



Selecting environmental factors to predict spatial distribution of soil organic carbon stocks, northwestern Iran

Kamal Khosravi Aqdam · Nafiseh Yaghmaeian Mahabadi · Hassan Ramezani · Salar Rezapour · Zohreh Mosleh

Received: 6 June 2021 / Accepted: 30 September 2021 / Published online: 12 October 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract Knowledge of environmental factors controlling soil organic carbon (SOC) stocks can help predict spatial distribution SOC stocks. So, this study was carried out to select the best environmental factors to model and estimate the spatial distribution of SOC stocks in northwestern Iran. Soil sampling was performed at 210 points by multiple conditioned Latin Hypercube method (cLHm) and SOC stocks were measured. Also, environmental factors, including terrain attributes, moisture index, and normalized difference vegetation index (NDVI), were calculated. SOC stocks were modeled using random forest (RF) and partial least squares regression (PLSR) models. Modeling SOC stocks by RF model showed that the efficient factors for estimating the SOC stocks were slope height (slph), terrain surface texture (texture),

standardized height (standh), elevation, relative slope position (rsp), and normalized height (normalh). Also, the PLSR model selected standardized height (standh), relative slope position (rsp), slope, and channel network base level (chnl base) to model SOC stocks. In both RF and PLSR methods, the standh and rsp factors were suitable parameters for estimating the SOC stocks. Predicting the spatial distribution of SOC stocks using environmental factors showed that the R^2 values for RF and PLSR models were 0.81 and 0.40, respectively. The result of this study showed that in areas with complex land features, terrain attributes can be good predictors for estimating SOC stocks. These predictors allow more accurate estimates of SOC stocks and contribute considerably to the effective application of land management strategies in arid and semiarid area.

K. Khosravi Aqdam (✉) · N. Yaghmaeian Mahabadi (✉) · H. Ramezani
Department of Soil Science, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran
e-mail: k.khosraviaqdam@gmail.com

N. Yaghmaeian Mahabadi
e-mail: yaghmaeian_na@guilan.ac.ir

S. Rezapour
Department of Soil Science, Faculty of Agriculture, Urmia University, Urmia, Iran

Z. Mosleh
Soil and Water Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

Keywords Moisture index · Partial least squares regression · Random forest · Terrain attributes

Introduction

Soil organic carbon (SOC) stocks are one of the most important properties of soil. It has a strong connection with soil behavior and production potentials, such as providing nutrients to plants, water retention, greenhouse gas retention, resistance against physical degradation, and yield. Therefore, its reduction can have detrimental effects on soil properties (Maia et al.,

2010; Venter et al., 2021). The effects of climate, soil characteristics, and management on SOC stock accumulation have been extensively investigated (Rabbi et al., 2015; Söderström et al., 2014). However, the relative importance of these factors remains unclear, mainly in the arid and semiarid zones (Sabetizade et al., 2021).

Sanderman et al. (2017) stated that environmental factors such as land use changes affect the amount of SOC stocks (Chakan et al., 2017). Therefore, environmental factors are useful tools for predicting SOC stocks (Dong et al., 2021). Among the environmental factors, topography is an important factor in the formation of soil in different climates. Topographic features, including elevation, slope, aspect, curvatures, and other dependent factors, are effective factors in controlling the movement and maintenance of the soil water. Therefore, it will have influences on most soil characteristics, including SOC stocks (Hu et al., 2018; Prietzel et al., 2016).

In areas with more topographic variation, a larger SOC stock variation is expected (Zhu et al., 2019). In addition, studying the relationships between climatic and environmental factors with the amount of SOC stocks in different regions can help us to predict SOC stocks. It can help us to simulate how the environmental changes affect soil carbon levels; therefore, modeling can be a useful tool in studying the SOC stocks using these parameters (Prichard et al., 2000).

To study the SOC stocks, the development of digital soil mapping (DSM) methods and their applications (McBratney et al., 2003) have created the ability to study the spatial distribution of SOC stocks using SCORPAN factors (e.g., soil, climate, organisms, material parent) (Bargaoui et al., 2019; Minasny et al., 2013). Many algorithms have been used for modeling SOC stocks, such as random forest (RF) model (Gomes et al., 2019; Hengl et al., 2015; Hounkpatin et al., 2018; Yang et al., 2016), super vector machine (SVM) model (Minasny et al., 2018; Ottoy et al., 2017; Wang et al., 2018), the models based on kriging (Gomes et al., 2019; Wang et al., 2018), and partial least squares regression (PLSR) model (Jiménez et al., 2019; Keskin et al., 2019; Zhu et al., 2019). The RF and PLSR methods are based on the well-known classification and regression. These models have been used in various digital soil mapping studies over the past decade (Behrens et al., 2019). Huang et al. (2018) showed that these models

predict the spatial distribution of soil properties using environmental factors with more accuracy.

Identifying suitable environmental factors for the SOC prediction model is still a challenging issue. Therefore, the aims of this study are the following: (1) modeling surface SOC stocks using environmental factors including terrain attributes, moisture index, and normalized difference vegetation index (NDVI); (2) selecting environmental factors using RF and PLSR models to achieve useful and effective environmental factors to optimize the model; and (3) evaluating the accuracy and comparing the efficiency of RF and PLSR models in modeling and estimating the spatial distribution of SOC stocks.

Materials and methods

Study area, land use, and sampling points

The study area is the northwest of Iran (Fig. 1A). It extends from latitudes of 45°52'00" N to 46°23'00" N and from longitudes of 36°24'00" E to 36°46'00" E with a total area of 1.14×10^3 km² (Fig. 1B). The study area has an average annual temperature of 12 °C with 350–450 mm of annual precipitation. Grasslands, gardens and irrigated farming, dry farming, and watercourse are major land uses (Fig. 1B). The elevation of this region ranges from 1311 to 2224 m. The slopes are from 2 to more than 60%. Also, this area has a variety of complex aspects. The soil orders of this region include Entisols and Inceptisols. Some of the highlands of this region are rock outcrops (Iranian soil and water institute, 1991).

For land use map, Landsat 8 satellite images were used with a spatial resolution of 30 m (Mohajane et al., 2018). Landsat 8 satellite images of the study area were downloaded from the earth explorer website (<https://earthexplorer.usgs.gov/>). Pre-processing, including atmospheric and radiometric calibrations, were performed in ENVI 5.3 software. To classify land uses, a maximum likelihood algorithm (Jensen, 2005) was employed by controlling 200 points in different land uses and 200 points in Google Earth software. Based on this, a land use map was obtained (Fig. 1B).

Multiple conditioned Latin Hypercube method (cLHM) was used to select the sampling points. Using this method, 210 points with a density of 0.184

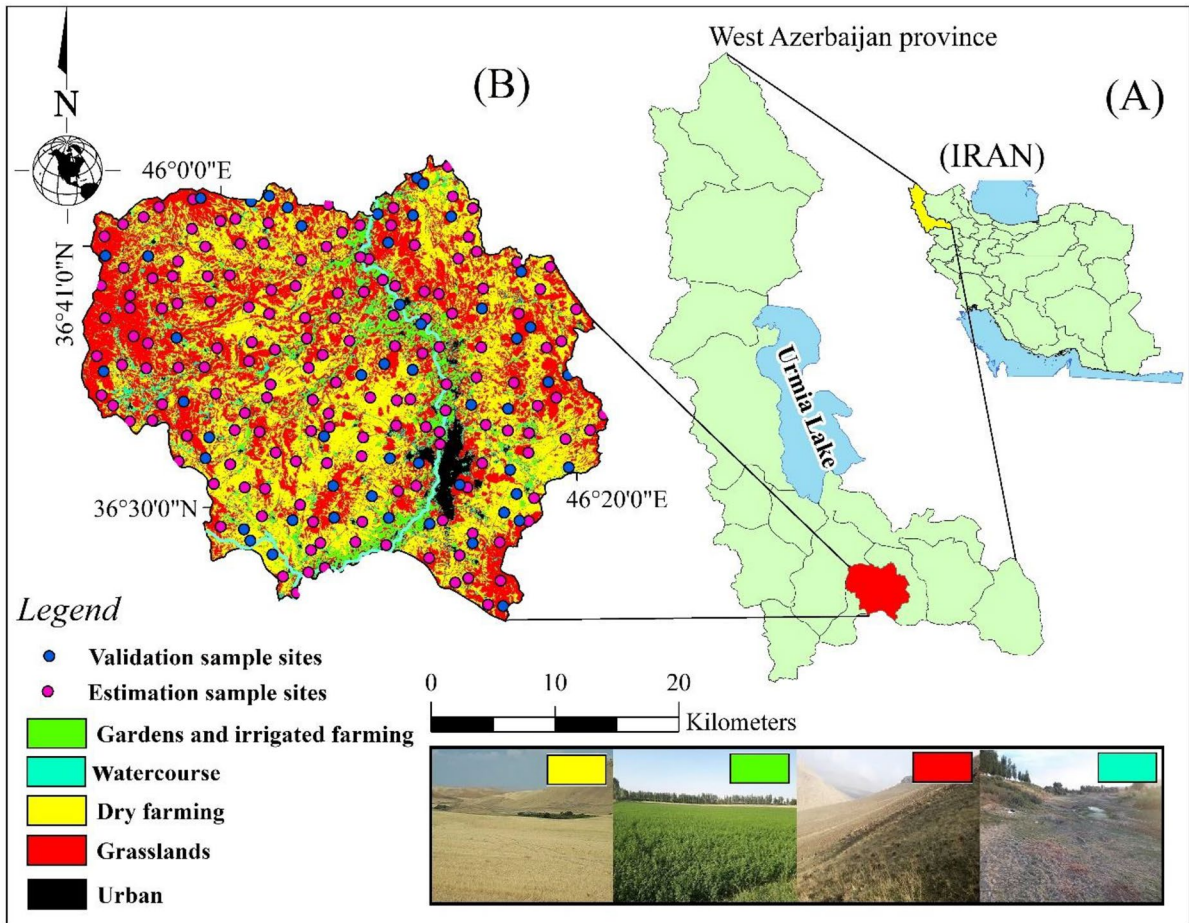


Fig. 1 Location of the study area in Iran and West Azerbaijan province (A), and locations of sampling points (B)

were identified for sampling (Fig. 1B) (Ließ, 2020; Minasny & McBratney, 2006; Minasny et al., 2013). Sampling points were identified by Montana 680 GPS-Garmin, and soil sampling was performed. All of the soil samples were collected from 8 June to 30 July 2019.

Laboratory analysis and calculation of SOC stocks

After sampling, the soil samples were air-dried and passed through a 2-mm sieve. Organic Carbon (OC) was measured using the Walkley–Black method (Nelson & Sommers, 1982). Some researchers have shown that the recovery of OC by the Walkley–Black method is nearly 76 percent, as OC exists in a reduced form in organic compounds, and it can be oxidized to

CO₂. However, mineral carbonates exist in oxidized forms and do not participate in oxidation and reduction reactions (Schumacher, 2002). To overcome this problem, 1.32 as a correction factor, is often used to adjust for the complete recovery of OC (1).

$$OC_{Corrected} = OC_{Measured} \times 1.32 \tag{1}$$

where $OC_{Corrected}$ is the measured organic carbon in the laboratory.

Soil bulk density was measured by the cylinder method (Klute & Page, 1986), because the gravels cannot hold the SOC stocks; therefore, gravels were removed, and the actual amount of soil was calculated (Tian et al., 2009). After removing the gravel, the equivalent soil depth was calculated by Eq. 2 (Ellert

et al., 2002). Finally, the amount of soil SOC stocks was obtained using Eq. 3 (Deng et al., 2014).

$$h_i = \frac{D \times Bd_{min}}{Bd_i} \quad (2)$$

$$SOC_{stocks} = OC_{Corrected} \times Bd_i \times h_i \times 10 \quad (3)$$

where h_i is the equivalent soil depth (m), D is the soil depth (0.3 m), Bd_{min} is minimum soil bulk density (gr/cm^3) in total samples (with removed gravel), and Bd_i is the measured soil bulk (gr/cm^3) density for i sample (with removed gravel).

Environmental factors including terrain attribute, vegetation, and moisture indices

Digital elevation model (DEM) of the study area, with 30×30 m² spatial resolution, was acquired from the earth explorer website (<https://earthexplorer.usgs.gov/>). Based on the DEM data, 23 terrain attributes

(Guo et al., 2019) were derived using SAGA GIS software (Conrad et al., 2015). All of these indicators are given in Table 1. The terrain attributes were divided into three groups, including local, regional, and combined attributes which were calculated based on fixed window and neighboring pixels, contributing area concepts, and local and regional attributes, respectively (Quinn et al., 1991).

To obtain the moisture index, the evaporation of MODIS products and precipitation data of TRMM products from the Giovanni website were used (<https://giovanni.gsfc.nasa.gov/>). These parameters were resampled to 30×30 m² by R-Studio software, which adopts the digital elevation model (DEM) data as a covariant. After that, the moisture index was calculated according to Ivanov's moisture formula by R-Studio software (Eq. 4) (Wang et al., 2019).

$$K = \frac{R}{E_0} \quad (4)$$

Table 1 The list of terrain attributes as predictors of SOC stock modeling derived from the DEM (Guo et al., 2019)

Classification	Symbol	Attribute	Unit
Original DEM	elevation	Digital elevation model	m
Local	slope	Slope	Degree
	aspect	Aspect	Degree
	hcurv	Plan curvature	m ⁻¹
	vcurv	Profile curvature	m ⁻¹
	ruggedness	Terrain ruggedness index	Non-dimensional
	convexity	Terrain surface convexity	Non-dimensional
	convergence	Convergence index	%
Regional	twi	Topographic wetness index	Non-dimensional
	chnl base	Channel network base level	m
	chnl alti	Vertical distance to channel network	m
	vall depth	Valley depth	m
	rsp	Relative slope position	Non-dimensional
	slph	Slope height	[0–1]
	normalh	Normalized height	%
	standh	Standardized height	m
	midslppst	Mid-slope position	Non-dimensional
	sink	Closed depressions	Non-dimensional
	texture	Terrain surface texture	Non-dimensional
	mrxbf	Multi-resolution valley bottom flatness	Non-dimensional
	mrrtf	Multi-resolution ridge top flatness	Non-dimensional
we	Wind effect	Non-dimensional	
waf	Effective air flow heights	Non-dimensional	
Combined	LS	Slope length (LS) factor	m

where E_0 is the evaporation, K is the moisture index, and R is the annual precipitation (mm).

After preparing Landsat 8 satellite images from the USGS website and performing pre-processing, including all corrections made to satellite image bands, NDVI was calculated by Red (R) and infrared (NIR) bands, according to Eq. 5 in ENVI 5.3 software (Zhao et al., 2014).

$$NDVI = \frac{NIR - R}{NIR + R} \tag{5}$$

Selecting environmental factors to predict spatial distribution of SOC stocks

In this study, at the first stage, Pearson’s correlation between SOC stocks and environmental factors was obtained. Then, SOC stocks were modeled by random forest (RF) (Gomes et al., 2019; Hounkpatin et al., 2018) and partial least squares regression (PLSR) (Jiménez et al., 2019; Keskin et al., 2019) models to select environmental factors for estimation spatial prediction. The RF and PLSR methods divided the data into two groups: test and train (train=170 data of 210 data and test=40 data of 210 data) (RCOLOR-Brewer & Liaw, 2018). To perform the RF model, the most important parameters to predict SOC stocks were identified with SAS JMP software. To perform the PLSR model at first, SmartPLS software was used to identify environmental factors for estimating SOC stocks. Then, the selected data using SmartPLS software were transferred to the Unscrambler software, and the major parameters were identified. Finally, the spatial distribution of SOC stocks was predicted by the RF and PLSR models in the R program.

Evaluation of spatial estimation methods

The SOC stocks data were obtained for the train and test sample sites from the assessment of spatial distribution

maps of the estimated SOC stocks. Different validation indices, including the root-mean-square error (RMSE), mean absolute deviation (MAD), coefficient of determination (R^2), and concordance (ρ_c) were used to interpret the measured and estimated values of SOC stocks using the following equations (Eqs. 6, 7, 8, and 9) (Kuhn & Johnson, 2013).

$$R^2 = 1 - \frac{\sum_{i=1}^n (Obs - Pred)^2}{\sum_{i=1}^n (Obs - \overline{Obs})^2} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Obs - Pred)^2}{n}} \tag{7}$$

$$MAD = \frac{\sum_{i=1}^n |Obs - Pred|}{n} \tag{8}$$

$$\rho_c = \frac{2\rho\sigma_{obs}\sigma_{pred}}{\sigma_{obs}^2 + \sigma_{pred}^2 + (\mu_{obs} - \mu_{pred})^2} \tag{9}$$

where Obs is the measured value, $Pred$ is the prediction value extracted from the model, \overline{Obs} is the average measured values, n is the number of sampling points, ρ is Pearson’s correlation coefficient between the predictions and observations, and μ_{Obs} and μ_{Pred} are the means of the predicted and observed values, respectively. σ_{Obs}^2 and σ_{Pred}^2 are the corresponding variances.

Results

Descriptive statistics of SOC stocks in different land uses

The summary statistical of SOC stocks has been shown for all land uses and each land use in Table 2. The results showed that the maximum, minimum,

Table 2 Descriptive statistics of SOC stocks (kg/m²) data

Type of land use	N	Maximum	Minimum	Mean	Median	Skewness	Kurtosis
All land use	210	4.5	0.51	2.7	2.57	-0.139	-0.746
Grasslands	96	4.5	1.28	3.23	1.28	-0.768	0.266
Dry farming	91	3.85	0.9	2.18	2.05	0.201	-0.248
Garden and irrigated farming	14	3.85	1.28	2.74	2.76	-0.228	-0.315
Watercourse	9	3.72	0.51	1.97	2.12	1.641	0.036

mean, median, skewness, and kurtosis values for SOC stocks were 4.5, 0.514, 2.7, 2.571, -0.139, and -0.746, respectively, in the total land uses in the study area (Table 2). Also, the amount of SOC stocks for grasslands was the highest. In this land use, the presence of natural vegetation has increased SOC stocks, and thus soil quality has been improved (Roose et al., 2005). As a result of the higher micro-organisms activity, SOC stocks were further accumulated (Hooper et al., 2000; Wang et al., 2019). The

maximum, minimum, mean, median, skewness, and kurtosis values for SOC stocks in grasslands were 4.5, 1.286, 3.239, 3.20, -0.768, and 0.266, respectively (Table 2). The lowest amount of SOC stocks is related to the watercourse. Probably, soil erosion in this land use reduced the amount of SOC stocks (Wang et al., 2010). The maximum, minimum, mean, median, skewness, and kurtosis values for SOC stocks in the watercourse were 3.729, 0.514, 1.977, 2.121, 1.641, and 0.036, respectively (Table 2).

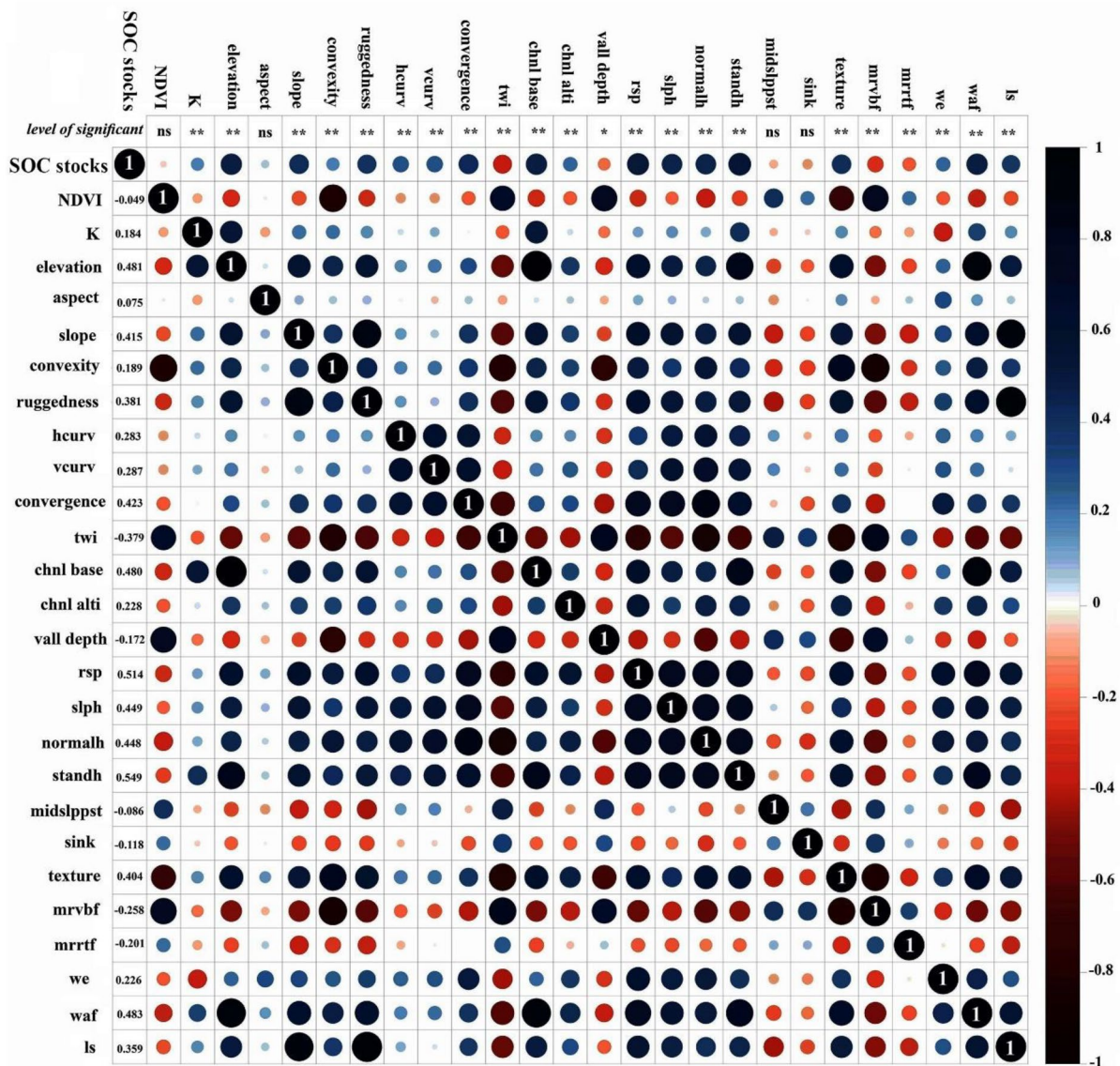
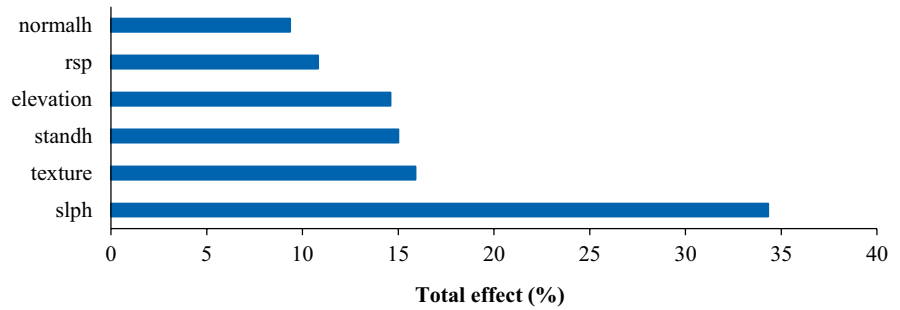


Fig. 2 Pearson’s correlation (P -value < 0.05 level) between SOC stocks with environmental factors (including vegetation index, moisture index, and terrain attributes)

Fig. 3 Total effect (%) of each environmental factors on estimating SOC stocks by RF model



Relative environmental factors with SOC stocks

Based on the Pearson correlation (p -value < 0.05 level), SOC stocks were not correlated with the NDVI, and did not show any correlation with the aspect in the local group. Also, SOC stocks were not correlated with the midslpst (mid-slope position),

and sink (closed depressions) in the regional group (Fig. 2). Rahmati et al. (2016) investigated the SOC stocks in the Lighvan watershed located in north-western Iran in four land uses, including barren lands, weak grasslands, irrigated lands, and dry farming using the ETM⁺ sensor. Their results revealed that remote sensing was an ineffective method in

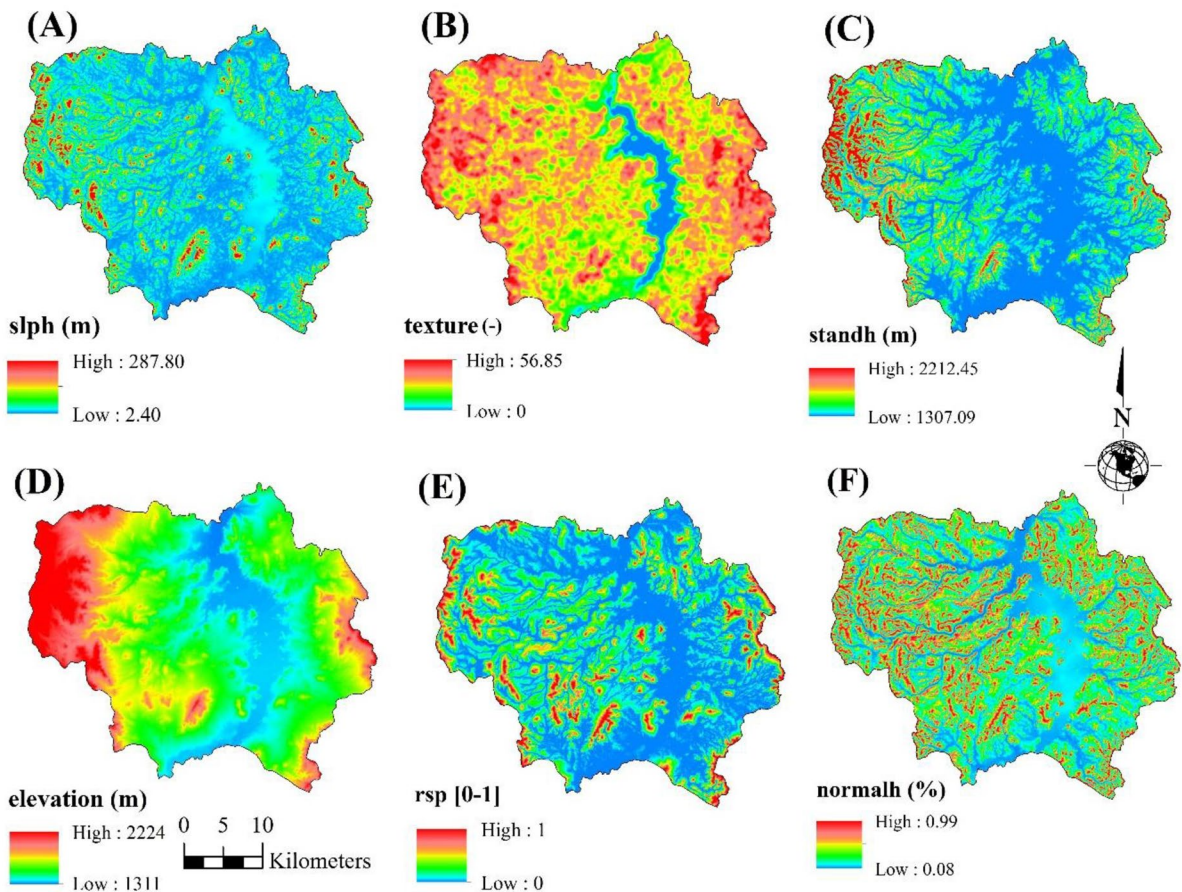


Fig. 4 Environmental factors selected by the RF model, slope height (A), terrain surface texture (B), standardized height (C), elevation (D), relative slope position (E), and normalized height (F)

estimating SOC in areas using vegetation cover. This result might attribute to the disturbance of vegetation in the spectral reflectance of OC.

Selecting environmental factors by RF model

The results of SOC stock modeling using the RF model showed that the environmental factors that have the greatest effect on the prediction of SOC stocks include standh, texture, slph, elevation, rsp, and normalh. The modeling results with these parameters showed that the total effects were slph, standh, texture, elevation, rsp, and normalh 34.32, 15.9, 15.1, 14.6, 10.81, and 9.36 (%), respectively (Fig. 3). Therefore, the total effect of the slph parameter was the highest value, and the total effect of normalh was the lowest value. The importance of the RF model in estimating the factors is represented in Fig. 4.

The estimated factors varied significantly: slph (2.4 to 287.8), texture (0 to 56.85), standh (1307.09 to 2212.45), elevation (1311 to 2224), rsp (0 to 1), and normalh (0.08 to 0.99). The highest value of these factors was in the west of the watershed, and the lowest value was in the middle of the studied watershed.

Selecting environmental factors by PLSR model

The analytical model shows the effects of the studied environmental factors in Fig. 5. In this diagram, each line has a path and direction, which is the path coefficient, or the standardized beta coefficient of the multiple regression model. Each coefficient represents the value of the effect of the independent variable on the dependent variable. Also, in path analysis, the unknown variable of error quantity (e^2) remains, and the sum of the coefficient of explanation and the

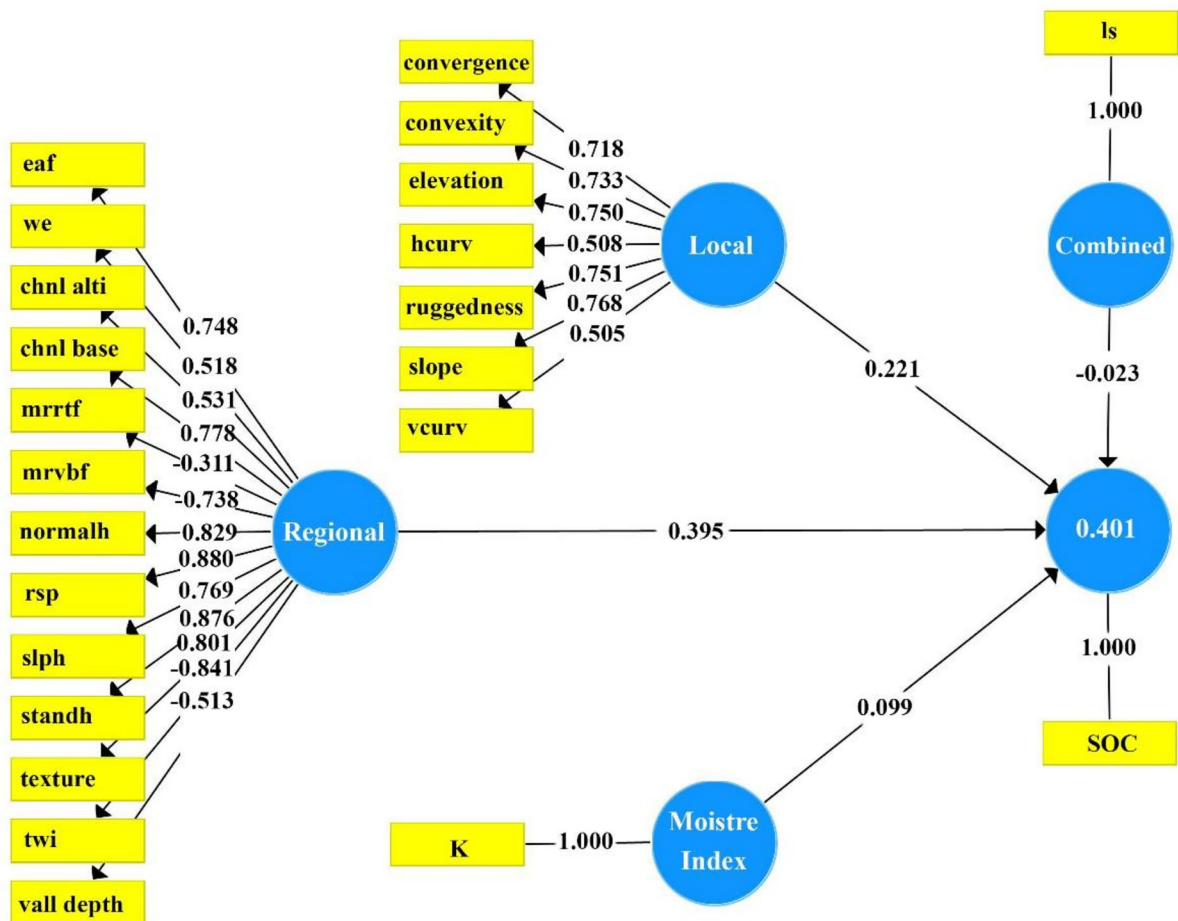


Fig. 5 PLS algorithm with all environmental factors (including moisture index, local, regional and combined attributes)

variable of error is equal to one ($R^2 + e^2 = 1$) (Norris et al., 2015). The results of SOC stock modeling from the PLS algorithm in SmartPLS software showed that the path coefficient of moisture index was 0.099 and terrain attributes including local, regional, and combination were 0.221, 0.395, and -0.023 , respectively (Fig. 5).

The purpose of factor analysis is summarizing the data in the form of more effective factors in the model (Harman, 1976). Factor analysis results showed that local parameters including slope, ruggedness, elevation, convexity, and convergence and regional parameters including rsp, standh, normalh, texture, chnl base, slph, and eaf had significant effects on SOC stocks (Table 3). So, modeling was carried out with these parameters. The results of modeling by selected parameters using factor analysis showed no change in the R^2 value (Fig. 6).

The results from PLSR analysis in Unscrambler software showed that among the selected parameters using factor analysis modeling in SmartPLS

software (Zhu et al., 2019), the four factors of standh, rsp, slope, and chnl base demonstrated 40% of SOC stock variations. Also, for test data, this relationship was 34% (Fig. 7). The values of the path coefficient parameter for standh, rsp, slope, and chnl base were 0.929, 0.885, 0.850, and 0.843, respectively (Fig. 6). Thus, among the selected parameters using factor analysis, they had a path coefficient of more than 0.840. Therefore, the spatial distribution of the PLSR model in R software was selected for SOC stocks using these four factors (Fig. 7).

After obtaining the main parameters of the PLSR model, the PLSR relationship for the selected parameters using these four factors was obtained (Table 4). These relationships with a 40% coefficient of determination (R^2) predict SOC stocks using the selected parameters by the PLSR model. The importance order of the PLSR model in terms of factor analysis is demonstrated in Fig. 8. There were the following variations in the parameter values: standh (1307.09 to 2212.45), rsp (0 to 1), slope (0 to 1.09%), and chnl

Table 3 The result of factor analysis

Parameters	SOC stocks (kg/m ²)	Moisture index	Terrain attributes		
			Local	Regional	Combined
SOC stocks (kg/m ²)	1.00				
K		1.00			
convergence			0.718		
convexity			0.713		
elevation			0.750		
hcurv			0.508		
ruggedness			0.751		
slope			0.768		
vcurv			0.505		
chnl alti				0.531	
chnl base				0.778	
eaf				0.748	
mrrtf				-0.311	
mrvmf				-0.738	
normalh				0.829	
rsp				0.880	
slph				0.769	
standh				0.876	
texture				0.801	
twi				-0.841	
vall depth				-0.513	
we				0.518	
ls					1.00

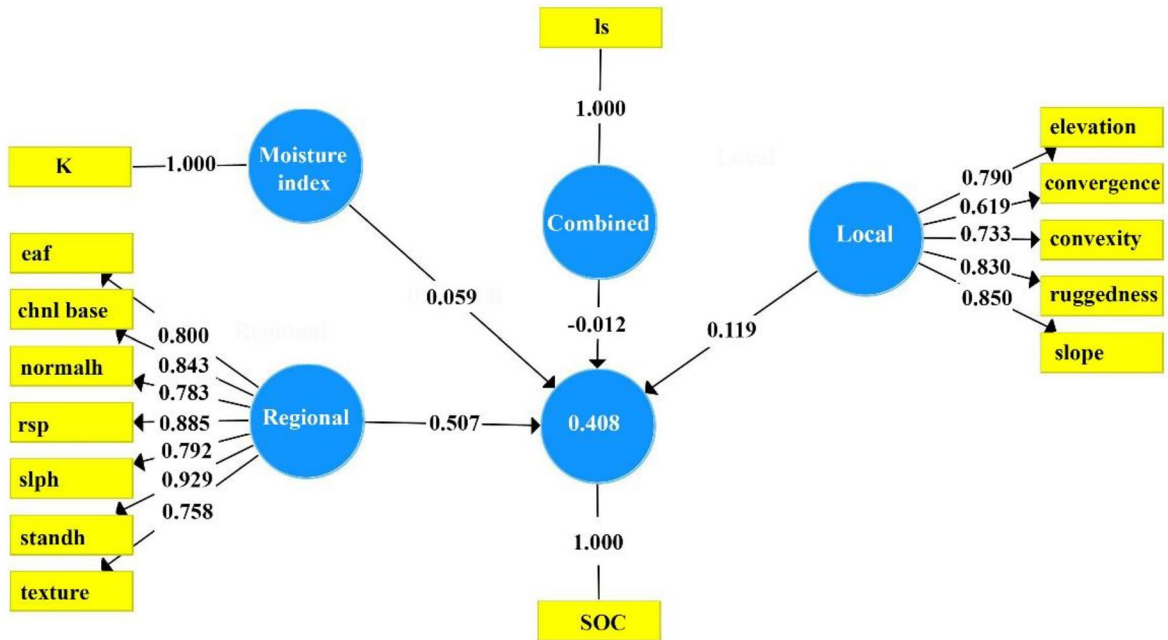


Fig. 6 PLS algorithm using selected environmental factors by factor analysis

base (1311 to 2224). The highest and lowest values are related to the northwest and middle of the watershed, respectively.

Spatial distribution of SOC stocks

The spatial distribution results of RF (Fig. 9A) and PLSR (Fig. 9B) models using training points (170 points) are presented. The R^2 values for RF and PLSR models are 0.81 and 0.40, respectively. Also, the accuracy criteria RMSE and MAE and ρ_c values for the RF model are better than the values of these parameters for the PLSR model (Table 5). The

difference in spatial variation is due to the difference in selecting the factors of these models to estimate the SOC stock distribution, but generally, the pattern of the SOC stock distribution using the RF and PLSR methods was similar.

The results of R^2 for the model validation using test points (40 points) of RF and PLSR methods are 0.76 and 0.34, respectively. Also, accuracy criteria RMSE, MAE, and ρ_c values for the RF model are better than the values of these parameters for the PLSR model (Table 5). Generally, it can be concluded that the RF method is a more suitable method than PLSR in estimating SOC stock distribution. Increasing elevation

Fig. 7 The relative importance of covariates for SOC stock prediction using the PLSR model



Table 4 PLS regression model of SOC stocks at 0–30 cm soil depth ($n=170$)

Model equation	R^2
$Y=9.6E-4 x_1+0.495 x_2+9.01 E-3 x_3+7.837 E-4 x_4-0.253$	0.40**

Y is SOC stocks (kg/m^2), x_1 , x_2 , x_3 , and x_4 are standh, rsp, slope, and chnl base respectively

and topography variation increases the rate of spatial changes in SOC stocks. As a key factor in soil formation, topography is a major factor which has a significant effect on soil properties. Therefore, it is expected that in areas with high topographic changes, the SOC stocks have greater changes (Zhu et al., 2019). The SOC stock distributions in the western and eastern regions were the highest amounts (Fig. 9). The increase in elevation has probably reduced anthropogenic activity because of the return of plant residues, and the accumulation of plant residues has increased

the amount of SOC stocks in these areas (Bonfatti et al., 2016).

Discussion

The benefits of SOC stock in agricultural development have been well known, and many models have been proposed to understand and predict SOC stock (Gurung et al., 2020). But what is important is to predict the amount of SOC stock using the most efficient

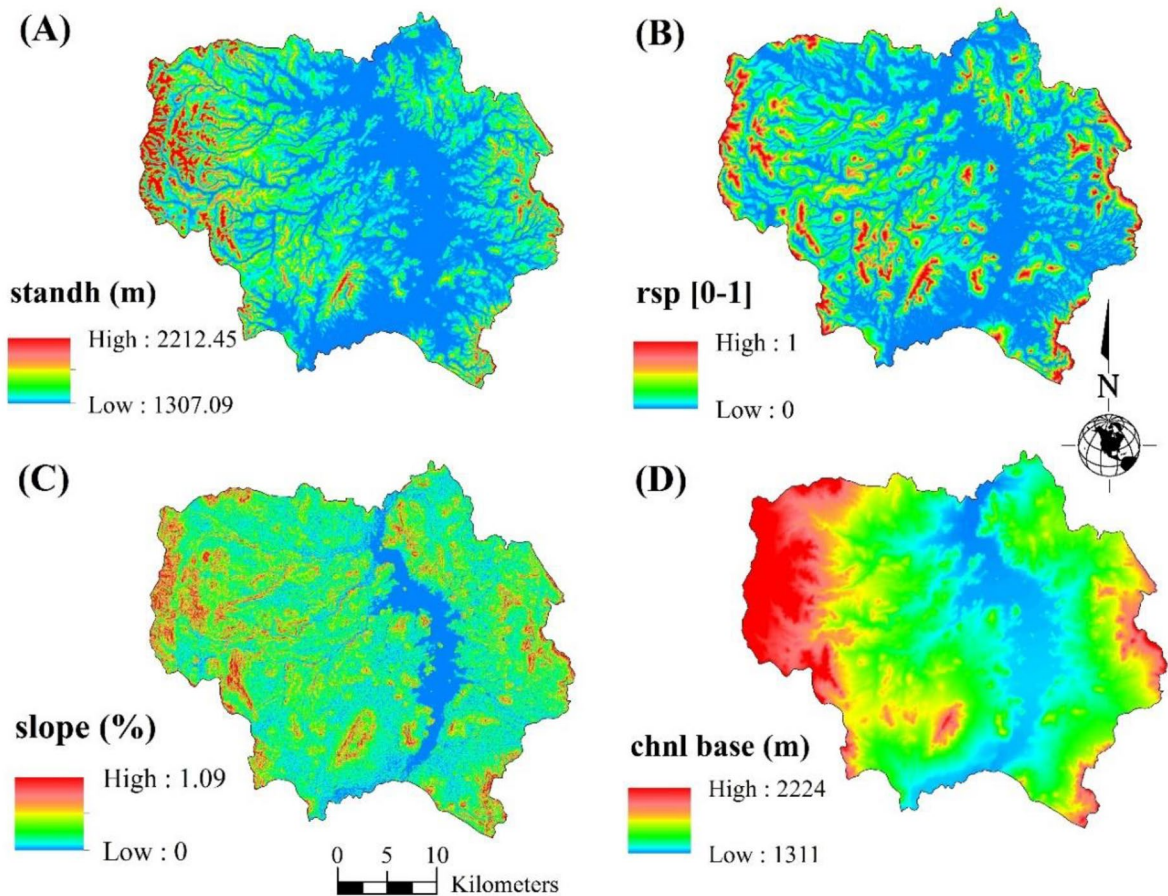


Fig. 8 Environmental factors selected by the PLSR model, standardized height (A), relative slope (B), slope (C), and channel network base level (D)

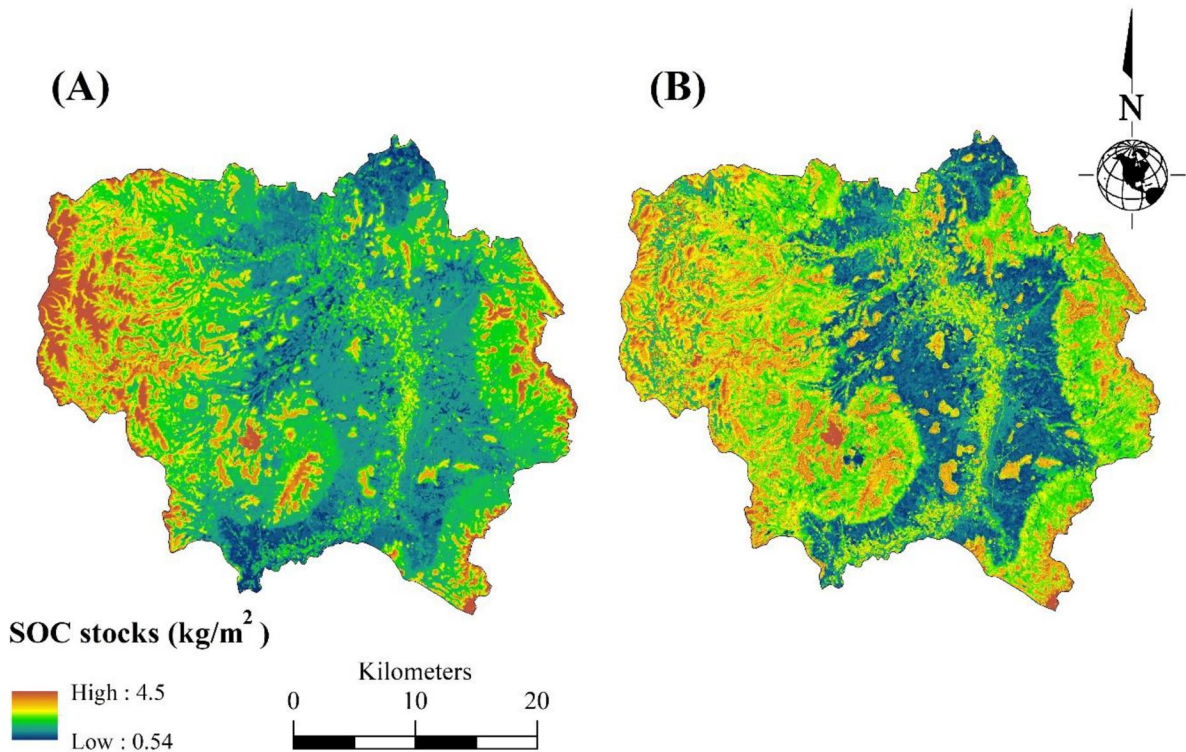


Fig. 9 The spatial predicted of SOC stocks at 0–30 cm soil depth using the RF (A) and PLSR (B) models

indicators. In this study, we tried to select the best environmental factors for predicting the amount of SOC stock using RF and PLSR models. The results showed that the prediction accuracy of these models to predict SOC stock varied (Tables 4 and 5). These differences in model prediction can be due to differences in the inconsistent state of nature, the nature of the model, and differences in the characteristics of the sampling points (Zhao & Li, 2017). Therefore, it is not possible to avouch which models are inefficient for predicting SOC stock, but it is clear that the accuracy of the predictive models varies (Gurung et al., 2020).

In this study, the results of the PLSR model by selecting the parameters using path analysis showed no change in the R^2 value (Table 3, Figs. 6, 7, and 8). We concluded that path analysis can be useful in recognizing the effects of variables on each other and prioritizing them in predicting the spatial variation of SOC stocks (Jiménez et al., 2019). Factor analysis using the studied indicators showed that the standh index had the maximum effect on SOC

stocks because in many areas the climate is controlled by topography variations (Gao et al., 2015). Probably, increasing the elevation affects the soil formation processes such as increasing clay, limestone leaching, and reducing soil acidity (Rhoton et al., 2006). Other chosen parameters for path analysis were the slope and relative slope position (rsp). These parameters significantly affect the amount of SOC stocks, because the following particles that are transferred to the lower areas by erosion accumulate

Table 5 Calibration and validation indices of SOC stocks (0–30 cm) predicted by RF and PLSR methods

Methods	R^2	RMSE (kg/m ²)	MAD (kg/m ²)	ρ_c
Training points ($n=170$)				
RF model	0.81	0.44	0.32	0.83
PLSR model	0.40	0.70	0.56	0.51
Testing points ($n=40$)				
RF model	0.76	0.42	0.30	0.74
PLSR model	0.34	0.73	0.60	0.44

at the foot slope and increase the amount of SOC stocks in these areas (Zhao & Li, 2017). Another effective factor in path analysis was the chnl base. The chnl base, with slope, plays an essential role in the movement of materials and erosion. Therefore, this parameter has a fundamental impact on the SOC stocks (Maerker et al., 2016; Schillaci et al., 2017; Shahini Shamsabadi et al., 2019).

In this study, some differences in the relative contribution of attributes were observed by the RF model. In the RF model, the main factor for estimating the amount of SOC stocks was slph; however, other parameters were selected to estimate the amount of SOC stocks including standh, elevation, rsp, and normalh (Fig. 3). The complex topography in this area may have led to the heterogeneity of SOC stock estimation because in areas with complex topography, there are many uncertainties in estimating SOC stocks. In fact, it is expected that in these areas, changes in the topographic pattern cause changes in slope-dependent parameters. As a result, it makes the different slope and rsp, making difficult the estimation of SOC stocks. However, a deep understanding of the spatial variation of SOC stocks and its effective factors has not yet been achieved (Zhu et al., 2019). But in these and similar areas, elevation, slope, and aspect with their related parameters are probably the main factors in controlling SOC stocks, because they cause changes in climate, hydrological, and environmental conditions (Qin et al., 2016). The changes in these conditions are related to the response of topography variations, and as a result, they will affect the SOC stocks (Zhao & Li, 2017). The texture is another factor that was selected by the RF model (Fig. 3). It showed the softness and roughness of the ground earth surface. By the elevation of this index, the amount of surface roughness probably increased, so it acts as a barrier against particle transfer (Iwahashi & Pike, 2007).

Conclusions

In this study, we aimed to select environmental factors to predict and estimate the spatial distribution of SOC stocks using RF and PLSR models. The overall results showed that the RF model was more accurate than the PLSR model in selecting suitable environmental factors for estimating SOC stocks. In

both RF and PLSR models, selected standh and rsp factors were effective parameters in estimating SOC stocks, which indicates that the standh and rsp play important roles in determining the amount of SOC stocks, and by entering these parameters and the other important factors in these models, the amount of SOC stocks can be easily obtained. Nevertheless, because of the complex relationships between SOC stocks and related environmental factors, more detailed studies are needed to find causal relationships and enhance the accuracy of SOC stock estimates. As the shortage of SOC stocks is a new threat to land degradation and a reduction in the agricultural production potential, simple ways need to be found to estimate SOC stocks.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Supplementary data associated with this article can be found, in the online version, at <https://earthexplorer.usgs.gov/> and <https://giovanni.gsfc.nasa.gov/giovanni/>.

References

- Bargaoui, Y. E., Walter, C., Michot, D., Saby, N. P., Vincent, S., & Lemerrier, B. (2019). Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma*, 356, 113907.
- Behrens, T., MacMillan, R. A., Rossel, R. A. V., Schmidt, K., & Lee, J. (2019). Teleconnections in spatial modelling. *Geoderma*, 354, 113854.
- Bonfatti, B. R., Hartemink, A. E., Giasson, E., Tornquist, C. G., & Adhikari, K. (2016). Digital mapping of soil carbon in a viticultural region of Southern Brazil. *Geoderma*, 261, 204–221.
- Chakan, A. A., Taghizadeh-Mehrjardi, R., Kerry, R., Kumar, S., Khordehbin, S., & Khanghah, S. Y. (2017). Spatial 3D distribution of soil organic carbon under different land use types. *Environmental Monitoring and Assessment*, 189(3), 131.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhrner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific Model Development*, 8(7), 1991–2007.
- Deng, L., Sweeney, S., & Shangguan, Z. (2014). Long-Term Effects of Natural Enclosure: Carbon Stocks, Sequestration Rates and Potential for Grassland Ecosystems in the Loess Plateau. *CLEAN—Soil Air Water*, 42(5), 617–625.
- Dong, Z., Wang, N., Liu, J., Xie, J., & Han, J. (2021). Combination of machine learning and VIRS for predicting soil organic matter. *Journal of Soils and Sediments*, 21(7), 2578–2588. <https://doi.org/10.1007/s11368-021-02977-0>

- Ellert, B. H., Janzen, H. H., & Entz, T. (2002). Assessment of a method to measure temporal change in soil carbon storage. *Soil Science Society of America Journal*, 66(5), 1687–1695.
- Gao, L., Shao, M., Peng, X., & She, D. (2015). Spatio-temporal variability and temporal stability of water contents distributed within soil profiles at a hillslope scale. *CATENA*, 132, 29–36.
- Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G., & Fernandes Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337–350.
- Guo, Z., Adhikari, K., Chellasamy, M., Greve, M. B., Owens, P. R., & Greve, M. H. (2019). Selection of terrain attributes and its scale dependency on soil organic carbon prediction. *Geoderma*, 340, 303–312.
- Gurung, R. B., Ogle, S. M., Breidt, F. J., Williams, S. A., & Parton, W. J. (2020). Bayesian calibration of the Day-Cent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. *Geoderma*, 376, 114529.
- Harman, H. H. (1976). *Modern factor analysis*. University of Chicago press.
- Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., de Jesus, J. M., Tamene, L., & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS One*, 10(6), e0125814.
- Hooper, D. U., Bigneil, D. E., Brown, V. K., & Brassard, L. (2000). Interactions between aboveground and below-ground biodiversity in terrestrial. *BioScience*, 50(12), 12.
- Houkpatin, O. K., de Hipt, F. O., Bossa, A. Y., Welp, G., & Amelung, W. (2018). Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). *CATENA*, 166, 298–309.
- Hu, P. L., Liu, S. J., Ye, Y. Y., Zhang, W., Wang, K. L., & Su, Y. R. (2018). Effects of environmental factors on soil organic carbon under natural or managed vegetation restoration. *Land Degradation & Development*, 29(3), 387–397.
- Huang, J., Minasny, B., McBratney, A. B., Padarian, J., & Triantafyllis, J. (2018). The location-and scale-specific correlation between temperature and soil carbon sequestration across the globe. *Science of the Total Environment*, 615, 540–548.
- Iranian soil and water institute. (1991). *Iranian soil map (1:1000.000)*. <http://www.swri.ir/>
- Iwahashi, J., & Pike, R. J. (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology*, 86(3–4), 409–440.
- Jensen, J. R. (2005). *Introductory digital image processing: A remote sensing perspective* (3rd ed.). Prentice-Hall Inc.
- Jiménez, J. G., Healy, M. G., & Daly, K. (2019). Effects of fertiliser on phosphorus pools in soils with contrasting organic matter content: A fractionation and path analysis study. *Geoderma*, 338, 128–135.
- Keskin, H., Grunwald, S., & Harris, W. G. (2019). Digital mapping of soil carbon fractions with machine learning. *Geoderma*, 339, 40–58.
- Klute, A., & Page, A.L. (1986). *Methods of soil analysis*. Part 1. Physical and mineralogical methods; Part 2. Chemical and microbiological properties. American Society of Agronomy, Inc.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). Springer.
- Ließ, M. (2020). At the interface between domain knowledge and statistical sampling theory: Conditional distribution based sampling for environmental survey (CODIBAS). *Catena*, 187, 104423.
- Maerker, M., Hochschild, V., Maca, V., & Vilimek, V. (2016). Stochastic assessment of landslides and debris flows in the Jemma basin, Blue Nile, Central Ethiopia. *Geografia Fisica e Dinamica Quaternaria*, 39, 51–58.
- Maia, S. M., Ogle, S. M., Cerri, C. C., & Cerri, C. E. (2010). Changes in soil organic carbon storage under different agricultural management systems in the Southwest Amazon Region of Brazil. *Soil and Tillage Research*, 106(2), 177–184.
- McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1–2), 3–52.
- Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, 32(9), 1378–1388.
- Minasny, B., McBratney, A. B., Malone, B. P., & Wheeler, I. (2013). Digital mapping of soil carbon. In *Advances in Agronomy* (vol. 118, pp. 1–47). Elsevier
- Minasny, B., Setiawan, B. I., Saptomo, S. K., & McBratney, A. B. (2018). Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma*, 313, 25–40.
- Mohajane, M., Essahlaoui, A., Oudija, F., Hafyani, M. E., Hmaidi, A. E., Ouali, A. E., Randazzo, G., & Teodoro, A. C. (2018). Land use/land cover (LULC) using landsat data series (MSS, TM, ETM+ and OLI) in Azrou Forest in the Central Middle Atlas of Morocco. *Environments*, 5(12), 131.
- Nelson, D. W., & Sommers, L. (1982). Total carbon, organic carbon, and organic matter. *Methods of soil analysis: Part 2 chemical and microbiological properties*, 9, 539–579.
- Norris, D., Brown, D., Moela, A. K., Selolo, T. C., Mabelebele, M., Ngambi, J. W., & Tyasi, T. L. (2015). Path coefficient and path analysis of body weight and biometric traits in indigenous goats. *Indian Journal of Animal Research*, 49(5), 573–578.
- Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., & Van Orshoven, J. (2017). Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecological Indicators*, 77, 139–150.
- Prichard, S. J., Peterson, D. L., & Hammer, R. D. (2000). Carbon distribution in subalpine forests and meadows of the Olympic Mountains, Washington. *Soil Science Society of America Journal*, 64(5), 1834–1845.
- Prietzl, J., Zimmermann, L., Schubert, A., & Christophel, D. (2016). Organic matter losses in German Alps forest soils since the 1970s most likely caused by warming. *Nature Geoscience*, 9(7), 543–548.
- Qin, Y., Feng, Q., Holden, N. M., & Cao, J. (2016). Variation in soil organic carbon by slope aspect in the middle of the Qilian Mountains in the upper Heihe River Basin, China. *CATENA*, 147, 308–314.

- Quinn, P. F. B. J., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, 5(1), 59–79.
- Rabbi, S. M. F., Tighe, M., Delgado-Baquerizo, M., Cowie, A., Robertson, F., Dalal, R., Page, K., Crawford, D., Wilson, B. R., Schwenke, G., McLeod, M., Badgery, W., Dang, Y. P., Bell, M., O’Leary, G., Liu, D. L., & Baldock, J. (2015). Climate and soil properties limit the positive effects of land use reversion on carbon storage in Eastern Australia. *Scientific Reports*, 5(1), 17866. <https://doi.org/10.1038/srep17866>
- Rahmati, M., Neyshabouri, M. R., Oskouei, M. M., Fard, A. F., & Ahmadi, A. (2016). Soil organic carbon prediction using remotely sensed data at Lighvan watershed, north-west of Iran. *Azarian Journal of Agriculture*, 3(2), 45–49.
- RColorBrewer, S., & Liaw, M. A. (2018). *Package ‘random-Forest.’* University of California.
- Rhoton, F., Emmerich, W., Goodrich, D., Miller, S., & McChesney, D. (2006). Soil geomorphological characteristics of a semiarid watershed. *Soil Science Society of America Journal*, 70(5), 1532–1540.
- Roose, E. J., Lal, R., Feller, C., & Barthes, B. (2005). *Soil erosion and carbon dynamics*. CRC Press.
- Sabetzade, M., Gorji, M., Roudier, P., Zolfaghari, A. A., & Keshavarzi, A. (2021). Combination of MIR spectroscopy and environmental covariates to predict soil organic carbon in a semi-arid region. *Catena*, 196, 104844.
- Sanderman, J., Hengl, T., & Fiske, G. J. (2017). Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences*, 114(36), 9575–9580.
- Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappie, M., Märker, M., & Saia, S. (2017). Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Science of the Total Environment*, 601, 821–832.
- Schumacher, B. A. (2002). *Methods for the determination of total organic carbon (TOC) in soils and sediments*. United States Environmental Protection Agency.
- Shahini Shamsabadi, M., Esfandiarpour-Borujeni, I., Shirani, H., & Salehi, M. H. (2019). Application of soil properties, auxiliary parameters, and their combination for prediction of soil classes using decision tree model. *Desert*, 24(1), 153–169.
- Söderström, B., Hedlund, K., Jackson, L. E., Kätterer, T., Lugato, E., Thomsen, I. K., & Jørgensen, H. B. (2014). What are the effects of agricultural management on soil organic carbon (SOC) stocks? *Environmental Evidence*, 3(1), 1–8. <https://doi.org/10.1186/2047-2382-3-2>
- Tian, G., Granato, T. C., Cox, A. E., Pietz, R. I., Carlson, C. R., Jr., & Abedin, Z. (2009). Soil carbon sequestration resulting from long-term application of biosolids for land reclamation. *Journal of Environmental Quality*, 38(1), 61–74.
- Venter, Z. S., Hawkins, H. J., Cramer, M. D., & Mills, A. J. (2021). Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. *Science of The Total Environment*, 771, 145384.
- Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., & Li Liu, D. (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of the Total Environment*, 630, 367–378.
- Wang, S., Fan, J., Zhong, H., Li, Y., Zhu, H., Qiao, Y., & Zhang, H. (2019). A multi-factor weighted regression approach for estimating the spatial distribution of soil organic carbon in grasslands. *CATENA*, 174, 248–258.
- Wang, Z., Govers, G., Steegen, A., Clymans, W., Van den Putte, A., Langhans, C., Merckx, R., & Van Oost, K. (2010). Catchment-scale carbon redistribution and delivery by water erosion in an intensively cultivated area. *Geomorphology*, 124(1–2), 65–74.
- Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., Yang, M., Zhao, Y. G., & Li, D. C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*, 60, 870–878.
- Zhao, M., Yue, T., Zhao, N., Sun, X., & Zhang, X. (2014). Combining LPJ-GUESS and HASM to simulate the spatial distribution of forest vegetation carbon stock in China. *Journal of Geographical Sciences*, 24(2), 249–268.
- Zhao, N., & Li, X. G. (2017). Effects of aspect–vegetation complex on soil nitrogen mineralization and microbial activity on the Tibetan Plateau. *CATENA*, 155, 1–9.
- Zhu, M., Feng, Q., Qin, Y., Cao, J., Zhang, M., Liu, W., Deo, R. C., Zhang, C., Li, R., & Li, B. (2019). The role of topography in shaping the spatial patterns of soil organic carbon. *CATENA*, 176, 296–305.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.