) and SMLR model (b) for predicting STN stocks in the 0–10 cm, 10–20 cm, and 20–30 cm soil layers.

3.3 the spatial distribution of Soil Organic Carbon and Soil Total Nitrogen Stocks

The model verification results showed that the simulation performance of the MLR model and the SMLR model was close. Therefore, the SMLR model with fewer predictors was selected as the final model to predict the spatial distribution of SOC and STN stocks at soil depths of 0–10 cm, 10–20 cm, and 20–30 cm. The higher values of SOC and STN stocks of each soil depth were located on the shady slope at high altitude (mainly covered by spruce forests and bush woods), and the lower values were located on the sunny slope at low altitude (mainly covered by steppes and sabina forests). In addition, the model simulation results showed that, with the increase in soil depth, the contents of SOC and STN both showed a significant decreasing trend (Fig. 5 and Fig. 6).



Legend

SOC (g/kg)

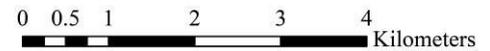
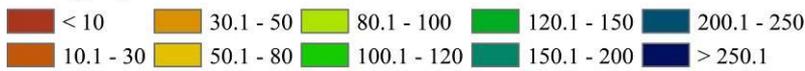
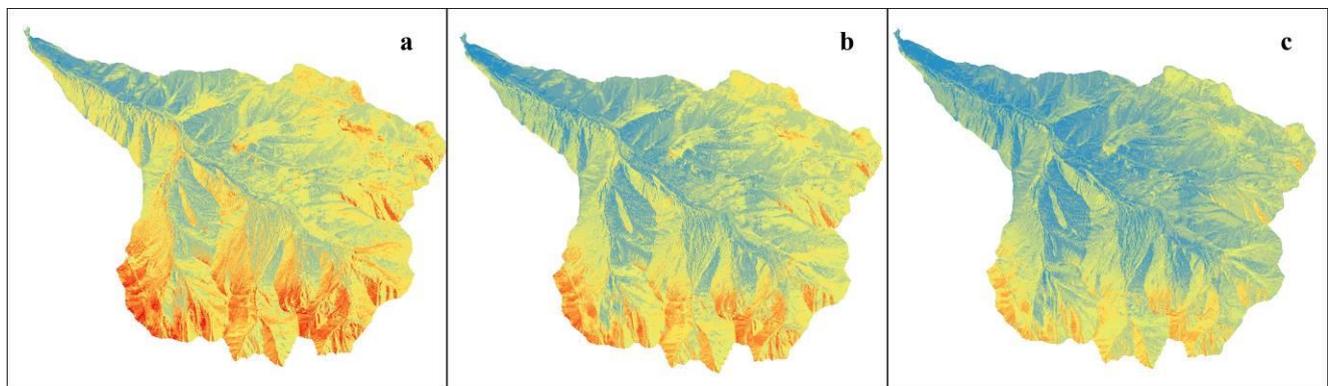


Fig. 5 Spatial distribution of the predicted SOC stocks calculated from SMLR models in the 0–10 cm (a), 10–20 cm (b), and 20–30 cm (c) soil layers.



Legend

STN (g/kg)



Fig. 6 Spatial distribution of the predicted STN stocks calculated from SMLR models in the 0–10 cm (a), 10–20 cm (b), and 20–30 cm (c) soil layers.

4 Discussion

4.1 Choices of Environmental Variables of Remote Sensing for SOC and STN Stocks Estimations

Using remote sensing technology to obtain various environmental variables combined with field sampling soil properties to estimate soil properties in an area is one of the main methods to obtain the spatial distribution of soil properties quickly and cost-effectively. Many studies have demonstrated the effectiveness of this estimation method (Yang et al., 2016; Wang et al., 2020). Although different researches used different modeling methods such as linear models (Razakamanarivo et al., 2011; Xie et al., 2021), nonlinear models (Mondal et al., 2017), and machine learning (Lin et al., 2020; Zhou et al., 2020), the selection of predictor variables was the common foundation of all such methods (Mirchooli et al., 2020). The selection of these predictors can be roughly divided into terrain-related predictors and vegetation-related predictors (Mirchooli et al., 2020; Wang et al., 2020). The terrain-related variables generally affect the SOC and STN stocks by changing the spatial combination of water and heat conditions and the accumulation and transport of soil materials (Keskin et al., 2019). In this study, solar radiation (SR), topographic wetness index (TWI), elevation, aspect, and slope were selected to explore the relationship between topographic characteristics and SOC and STN stocks. The results of the study indicated that aspect, elevation, and solar radiation were the main topographical factors affecting SOC and STN stocks. Vegetation-related variables are mainly used to explore the influence of different vegetation growth status on the spatial distribution of SOC and STN stocks. The vegetation growth status can be obtained by calculating the vegetation index through the spectral information of different bands. In our research, four vegetation indices, NDVI, SMAVI, PVI, and RVI, were selected to reflect the vegetation growth status. The results of the study found that SMAVI and PVI had more significant effects on the spatial distribution of SOC and STN stocks. However, the importance of predictive variables may vary if the study area is different or the prediction model is different. Yang et al.(2016) concluded that Landsat Thematic Mapper (TM) band 3, mean annual precipitation, NDVI, aspect, and elevation were the five most important predictors by using boosted regression tree and random forest models to map the distribution of SOC stocks in Qilian Mountains in western China. Wang et al.(2020) found that NDVI and SMAVI were the key environmental variables to estimate the SOC and STN stocks of topsoil in the forest-dominated areas in northeastern China, and the green band reflectance, enhanced vegetation index (EVI), renormalization difference vegetation index (RDVI), and difference vegetation index (DVI) also significantly affected the spatial variability of the SOC and STN stocks. Bangroo et al.(2020) concluded that elevation, slope, and NDVI were the major factors driving spatial variation in SOC and STN stocks using regression kriging in the North Kashmir forest of the Himalayas. Gholizadeh et al.(2018) found that the soil-adjusted total vegetation index (SATVI) and green normalized difference vegetation index (GNDVI) provided the strongest correlations with SOC stocks at four agricultural sites in the Czech Republic. The reason for these differences may be different soil-forming factors affecting soil development in different regions.

4.2 Choices of Spatial Resolution for Remote Sensing with Respect to SOC and STN Stock Estimations

According to the literature, using remote sensing data with different spatial resolutions can obtain different results with respect to estimations of SOC and STN stocks. Therefore, the spatial resolution of remote sensing data restricts the accuracy of SOC and STN stock estimations when based on remote sensing technology. Remote sensing data with high spatial resolution may obtain a more detailed mapping of soil properties. However, a higher spatial resolution results in a higher cost of acquiring remote sensing data and field sampling data. Therefore, when selecting the spatial resolution of remote sensing data for mapping of soil properties, we should focus on whether the soil properties in the study area have high spatial heterogeneity. In areas with fewer vegetation types and a simple topography (Keskin et al., 2019), remote sensing images with coarser spatial resolution can be used, whereas, in mountainous areas with diverse vegetation types and complex topography, remote sensing data with a higher spatial resolution should be selected for SOC and STN stock mapping (Gholizadeh et al., 2018; Wang et al., 2020). For example, although the research area of this project was only 12.8 km², its diverse vegetation types, large elevation ranges, and obvious topographical undulations led to the SOC and STN stocks having high spatial heterogeneity. Obviously, this spatial heterogeneity could not be described using coarse-resolution remote sensing data. In this case, higher-spatial-resolution remote sensing data (such as 2 m × 2 m) should be used to obtain more accurate soil property maps.

5 Conclusions

In this study, using a multiple linear regression model and stepwise multiple linear regression model, we estimated soil organic carbon and soil total nitrogen stocks in the Qilian Mountains on the northeastern edge of the Qinghai–Tibet Plateau, western China. The main purpose was to evaluate the potential of a linear regression model based on the combination of high-spatial-resolution multispectral remote sensing data and field sampling data for digital soil mapping in alpine mountainous areas, identify the key environmental variables for soil organic carbon

and soil total nitrogen stock estimation, and analyze the feasibility of stratified prediction for topsoil (0–30 cm). The results indicated that both models showed good performance. Thus, the combination of high-spatial-resolution multispectral remote sensing data and field sampling data is sufficient for digital soil mapping in alpine mountainous areas. In this study area, MSAVI, PVI, aspect, elevation, and solar radiation were the key environmental variables affecting the spatial distribution of soil organic carbon and soil total nitrogen. In topsoil, remote sensing technology could be used to predict the soil properties in layers; however, as the soil depth increased, the performance of the linear regression model gradually decreased. As the soil properties of topsoil often vary greatly with depth, future research should focus on the use of remote sensing technology to estimate soil properties in layers.

Acknowledgments: The authors are grateful to Huanghua Peng, Yang Wang, Xianglin Zheng, YiYue Liu, Wenying Ma, Bei Liang and Wenjuan Li for help during field and laboratory works.

Author Contributions: Conceptualization, Chao Wang and Chuanyan Zhao; Data curation, Chao Wang; Formal analysis, Chao Wang and Kaiming Li; Funding acquisition, Chuanyan Zhao; Investigation, Chao Wang and Shouzhang Peng; Methodology, Kaiming Li; Project administration, Chuanyan Zhao; Validation, Kaiming Li; Writing – original draft, Chao Wang; Writing – review & editing, Shouzhang Peng and Ying Wang. All authors read and agreed to the published version of the manuscript.

Funding: This research was funded by the Doctoral Research Fund of Lanzhou City University (Grant No. LZCU-BS2019-41), the Creative Research Groups of Gansu Province (Grant No. 20JR5RA121), and the National Natural Science Foundation of China (Grant No. 91025015).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

- Angelopoulou T, Tziolas N, Balafoutis A, et al. Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing*, 2019, 11(6): 676.
- Bangroo S A, Najjar G R, Achin E, et al. Application of predictor variables in spatial quantification of soil organic carbon and total nitrogen using regression kriging in the north kashmir forest himalayas. *CATENA*, 2020, 193: 104632.
- Brevik E C, Calzolari C, Miller B A, et al. Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma*, 2016, 264: 256-274.
- Cardoso E J B N, Vasconcellos R L F, Bini D, et al. Soil health: Looking for suitable indicators. What should be considered to assess the effects of use and management on soil health? *Scientia Agricola*, 2013, 70(4): 274-289.
- De Vos B, Lettens S, Muys B, et al. Walkley–black analysis of forest soil organic carbon: Recovery, limitations and uncertainty. *Soil Use and Management*, 2007, 23(3): 221-229.
- Deb S, Bhadoria P B S, Mandal B, et al. Soil organic carbon: Towards better soil health, productivity and climate change mitigation. *Climate change and environmental sustainability*, 2015, 3(1): 26-34.
- Elvidge C D, Chen Z. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, 1995, 54(1): 38-48.
- Florinsky I V. The dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). *Eurasian Soil Science*, 2012, 45(4): 445-451.
- Forkuor G, Hounkpatin O K, Welp G, et al. High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: A comparison of machine learning and multiple linear regression models. *PloS one*, 2017, 12(1): e0170478.
- Gholizadeh A, Žižala D, Saberioon M, et al. Soil organic carbon and texture retrieving and mapping using proximal, airborne and sentinel-2 spectral imaging. *Remote Sensing of Environment*, 2018, 218: 89-103.
- Gomez C, Viscarra Rossel R A, Mcbratney A B. Soil organic carbon prediction by hyperspectral remote sensing and field vis-nir spectroscopy: An australian case study. *Geoderma*, 2008, 146(3): 403-411.
- Grabs T, Seibert J, Bishop K, et al. Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. *Journal of Hydrology*, 2009, 373(1-2): 15-23.
- Jenny H. *Factors of soil formation: A system of quantitative pedology*. Chelmsford, MA, USA: Courier Corporation, 1994.
- Johnson D L, Schaeztl R J. Differing views of soil and pedogenesis by two masters: Darwin and dokuchaev. *Geoderma*, 2015, 237-238: 176-189.

- Jordan C F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 1969, 50(4): 663-666.
- Kalambukattu J G, Kumar S, Raj R A. Digital soil mapping in a himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. *Environmental earth sciences*, 2018, 77(5): 1-14.
- Keskin H, Grunwald S, Harris W G. Digital mapping of soil carbon fractions with machine learning. *Geoderma*, 2019, 339: 40-58.
- Li H, Peng J, Yanxu L, et al. Urbanization impact on landscape patterns in beijing city, china: A spatial heterogeneity perspective. *Ecological Indicators*, 2017, 82: 50-60.
- Lin L, Gao Z, Liu X. Estimation of soil total nitrogen using the synthetic color learning machine (sclm) method and hyperspectral data. *Geoderma*, 2020, 380: 114664.
- Luo Z, Feng W, Luo Y, et al. Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions. *Global Change Biology*, 2017, 23(10): 4430-4439.
- MAEstre F T, Cortina J. Spatial patterns of surface soil properties and vegetation in a mediterranean semi-arid steppe. *Plant and soil*, 2002, 241(2): 279-291.
- Mirchooli F, Kiani-Harchegani M, Khaledi Darvishan A, et al. Spatial distribution dependency of soil organic carbon content to important environmental variables. *Ecological Indicators*, 2020, 116: 106473.
- Mondal A, Khare D, Kundu S, et al. Spatial soil organic carbon (soc) prediction by regression kriging using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Science*, 2017, 20(1): 61-70.
- Mulder V, De Bruin S, Schaepman M E, et al. The use of remote sensing in soil and terrain mapping—a review. *Geoderma*, 2011, 162(1-2): 1-19.
- Oldfield E E, Bradford M A, Wood S A. Global meta-analysis of the relationship between soil organic matter and crop yields. *Soil*, 2019, 5(1): 15-32.
- Panda S S, Ames D P, Panigrahi S. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2010, 2(3): 673-696.
- Portell X, Pot V, Garnier P, et al. Microscale heterogeneity of the spatial distribution of organic matter can promote bacterial biodiversity in soils: Insights from computer simulations. *Frontiers in microbiology*, 2018, 9: 1583.
- Prueger J H, Gish T J, McConnell L L, et al. Solar radiation, relative humidity, and soil water effects on metolachlor volatilization. *Environmental science & technology*, 2005, 39(14): 5219-5226.
- Qi J, Chehbouni A, Huete A R, et al. A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 1994, 48(2): 119-126.
- Razakamanarivo R H, Grinand C, Razafindrakoto M A, et al. Mapping organic carbon stocks in eucalyptus plantations of the central highlands of madagascar: A multiple regression approach. *Geoderma*, 2011, 162(3): 335-346.
- Richardson A J, Wiegand C. Distinguishing vegetation from soil background information. *Photogrammetric engineering and remote sensing*, 1977, 43(12): 1541-1552.
- Ruiz Sinoga J D, Pariente S, Diaz A R, et al. Variability of relationships between soil organic carbon and some soil properties in mediterranean rangelands under different climatic conditions (south of spain). *CATENA*, 2012, 94: 17-25.
- Scull P, Franklin J, Chadwick O A, et al. Predictive soil mapping: A review. *Progress in Physical Geography: Earth and Environment*, 2003, 27(2): 171-197.
- Shank G C, Evans A, Yamashita Y, et al. Solar radiation-enhanced dissolution of particulate organic matter from coastal marine sediments. *Limnology and Oceanography*, 2011, 56(2): 577-588.
- Sparks D L, Page A, Helmke P, et al. *Methods of soil analysis, part 3: Chemical methods*. Hoboken, NJ, USA: John Wiley & Sons, 2020.
- Tucker C J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 1979, 8(2): 127-150.
- Turner M G, Chapin F S. Causes and consequences of spatial heterogeneity in ecosystem function. *Ecosystem function in heterogeneous landscapes*. Cham, Switzerland; Springer. 2005: 9-30.
- Walkley A, Black I A. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil science*, 1934, 37(1): 29-38.
- Wang C, Zhao C, Xu Z, et al. Effect of vegetation on soil water retention and storage in a semi-arid alpine forest catchment. *Journal of Arid Land*, 2013, 5(2): 207-219.
- Wang S, Wang X, Ouyang Z. Effects of land use, climate, topography and soil properties on regional soil organic carbon and total nitrogen in the upstream watershed of miyun reservoir, north china. *Journal of Environmental Sciences*, 2012, 24(3): 387-395.
- Wang S, Zhuang Q, Jin X, et al. Predicting soil organic carbon and soil nitrogen stocks in topsoil of forest ecosystems in northeastern china using remote sensing data. *Remote Sensing*, 2020, 12(7): 1115.
- Xie X, Wu T, Zhu M, et al. Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. *Ecological Indicators*, 2021, 120: 106925.
- Xue J, Su B. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017, 2017: 1353691.
- Yang R-M, Zhang G-L, Liu F, et al. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*, 2016, 60: 870-878.
- Zheng X, Zhao C, Peng S, et al. Soil co₂ efflux along an elevation gradient in qinghai spruce forests in the upper reaches of the heihe river, northwest china. *Environmental earth sciences*, 2014, 71(5): 2065-2076.
- Zhou T, Geng Y, Chen J, et al. Mapping soil organic carbon content using multi-source remote sensing variables in the heihe river basin in china. *Ecological Indicators*, 2020, 114: 106288.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [sampleall.xlsx](#)