Science of the Total Environment xxx (xxxx) xxx

Contents lists available at ScienceDirect

Science of the Total Environment

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

Investigating the spatio-temporal variability of soil organic carbon stocks in different ecosystems of China

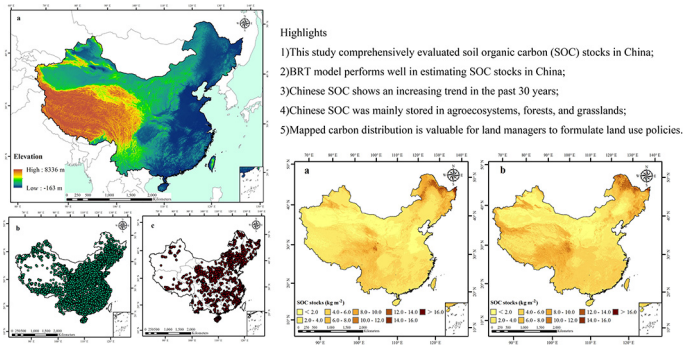
Shuai Wang [a,b], Li Xu [b], Qianlai Zhuang [c], Nianpeng He [b,d,e,*]

[a] College of Land and Environment, Shenyang Agricultural University, Shenyang, Liaoning Province 110866, China
[b] Key Laboratory Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
[c] Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN 47907, USA
[d] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[e] Key Laboratory of Vegetation Ecology, Ministry of Education, Northeast Normal University, Changchun 100024, China

HIGHLIGHTS

- This study comprehensively evaluated soil organic carbon (SOC) stocks in China.
- BRT model performs well in estimating SOC stocks in China.
- Chinese SOC shows an increasing trend in the past 30 years.
- Chinese SOC was mainly stored in agroecosystems, forests, and grasslands.
- Mapped carbon distribution is valuable for land managers to formulate land use policies.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 September 2020
Received in revised form 6 November 2020
Accepted 7 November 2020
Available online xxxx

Editor: Jay Gan

Keywords:
Soil organic carbon stocks
Climate variables
Spatial variation
Boosted regression trees

ABSTRACT

Soil organic carbon (SOC) significantly influences soil fertility, soil water holding capacity, and plant productivity. In this study, we applied two boosted regression tree (BRT) models to map SOC stocks across China in the 1980s and the 2010s. The models incorporated nine environmental variables (climate, topography, and biology) and 8897 (in the 1980s) and 4534 (in the 2010s) topsoil (0–20 cm) samples. During the two study periods, 20% of the soil samples were randomly selected for model testing, and the remaining samples were used as a training set to construct the models. The verification results showed that incorporating climate environment variables significantly improved the model prediction in both study periods. Mean annual temperature, mean annual precipitation, elevation, and the normalized difference vegetation index were the dominant environmental factors affecting the spatial distribution of SOC stocks. The full-variable model predicted similar spatial distributions of SOC stocks for the 1980s and the 2010s. SOC stocks in China showed an increasing trend over the past 30 years, from 3.9 kg m^{-2} in the 1980s to 4.6 kg m^{-2} in the 2010s. In both periods, topsoil SOC stocks were mainly stored in agroecosystems, forests, and grasslands in the 1980s, with values of 9.5, 12.0, and 11.4 Pg C, respectively. Our study provides reliable information on Chain's carbon distribution, which can be used by land managers and the national government to formulate relevant land use and carbon sequestration policies.

1. Introduction

Soil organic carbon (SOC) is the largest carbon pool in terrestrial ecosystems, accounting for more than 80% of the global terrestrial ecosystem carbon (Batjes, 1996). The soil decomposition of SOC affects

* Corresponding author at: Institute of Geographical Sciences and Natural Resources Research, CAS, 11A, Datun Road, Chaoyang District, Beijing 100101, China.
E-mail address: henp@igsnrr.ac.cn (N. He).

https://doi.org/10.1016/j.scitotenv.2020.143644
0048-9697/

atmospheric CO_2 concentrations and subsequently impacts the global climate system (Post et al., 1982; Lal, 2004). Increasing the stocks of SOC is considered one of the most economical and effective ways of balancing atmospheric CO_2 concentrations (Powers and Schlesinger, 2002; Pan et al., 2010). Accurate estimations of regional SOC stocks and their variability are necessary for understanding the carbon cycle and global climate change as well as improving carbon management strategies in terrestrial ecosystems.

Chinese terrestrial ecosystems and their carbon dynamics significantly influence the global carbon budget (Xu et al., 2019). To mitigate global climate change and protect the ecological environment, the Chinese government launched a series of major ecological protection projects since the 1980s, such as the Three-North Shelterbelt Project, Yangtze River Shelterbelt Project, Pearl River Shelterbelt Project, Northern Forest Protection Project, and Natural Forest Protection Project. In addition, a series of new agricultural fertilization and production technologies have been implemented to increase crop yield and protect farmland soils. These engineering measures can increase the carbon storage of terrestrial ecosystems to effectively cope with the challenges of global climate change (Wang et al., 2020a). Most studies have focused on the changes in carbon storage of terrestrial ecosystems—particularly with regard to changes in vegetation carbon (Cambule et al., 2014; Elbasiouny et al., 2014). However, few studies to date have assessed the changes in terrestrial SOC stocks. Some studies have shown that soil has a similar or higher carbon sequestration capacity to that of vegetation. Therefore, accurate assessments of China's SOC stocks over the past 30 years as well as the SOC stock variability of major ecosystems (forests, grasslands, farmland, wetlands, etc.,) are necessary for determining the CO_2 sequestration capacity of China's terrestrial ecosystems. These findings can also elucidate the impacts of various ecological protection measures on China's carbon fixation capacity (Xu et al., 2019; Wang et al., 2020a).

The spatial-temporal variability of soil attributes is influenced by the interaction between natural and human factors, which can be classified into five major elements: climate, terrain, parent material, time, and biology (McBratney et al., 2003). However, the spatial prediction of soil properties on regional scales is complicated by the high number of influencing factors. Digital soil mapping (DSM) technology is a fast and effective method for predicting the spatial distribution of SOC stocks across large areas based on a small number of sampling data points and major environmental covariates (McBratney et al., 2003; Minasny et al., 2006; Saby et al., 2010; Krishnan et al., 2007; Chen et al., 2018). Among the different DSM technologies, tree-based models are widely used to predict soil properties, such as soil salt, pH, SOC, soil total nitrogen, and soil texture (Müller et al., 2013; Cheong et al., 2014; Padarian et al., 2020). Compared with traditional methods, tree-based models have better performance and effectiveness (Wang et al., 2018). Moreover, the most reliable prediction model is included in the DSM toolbox. The boosted regression tree (BRT) model can avoid the problem of transition fitting and effectively deal with nonlinear and complex problems (Cheong et al., 2014; Wang et al., 2020a). Thus, the BRT model is widely used in remote sensing science, epidemiology, ecology, and fishery sciences (Müller et al., 2013; Wang et al., 2018, 2020a). However, few studies have applied the BRT model to assess the spatio-temporal changes in SOC stocks.

A soil field monitoring database for the 1980s and the 2010s was constructed using data from the second soil survey in China (1979–1985) and published literature data of SOC stocks (0–20 cm) during the 2010s (2004–2014). The database covers the main ecosystems in China, such as forests, grasslands, farmland, and wetlands. In this study, we applied this data to 1) comprehensively evaluate the SOC stocks in China in the 1980s and the 2010s and estimate the changes in SOC stocks between these two periods, 2) quantify the impact of environmental variables on SOC stock variability, and 3) map soil carbon stocks during the 1980s to the 2010s.

2. Materials and methods

2.1. Study area

China is located in East Asia and west of the Pacific Ocean (Fig. 1a). The country spans from the center of the Heilongjiang River (near the Mohe River) in the north to zengmuyinsha in the Nansha Islands in the south. It also spans the Pamir Plateau in the west to the confluence of the Heilongjiang and Wusuli Rivers in the east. China has a land area of 9.6 million square kilometers and a land boundary of more than 20,000 km. China's terrain is complex and diverse, with relatively high altitudes in the west and low altitudes in the east. The country's terrain consists of 33.3% mountains, 26% plateaus, 18.8% basins, 12% plains, and 9.9% hills. The terrain from west to east consists of three steps. In the west, the Qinghai Tibet Plateau is the first step, with the highest terrain at altitudes of more than 4000 m. The plateau is bounded by the second step, which consists of the Kunlun Mountains, Qilian Mountains, and Hengduan Mountains. In the east, the Qinghai Tibet Plateau has an altitude of 1000–2000 m and is also considered the second step between the Daxingan Mountains, Taihang Mountains, Wushan Mountains, and Xuefeng Mountains; this region is also composed of plateaus and basins. The broad plains and hills in eastern China are the third step. This geomorphic pattern was formed by the Yanshan movement during the Mesozoic era.

2.2. Soil data sources

SOC content and soil depth were recorded to determine the soil genetic layers without standardization. To establish a database with soil depth consistency, we first standardized the profile of SOC content. In theory, the vertical distribution of SOC can be characterized by a function and extrapolated at depths (Wang et al., 2018). Therefore, we developed an empirical relationship based on the long-term monitoring data of 74 typical terrestrial ecosystems from the China Ecosystem Research Network. We found that the vertical distribution of SOC in both artificial (agricultural) and natural (forest and grassland) ecosystems is well described by exponential functions and power functions (Chai et al., 2015). Based on the empirical relationship between SOC content and soil depth, we obtained the SOC content for every 20 cm soil layer. The detailed process is described in Xu et al. (2015).

2.2.1. Soil sample collection in the 1980s

The data used in this study covers the 1980s and the 2010s. The data for the 1980s were mainly obtained from the second soil census (1979–1985) in China, including national (volume 1–6), provincial (municipalities directly under the central government), and local soil records. The geographical information was extracted from a digital map (https://www.tianditu.gov.cn/) (Yang et al., 2014). A total of 8897 soil profiles were collected, including the geographical location, soil type, thickness, and organic matter content of each profile (Fig. 1b).

2.2.2. Soil sample collection in the 2010s

The 2010s soil data were obtained from two main sources: 1) actual monitoring data of the experimental group and the related experimental groups from the field study, and 2) officially published data selected from the ISI database (http://apps.webofknowledge.com) and China Knowledge Network (http://www.cnki.net) from 2004 to 2014. We used “soil organic carbon density” and “soil organic carbon storage” as key words in the web search. The selection of literature data followed the subsequent criteria: 1) SOC density data must be the actual monitoring data at the sample plot, excluding model fitting and statistical data; 2) field sampling must be conducted after 2000; and 3) the sampling method of SOC density should be comparable. Consequently, 4534 surface soil (0–20 cm) samples were collected (Fig. 1c) covering major ecosystems in China—including forests, grasslands, farmland, and wetlands. We adopted the same method as that used for the 1980s data

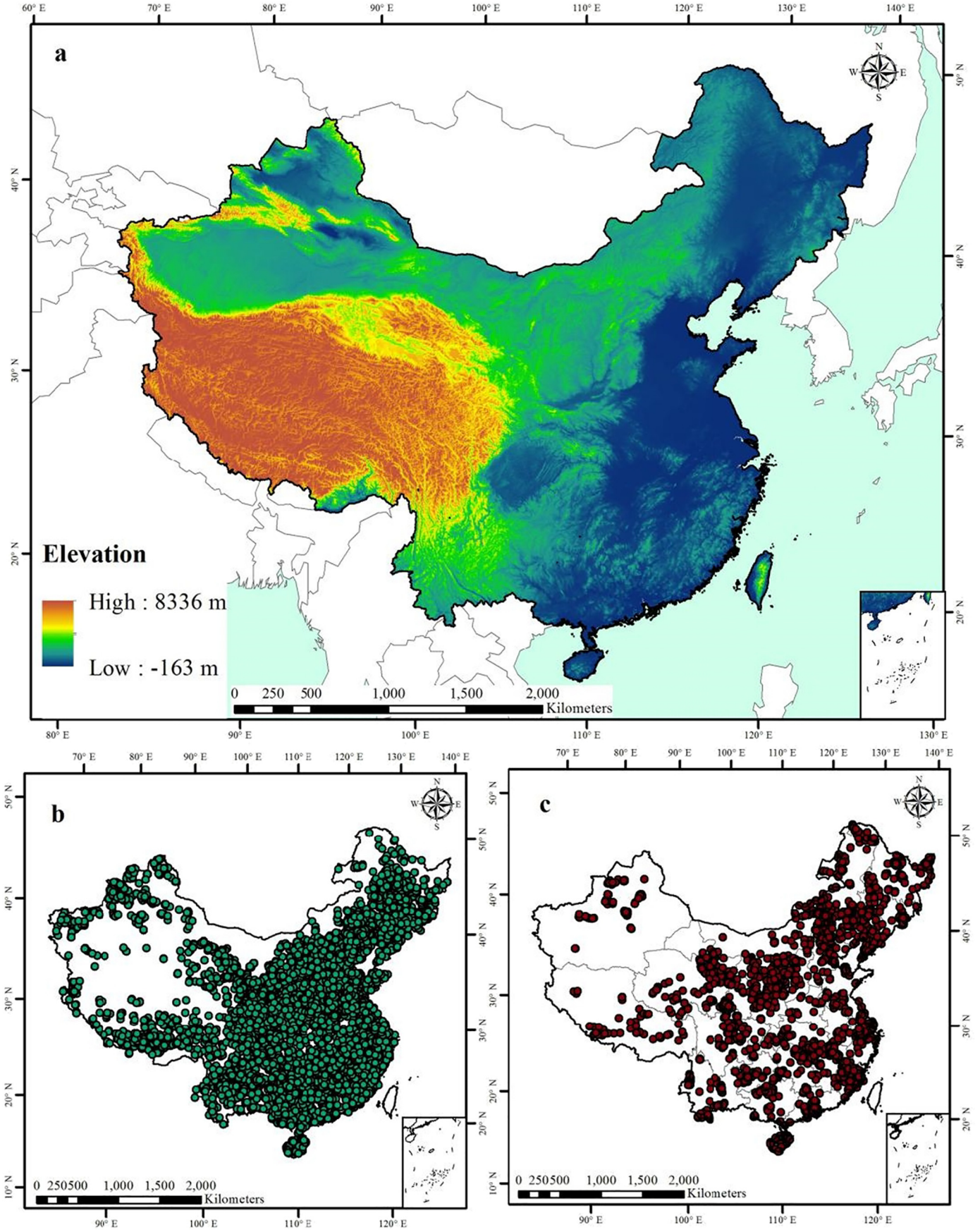


Fig. 1. Location of sampling sites in the1980s (b) and the 2010s (c) overlaid on a digital elevation model of the study area (a) .

to approximately extract the coordinate information from a digital map (https://www.tianditu.gov.cn/) through the description of the published studies.

2.3. Environmental data sources

Nine environmental variables were selected to represent climate, topography, and biology. In addition, as the data were obtained from different departments and platforms, we resampled them using ArcGIS 10.2 software to generate grid data at 90 m spatial resolution. High-precision and high-quality environmental variables were key to accurately predicting surface SOC stocks. The 90 m spatial resolution environmental data were used to meet the needs for national soil attribute spatial prediction research.

2.3.1. Climatic data

Climatic data from the 1980s and the 2010s were obtained from the Resource and Environment Data Cloud Platform (http://www.resdc.cn/). The dataset included mean annual temperature (MAT) and mean annual precipitation (MAP), which was based on the daily observational data of more than 2400 meteorological stations in China, and was generated through sorting, calculation, and spatial interpolation processing. The interpolation of climate variables was based on the ANUSPLIN interpolation software of Australia. ANUSPLIN is a tool for analyzing and interpolating multivariate data by using a smooth spline function. The tool was used to provide statistical analysis and data diagnosis and analyze the spatial distribution of data (Hutchinson, 1998). Finally, the spatial resolution of the two periods of climate data was 1000 × 1000 m, which was then resampled to 90 × 90 m using ArcGIS 10.2 software.

2.3.2. Topographic data

The Shuttle Radar Topography Mission and Digital Elevation Model (SRTM DEM) in the two periods with spatial resolutions of 90 × 90 m were downloaded from the Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn). The elevation (ELE), slope gradient (SG), slope aspect (SA), profile curvature (PG), and two secondary terrain variables—including topographic wetness index (TWI) and catchment area (CA)—were obtained from SRTM DEM. All terrain variables were generated based on SRTM data in ArcGIS 10.2 and the system for automated geoscientific analysis (SAGA) and geographic information system (GIS) software.

2.3.3. Biological data

The normalized difference vegetation index (NDVI) reflects the vegetation coverage status of the land surface. A negative NDVI indicates water and snow coverage on the ground, NDVI = 0 indicates the presence of rock or bare soil, and positive values indicate vegetation coverage. NDVI reflects the background influence of plant canopy, such as soil, wet ground, snow, dead leaves, and roughness, and is also related to vegetation coverage. NDVI data with resolutions of 250 m, 500 m, and 1000 m were obtained from NASA's official website, and users could select the data according to the specific application purposes. In addition, NDVI data for the two periods (the 1980s and the 2010s) were obtained from the Resource and Environment Data Cloud Platform (http://www.resdc.cn/) at a spatial resolution of 1 × 1 km in this study.

2.4. Model prediction and its uncertainty

The BRT model proposed by Friedman et al. (2000) was used to map the SOC stocks of topsoil (0–20 cm) in China during the two periods. The BRT model is usually composed of two parts: regression tree and boosting (Elith et al., 2008). The regression tree model uses binary segmentation technology to fit a simple model to each result (Cheong et al., 2014). On the basis of the decision tree algorithm, the model uses the mean value of the target variables of all samples of each leaf node to predict numerical variables (Pouteau et al., 2011). Boosting refers to integrating multiple models, and each model enhances the overall effect (Elith et al., 2008; Vaysse and Lagacherie, 2015). The boosting algorithm uses an iterative method to gradually add trees to develop the final model (Cheong et al., 2014). The gradient boosting technology and multi-data optimization technology makes the BRT model an ideal tool for our study.

In this study, a BRT model was built using the GEM package (Ridgeway, 2007) running in R language environment (R Development Core Team, 2013). Users must set four parameters in the process of building the BRT model: learning rate (LR), tree complexity (TC), bag fraction (BF), and number of trees (NT) (Cheong et al., 2014). LR represents the contribution of each tree in the model to the BRT model (Pouteau et al., 2011). TC is used to control the interaction between variables and the size of the tree (Vaysse and Lagacherie, 2015). The BF setting is mainly used for the proportion of the training set (Pouteau et al., 2011), and NT is determined by the combination of LR and TC (Pouteau et al., 2011). We finally tested different combinations of the four parameters to obtain the minimum prediction error in both periods. For the best BRT model predictions, LR, TC, BF, and NT were set to 0.025, 9, 0.70, and 1500 in 1980, and 0.025, 9, 0.75, and 2500 in 2010, respectively.

The BRT model was iterated 100 times. Each iteration used different combinations of SOC observations and environmental variables in the two periods. It produced 100 SOC forecasts, and the average of 100 predictions was considered as the final forecast. Its standard deviation (SD) was used as the model uncertainty associated with the prediction. Similarly, to evaluate the importance of environmental variables in predicting SOC stocks in both periods, the relative importance of all iterative environmental variables in the BRT model was obtained and normalized to 100% to compare each variable.

2.5. Statistical analysis

SOC stocks and environmental variables during the two periods (the 1980s and the 2010s) were analyzed using SPSS 16.0. The Pearson correlation coefficient was used to express the degree of linear correlation between variables. The *p* value was used to detect significant differences between variables, and the skewness coefficients were used to characterize the degree of asymmetry of the probability distribution density curve relative to the average value.

2.6. Model evaluation

To evaluate the spatial prediction of SOC stocks by the two BRT models in the two periods, we used the subset feature tool in ArcGIS10.2 to randomly select 20% of the sampling sites (1780 in the 1980s vs. 1537 in the 2010s) as the independent verification set; the remaining 80% of the sampling sites were used as the training set for constructing the BRT model. In addition, four performance verification indicators—mean error (ME), root mean square error (RMSE), coefficient of determination (R^2), and Lin's concordance correlation coefficient (LCCC) (Lin, 1989)—were calculated to evaluate the model:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(a_i - b_i) \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - b_i)^2} \tag{6}$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(a_i - \overline{b}_i\right)^2}{\sum_{i=1}^{n}\left(b_i - \overline{b}_i\right)^2} \tag{7}$$

$$LCCC = \frac{2r\sigma_a\sigma_b}{\sigma_a^2 + \sigma_b^2 + \left(\overline{a} + \overline{b}\right)^2} \quad (8)$$

where a_i, b_i, $\overline{a}$, $\overline{b}$, σ_a, and σ_b represent the variance of the predicted value, observed value, average predicted value, average observed value, and variance of the predicted value and observed value, respectively. n represents the number of samples, and r represents the correlation coefficient between the predicted and observed values.

ME was used to evaluate the deviation degree of the predicted value, and the prediction is stronger when the value is closer to 0. RMSE was used to evaluate the overall accuracy of the prediction, and a smaller value infers a higher prediction accuracy of the model. R^2 was used to evaluate the goodness of fit of the model, and a value closer to 1 indicates a higher reference value of the model. LCCC is used to measure the degree of a 1:1 linear distribution between the predicted and measured values. Therefore, values closer to 1 indicate a higher coincidence between the predicted and observed values and infers a stronger prediction ability of the model.

3. Results

3.1. Exploratory variable analysis

The boxplot of SOC stocks and environmental variables at sampling sites in the two periods is shown in Fig. S1. In the 1980s, SOC stocks ranged from 0.1 to 57.4 kg m^{-2}, with an average of 3.4 ± 2.6 kg m^{-2}. Correspondingly, the average value of SOC stocks was 4.3 ± 3.3 kg m^{-2} in the 2010s. In addition, the skewness coefficients of the 1980s and the 2010s were 3.19 and 1.61, respectively (Table S1). SOC stocks were positively correlated with ELE, SG, MAP, and NDVI in both periods (Table 1). Correspondingly, SOC stocks were negatively correlated with CA, TWI, and MAT in both periods. Surprisingly, the correlation between climate variables and SOC stocks was significant in both periods.

3.2. Model performance

To obtain the best prediction model, we compared the spatial prediction performance of the two BRT models with different combinations of variables in the verification (1780 in the 1980s and 1537 in the 2010s). The full-variable model included climate variables with lower ME and RMSE and higher R^2 and LCCC in both periods (Table 2). It is important to note that the R^2 of the full-variable model in the two periods was 0.53 and 0.57, respectively. Although we did not conduct a wall-to-wall comparison with other studies, we consider this result reasonable due to China's vast landscape and complex topography and climate compared with that of smaller areas (Cambule et al., 2014). The absolute residuals of all training sites of Model A (only topographic variables and NDVI) and Model B (NDVI + topographic variable + climatic variable) in the two periods showed mostly negative values in both models (Fig. 2), which suggests that Model A deviated more from the measured values. The difference between Model A and Model B estimates was 1.44 ± 1.34 kg m^{-2} in the 1980s and 0.89 ± 1.40 kg m^{-2} in the 2010s (Fig. S2), indicating that the addition of climate variables significantly improved BRT prediction.

Table 2
Predictive quality of three boosted regression trees (BRT) models for SOC stocks in the topsoil (0–20 cm) during the two periods.

Year	Item	ME	RMSE	R^2	LCCC
1980s	Model A	0.03	0.27	0.42	0.59
	Model B	0.001	0.21	0.53	0.78
2010	Model A	0.015	0.31	0.39	0.54
	Model B	0.001	0.19	0.57	0.82

Notes: Model A, only topographic variables and NDVI; Model B, full variables model (NDVI+ topographic variable + climatic variable); ME, the mean error; RMSE, the root mean squared error; R^2, the coefficient of determination; LCCC, Lin's concordance correlation coefficient.

The BRT model was iterated 100 times, and the average SD of the model results was used to evaluate the model uncertainty in the two periods (Fig. 3a and b). The results showed that the BRT model had low uncertainty in both periods. The average SD of SOC stocks in the 1980s and the 2010s were 0.17 ± 0.11 kg m^{-2} and 0.40 ± 0.20 kg m^{-2}, respectively (Fig. 3a and b).

3.3. Importance of environmental variables

To identify the key controls for the model prediction, we iterated the full-variable model—Model B—100 times and calculated the average relative importance (RI) of each environmental variable and weighted them to 100%. We found that MAT, MAP, TWI, ELE, and NDVI were the main environmental variables affecting the spatial variability of SOC stocks in the 1980s (Fig. 4), accounting for 80.6% of the total RI. Correspondingly, the main environmental variables affecting the spatial variability of SOC stocks in the 2010s were MAT, MAP, ELE, and NDVI, accounting for 74.3% of the RI. In addition, we found that climate

Table 1
Relationships between the observed SOC stocks (kg m^{-2}) with all environment variables in the 1980s and the 2010s.

Year	Property	SOC stocks	ELE	SG	SA	PC	CA	TWI	MAP	MAT
1980s	ELE	0.18**								
	SG	0.23**	0.51**							
	SA	−0.08	−0.32**	−0.06*						
	PC	−0.05	0.47	0.31**	−0.03					
	CA	−0.12**	−0.18**	−0.33**	−0.13	0.38**				
	TWI	−0.26**	−0.49**	−0.68**	0.11	−0.17	0.32**			
	MAP	0.66**	0.40**	−0.02	0.15	−0.07	−0.05	−0.13**		
	MAT	−0.33**	−0.61**	−0.16**	−0.17**	0.09	−0.09	−0.03	0.82**	
	NDVI	0.12**	−0.45**	−0.07*	−0.03	−0.14	−0.02	−0.25**	0.38**	0.33**
2010	ELE	0.21**								
	SG	0.20**	0.36**							
	SA	0.07*	0.03	0.06*						
	PC	0.06	0.06*	0.20**	0.07					
	CA	−0.22*	−0.18**	−0.29**	0.11	0.35**				
	TWI	−0.21**	−0.39**	−0.67**	−0.14	−0.08	0.51**			
	MAP	0.51**	0.33**	−0.17**	0.12	0.09*	−0.31**	−0.19**		
	MAT	−0.27**	−0.50**	0.09	−0.02	−0.10	−0.13**	−0.07*	0.77**	
	NDVI	0.33**	0.18**	0.20*	−0.04	−0.12	−0.19**	−0.20*	0.36**	0.18**

Note: ELE, elevation; SG, slope gradient; SA, slope aspect; PU, profile curvature; CA, catchment area; TWI, topographic wetness index; MAP, mean annual precipitation; MAT, mean annual temperature; NDVI, Normalized Difference Vegetation Index.

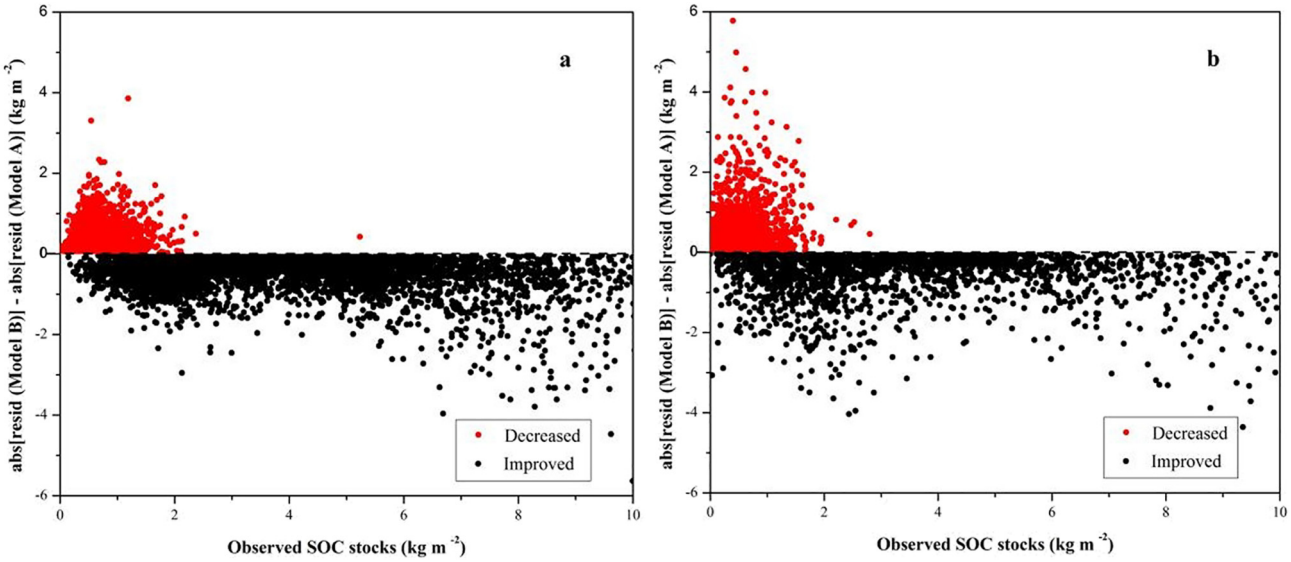


Fig. 2. Differences in absolute residual values between Model A model and Model B in the 1980s (a) and the 2010s (b). Model A (only topographic variables and NDVI) and Model B (NDVI+ topographic variable + climatic variable).

variables were the main environmental factors influencing the spatial distribution of SOC stocks in the 1980s and 2010s, accounting for 43.0% and 50.9% of the RI, respectively.

3.4. Spatial variation of SOC stocks

Full-variable models that included climate variables showed the strongest predictions. To ensure the stability of the prediction, we iterated the model 100 times and calculated the average prediction value in ArcgGIS10.2 to obtain the final prediction. The full-variable model predictions of the spatial distribution pattern of SOC stocks were similar for the 1980s and 2010s (Fig. 3c and d). High SOC stocks were concentrated in alpine and grassland areas that were less affected by human activities, such as the Greater Khingan Mountains and the Qinghai Tibet Plateau. However, low SOC stocks occurred in regions of high anthropogenic impact, such as the central plain and eastern coastal areas. To examine the changes in the spatial variation of SOC stocks over the past 20 years in China, we calculated the difference between the two spatial SOC stock maps of the two periods and classified the spatial SOC variability into six levels. The areas with reduced SOC stocks were mainly distributed in the central and northwest areas of Northeast China (Fig. 5), which are the main commodity grain bases in the country. To ensure China's food security, cultivation is continually conducted throughout the year. SOC stocks increased in 64% of the total area and were mainly distributed in the Greater Khingan Mountains and Qinghai Tibet Plateau. In general, SOC stocks in China showed an increasing trend in the past 20 years, from 3.9 kg m^{-2} in the 1980s to 4.6 kg m^{-2} in the 2010s.

We examined the changes in SOC stocks in various ecosystems over the past 30 years to further clarify the temporal changes in SOC stocks in China (Table 3). In the 1980s, SOC stocks in the topsoil were mainly stored in agroecosystems, forests, and grasslands, accounting for 9.5, 12.0, and 11.4 Pg C, respectively; this is equivalent to 84.6% of China's total carbon reservoirs and 82.3% of China's total area. In the 2010s, topsoil SOC stocks were also mainly stored in agroecosystems, forests, and grasslands ecosystems, accounting for 10.9, 8.5, and 12.3 Pg C, respectively. Topsoil SOC stocks in China increased by 6.2 Pg C over the past 20 years, mainly from wetlands, settlement ecosystems, and agroecosystems. The most unexpected decrease in topsoil SOC stocks occurred in forest ecosystems, with a reduction of 3.47 Pg C, which may be attributed to large-scale deforestation during rapid economic development of the past 20 years.

4. Discussion

4.1 . Effects of environmental factors on SOC stocks

Climate variables are key to mapping the spatial distribution of SOC stocks at regional scales (Thornton et al., 2009; Podwojewski et al., 2011; Follett et al., 2012; Adhikari et al., 2014). In a typical mountain landscape of the cold temperate zone in Japan, Li et al. (2010) concluded that vegetation was the main source of SOC, and climate significantly affected the variation of SOC in natural ecosystems. They identified a significant correlation between climate factors and SOC, which was also confirmed in our study (Table 1). Follett et al. (2015) investigated 14 sites on the Great Plains of the United States to determine the sensitivity of SOC to climate gradients (temperature and precipitation) and land use change (nature, conservation, and reserve programs). They found that climate factors were the main environmental variables affecting the spatial variability of SOC. In the growing region of Australia, Luo et al. (2017) assessed the direct and indirect effects of climate, soil properties, carbon input, and the soil carbon pool (a total of 17 variables) on the change rate of SOC, and found that MAT was the main environmental factor affecting the spatial variation of SOC stocks. It should be noted that the sample points in this study may not be representative because our data were obtained from different departments/platforms and different studies. Therefore, some environmental variables may be overemphasized in the modeling process, which may enhance the uncertainty of the model estimates.

MAT and MAP are dominant climatic variables and are therefore widely used in the spatial simulation of SOC stocks. Vegetation productivity and microbial decomposition and transformation are affected by changes in rainfall and temperature (Jobbágy and Jackson, 2000; Chaminade, 2005; Yimer et al., 2006; Saiz et al., 2012; Li et al., 2010), which subsequently impacts SOC stocks. For instance, SOC in natural ecosystems was found to decrease exponentially with increasing temperature (Willaarts et al., 2016). Meersmans et al. (2012) mapped SOC across France at a resolution of 250 m by considering the effects of land use, soil type, climate, and agricultural management; the total SOC stocks in France was approximately 3.7 ± 1.3 Pg C, and precipitation patterns controlled the overall spatial distribution of SOC.

Topography is another important factor in soil formation, which not only controls the redistribution of water and heat, but also affects soil properties, material, and energy flow in ecosystems (McBratney et al., 2003; Martin et al., 2011; Yang et al., 2016; Wang et al., 2016,

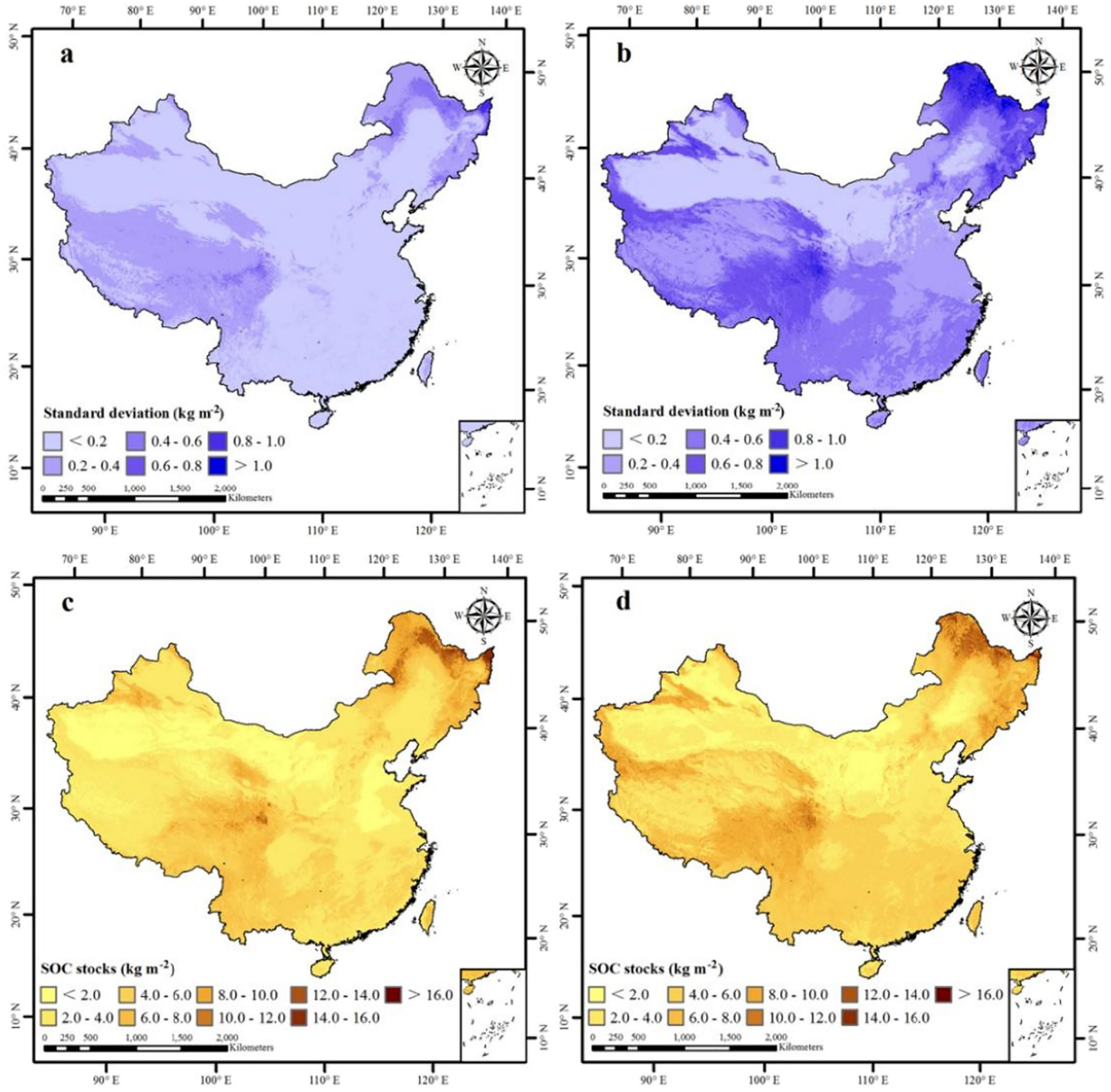


Fig. 3. Standard deviation and spatial distribution maps of SOC stocks predicted by full variables BRT model in the 1980s (a, c) and the 2010s (b, d).

2018). Altitude—an important terrain variable—indirectly affects the distribution of SOC stocks by altering bioclimatic factors, such as rainfall and temperature (Wang et al., 2000; Yang et al., 2016). However, the southwest of Liaoning Province is mountainous and hilly, and the slope position as well as other topographical factors often induce geological subsidence and soil erosion, resulting in low SOC storage in these areas.

4.2. Estimates of SOC stocks and their changes

Except for forest ecosystems, SOC stocks in the topsoil of all other ecosystems showed an increasing trend during the study period (1980s–2010s). Agricultural ecosystems have undergone fertilization, straw returning, and other management practices to increase crop yield, resulting in higher SOC stocks during this period (Xu et al., 2019). Desert ecosystems are usually the weakest link in China's economic development. However, SOC stocks have increased in recent years due to the planting of more suitable plants in arid regions (Wang et al., 2011). China also implemented the protection of 2.32×10^5 km^2 or 41% of the country's total wetland area (Yang, 2014), which subsequently enhanced topsoil SOC stocks in wetland ecosystems (Ren et al., 2020).

As for the settlement ecosystems, rapid economic development has significantly promoted urban expansion. These ecosystems have increased from 269,000 km^2 in the 1980s to 1.17×10^7 km^2 in the 2010s (Table 3). As a result, a large area of cultivated land has been replaced by permanent structures (Wang et al., 2018; Zhao et al., 2018). In addition, finance from the government and the income of local

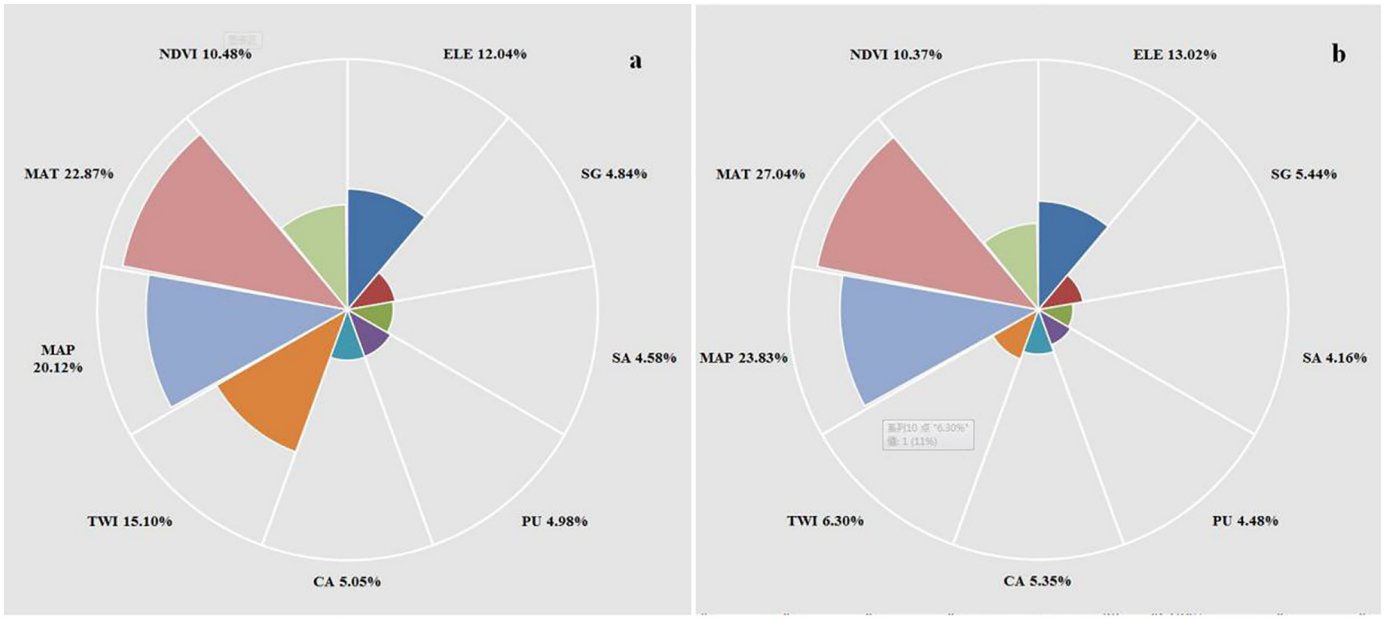


Fig. 4. Relative importance (RI) of each variable as determined from 100 runs of the boosted regression trees model, which are showed in a decreasing order and normalized to 100%. a) RI of SOC stocks in the 1980s and b) RI of SOC stocks in the 2010s. ELE, elevation; SG, slope gradient; SA, slope aspect; PU, profile curvature; CA, catchment area; TWI, topographic wetness index; MAP, mean annual precipitation; MAT, mean annual temperature; NDVI, Normalized Difference Vegetation Index.

inhabitants have also greatly improved during this period (Wang et al., 2018). Moreover, urban management workers regularly manage and protect vegetation in the region, which contributed to the increase in topsoil SOC stocks during this period (Zhang et al., 2012).

Forests experienced a decrease in SOC stocks by 3.47 Pg C over the past 30 years in response to deforestation (Table 3). The forests are mainly distributed in the central, northeast, and southwest regions of China, and these areas experienced rapid urbanization during the study period. As a result, human activities severely degraded surface soils, leading to a decline in topsoil organic carbon. Wang et al. (2020a) concluded that rapid urbanization and altered land use patterns were the main causes for the sharp decrease in topsoil SOC stocks in Dalian in Northeast China. Similarly, Bronson et al. (2004) also attributed carbon loss to urbanization and land use in West Texas in the United States.

SOC stocks in the northeast and southwest forest ecosystems and high-altitude grassland ecosystems showed an increasing trend due to the dense vegetation and relatively low human disturbances. In addition, some areas in central China showed an increase in SOC, which may be due to the implementation of a national policy to return farmland to forests and grassland. Wang et al. (2020b) used nine remotely sensed environmental variables and a BRT model to predict the spatial variation of topsoil SOC stocks in Liaoning Province in Northeast China. They concluded that the increase in forest SOC stocks was closely

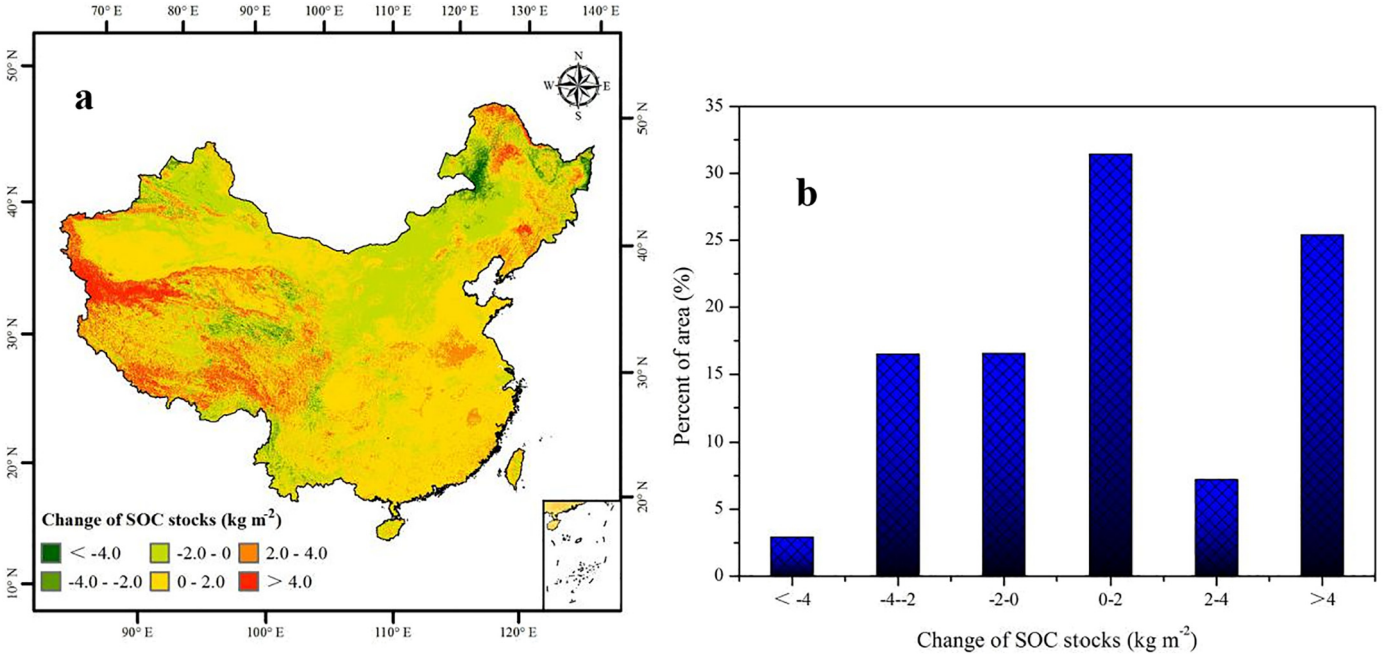


Fig. 5. Spatial distributions (a) and area percentages (b) of SOC change between the 1980s and the 2010s.

Table 3
Predicted soil organic carbon in different ecosystems derived for 20 cm depth in the two periods.

Ecosystem	1980s			2010s			Chang (2010s–1980s)	
	Area ($\times 10^4$ km^2)	SOCD (kg m^{-2})	SOC stocks (Pg C)	Area ($\times 10^4$ km^2)	SOCD (kg m^{-2})	SOC stocks (Pg C)	Area ($\times 10^4$ km^2)	SOC stocks (Pg C)
Agroecosystem	279.29	3.39	9.47	237.6	4.57	10.86	−41.69	1.39
Forest ecosystem	248.74	4.81	11.96	165.8	5.12	8.49	−82.94	−3.47
Grassland ecosystem	265.02	4.30	11.40	247.92	4.94	12.25	−17.1	0.85
Wetland ecosystem	46.79	4.14	1.94	90.52	4.76	4.31	43.73	2.37
Settlement ecosystem	26.9	2.97	0.80	117.21	3.70	4.34	90.31	3.54
Desert ecosystem	51.85	2.89	1.50	51.24	3.35	1.72	−0.61	0.22
Other ecosystems	44.81	3.85	1.73	53.1	5.71	3.03	8.29	1.3
Total	963.41	–	38.80	963.41	–	45.00	–	6.2

related to the government-implemented policy to return farmland to forest and grassland over many years. Zhao et al. (2018) identified an increase in farmland SOC stocks in China from 1980 to 2011 in response to rapid changes in management methods caused by economic and policy incentives, such as fertilization, tillage, and residual treatment. Similar results have also been observed in previous studies (Wang et al., 2018; Xu et al., 2019; Wang et al., 2020a and b).

4.3. Study limitations

The high R^2 and LCCC and low ME and RMSE suggest that the BRT models accurately predicted the spatial distribution of SOC stocks in this study. However, several uncertainties in the model analysis are evident. First, the soil data for the 1980s were obtained from the second national soil survey, and the variable levels of sampling personnel may have caused sampling or measurement errors. Second, the data from the 2010s were obtained from different departments and span multiple studies. Moreover, some of the data may not be representative of certain types of ecosystems, resulting in regional extrapolation errors. Third, the environmental data were obtained from different platforms with different data accuracies; the data therefore required resampling, which likely introduced data errors. Fourth, the impacts of extreme climate events, such as floods and drought, on SOC stocks were not considered in these two periods. Finally, this study is limited to the estimation of SOC stocks in the topsoil (0–20 cm), and thus SOC storage in deep soils must also be assessed. Next, we use process-based biogeochemistry models to adequately map SOC distributions in China.

5. Conclusions

In this study, we applied BRT models to predict the spatio-temporal variations of SOC stocks in China and identify the main environmental factors affecting its distribution. The BRT models could accurately predict the spatial distribution of SOC stocks, with high R^2 and LCCC and low ME and RMSE values. In addition, SOC stocks showed similar spatial distribution characteristics in the two study periods (1980s and 2010s). SOC stocks increased from 3.9 kgm^{-2} in the 1980s to 4.6 kgm^{-2} in the 2010s, with a total increase of 6.2 Pg C. SOC stocks in the Daxinganling area and Qinghai Tibet Plateau were higher than those in central and eastern coastal areas. Climate factors were the dominant environmental variables affecting the spatial distribution of SOC stocks. The results of our study are highly useful for the development of ecological restoration, soil and water conservation, land use planning, and environmental management projects in China.

CRediT authorship contribution statement

Shuai Wang: Conceptualization, Investigation, Methodology, Writing - original draft. **Li Xu:** Data curation, Methodology. **Qianlai Zhuang:** Writing - review & editing. **Nianpeng He**: Funding acquisition, Investigation, review & editing.

Declaration of competing interest

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

Acknowledgements

This research was partially supported from the Chinese Academy of Sciences Strategic Priority Research Program (XDA19020302), National Natural Science Foundation of China (No. 31988102, No. 31870437, No. 31961143022), and National Key Research and Development Program of China (2017YFA0604803). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.143644.

References

Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS One 9, e105519.

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47 (2), 151–163.

Bronson, K., Zobeck, T., Chua, T.T., Acosta-Martinez, V., van Pelt, R.S., Booker, J.D., 2004. Carbon and nitrogen pools of southern high plains cropland and grassland soils. Soil Sci. Soc. Am. J. 68, 1695–1704.

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2014. Soil organic carbon stocks in the Limpopo National Park, Mozambique: amount, spatial distribution and uncertainty. Geoderma 213, 46–56.

Chai, H., Yu, G.R., He, N.P., Wen, D., Li, J., Fang, J.P., 2015. Vertical distribution of soil carbon, nitrogen, and phosphorus in typical Chinese terrestrial ecosystems. Chin. Geogr. Sci. 25 (3). https://doi.org/10.1007/s11769-015-0756-z.

Chaminade, G., 2005. Topography, soil carbon-nitrogen ratio and vegetation in boreal coniferous forests at the landscape level In: A Master of Science Thesis in Soil Sciences at the Department of Forest Soils at the Swedish University of Agricultural Sciences.

Chen, S., Martin, M.P., Saby, N.P.A., Walter, C., Angers, D.A., Arrouays, D., 2018. Fine resolution map of top- and subsoil carbon sequestration potential in France. Sci. Total Environ. 630 (15), 389–400.

Cheong, Y.L., Leitão, P.J., Lakes, T., 2014. Assessment of land use factors associated with dengue cases in Malaysia using boosted regression trees. Spat. Spatio-temporal Epidemiol. 10, 75–84.

Elbasiouny, H., Abowaly, M., Abu_Alkheir, A., Gad, A., 2014. Spatial variation of soil carbon and nitrogen pools by using ordinary Kriging method in an area of north Nile Delta, Egypt. Catena 113, 70–78.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.

Follett, R.F., Stewart, C.E., Pruessner, E.G., Kimble, J.M., 2012. Effects of climate change on soil carbon and nitrogen storage in the US Great Plains. J. Soil Water Conserv. 67 (5), 331–342.

Follett, R.F., Stewart, C.E., Pruessner, E.G., Kimble, J.M., 2015. Great plains climate and land-use effects on soil organic carbon. Soil Sci. Soc. Am. J. 79 (1), 261–271.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. Ann. Stat. 28 (2), 337–407.

Hutchinson, M.F., 1998. Inerpolation of rainfall data with thin plate smoothing splines I. Two dimensional smoothing of data with short range correlation. Geographic Information Decision Analysis 2, 153–167.

Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–436.
Krishnan, P., Bourgeon, G., Lo Seen, D., Nair, K.M., Prasanna, R., Srinivas, S., Muthusankar, G., Dufy, L., Ramesh, B.R., 2007. Organic carbon stock map for soils of southern India: a multifactorial approach. Curr. Sci. 93 (93), 706–710.
Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623–1627.
Li, P., Wang, Q., Endo, T., Zhao, X., Kakubari, Y., 2010. Soil organic carbon stock is closely related to aboveground vegetation properties in cold-temperate mountainous forests. Geoderma 154 (3–4), 407–415.
Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.
Luo, Z., Feng, W., Luo, Y., Baldock, J., Wang, E., 2017. Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions. Glob. Chang. Biol. 23 (4430), 4439.
Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., 2011. Spatial distribution of soil organic carbon stocks in France. Biogeosciences 8, 1053–1065.
McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
Meersmans, J., Martin, M.P., Lacarce, E., De Baets, S., Jolivet, C., Boulonne, L., et al., 2012. A high resolution map of french soil organic carbon. Agron. Sustain. Dev. 32, 841–851.
Minasny, B., McBratney, A.B., Mendonça-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Aust. J. Soil Res. 44, 233–244.
Müller, D., Leitão, P.J., Sikor, T., 2013. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. Agric. Syst. 117, 66–77.
Padarian, J., Minasny, B., Mcbratney, A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools. Soil 6 (1), 35–52.
Pan, G., Xu, X., Smith, P., Pan, W., Lal, R., 2010. An increase in topsoil SOC stock of China's croplands between 1985 and 2006 revealed by soil monitoring. Agric. Ecosyst. Environ. 136 (1-2), 133–138.
Podwojewski, P., Poulenard, J., Nguyet, M.L., de Rouw, A., Nguyen, V.T., Pham, Q.H., Tran, D.T., 2011. Climate and vegetation determine soil organic matter status in an alpine inner-tropical soil catena in the Fan Si Pan Mountain, Vietnam. Catena 87, 226–239.
Post, W.M., Emanuel, W.r., Zinke, P.J., et al., 1982. Soil carbon pools and word life zones. Nature 298 (8), 825–835.
Pouteau, R., Rambal, S., Ratte, J.P., Gogé, F., Joffre, R., Winkel, T., 2011. Downscaling MODIS-derived maps using GIS and boosted regression trees: the case of frost occurrence over the arid Andean highlands of Bolivia. Remote Sens. Environ. 115, 117–129.
Powers, J.P., Schlesinger, W.H., 2002. Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica. Geoderma 109, 165–190.
R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://www.rproject.org/.
Ren, Li, Mao, Wang, Chen, 2020. Investigating spatial and vertical patterns of wetland soil organic carbon concentrations in china's western songnen plain by comparing different algorithms. Sustainability 12 (3), 932.
Ridgeway, G., 2007. Gbm: generalized boosted regression models, R package version 1.6-3. http://132.180.15.2/math/statlib/R/CRAN/doc/packages/gbm.pdf.
Saby, N.P.A., Arrouays, D., Antoni, V., Lemercier, B.F., Schvartz, C., 2010. Changes in soil organic carbon in a mountainous french region, 1990-2004. Soil Use Manag. 24 (3), 254–262.
Saiz, G., Bird, M.I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T.R., Veenendaal, E., Djagbletey, G., Hien, F., Compaore, H., Diallo, A., Lloyd, J., 2012. Variation in soil carbon stocks and their determinants across a precipitation gradient in West Africa. Glob. Chang. Biol. 18 (5), 1670–1683.
Thornton, P.E., Doney, S.C., Lindsay, K., Moore, J.K., Mahowald, N.M., Randerson, J.T., Fung, I.Y., Lamarque, J.F., Feddema, J.J., Lee, Y.H., 2009. Carbon–nitrogen interactions regulate climate–carbon cycle feedbacks: results from an atmosphere–ocean general circulation model. Biogeosciences 6 (10), 2099–2120.
Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Reg. 4, 20–30.
Wang, J.R., Letchford, T., Comeau, P., Kimmins, J.P., 2000. Above-and below-ground biomass and nutrient distribution of a paper birch and subalpine fir mixed-species stand in the Sub-Boreal Spruce zone of British Columbia. For.Ecol. Manage. 130 (1), 17–26.
Wang, S., Zhuang, Q., Jia, S., Jin, X., Wang, Q., 2018. Spatial variations of soil organic carbon stocks in a coastal hilly area of China. Geoderma 314, 8–19.
Wang, S., Adhikari, K., Zhuang, Q., Gu, H., Jin, X., 2020a. Impacts of urbanization on soil organic carbon stocks in the northeast coastal agricultural areas of China. Sci. Total Environ. 721, 137814.
Wang, S., Zhuang, Q., Jin, X., Yang, Z., Liu, H., 2020b. Predicting soil organic carbon and soil nitrogen stocks in topsoil of forest ecosystems in northeastern China using remote sensing data. Remote Sens. 12, 1115.
Wang, Y., Fu, B., Lü, Y., Chen, L., 2011. Effects of vegetation restoration on soil organic carbon sequestration at multiple scales in semi-arid loess plateau, China. Catena 85 (1), 58–66.
Wang, S., Wang, Q., Adhikari, K., Jia, S., Jin, X., Liu, H., 2016. Spatial-Temporal changesof soil organic carbon content in wafangdian China. Sustainability 8 (11), 1154.
Willaarts, B.A., Oyonarte, C., MuñozRojas, M., Ibáñez, J.J., Aguilera, P.A., Willaarts, B.A., 2016. Environmental factors controlling soil organic carbon stocks in two contrasting Mediterranean climatic areas of southern Spain. Land Degrad. Dev. 27, 603–611.
Xu, L., He, N.P., Yu, G.R., Wen, D., Gao, Y., He, H.L., 2015. Differences in pedotransfer functions of bulk density lead to high uncertainty in soil organic carbon estimation at regional scales: evidence from Chinese terrestrial ecosystems. J. Geophys. Res. Biogeosci. 120, 1567–1575.
Xu, L., Yu, G.R., He, N.P., 2019. Increased soil organic carbon storage in Chinese terrestrial ecosystems from the 1980s to the 2010s. J. Geogr. Sci. 29 (1), 49–66.
Yang, Y.F., 2014. Problems existing in wetland protection in China and suggestions for solution. Wetland Science & Management 10 (04), 26–29.
Yang, Y., Li, P., Ding, J., Zhao, X., Ma, W., Ji, C., et al., 2014. Increased topsoil carbon stock across china's forests. Glob. Chang. Biol. 20 (8), 2687–2696.
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 mappingtopsoil organic carbon concentration in an alpine ecosystem. Ecol. Indic. 60, 870–878.
Yimer, F., Ledin, S., Abdelkadir, A., 2006. Soil property variations in relation to topographic aspect and vegetation community in the south-eastern highlands of Ethiopia. For. Ecol. Manag. 232 (1), 90–99.
Zhang, W., Sheng, H., Qian, Y., Chang, Y., Dai, S., 2012. Carbon storage in urban green space. Journal of Southern Agriculture 43 (11), 1712–1717.
Zhao, Y., Wang, M., Hu, S., Zhang, X., Ouyang, Z., Zhang, G., et al., 2018. Economics- and policy-driven organic carbon input enhancement dominates soil organic carbon accumulation in chinese croplands. PNAS 115 (16), 4045.