Impacts of urbanization on soil organic carbon stocks in the northeast coastal agricultural areas of China

Shuai Wang [a,b,c], Kabindra Adhikari [d], Qianlai Zhuang [c], Hanlong Gu [a,*], Xinxin Jin [a,*]

[a] College of Land and Environment, Shenyang Agricultural University, Shenyang 110866, Liaoning Province, China
[b] Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
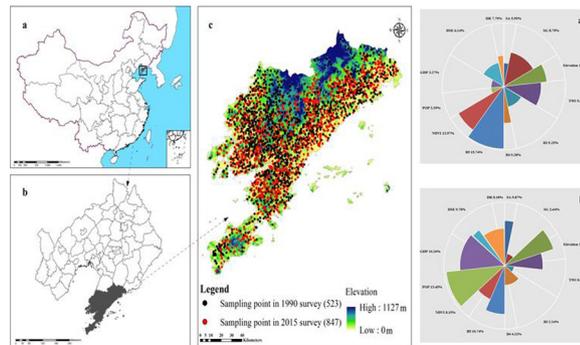[c] Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA
[d] USDA-ARS, Grassland, Soil and Water Research Laboratory, Temple, TX 76502, USA

HIGHLIGHTS

- Rapid urbanization led to drastic land-use change, decreasing SOC stocks in northeast coastal agricultural areas of China
- Inclusion of socioeconomic factors significantly improved the prediction accuracy of topsoil SOC stocks.
- Both environmental and urban-specific variables shall be used to predict the spatial-temporal distribution of SOC stocks in the study region.

GRAPHICAL ABSTRACT

ABSTRACT

Dynamic changes in soil organic carbon pools have significant impacts on regional and global carbon balance. Due to rapid development in urbanized areas, the land use changes dramatically, impacting soil organic carbon (SOC) stocks in topsoil. This study aimed to document the impacts of urbanization on SOC stocks in a rapidly urbanized area from northeastern China. A total of 12 auxiliary variables were as SOC predictors including elevation, slope aspect, slope gradient, topographic wetness index, Landsat TM band3, Landsat TM band4, Landsat TM5, and normalized difference vegetation index. Urban-specific variables including population (POP), gross domestic product (GDP), distance to the socio-economic center, and distance to the roads are also considered. A set of 523 (in 1990) and 847 (in 2015) top soil samples with SOC measurement were collected. Two random forest (RF) models, one with all auxiliary variables except urban-specific variable (MA) and the other with all auxiliary variables (MB) were used to map the spatial distribution of SOC stocks in the two periods. Ten-fold cross-validation was conducted to evaluate the performance of RF models. We find that the full auxiliary variables model had a better performance for the both periods. POP and GDP were key auxiliary variables affecting spatial variability of SOC stocks in 2015. Over a 25-year period, SOC stocks decreased from 2.77 ± 1.09 kg m^{-2} to 2.16 ± 0.93 kg m^{-2}, resulting in 3.78 Tg SOC loss in this region. Rapid urbanization led to drastic land-use change, which was the main reason for the decrease of SOC stocks. Additionally, urban-specific variables should be

* Corresponding authors at: College of Land and Environment, Shenyang Agricultural University, No. 120 Dongling Road, Shenhe District, Shenyang, Liaoning Province 110866, China.
E-mail addresses: allenguhan@126.com (H. Gu), jinxinxin0218@163.com (X. Jin).

used as the main auxiliary variables when predicting SOC stocks in the areas that experience rapid urbanization. We believe that accurate prediction and mapping of SOC stocks will help manage land use and facilitate soil quality assessment so as to increase soil carbon sequestration in the region.

1. Introduction

Soils store large amounts of organic carbon (SOC) in global terrestrial ecosystems (Batjes, 1996). Increasing population and economic development have led to an exponential increase in land use changes (Liu et al., 2016; Stumpf et al., 2018), resulting in large changes in soil carbon stocks, impacting the global climate (Lal, 2004; Wang et al., 2018a). Over the past 40 years, changes in land use and land management had a profound impact on organic carbon storage in agricultural soils. SOC temporal and spatial dynamics in topsoil are a major concern (Xia et al., 2017).

With a rapid development of urbanization and increase of human population, the urban land use pattern has changed dramatically, significantly affecting the soil environment around the city and its surroundings (Liu et al., 2016; Vasenev et al., 2018). The urban population is highly concentrated, creating an imbalance of nitrogen and phosphorus in the urban and surrounding soils, resulting in soil eutrophication (Muñoz-Rojas et al., 2015; Adhikari et al., 2019). Intense urban industrialization and agricultural production and other human activities pollute urban soils. The construction of factories, houses and roads in cities altered soil pore distribution, profile structure and conditions of soil water, heat, gas and nutrients (Wiesmeier et al., 2011). Urbanization has direct and indirect effects on SOC (Zhao et al., 2015; Stumpf et al., 2018).

Agricultural SOC has both natural and human attributes, and has a large spatial variability (Kaushal et al., 2006; Adhikari et al., 2014; Chaminade, 2005; Were et al., 2015). On medium and large spatial scales, SOC distribution is consistent with the horizontal pattern of soil types, deflecting with topography and vegetation, showing a strong climatic and geographic zonality (Don et al., 2007; Bartholomeus et al., 2011). However, on small scales as in agriculture lands, the pattern shows a strong spatial heterogeneity due to the influence of local topography, soil types, land use patterns and management measures (Dalal et al., 2011; Wiesmeier et al., 2011). Previous studies have shown that natural soils exhibit complex changes under the interactive effects of changes in climate, altitude, and land cover and are greatly influenced by soil types and land use patterns (Vågen and Winowiecki, 2013). Warming temperature stimulates soil biological activities, promoting SOC decomposition (Phachomphon et al., 2010; Mishra and Riley, 2012). Changes in land use and land cover and tillage practices not only affect the SOC input in soils but also regulate the decomposition of SOC through changing microclimate and soil conditions, thus altering regional carbon storage (Muñoz-Rojas et al., 2015). In addition, the spatial and temporal variations of soil properties, and changes in landscape properties are important driving factors to regional soil ecological processes, enhancing the spatial heterogeneity of SOC in farmland (Muñoz-Rojas et al., 2015; Liu et al., 2016).

To date, the existing regional SOC studies are mostly concentrated on natural soils or in agricultural areas. There are only few studies to characterize SOC stocks and their changes in rapidly urbanized areas (Vasenev et al., 2014). Most studies focused on the effects of changes in physical geographical environment, land use and land cover and farm management on SOC changes at field scales (Adhikari and Hartemink, 2015; Edmondson et al., 2014; Phachomphon et al., 2010). There are less studies focusing on the role of population growth, social and economic development in SOC changes in rapidly urbanized areas. Therefore, identifying key factors affecting spatial-temporal SOC changes in such areas is needed.

Traditional methods to predict the spatial distribution of SOC stocks are to estimate the average SOC stocks for each soil type or land use type, and then assign them to each mapping unit (Wang et al., 2018a). However, to adequately quantify SOC for a region, a large number of samples covering the entire study area needs to be collected, which is time-consuming, laborious and prone to errors and omissions (McBratney et al., 2003; Hengl et al., 2015; Were et al., 2015; Wang et al., 2018b; Ramcharan et al., 2018). To overcome these problems, digital soil mapping (DSM) techniques (McBratney et al., 2003) to study the spatial variability of SOC stocks is efficient and robust in providing more accurate prediction based on sparse data and environmental variables (McBratney et al., 2003; Minasny et al., 2013; Brus et al., 2016). Most DSM studies used the 'scorpan' approach (McBratney et al., 2003) by considering environmental variables (i.e., soil, climate, organisms, topography, parent material, age and space) to model SOC spatial variability (McBratney et al., 2003; Adhikari et al., 2014; Hengl et al., 2015; Xia et al., 2017; Wang et al., 2018a, 2018b). Numerous DSM techniques have been developed to predict and map SOC distribution, including linear mixed models (Stewart et al., 2007), multiple linear regression (Meersmans et al., 2008), rule-based regression (Adhikari et al., 2014), support vector machines (Morellos et al., 2016), artificial neural networks (Were et al., 2015), random forest (RF) model (Were et al., 2015), and boosted regression tree (BRT) model (Wang et al., 2018b). Among different DSM techniques, tree-based algorithms such as random forest (RF) have been popular in SOC mapping (Wiesmeier et al., 2012; Hengl et al., 2015; Reyes Rojas et al., 2018; Ramcharan et al., 2018).

This study used two RF models considering two different datasets of auxiliary variables to predict and map the spatial variability of SOC stocks in rapidly urbanized areas along the northeastern coast of China during 1990–2015. Our specific objects were:

(i) to predict and map SOC stocks for two specific time periods (1990 and 2015) with associated uncertainty;
(ii) to investigate the potential urbanization processes as indicators for changes in SOC stocks;
(iii) to quantify the changes in SOC stocks during 1990–2015.

2. Materials and methods

2.1. Site description

This study was conducted in Dalian (39.02°–39.07° N, 21.73°–121.82° E), which is located in the eastern coast of Eurasia and the southernmost tip of Liaodong Peninsula in northeastern China. It covers an area of about 13,000 km^2, 68.3% of which falls under agricultural land, 18.2% as forest land and the remaining area mainly comprised of urban land expanded during the reform and opening-up period due to rapid urbanization (Wang et al., 2018b). The altitude in the study area ranging from 0 m to 1127 m above sea level and it gradually increases from northwest to southeast. The region has four distinct seasons (Spring, Summer, Autumn and Winter) with a humid warm temperate continental monsoon climate. The annual average rainfall is between 550 mm and 800 mm, and is mainly concentrated in summer as heavy rains. The average annual temperature is about 10.5 °C. According to the World Reference Base for Soil Resources (WRB) (IUSS Working Group, 2014), the dominant soil types are Cambisols (51% of the area)

and Luvisols (32% of the area). According to the classification of land use status in the third national land survey of China (Ministry of Natural Resources of China, 2017), the main land types in the study area are cultivation land, forest land, grassland, urban land, they collectively account for 91% of the total area of the study area. The main crops are corn, rice, soybean, apple, and cherry. The natural vegetation is dominated by deciduous broad-leaved forest, and the main tree species are *quercus* spp., *Robinia pseudoacacia*, willow, pine, Chinese fir, cypress and ginkgo.

Dalian is rich in hydrothermal resources and has a long history of agricultural reclamation, showing different land use patterns. As one of the pioneers of China's reform and opening up, its population continues to grow, urbanization, and industrialization processes are also rapid. Consequently, farmland is segmented, and landscape has a high degree of spatial heterogeneity. This region has become an ideal case study area to test the hypothesis that "urbanization works as a dominant factor influencing SOC change".

2.1.1. Soil survey data from 1990

With the implementation of China's economic reform policy in the 1980s, Dalian has achieved unprecedented growth in urbanization and economic construction developments. With the acceleration of urbanization and continuous population growth, a large area of forests, grasslands and wetlands were converted to urban land and farmlands. Therefore, this study selected 1990 as the base year to investigate urbanization impacts on SOC distribution.

Typical soil profile description datasets were obtained from the Second National Soil Census Office of China (Office of Soil Survey in Liaoning Province (OSSLP), 1990) including soil physical and chemical properties, climate, parent material, and topographic information. However, our study only focused on the SOC stocks in topsoil (0–30 cm soil depth). A total of 523 soil observations (231 from cultivated land, 187 from grassland, and 105 from forest land) with SOC and bulk density (BD) measurements were obtained, covering different elevation gradients, parent material types, soil types and land use types in the whole study area (Fig. 1c). For samples without soil bulk density measurements, it was calculated using the following pedotransfer function specific to the study area (Wang et al., 2018a):

$$BD = 1.46 - 0.09 * \sqrt{SOC} \quad \left(R^2 = 0.78, p < 0.001\right) \quad (1)$$

2.1.2. Soil sampling in 2015

Due to the rapid urbanization and economic development in the past decades, land use patterns in the study area have changed drastically. It was unrealistic to carry out in-situ sampling following the 1990. Therefore, we designed a purposive sampling method and obtained 847 new topsoil samples (359 from cultivated land, 247 grassland, and 241 from forest land) in 2015, representing variations of slope gradient, slope aspect, elevation, land use type, soil type and parent material in the study area (Fig. 1c). A fuzzy C-means clustering algorithm was employed to group the study area into 54 different categories based on soil, topography, land use and parent material, and from each category, 10–20 sampling points were selected. Details of the sampling design can be found in Wang et al. (2018a). To record the geographic coordinates at each sampling site, a hand-held global positioning system (GPS) was used. Soil samples were collected and properly mixed in plastic cloth or
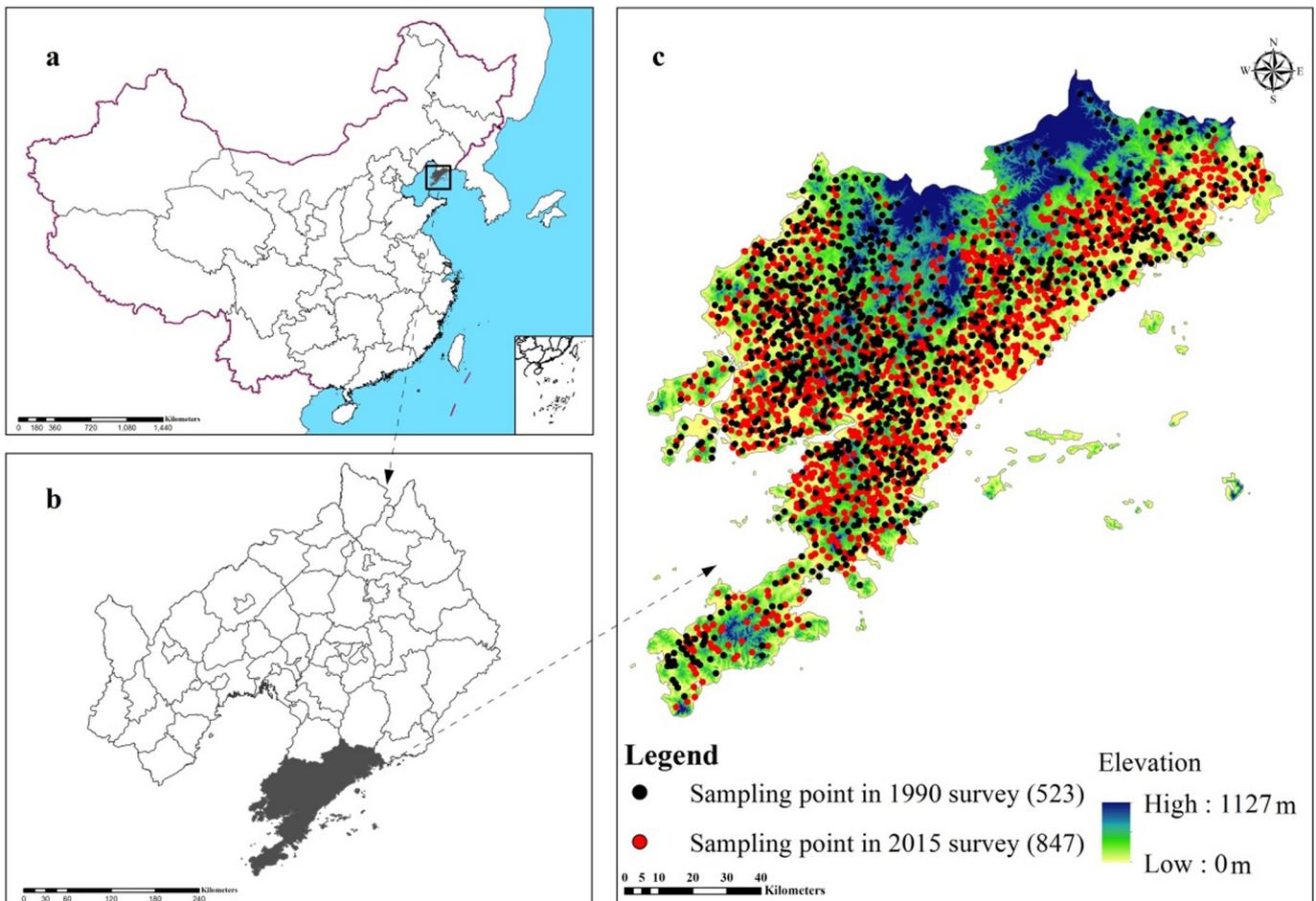

Fig. 1. Soil sampling point in two periods (1990 vs. 2015) overlaid on a digital elevation model of the study area (c) in Liaoning Province (b) of China (a).

wooden tray before removing unnecessary litter, grass roots, insect shells, gravel and other impurities. Subsequently, 1 kg subsample from each site was retained and packed in cloth bags for subsequent laboratory analysis. A wet oxidation method was used to determine SOC content (% mass) in the Analysis and Testing Center of Shenyang Agricultural University, Shenyang, Liaoning Provence, China. From the selected (847) sites, a 100 cm^3 undisturbed soil cores were collected to determine soil BD by oven drying method (105 °C for 48 h).

2.2. Auxiliary variables as SOC predictors

2.2.1. Traditional variables

We selected some traditional and commonly used topographic variables such as elevation, slope gradient, slope aspect, and topographic wetness index (TWI); and remote sensing data such as Normalized Difference Vegetation Index (NDVI), and other band ratios to predict of SOC stocks in the study area. Elevation is widely considered as an important topographic variable affecting the distribution of soil properties in landscapes. Slope gradient, slope aspect, and TWI have important effects on regional soil erosion and are closely related to spatial variability of soil properties including SOC distribution. These topographic variables were acquired from a 90-m gridded Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) downloaded from the United States Geological Survey (USGS, Reston, VA, USA).

Remote sensing data of 1990 and 2015 were retrieved from Landsat 5 and Landsat 8 satellites imageries at 30-m resolution. The imageries were downloaded from the USGS (https://www.usgs.gov/) for the growing season (July–September) in 1990, and 2015 with cloud cover <10%. ENVI 4.7 software was used to process satellite imageries following radiation and geometric correction, image finishing, projection transformation, mosaicked and trimming of the images. From the imageries, three selected bands-visible red band (B3, 0.63–0.69 m), near infrared band (B4, 0.78–0.90 m) and short-wave infrared band (B5, 1.55–1.75 m) representing vegetation growth, coverage and biomass were derived; NDVI were calculated as.

$$NDVI = (B4-B3)/(B4+B3) \tag{2}$$

2.2.2. Urban-specific variables

We selected four socio-economic factors or urban-specific variables that reflect the process of urbanization in Dalian, and they were population (POP), gross domestic product (GDP), distance to the socio-economic center (DSE) or hot-spots, and distance to roads (DR). The spatial datasets of POP and GDP were based on the statistical data in all townships of the city. Firstly, the distribution weights of GDP, and POP at different land use type, night light brightness and residential density were calculated. Then the total weight of each county administrative unit was derived on the basis of standardization of the above three aspects. Finally, on the basis of calculating the proportion of unit weight GDP and POP of each town-level administrative unit, the grid space calculation is used to combine the population on unit weight with the total weight distribution map to make the population spatialized. The specific methods can be found in the study of Liu et al. (2005). The socio-economic center, roads, land use type, night light

brightness and residential density data were downloaded from Institute of Geographic Sciences and Resources, Chinese Academy of Sciences, Resource and Environment Cloud Data Platform (http://www.resdc.cn/). The 1-km grid space calculation was conducted in ArcGIS 10.2 to combine the POP number on the unit weight with the GDP and total weight distribution map:

$$POP_{ij} = POP \times (Z_{ij}/Z) \tag{3}$$

$$GDP_{ij} = GDP \times (Z_{ij}/Z) \tag{4}$$

where POP_{ij} and GDP_{ij} are grid unit values; POP and GDP are statistical values of the township administrative unit where the grid unit is located; Z_{ij} is the total weight of land use type, night light brightness and residential density of the grid unit; Z is the total weight of land use type, night light brightness and residential density of the county administrative unit where the grid unit is located.

In addition, since the socio-economic variables based on the primary census (e.g., POP and GDP) are only at the township level, the accuracy of the data may not be high for mapping. We chose two additional variables to better reflect the process of urbanization, including DSE and DR. The city's socio-economic centers and roads are the main factors to attract migrants which bring many benefits such as transportation and navigation, water supply, entertainment, natural protection and landscape aesthetics. All urban-specific-related variables were calculated at a spatial resolution of 30 m using ArcGIS 10.2.

2.3. Prediction model

Random forest is a machine learning algorithm based on classification and regression tree (CART) proposed by Leo Breiman in 2001 (Grimm et al., 2008). This method combines bootstrapping and feature random selection (Breiman, 2001) during the prediction. The advantages of RF are embodied in two aspects: (1) when training each tree, a subset of the training samples is selected randomly for training and the error evaluation is carried out with out-of-bag (OOB) data (Wiesmeier et al., 2011); (2) at each node, a subset of all features is randomly selected to calculate the optimal segmentation method. Because RF is a model based on random method, which contains multiple classification and regression trees, it ensures the diversity and stability of the model, and can be used to solve the related problems of classification and regression (Were et al., 2015). Therefore, it is widely used in DSM research.

RF model requires three parameters defined by the user, namely the number of trees (ntree), the number of random samples as candidate variables (mtry) at each split, and the minimum size of the terminal node (node) (Yang et al., 2016). The default value of ntree is generally set to 500, but we tested ntree set as 500, 1000, 1500, 2000, and obtained the best prediction when ntree was set to 1500. Similarly, mtry, which is a very sensitive parameter, determines the strength of each tree in the model and the correlation between them. As it decreases, the intensity and correlation of each tree weakens. For the node, we use its default value, setting it to 5. The relative importance (RI) of the prediction variables generated in RF model was estimated according to the average decline in the prediction accuracy when the variable

Table 1
Summary statistics of SOC stocks (kg m^{-2}) under different land use patterns during two periods.

Year	Land use Pattern	Number of samples	Min.	Median	Mean	Max.	SD	CVs (%)	Skewness	Kurtosis
1990	Cultivated	231	0.44	2.10	2.04	4.50	1.16	57.13	0.34	1.09
	Grassland	187	0.33	2.26	2.17	4.98	0.94	43.53	0.46	0.96
	Forest	105	0.51	2.36	2.28	4.41	1.00	43.97	0.38	0.92
2015	Cultivated	359	0.14	1.43	1.34	3.09	0.71	52.90	0.46	0.89
	Grassland	247	0.50	2.21	2.06	3.66	0.74	35.77	0.51	0.92
	Forest	241	0.38	2.29	2.24	4.10	0.68	30.54	0.37	0.98

Note: Min., minimum; Max., maximum; SOC, soil organic carbon; SD, standard deviation; CV, coefficient variation.

Table 2
Pearson's correlation coefficients between ln-transformed soil organic carbon (SOC) stocks with all predictors in 1990 and 2015 surveys based on collected samples.

Property	lnSOC	ELE	SA	SG	TWI	B3	B4	B5	NDVI	POP	GDP	DSE
1990												
ELE	0.47**											
SA	−0.07	−0.21**										
SG	−0.28**	−0.57**	0.46**									
TWI	0.08*	0.18**	−0.14**	−0.36**								
B3	−0.32**	−0.17**	0.06	0.09	0.04							
B4	0.46**	0.07	−0.08	−0.08	0.07	0.34**						
B5	−0.35**	−0.06	−0.06	−0.06	0.15**	0.68**	0.17**					
NDVI	0.61**	0.25**	−0.07**	−0.19**	0.06	−0.59**	0.39**	−0.35**				
POP	−0.22**	−0.24**	−0.16*	−0.23**	−0.27**	−0.21**	−0.26**	−0.19**	−0.43**			
GDP	−0.09*	−0.13**	−0.09	−0.13*	0.09	−0.17	−0.11	−0.15	−0.13	0.57**		
DSE	−0.13*	−0.12**	−0.14*	−0.12**	0.18	0.07	0.05	0.08	0.09	−0.45**	−0.34**	
DR	0.05	−0.17**	−0.21**	0.07	−0.15*	−0.13	−0.09	−0.14*	−0.11*	−0.36**	−0.41**	0.43**
2015												
ELE	−0.71**											
SA	−0.21**	0.15**										
SG	−0.64**	0.58**	0.18**									
TWI	0.55**	−0.46**	−0.34**	−0.71**								
B3	0.23**	−0.24**	−0.05	−0.18**	0.18**							
B4	0.03	−0.04	−0.08	−0.07	0.06	0.35**						
B5	0.08*	−0.11**	−0.06	−0.11*	0.09*	0.72**	0.14**					
NDVI	0.26**	0.25**	−0.03	0.23**	−0.21**	−0.63**	0.23**	−0.57**				
POP	−0.76**	−0.21**	−0.22**	−0.19*	−0.23**	−0.19**	−0.28**	−0.18**	−0.34**			
GDP	−0.55**	−0.08	−0.08	−0.13**	0.14*	0.06	−0.07	−0.13*	−0.16	0.48**		
DSE	−0.57**	−0.16**	−0.13*	−0.14*	0.12	0.14	0.23*	0.08	0.11	−0.53**	−0.37**	
DR	0.37**	−0.23**	−0.15*	0.07	−0.17*	−0.16	−0.11	0.08	0.11	−0.44**	−0.33**	0.39**

Note: Significant relationship between two variables with $p < 0.05$ shown as "*"; $p < 0.01$ shown as "**". ELE, elevation (m); SA, slope aspect (degree); SG, slope gradient (degree); TWI, topographic wetness index; B3, Landsat TM band 3 (digital number); B4, Landsat TM band 4 (digital number); B5, Landsat TM band 5 (digital number); NDVI, Normalized Difference Vegetation Index; POP, population (Person/km^2); GDP, gross domestic product (10,000 yuan / km^2); DSE, distance to the socio-economic center (km); DR, distance to the roads (km).

was removed from the pool (Yang et al., 2016). RF model was generated in R statistics software version 3.2.2 (R Development Core Team, 2013) using "randomForest" package. One-hundred model iterations were conducted and their average was reported as the final prediction and, their standard deviation as prediction uncertainty.

2.4. Calculation of SOC stocks

This study focused on the spatial-temporal change of SOC stocks in topsoil, and was calculated following Eq. (5). The stocks were calculated for two different periods, i.e., for 1990 and for 2015:

$$SOCD = \sum_{i=1}^{k} SOC_c \times BD_i \times LT_i \times (1 - V_i) \tag{5}$$

where $SOCD$ (kg m^{-2}), SOC_c (g kg^{-1}), BD$_i$ (g cm^{-3}), LT_i (m) and V_i represent SOC density, SOC content, soil bulk density, layer thickness, and the volume fraction of fragments >2 mm diameter in soil layer i, respectively.

2.5. Model validation

A 10 fold cross-validation technique was used to evaluate the predictive performance of RF models. Four classical quality evaluation indices, namely absolute prediction error (MAE), root mean square error (RMSE), determination coefficient (R^2), and Lin's concordance correlation coefficient (LCCC) (Lin, 1989) were calculated to evaluate the accuracy (Eqs. (6)–(9)).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |a_i - b_i| \tag{6}$$

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

$$R^2 = \frac{\sum_{i=1}^{n} (b_i - \overline{a}_i)^{2^n}}{\sum_{i=1}^{n} (a_i - \overline{a}_i)^2} \tag{8}$$

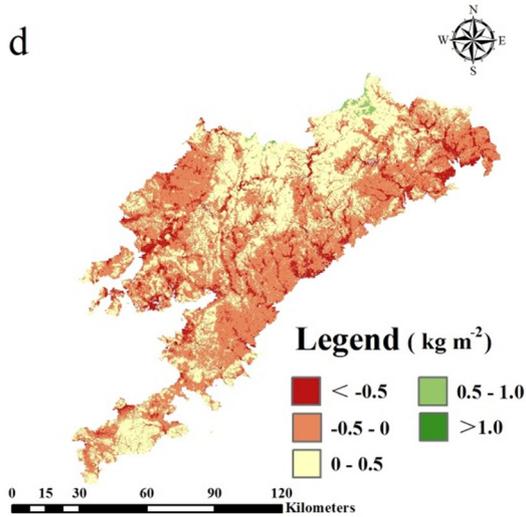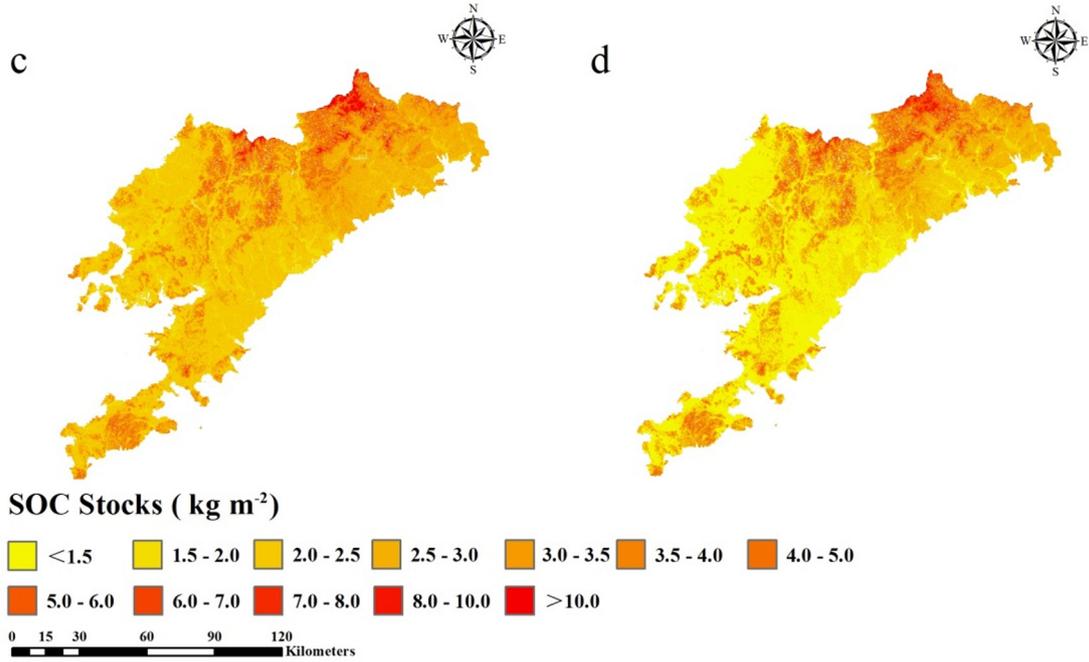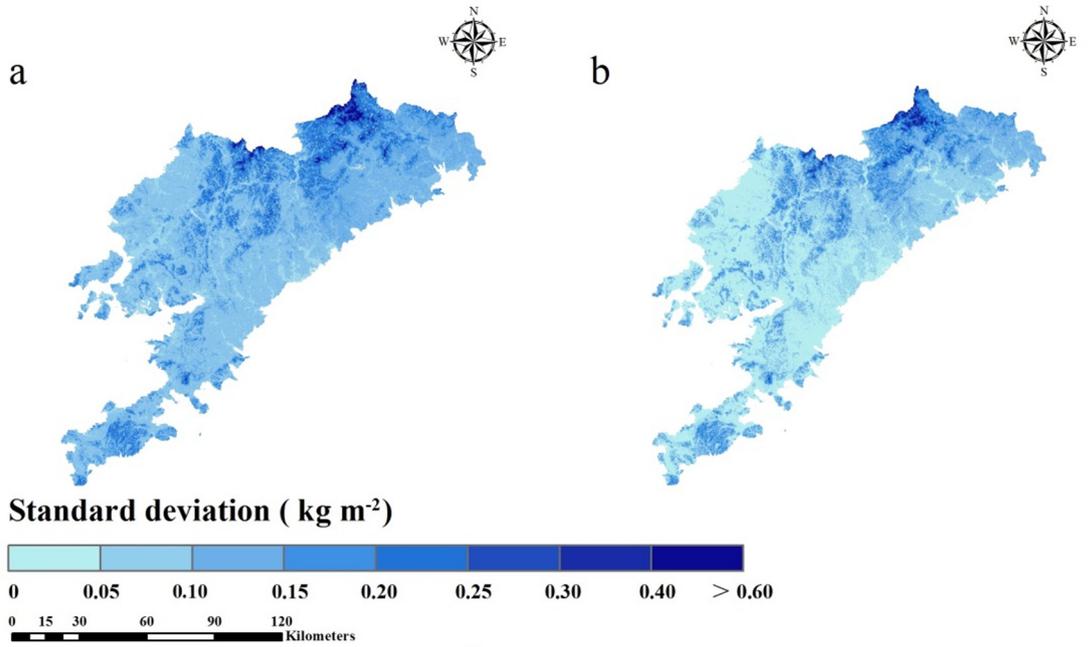$$LCCC = \frac{2r\partial_a\partial_b}{\partial_a^2 + \partial_b^2 + (\overline{a} + \overline{b})^2} \tag{9}$$

where a_i and b_i were the predicted and observed SOC density values at site i; ∂_a and ∂_b were the variances of the predicted and observed values; r is the Pearson's correlation coefficient between the predicted and observed values; n is the number of samples.

Table 3
Summary statistics of the predictive quality of random forest (RF) in 1990 and 2015 for soil organic carbon (SOC) stocks prediction with 100 runs.

Year	Model	Index	Min	Median	Mean	Max	SD
1990	Model A	MAE	0.33	0.34	0.34	0.35	0.13
		RMSE	0.47	0.48	0.49	0.51	0.26
		R^2	0.43	0.53	0.52	0.60	0.07
		LCCC	0.64	0.69	0.69	0.71	0.05
	Model B	MAE	0.20	0.20	0.20	0.21	0.08
		RMSE	0.26	0.26	0.26	0.27	0.11
		R^2	0.49	0.54	0.54	0.58	0.13
		LCCC	0.65	0.69	0.68	0.70	0.03
2015	Model A	MAE	0.41	0.42	0.41	0.43	0.18
		RMSE	0.58	0.60	0.61	0.64	0.31
		R^2	0.42	0.53	0.52	0.60	0.09
		LCCC	0.64	0.69	0.69	0.71	0.06
	Model B	MAE	0.24	0.25	0.24	0.25	0.11
		RMSE	0.30	0.30	0.32	0.31	0.09
		R^2	0.54	0.61	0.61	0.65	0.04
		LCCC	0.73	0.77	0.76	0.78	0.02

Note: SOC, soil organic carbon; Model A included only traditional variables; Model B included all variables (traditional variables and urban-specific variables); MAE, the mean absolute error; RMSE, the root mean squared error; R^2, the coefficient of determination; LCCC, Lin's concordance correlation coefficient.

Standard deviation (kg m⁻²)

0   0.05   0.10   0.15   0.20   0.25   0.30   0.40   > 0.60

0 15 30   60   90   120 Kilometers

SOC Stocks (kg m⁻²)

<1.5   1.5 - 2.0   2.0 - 2.5   2.5 - 3.0   3.0 - 3.5   3.5 - 4.0   4.0 - 5.0

5.0 - 6.0   6.0 - 7.0   7.0 - 8.0   8.0 - 10.0   >10.0

0 15 30   60   90   120 Kilometers

Legend (kg m⁻²)

< -0.5   0.5 - 1.0
-0.5 - 0   >1.0
0 - 0.5

0 15 30   60   90   120 Kilometers

3. Results

3.1. Exploratory statistics

As a result of rapid urbanization during the past few decades, the land use pattern in the study area has changed dramatically. Therefore, we provided the summary statistics of the SOC stocks of the two periods according to different land use type (Table 1). In 1990, the SOC stock of cultivated land ranged from 0.44 kg m^{-2} to 4.50 kg m^{-2}, with an average of 2.04 kg m^{-2}. Correspondingly, its average SOC stocks in 2015 declined to 1.34 kg m^{-2}. By comparing the land use distribution patterns of the two periods, it was found that the SOC stocks of the cultivated land in particular, were decreasing in 2015. In 1990, the standard deviation (SD) and coefficient of variation (CVs) of SOC stocks were 1.16 kg m^{-2}, and 57.1%, respectively. In 2015, the corresponding SD and CVs were 0.71 kgm^{-2} and 52.9%. However, the SOC stocks of grasslands and forests did not change significantly in 1990 and 2015. We also observed an approximate lognormal (ln) distribution of SOC stocks in each land use type in the both periods (Table 1).

The Pearson's correlation between ln-transformed SOC stocks and the auxiliary variables for both periods is shown in Table 2. SOC stocks were significantly correlated with POP for both periods, and the correlation coefficient in 2015 (−0.76) was greater than that in 1990 (−0.22). In addition, the correlation coefficients of urban specific variables such as GDP, DSE and DR with SOC stocks in 2015 (−0.55, −0.57 and 0.37) were higher than those in 1990. Similarly, there was a significant and higher correlation between remote sensing image variables (B3, B4, and B4) and SOC stocks in 1990 compared to that in 2015.

In order to alleviate multi-collinearity, we adopted the stepwise linear regression method to reduce the closely related predictive variables. The coefficients of variance expansion of all environmental variables in 1990 was no >5, while in 2015 it was no >4. The results showed that the coefficients of variance expansion of all covariates was <10 (Wang et al., 2018a), suggesting no multi-collinearity in SOC stocks prediction and modeling in the both periods.

3.2. Performance evaluation and uncertainty

A 10-fold cross-validation technique was selected to evaluate the predictive performance of two RF models with 100 iterations, and results are summarized in Table 3. During two periods, Model B including all variables (traditional variables, and urban-specific variables) showed a better prediction compared to Model A that only included traditional variables as SOC predictors.

The full variable model (Model B) predictions had lower uncertainties with a mean SD of 0.17, and 0.08 kg m^{-2} for 1990, and 2015, respectively, compared to Model A (Fig. 2a and b). The results showed that the RF of the full variable model produced lower MAE and RMSE, but higher R^2 and LCCC in the both periods (Table 3). The results also verified that the RF model had an excellent prediction performance on SOC stocks distribution in both periods.

3.3. Changes in SOC stocks

The final predicted SOC stocks maps for 1990, and 2015 represented a mean of 100 individual maps from 100 model iterations (Fig. 2), and the best predicted maps for 1990, and 2015 were the ones derived from Model B (full variable model). We also observe that there was a substantial variability among 100 different models for both periods. For the best predicted maps, average SOC stocks were higher in 1990 (2.77 ± 1.09 kg m^{-2}) than that in 2015 (2.16 ± 0.93 kg m^{-2}) (Fig. 2c and d).

In order to further explore the spatial characteristics of SOC stocks, we reported SOC stocks of different soil groups in 1990 and 2015 (Table 4). Luvisols, and Cambisols together accounted for 91% total SOC stocks in 1990, and 92% in 2015, respectively. (3% vs. 3%) and Fluvisols (5% vs. 4%) were the other soil types that contained a large amount of SOC stocks in 1990, and 2015, respectively. Overall, except for the aggregation effect of SOC stocks in the central and northwestern forest areas of Dalian, SOC stocks in other areas showed a decreasing trend during 25 years.

3.4. Importance of the variables

Variable importance revealed different dominating environmental variables on the spatial distribution of SOC stocks in 1990, and in 2015 (Fig. 3). We found that urban-specific variables were the primary variables influencing SOC stocks in 2015, and RS indices in 1990. In 2015, the urban-specific variables (RI >42%) were important indicators of spatial variability of topsoil SOC stocks in rapidly urbanized areas. In order to further explore the effect of urbanization on SOC stocks, we analyzed the variation of SOC stocks caused by changes in land use pattern in the process of urbanization (Table 5 and Fig. 4). With the increase of cultivated land reclamation years, the decrease of SOC stocks is the largest, followed by the conversion of grassland to cultivated land, accounting for 82% and 15% of the decrease under the change of major land use patterns in the region. The main way to increase is to convert cultivated land to forest land, and the SOC stocks increased by 0.11 Tg in 25 years (Table 5).

4. Discussion

4.1. Model performance

A full-variable model (model with traditional variables and urban-specific variables), showed a better predictive performance, with lower MAE and RMSE, and higher R^2 and LUCC in the both periods compared to the model A that excluded urban-specific variables in the prediction. Wang et al. (2018c) reported a higher performance when all environmental variables were selected compared to a model that excluded reclamation years in SOC prediction in Northeast China. Although the sampling method, variable selection and validation method were different from the previous SOC mapping studies, our prediction evaluation results were comparable. For example, Yang et al. (2016) used a RF model to predict SOC in Qilian region of Northwest China, and could explain 68% of spatial variance of SOC in the region. Were et al. (2015) developed an RF model to map the SOC stocks across a montane landscape and found that the model could account for 52% of the spatial variations of SOC stocks. However, few similar studies had lower performance results compared to the present study. For example, while predicting SOC stocks of the Dano catchment (Southwest Burkina Faso) using RF model, Hounkpatin et al. (2018) could only explain 14% of SOC stocks variability.

4.2. Estimates of SOC stocks and their changes

During the two periods, the spatial distribution SOC stocks of forest areas in the southwest and northeast of Dalian showed an increasing trend, while that of agricultural areas in the middle of Dalian showed a decreasing trend. SOC stocks were often concentrated in areas with better vegetation coverage and rich biomass, which has a great uncertainty for the areas with strong human disturbance. In 1990, we found a high correlation between regional SOC stock and vegetation variables, especially NDVI, which had been confirmed by previous studies. In Seoul National Forest Park of South Korea, Bae and Ryu (2015) analyzed

Fig. 2. Standard deviation and spatial distribution of soil organic carbon (SOC) predicted by full variable model (model that used traditional as well as urban-specific variables) using random forest model in 1990 (a, c) and 2015 (b, d), and spatial distributions of soil organic carbon (SOC) stocks change between the1990 and 2015 surveys (f).

Table 4
Soil organic carbon stocks in the topsoil (0—30) depth according to the World Reference Base for Soil Resources (WRB) in 1990 and 2015.

Soil groups	Area (km^2)	Average SOC stock (kg m^{-2})		SOC stock (Tg)	
		1990	2015	1990	2015
Anthrosols	565	1.95	1.76	1.10	0.99
Luvisols	4272	3.63	3.45	15.51	14.74
Cambisols	6804	3.37	3.00	22.93	20.41
Gleysols	40	1.81	1.6	0.07	0.06
Solonetz	216	1.82	1.56	0.39	0.34
Histosols	4	4.19	3.56	0.02	0.01
Phaezems	15	4.16	4.05	0.06	0.06
Fluvisols	1321	1.48	1.2	1.96	1.59
Sum	13,237	–	–	42.04	38.21

the soil sampling data of 2003, and 2013 combined with NDVI of the two periods, and found the main reasons for the change of land use, the expansion of plant area and the growth of plants. In addition to vegetation related variables, SOC stocks were closely related to urbanization-related variables in 2015, especially POP variables. Recent studies have shown that urbanization-related variables could be used as key environmental variables to effectively predict SOC stocks (Vasenev et al., 2014, 2018).

In the past 25 years, the decrease of SOC stocks was mainly in the areas that have experienced transitions from grassland to cultivated land, from forest land to cultivated land, and from cultivated land to cultivated land, which decreased by 0.61Tg, 0.01Tg and 3.40Tg, respectively. Those areas were mainly distributed in the central area of Dalian, which were the traditional agricultural area. In addition, these areas are relatively flat, and are close to the town for development and utilization, which have been strongly disturbed by human activities. As a result, land use transformation has been rapid, and the topsoil SOC stocks in this area showed a downward trend. In Wafangdian District of Dalian, Wang et al. (2018c) concluded that land reclamation and related human activities were the main reasons for the decrease of topsoil SOC stocks. Similar conclusions have been obtained in Bae and Ryu (2015).

Throughout the whole study area, the SOC stocks showed an increasing trend in the northeast and southwest forest areas, which was due to the dense vegetation coverage in those areas and little human interference. In addition, some parts in the middle of the study area were also showing an increasing phenomenon, which was related to the current national policy of returning farmland to forest or grass, resulting in the increasing trend of topsoil SOC stocks. Wang et al. (2018c) selected 9 main environmental variables influencing the spatial variability of

SOC stocks, and combined with a BRT model to predict the topsoil SOC stocks in Wafangdian of Northeast China. Based on their results, the increase of forest SOC stocks in the northeast of the study area was closely related to the policy of returning farmland to forest and grass implemented by the government for many years.

Of all soil groups, SOC stocks were mainly stored in Luvisols, and Cambisols during the two periods. Luvisols refers to the soil with obvious clay leaching and deposition, which is fully leached by lime under a moist soil condition (Schöning et al., 2006). However, due to the influence of regional climate, the degree of soil leaching is limited. In addition, farmers often use a large amount of fertilizer (Jacinthe et al., 2002) in the relatively flat farming areas in the middle of Dalian. Although the distribution of farming areas was not the most extensive, its SOC stocks were the largest in all soil groups.

Cambisols has the most widely distributed area, accounting for about 51% of the whole study area, and the SOC stocks accounts for about 52% of the total stocks during the two periods. Therefore, in order to accurately and reasonably manage Cambisols in the study area, appropriate land management practices should be followed considering the large size of SOC stocks in this particular soil.

4.3. Effects of urban-specific variables on SOC stocks

For the both periods, general spatial patterns showed a higher SOC stocks level in the northeast but lower level in the southwest of Dalian City. There were abundant SOC stocks in areas with dense forest. The spatial pattern of SOC stocks in 1990 was closely related to remote sensing-related variables (Table 2) as these variables mostly represent vegetation growth, coverage and biomass, and human impact on nature was relatively small during this period. Similar results were also reported by Wang et al. (2018b). On the contrary, this pattern was closely related to GDP and POP in 2015 as both variables were more related to SOC in 2015 compared to the same variables in 1990 (Table 2). This conclusion agreed with the findings of previous studies (Liu et al., 2016; Xia et al., 2017; Vasenev et al., 2014, 2018) verifying that the urban-specific information could be used to map SOC stocks distribution in a rapidly urbanized regions with intense human activity. Vasenev et al. (2018) pointed out that such indicators were outstanding predictors of SOC stocks, which could indirectly affect soil microbial acquisition, thereby affecting the level of SOC stocks.

Previous studies have revealed that rapid urbanization significantly impacts the spatial-temporal changes in SOC stocks distribution (Liu et al., 2016; Xia et al., 2017; Vasenev et al., 2018). We also found that the rapid urbanization in Dalian during the past decades has led to
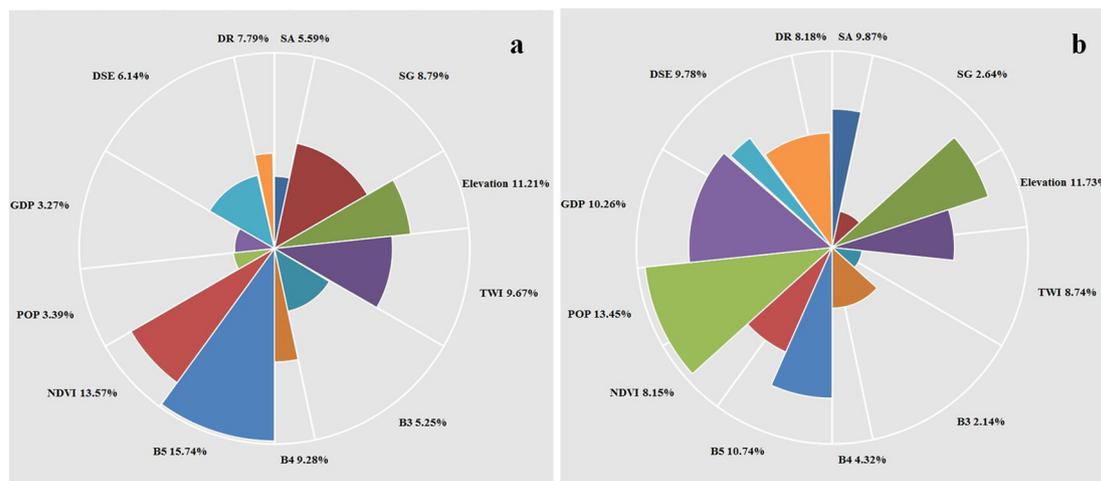
Fig. 3. Relative importance of variable used as soil organic carbon stocks prediction in 1990 (a), and 2015 (b) surveys, which are normalized to 100%. SA, slope aspect; SG, slope gradient; TWI, topographic wetness index; B3, Landsat TM band 3; B4, Landsat TM band 4; B5, Landsat TM band 5; NDVI, Normalized Difference Vegetation Index; POP, population; GDP, gross domestic product; DSE, distance to the socio-economic center; DR, distance to the roads.

Table 5
Change of soil organic carbon (SOC) stocks under major land use patterns during 1990–2015.

Major land use types	Area (km^2)	SOC stocks (Tg)		Change (Tg)
		1982	2015	
Cultivation-cultivation (C-C)	6081.55	22.79	19.39	−3.40
Cultivation-forest (C-F)	673.77	1.78	1.89	0.11
Cultivation-grassland (C-G)	208.98	0.54	0.57	0.03
Grassland-grassland (G-G)	831.43	2.33	2.24	−0.09
Grassland-cultivation (G-C)	2107.27	7.58	6.97	−0.61
Grassland-forest (G-F)	311.19	1.27	1.41	0.14
Forest-forest (F-F)	1234.03	3.24	3.32	0.08
Forest-cultivation (F-C)	18.33	0.05	0.04	−0.01
Forest-grassland (F-G)	28.5	0.10	0.07	−0.03
Sum	11,495.05	39.68	35.9	−3.78

dramatic changes in land use patterns substantially affecting SOC stocks distribution. Xia et al. (2017) reported a decreasing trend in SOC stocks in Northeast China, mainly due to the rapid development and urbanization caused by changes in land use patterns. Our study found a slight increase in SOC stocks in the central and northwestern mountainous areas. In other areas, especially in coastal plain farming areas, there

has been a decline (Fig. 3a). SOC stocks decreased (−5 to 0 kg m^{-2}) mainly in the eastern and coastal plains of the study area that accounted for about 52% of the total area (Fig. 3b). The maximum increase (>1.0 kg m^{-2}) occurred in the northwest mountainous area. The area with SOC stocks increased accounted for 36% of the total study area and was mainly distributed in the central and northwestern mountainous areas.

The RI of urban-specific variables (POP and GDP) was lower in 1990. The weaker influence of urban-specific variables in 1990 could be attributed to the slow urbanization process, and less pressure on land use with optimal land use management. With the development of China's reform and opening-up policy, the scale of development in Dalian has been expanding with rapid population growth and dense road networks. Rapid urbanization not only changes the land use pattern, but also changes the properties of topsoil, thus changing the characteristics of soil carbon. As observed in the present study, a significant impact of rapid urbanization on regional SOC stocks and balance has also been reported (Wiesmeier et al., 2012; Edmondson et al., 2014; Liu et al., 2016; Xia et al., 2017; Vasenev et al., 2018; Stumpf et al., 2018).

Changes in land use patterns will lead to changes in soil environment, such as soil properties and water balance, which indirectly affect the rate of accumulation and decomposition of soil organic matter
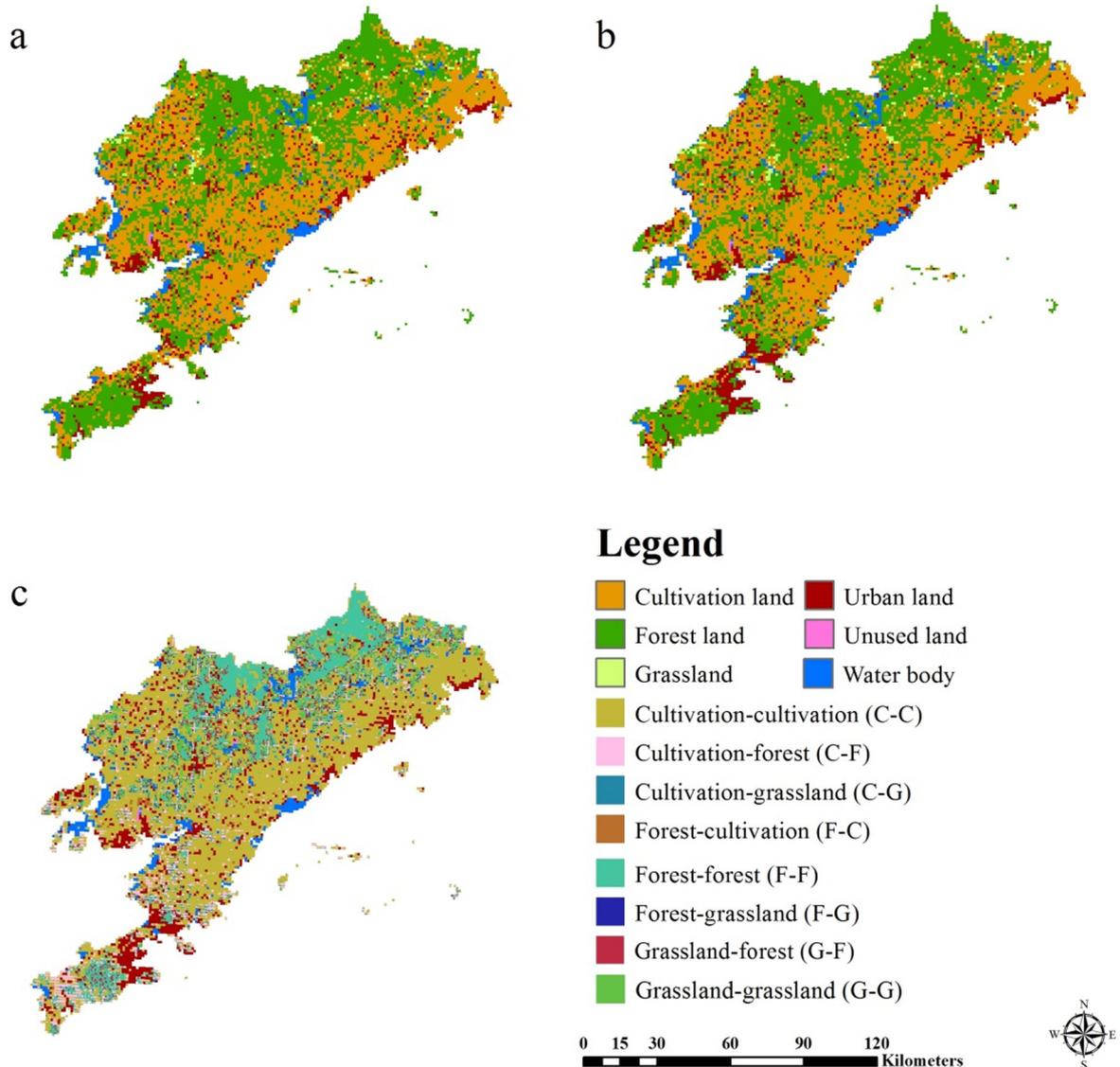
Fig. 4. Land use classification maps in 1990 (a), and 2015 (b) and their change map (c).

(Muñoz-Rojas et al., 2015; Liu et al., 2016; Adhikari et al., 2019). In Jiangsu Province of China, Zhao et al. (2015) reported the effects of land use change on soil organic matter. When cultivated land was converted to forest, and to grassland during the past 25-year, SOC stocks increased by 0.11 Tg, and 0.03 Tg, respectively (Table 5). In contrast, the amount of SOC stocks decreased by 3.40 Tg in 2015 when cultivated land remained unchanged. Overall, from 1990 to 2015 (since the reform and opening up of China), the scale and speed of soil utilization has been increased, and SOC stocks had decreased by 3.78 Tg in total. However, during this period, the Chinese government announced the implementation of the policy of returning farmland to forestry and grassland, which resulted in the accumulation of litter in soil, and the increase of underground root biomass, resulting in the increase of SOC (Zhao et al., 2015; Wang et al., 2018c). These results indicated that land use changes play a crucial role in SOC stocks changes. In addition, conversion of land use to forests have increased soil microorganisms and biological activity, accelerated the turnover of organic matter, deepened the rooting system of trees, and facilitated the accumulation of soil organic carbon (Zhao et al., 2015; Wang et al., 2018c; Vasenev et al., 2018).

4.4. Uncertainties in the present study

Apart from the RF model uncertainty associated with this prediction, we believed that there might be other sources as well. Firstly, POP and GDP data were only collected at 1-km grid, which might have reduced our prediction accuracy. The data for the rest of auxiliary variables are at 90 m resolution, while POP and GDP data are at 1 km resolution. Therefore, it is necessary to resample the 90-m grid data to 1 km in our modeling. Secondly, soil samples in 1990 were collected from the historical data (Second National Soil Survey in China). Limited by the personnel and equipment at that time, sampling and experimental errors might have occurred. Thirdly, due to the lack of BD data of some soil profiles in 1990, we used the pedotransfer functions to estimate BD. However, PTFs cannot accurately estimate the actual bulk density due to prevailing land use complexity, and variations in soil types and conditions in the study area, which would further increase prediction errors. Fourthly, with urbanization development, urban built-up areas were expanding. Due to some limitation, we only collected soil data from the cultivated, forest, and grassland, and not from the urban built-up areas. Therefore, the topsoil SOC stocks in urban built-up areas might be over- or under-estimated by our model. Finally, our study predicted SOC stock only for the topsoil depth, which might have underestimated SOC stocks, as deeper soils also contain SOC.

5. Conclusion

Two RF models with both traditional and urban-specific variables and traditional variables only were used to predict the spatial-temporal distribution of SOC stocks in a rapidly urbanized city of Dalian, Liaoning, China. Adding urban-specific variables improved the accuracy of topsoil SOC stocks prediction. The mean SOC stocks in the study area were 2.17 kg m^{-2} (± 1.09), and 2.16 kg m^{-2} (± 0.93) for 1990, and for 2015, respectively. The full variable model explained almost 54%, and 61% of the spatial variance of SOC stocks in northeastern coastal agroecosystems in 1990 and 2015, respectively. The spatial distribution of SOC stocks was well explained by urban-specific-related variables in 2015. These variables were important and practical indicators reflecting topsoil SOC stocks of rapidly urbanized areas. Therefore, future SOC mapping studies, especially in areas with rapid economic development and intense land use change, urban-specific-related variable should be selected as the main environmental variable. We believe that our prediction and the SOC stock maps thus generated will have impacts in urban environmental planning and government decision-making in our study region.

CRediT authorship contribution statement

Shuai Wang: Conceptualization, Investigation, Methodology, Writing - original draft. **Kabindra Adhikari:** Writing - review & editing. **Qianlai Zhuang:** Writing - review & editing. **Hanlong Gu:** Funding acquisition, Investigation. **Xinxin Jin:** Data curation, Methodology.

Acknowledgments

Declaration of competing interest

The authors declare no conflict of interest.

References

Adhikari, K., Hartemink, A.E., 2015. Digital mapping of topsoil carbon content and changes in the Driftless Area of Wisconsin, USA. Soil Sci. Soc. Am. J. 79, 155–164.

Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS One 9 (8), e105519.

Adhikari, K., Owens, P.R., Libohova, Z., Miller, D.M., Wills, S.A., Nemecek, J., 2019. Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. Sci. Total Environ. 667, 833–845.

Bae, J., Ryu, Y., 2015. Land use and land cover changes explain spatial and temporal variations of the soil organic carbon stocks in a constructed urban park. Landsc. Urban Plan. 136, 57–67.

Bartholomeus, H., Kooistra, L., Stevens, A., van Leeuwen, M., van Wesemael, B., Ben-Dor, E., Tychon, B., 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. Int. J. Appl. Earth Obs. 13 (1), 81–88.

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

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brus, D.J., Yang, R.M., Zhang, G.L., 2016. Three-dimensional geostatistical modeling of soil organic carbon: a case study in the Qilian Mountains, China. Catena 141, 46–55.

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

Dalal, R.C., Allen, D.E., Wang, W.J., Reeves, S., Gibson, I., 2011. Organic carbon and total nitrogen stocks in a Vertisol following 40 years of no-tillage, crop residue retention and nitrogen fertilisation. Soil Till. Res. 112 (2), 133–139.

Don, A., Schumacher, J., Scherer-Lorenzen, M., Scholten, T., Schulze, E.D., 2007. Spatial and vertical variation of soil carbon at two grassland sites—implications for measuring soil carbon stocks. Geoderma 141 (3–4), 272–282.

Edmondson, J.L., Davies, Z.G., McCormack, S.A., Gaston, K.J., Leake, J.R., 2014. Land-cover effects on soil organic carbon stocks in a European city. Sci. Total Environ. 472, 444–453.

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using Random Forests analysis. Geoderma 146 (1–2), 102–113.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10, e0125814.

Hounkpatin, 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.

IUSS Working Group, 2014. World Reference Base for Soil Resources 2014 International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome.

Jacinthe, P.A., Lal, R., Kimble, J.M., 2002. Effects of wheat residue fertilization on accumulation and biochemical attributes of organic carbon in a central Ohio Luvisol. Soil Sci. 167 (11), 750–758.

Kaushal, S.S., Lewis, J.W.M., McCutchan, J.J.H., 2006. Land use change and nitrogen enrichment of a Rocky Mountain watershed. Ecol. Appl. 16 (1), 299–312.

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623–1627.

Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.

Liu, H., Jiang, D., Yang, X., Luo, C., 2005. Spatialization approach to 1 km grid GDP supported by remote sensing. Geo-information Sci. 2, 026.

Liu, X., Li, T., Zhang, S., Jia, Y., Li, Y., Xu, X., 2016. The role of land use, construction and road on terrestrial carbon stocks in a newly urbanized area of western Chengdu, China. Lands. Urban Plan. 147, 88–95.

McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., Van Molle, M., 2008. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). Geoderma 143 (1–2), 1–13.

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. Advances in Agronomy. 118. Academic Press, pp. 1–47.

Mishra, U., Riley, W.J., 2012. Alaskan soil carbon stocks: spatial variability and dependence on environmental factors. Biogeosciences 9 (9), 3637–3645.

Ministry of Natural Resources of China, 2017. The classification of land use status in the third national land survey of China (GB/T21010-2017). The State Council of China, Beijing, China http://www.gov.cn/.

Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Mouazen, A.M., 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst. Eng. 152, 104–116.

Muñoz-Rojas, M., Jordán, A., Zavala, L.M., De la Rosa, D., Abd-Elmabod, S.K., Anaya-Romero, M., 2015. Impact of land use and land cover changes on organic carbon stocks in Mediterranean soils (1956–2007). Land Degrad. Dev. 26 (2), 168–179.

Office of Soil Survey in Liaoning Province (OSSLP), 1990. The Soils of Liaoning Province. Agriculture Press, Beijing, China, pp. 57–167 (in Chinese with English abstract). http://csdata.org/p/7/1/.

Phachomphon, K., Dlamini, P., Chaplot, V., 2010. Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables. Geoderma 155 (3–4), 372–380.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://www.rproject.org/.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Sci. Soc. Am. J. 82 (1), 186–201.

Reyes Rojas, L.A., Adhikari, K., Ventura, S.J., 2018. Projecting soil organic carbon distribution in central Chile under future climate scenarios. J. Environ. Qual. 47 (4), 735–745.

Schöning, I., Totsche, K.U., Kögel-Knabner, I., 2006. Small scale spatial variability of organic carbon stocks in litter and solum of a forested Luvisol. Geoderma 136 (3–4), 631–642.

Stewart, C.E., Paustian, K., Conant, R.T., Plante, A.F., Six, J., 2007. Soil carbon saturation: concept, evidence and evaluation. Biogeochemistry 86 (1), 19–31.

Stumpf, F., Keller, A., Schmidt, K., Mayr, A., Gubler, A., Schaepman, M., 2018. Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. Agric. Ecosyst. Environ. 258, 129–142.

Vågen, T.G., Winowiecki, L.A., 2013. Mapping of soil organic carbon stocks for spatially explicit assessments of climate change mitigation potential. Environ. Res. Lett. 8 (1), 015011.

Vasenev, V.I., Stoorvogel, J.J., Vasenev, I.I., Valentini, R., 2014. How to map soil organic carbon stocks in highly urbanized regions? Geoderma 226, 103–115.

Vasenev, V.I., Stoorvogel, J.J., Leemans, R., Valentini, R., Hajiaghayeva, R.A., 2018. Projection of urban expansion and related changes in soil carbon stocks in the Moscow Region. J. Clean. Prod. 170, 902–914.

Wang, S., Adhikari, K., Wang, Q., Jin, X., Li, H., 2018a. Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C: N ratio from the northeastern coastal agroecosystems in China. Ecol. Indic. 84, 263–272.

Wang, S., Jin, X., Adhikari, K., Li, W., Yu, M., Bian, Z., Wang, Q., 2018b. Mapping total soil nitrogen from a site in northeastern China. Catena 166, 134–146.

Wang, S., Zhuang, Q., Jia, S., Jin, X., Wang, Q., 2018c. Spatial variations of soil organic carbon stocks in a coastal hilly area of China. Geoderma 314, 8–19.

Were, K., Bui, D.T., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394–403.

Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil 340, 7–24.

Wiesmeier, M., Spörlein, P., Geuß, U., Hangen, E., Haug, S., Reischl, A., Kögel-Knabner, I., 2012. Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling depth. Glob. Chang. Biol. 18 (7), 2233–2245.

Xia, X., Yang, Z., Xue, Y., Shao, X., Yu, T., Hou, Q., 2017. Spatial analysis of land use change effect on soil organic carbon stocks in the eastern regions of China between 1980 and 2000. Geosci. Front. 8 (3), 597–603.

Yang, R.M., Zhang, G.L., Liu, F., Lu, Y.Y., Yang, F., Yang, F., Li, D.C., 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecol. Indic. 60, 870–878.

Zhao, M.S., Zhang, G.L., Wu, Y.J., Li, D.C., Zhao, Y.G., 2015. Driving forces of soil organic matter change in Jiangsu Province of China. Soil Use Manag. 31 (4), 440–449.