Geoderma 305 (2017) 250–263

Contents lists available at ScienceDirect

Geoderma

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

Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China

Shuai Wang[a,b], Qianlai Zhuang[b], Qiubing Wang[a,*], Xinxin Jin[a], Chunlan Han[a]

[a] College of Land and Environment, Shenyang Agricultural University, Shenyang, Liaoning Province 110866, China
[b] Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Digital soil mapping
Soil organic carbon
Soil total nitrogen
Boosted regression tree
Environmental variables

ABSTRACT

Estimation of carbon and nitrogen stocks is important for quantifying carbon and nitrogen sequestration as well as greenhouse gas emissions and inventorying national carbon and nitrogen balances. For Liaoning province of China, we estimated the vertical distribution of soil organic carbon (SOC), soil total nitrogen (STN), bulk density (BD), and mapped their spatial distribution at five standard soil depth intervals (0–5, 5–15, 15–30, 30–60 and 60–100 cm) using nine environmental variables as predictors including precipitation, temperature, land use, elevation, system for automated geoscientific analyses (SAGA) wetness index, and Normalized Difference Vegetation Index (NDVI). The highest average contents of SOC and STN were 15.2 g kg^{-1} and 1.6 g kg^{-1} in the 0–5 cm soil layer, and 1.5 g kg^{-1} SOC and 0.4 g kg^{-1} STN in the 60–100 cm soil layer, respectively. The prediction precision for SOC, STN and BD all decreased with soil depth. Average SOC and STN stocks for 0–30 cm were 3.1 kg m^{-2} and 0.5 kg m^{-2}, respectively. For the top 1 m, SOC and STN were 4.5 kg m^{-2} and 0.9 kg m^{-2}, respectively. In total, the soils stored approximately 588 Tg SOC and 128 Tg STN within the top 1 m. The soils under forest had the highest amount of carbon (356 Tg) and nitrogen (58 Tg) followed by agriculture and wetland that contributed 34% and 48% of the total stock, respectively. > 91% of the total SOC and STN stocks were in Argosols and Cambosols. We adopted a digital soil mapping method to map the spatial distribution of SOC and STN stocks and predict their uncertainties. The estimation was validated with a 10-fold cross-validation procedure. The data and high-resolution maps from this study can be used for future soil carbon and nitrogen assessment and inventorying.

1. Introduction

Soil organic carbon (SOC) and soil total nitrogen (STN), the essential nutrients for plant growth are major components of the global carbon and nitrogen cycles (Batjes, 1996; Reeves, 1997; Quilchano et al., 2008; Liu et al., 2012; Wang et al., 2016). SOC and STN affect the concentration of greenhouse gases in the atmosphere and global climate change, and become one of the foci of global climate change research in recent years (Powers and Schlesinger, 2002; Bronson et al., 2004; Lal, 2004; Stockmann et al., 2013; Lehmann and Kleber, 2015). Accurate quantification of SOC and STN stocks are important for assessing the C and N sink capacity of soils and the change rate of SOC and STN (L. Yang et al., 2016). Spatially explicit information of SOC and STN thus plays a crucial role in global carbon and nitrogen cycling studies (Batjes, 1996; Minasny et al., 2013; Adhikari et al., 2014; Minasny et al., 2017).

A simple approach to predict the spatial distribution of SOC and STN stocks is to allocate the average SOC and STN stocks to each map unit of soil type or land-use type (Batjes, 1996; Arrouays et al., 2001; Bernoux et al., 2002). However, this approach results in constant values within each map unit which cannot show the large spatial heterogeneity of SOC and STN in each map unit and the error of estimates for the values of average SOC and STN from few SOC and STN data points (Adhikari et al., 2014; L. Yang et al., 2016). To overcome these problems, digital soil mapping (DSM) technology is considered as an appropriate and useful method to produce more detailed spatial variations of SOC and STN stocks with assistance of auxiliary environmental variables (Minasny et al., 2013; Martin et al., 2014; Adhikari et al., 2014; Mulder et al., 2016). Based on DSM technology, Adhikari et al. (2014) combined regression kriging with the equal area spline to model continuous depth function of SOC and bulk density (BD) and calculated the soils stored approximately 570 Tg carbon within the top 1 m in Denmark. In mainland France, Mulder et al. (2016) estimated 1 m deep soil of SOC by applying regression tree and spline function. Therefore, DSM methods provide a rapid and inexpensive way to estimate the spatial distribution of SOC and STN stocks over a large area from a

* Corresponding author at: College of Land and Environment, Shenyang Agricultural University, No. 120 Dongling Road, Shenhe District, Shenyang, Liaoning Province 110866, China.
E-mail address: wangqbsy@yahoo.com (Q. Wang).

http://dx.doi.org/10.1016/j.geoderma.2017.05.048
Received 30 December 2016; Received in revised form 13 May 2017; Accepted 27 May 2017

limited number of sample sites and environmental covariates (McBratney et al., 2003; Minasny et al., 2013; Karunaratne et al., 2014).

Many DSM methods have been used in estimating SOC and STN (Minasny et al., 2013; Cambule et al., 2014; Wang et al., 2016). However, tree-based model had superior predictive performance comparing to other methods using average values by soil map (Carslaw and Taylor, 2009; R. Yang et al., 2016a; Wang et al., 2016). Furthermore, tree-based models had several advantages that are flexible to deal with linear, polynomial, exponential, logistic, periodic, or general nonlinear problems, and improved prediction effectively (Müller et al., 2013; Cheong et al., 2014; Heung et al., 2016). Therefore, tree-based development model of boosted regression trees (BRT) has an excellent predictive performance, combining many simple tree models to form an efficient prediction model compared with the traditional single tree model (Carslaw and Taylor, 2009; Kempen et al., 2011; L. Yang et al., 2016).

Continuous distributions of SOC and STN in vertical dimension are necessary to investigate stocks and controls of SOC and STN (Müller et al., 2013), and also important for understanding the role of SOC and STN in the global C and N cycles and quantifying the environmental controls on SOC and STN distribution (Wang et al., 2004; Adhikari et al., 2014). However, information of SOC and STN stocks at certain depths are sometimes missing. Thereby, soil depth functions are needed to interpolate continuous distributions of SOC and STN (Arrouays et al., 2014). Many soil depth functions have been developed, such as statistical depth functions (Zuo and Serfling, 2000), phytolith depth functions (Hart and Humphreys, 2003), exponential depth function (Meersmans et al., 2009), sigmoid depth function (Zhang et al., 2017), and splines function (Bishop et al., 1999). Among them splines function is now the most frequently used and recommended by GlobalSoilMap specifications (Arrouays et al., 2014).

In the present study, we mapped SOC and STN stocks within the upper 1 m depth in the northeast China (Fig. 1) using equal-area spline profile function and boosted regression tree methods. Our specific objectives are:

(i) Constructing soil depth functions to describe vertical distribution of SOC, STN and BD;
(ii) Deriving predictive models to map the soil depth functions across the study area;
(iii) Evaluating the importance of environmental covariates in controlling the spatial patterns of SOC and STN;
(iv) Producing the spatial distribution of SOC and STN stocks in Liaoning Province of China.

2. Materials and methods

2.1. Study area

Our study is Liaoning Province (118°–125°E, 38°–43°N) in northeast China (Fig. 1). It covers an area about 140,000 km^2, and had a population of 43.75 million in year of 2010 (Bureau of Statistics Liaoning Province, 2011). Liaoning Province is an economically important province in China. It is ranked seventh among all Chinese provinces in terms of the gross domestic product, which was 301 billion US dollars in 2010 (Bureau of Statistics Liaoning Province, 2011). The altitude of the study area ranges from 0 m to 1332 m above sea level. The region is dominated by a warm temperate continental monsoon climate, four distinct seasons. The mean annual precipitation ranges from 600 mm in the northeast to 1100 mm in the southwest, 60–70% of which falls between June and August, mostly in high intensity rainstorms. The mean annual temperature is 7–11 °C, with the coldest temperature of − 40 °C in January and the hottest temperature of 30 °C in July.

According to Chinese Soil Taxonomy (Cooperative Research Group on Chinese Soil Taxonomy, 2001), Argosols and Cambosols are the main soil types (32% and 51%), distributing in the Low Mountain and hilly areas of the northeast and southwest of Liaoning Province. The main land use in Liaoning Province can be classified as agriculture, forest, grassland and wetland. Agriculture areas are mainly distributed in the Liaohe plain area of central Liaoning, accounting for 43.3% of the study area, the main economic crop are corn, rice, sweet potatoes, apples, and cherries (Bureau of Statistics Liaoning Province, 2011).

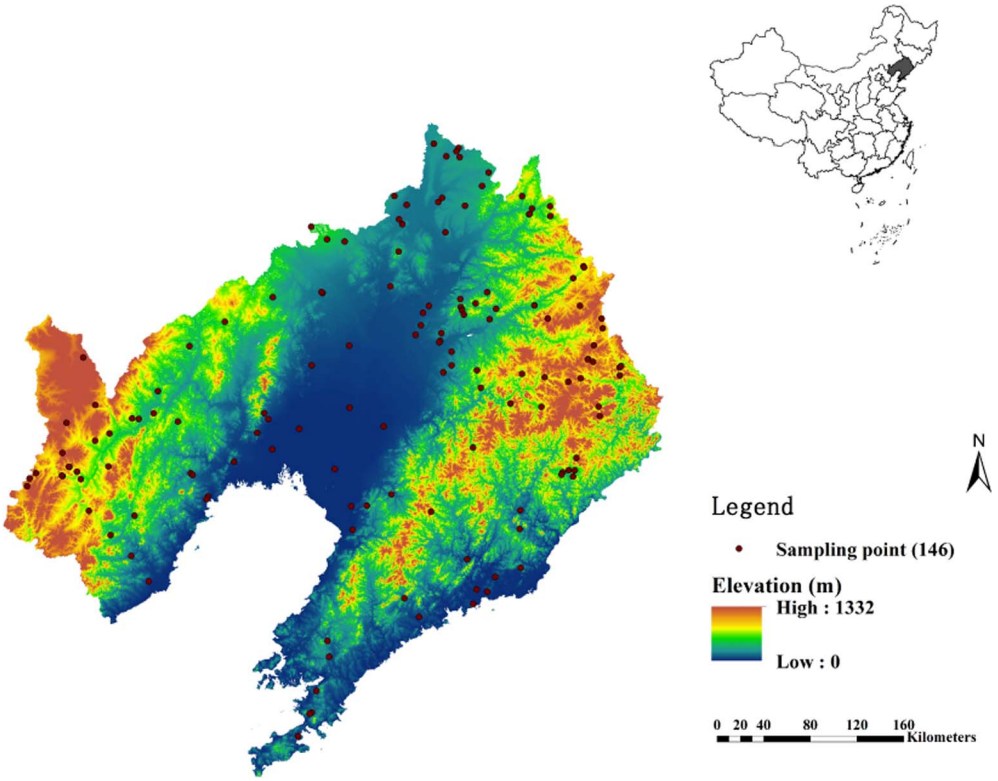

Fig. 1. Location of the study area and 146 sampling sites, which are superimposed on a 90-m resolution DEM.

2.2. Field sampling and laboratory analysis

Extensive field sampling is very labor intensive and costly. A proper sampling strategy is needed for characterizing the SOC and STN with high spatial variability (L. Yang et al., 2016; R. Yang et al., 2016a, 2016b) at such a large scale of our study area. In this study, a purposive sampling strategy was adopted. Based on the pedogenesis of SOC and STN of the study area, soil types, land use and terrain conditions were the main environmental factors considered in designing samples (Hudson, 1992). The study area was first stratified using the combination of soil type and landuse type (Yang et al., 2013). Thirty-two soil-land-use units were obtained using fuzzy c-means classification method based on eight soil subgroups from the Second National Soil Survey and five land-use types including agriculture, forest, grassland, wetland, and other types (e.g., urban area, industry, roads). Then within each soil-land-use type unit, four or five samples were taken at different landform positions by local soil experts. Finally, a total of 146 field observations were made in the field (Fig. 1). A soil pit down to a depth of 1.2 m or to bedrock was dug at each site. A local soil expert determined the soil horizons based on pedogenesis and major soil characteristics from a soil profile in the field, and then classified the soil according to the Chinese Soil Taxonomy (Chinese Soil Taxonomy Research Group, 2001). Soil surveys were conducted in 2013 and produced 146 soil sampling profiles. Samples of each horizon were collected in the following way. Soil samples from each location were a composite of five soil cores taken from four corners, and the center of a 1 m × 1 m square area (Cooperative Research Group on Chinese Soil Taxonomy, 2001). From the composite sample, 1 kg subsample was collected for laboratory analysis. To estimate dry bulk density, 100 cm^3 of undisturbed soil cores were collected from lower to upper soil layers.

In the laboratory, the samples were air dried and passed through a 2-mm sieve. SOC content was determined by wet oxidation method using potassium dichromate in acid medium followed by redox titration (Nelson and Sommers, 1996). Compared with dry combustion method, wet oxidation could oxidize about 90% of organic matter, so measure values were multiplied by a correction coefficient of 1.1 (Kalembasa and Jenkinson, 1973; Soon and Abboud, 1991). The 100 cm^3 cores were dried for 48 h at 105 °C for bulk density measurement. STN content was determined by dry combustion using a Vario EL III elemental analyzer (Elementar Analysen systeme GmbH, Hanau, Germany). STN and SOC contents were not measured together using the dry combustion method, but we revised the measurement of organic carbon according to the correction coefficient. The prediction and analysis of SOC and STN were analyzed separately.

2.3. Environmental variables

A suite of 9 environmental covariates representing topographic, climatic, Landsat thematic mapper (TM) variables and landuse were used as predictors for this study. Environmental variables were generated and transferred to raster layers using ArcGIS 10.2 (ESRI Inc., USA). Due to the widespread extent of the data and the low computational efficiency, predictors at a 90 m resolution were used in our study. All environmental covariates were scaled because distance-based learners, such as R software (R Development Core Team, 2013), require covariates to have a similar range in values. The relationships among SOC, STN and BD with the environmental variables are presented in Table 2.

2.3.1. Topographic variables

Terrain attributes are the most extensively used environmental predictors in digital soil mapping (Sculla et al., 2003). In our study, five topographic variables including elevation, slope, ground roughness, profile curvature, and topographic wetness index (TWI) were derived from a 90 m Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM). Elevation has been widely claimed as the vital factor to the soil nutrient distribution (Dorji et al., 2014; R. Yang et al., 2016a) and also as an important environmental variable in the prediction of soil properties. Slope, ground roughness and, profile curvature and TWI impact the soil erosion of region and are closely related to spatial variation of soil nutrients (Odeh et al., 1995). The four variables except TWI were generated using ArcGIS 10.2 (ESRI Inc., USA). TWI was generated using SAGA GIS (Olaya, 2004). This is because SAGA topographic wetness index is based on a modified catchment area, and it tends to predict a more realistic and higher potential soil wetness than conventional topographic wetness index (L. Yang et al., 2016).

2.3.2. Climatic variables

Temperature affects the accumulation rate of carbon and nitrogen in soil, and temperature variables are widely used in prediction of SOC and STN (Jobbágy and Jackson, 2000; Follett et al., 2012; R. Yang et al., 2016a; Wang et al., 2016). Mean annual temperature (MAT) and mean annual precipitation (MAP) over thirty years (1982–2010) of the study area was obtained from the Chinese Academy of Agricultural Sciences, Beijing. The data were interpolated from meteorological stations in China and were adjusted with the elevation data. The climate data originally had a cell size of 1000 m and were resampled to 90 m using a nearest neighbor strategy.

2.3.3. Landsat TM

Landsat 5 TM imagery was produced using 163 images downloaded from the United States Geological Survey (USGS), which was acquired from July to September (growing season) in 2013 with cloud cover < 10%. The polynomial geometric precision correction method was used to relief-correct the images, then the images were mosaicked and trimmed to cover the study area (Toutin, 2002). The visible-red band 3 (B3, 0.63–0.69 μm), near-infrared band 4 (B4, 0.78–0.90 μm) and short-wave infrared band 5 (B5, 1.55–1.75 μm) were collected to construct models. Malone et al. (2009) reported that the three bands represent vegetation growth, coverage and biomass, respectively. Normalized Difference Vegetation Index (NDVI) was determined by using bands B3 and B4, which was calculated as:

$$NDVI = (B4 - B3)/(B4 + B3) \quad (1)$$

2.3.4. Landuse data

Landuse map was a raster map created from Landsat images from 2013. Landuse types were grouped into agriculture, forest, grassland, wetland, and other types (e.g., urban area, Industry, Roads). In our model, landuse types of agriculture area, forest area, grassland area, wetland area were converted to numerical values 1, 2, 3, and 4, respectively. The correlation between target variables and land-use types was tested.

2.4. Model development

2.4.1. Boosted regression tree and uncertainty

In order to predict the spatial distribution of 1 m depth for SOC and TN stocks in Liaoning Province, we applied BRT model, a machine learning algorithm developed by Friedman et al. (2000). BRT combines two powerful statistical techniques: boosting and regression trees. The boosting algorithm uses an iterative method for developing a final model and progressively added trees to the model (Müller et al., 2013). Regression trees analyze the response variable with a set of predictor variables and apply a binary split to fit a simple model to each resulting section (Cheong et al., 2014; L. Yang et al., 2016). The data were further split so that the split-point achieves the best model fit (R. Yang et al., 2016a). BRT relies on stochastic gradient boosting, which allows for more accurate and faster computations through numerical optimization and regularization (Friedman, 2001). Compared with other data mining methods, BRT model emerges a high predictive accuracy and good interpretability of resulting input-output relationships (Friedman, 2001; Müller et al., 2013).

BRT model was operated in R version 3.2.2 (R Development Core Team, 2013) by using the package "dismo" version 0.8-17 (Hijmans et al., 2013) and package "gbm" version 2.1 (Ridgeway, 2007; Elith et al., 2008). The fitting of a BRT model is controlled by specification of four parameters: the learning rate (LR), tree complexity (TC), bag fraction (BF) and number of trees (NT). LR determines the contribution of each tree to the growing model (Müller et al., 2013). TC controls the size of trees and whether interactions between variables should be considered (L. Yang et al., 2016). BF sets the proportion of observations used in selecting variables (Zhang et al., 2016). NT is set based on the combination of LR and TC (Wang et al., 2016). We tested for several combinations of the LR (0.025, 0.05 and 0.1), TC (3, 4, 6, 8, 9 and 10), BF (0.55–0.75) and NT (500, 1000 and 1250) parameters. The final optimal values of LR, TC, NT and BF were set as 0.025, 9, 0.75 and 100 (Hastie et al., 2009; Wang et al., 2016).

The uncertainty associated with the BRT prediction was evaluated with standard deviation (SD) derived from running the model 100 times and the maps were generated as an indicator of SOC, SNT and BD prediction uncertainties. To obtain the relative importance (RI) of each predictor, the BRT model was repeated for 100 iterations. The RI of variables were measured based on the number of times a variable was selected for modeling and weighted by the square improvement to each split and averaged across all trees (Minasny et al., 2006). The RI of each variable was then scaled so that the sum added to 100 as percentages. A higher percentage of a variable indicated a stronger relative importance of this variable on the response (Pouteau et al., 2011; Wang et al., 2016).

2.4.2. Equal-area spline profile function

The existing studies on SOC and STN stocks are mostly concentrated on aspects in terms of two tendencies, one is to map SOC stock according to standard depths corresponding more or less to Intergovernmental Panel on Climate Change (IPCC) guidelines (often 30 cm or 1 m) (Meersmans et al., 2011; Ballabio et al., 2016; Minasny et al., 2017), the other aspect is to use the standard depths of GlobalSoilMap specifications (Arrouays et al., 2014; Minasny et al., 2013) that are 0–5, 5–15, 15–30, 30–60, 60–100, 100–200 cm. However, the standard depths of GlobalSoilMap specifications are more and more used at various levels from small regions to countries, continents and even the world (Poggio and Gimona, 2014; Vaysse and Lagacherie, 2015; Mulder et al., 2016).

We employed the equal-area spline algorithm a generalization of the algorithm developed by Bishop et al. (1999). The CSIRO SplineTool V2 (ASRIS, 2011) was used to convert the horizon-based values to the standard depths of GlobalSoilMap specifications (i.e. 0–5, 5–15, 15–30, 30–60 and 60–100 cm) (Arrouays et al., 2014). For a given soil profile, and a given soil property, the boundaries of the n layers are given by $x_0 < x_1, \ldots < x_n$. The soil property values, y_i ($i = 1, \ldots n$) therefore mathematically modeled as:

$$y_i = \overline{f_i} + e_i \tag{2}$$

where $\overline{f_i} = \int_{x_{i-1}}^{x_i} f(x)dx \big/ (x_i x_{i-1})$ is the mean value of $f(x)$ over the interval (x_{i-1}, x_i) and the measurement errors e_i are assumed independent, with mean 0 and common variance σ^2. Eventually, $f(x)$ represents a spline function, which can be determined by minimizing D:

$$D = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{f_i})^2 + \lambda \int_{x_o}^{x_n} [f^m(x)^2]d_x \tag{3}$$

The first term on the right side of Eq. (2) depicts the fit of the spline to the data; the second term measures the roughness of function $f'(x)$, expressed by its first derivative $f'(x)$. The λ parameter controls the trade-off between the fit and the roughness of the spline.

In our study, the raw carbon and nitrogen data displayed a log-normal distribution and subsequently was log-transformed prior to fitting the splines. The data for BD did not require any transformation. As the fitting quality of splines to profile attribute data depends on a smoothing parameter λ, we tested five values (0.0001, 0.001, 0.01, 0.1 and 1) for SOC, STN and BD data from all the profiles and selected 0.1 values that showed the best fit for all the profiles using the root mean square. Furthermore, once the depth function of SOC, STN and BD was modeled, a weighted-average value of these properties was derived for five soil depths (0–5, 5–15, 15–30, 30–60, and 60–100 cm) based on the GlobalSoilMap specifications (Arrouays et al., 2014).

2.4.3. Statistical analyses

We used SPSS 22.0 software for statistical analysis (Marr-Lyon et al., 2012). The descriptive statistics among SOC, STN and BD data with environmental variables were presented. Pearson correlation was also used to relate the dependent variable of SOC and STN contents to independent quantitative variables. P-value was applied to test for the significance and normal distribution between the variables.

2.4.4. SOC and STN stocks

This study analyzed the spatial distribution of SOC and STN stocks. For an individual profile with k layers (within first meter), the equation of Batjes (1996) was used to calculate the density of soil organic carbon (SOC) and soil total nitrogen (STN) in the whole soil profile:

$$\mathrm{SOC_{density}} = \sum_{i=1}^{k} \mathrm{SOC_{content}} = \sum_{i=1}^{k} \mathrm{SOC_{concentration}} \times \mathrm{BD}_i \times \mathrm{D}_i \times (1 - \mathrm{S}_i) \tag{5}$$

$$\mathrm{STN_{density}} = \sum_{i=1}^{k} \mathrm{STN_{content}} = \sum_{i=1}^{k} \mathrm{STN_{concentration}} \times \mathrm{BD}_i \times \mathrm{D}_i \times (1 - \mathrm{S}_i) \tag{6}$$

where $\mathrm{SOC_{density}}$ and $\mathrm{STN_{density}}$ are SOC and STN density of whole soil profile (kg m^{-2}), $\mathrm{SOC_{content}}$ and $\mathrm{STN_{content}}$ are SOC and STN contents (kg m^{-2}), BD_i is the bulk density (g cm^{-3}), $\mathrm{SOC_{concentration}}$ and $\mathrm{STN_{concentration}}$ are the SOC and STN concentration (g kg^{-1}), D_i is the thickness (m), S_i is the volume fraction of fragments > 2 mm, and i represents a specific soil horizon.

2.4.5. Model validation

Accuracy assessment was evaluated using the absolute prediction error (MAE), root mean square error (RMSE), coefficient of determination (R^2) and Lin's concordance correlation coefficient (LCCC) (Lin, 1989). They were calculated using 10-fold cross-validation procedure as described in the R version 3.2.2 (R Development Core Team, 2013). Cross-validation is a useful approach for model identification and assessment of model performance (Müller et al., 2013; Wang et al., 2016). These indices were calculated as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{8}$$

$$\mathrm{R^2} = \frac{\sum_{i=1}^{n}(P_i - \overline{O}_i)^2}{\sum_{i=1}^{n}(P_i - \overline{O}_i)^2} \tag{9}$$

$$\mathrm{LCCC} = \frac{2r\partial_O\partial_P}{\partial_O^2 + \partial_P^2 + (\overline{O} + \overline{P})^2} \tag{10}$$

where O_i and P_i are the observed and predicted SOC, STN and BD values at site i, respectively; n is the number of samples; ∂_O and ∂_P are the variances of observed and predicted values; and r is the Pearson correlation coefficient between the observed and predicted values.

Table 1
Descriptive statistics of soil organic carbon content (g kg^{-1}), soil total nitrogen content (g kg^{-1}) and bulk density (BD) (g cm^{-3}) data used in this study.

Depth (cm)	Parameters	Min.	Max.	Range	Mean	SD	Ske.	Kur.
0–5	SOC	0.50	164.02	163.53	17.65	24.15	0.19	1.09
	STN	0.29	8.85	8.56	1.32	1.33	0.47	0.99
	BD	0.28	1.68	1.40	1.27	0.28	−1.20	1.42
5–15	SOC	0.68	172.43	171.75	16.1	23.45	0.26	1.32
	STN	0.29	9.78	9.48	1.68	1.38	0.74	2.06
	BD	0.40	1.68	1.28	1.28	0.25	−1.29	1.77
15–30	SOC	0.02	235.10	235.07	9.50	26.67	0.39	4.79
	STN	0.18	15.18	15.00	1.36	1.75	1.23	4.35
	BD	0.51	1.65	1.14	1.36	0.30	−1.33	1.41
30–60	SOC	0.02	242.26	242.24	9.20	31.56	0.46	3.04
	STN	0.12	16.61	16.49	1.22	2.08	1.15	4.40
	BD	0.56	1.72	1.16	1.46	0.32	−0.79	0.38
60–100	SOC	0.02	230.44	230.43	7.42	27.60	0.74	1.67
	STN	0.01	11.02	11.01	0.82	1.25	1.20	2.74
	BD	0.83	1.61	0.78	1.52	0.34	−1.26	0.95

Note: Min., minimum; Max., maximum; SD, standard deviation; Ske., skewness; Kur., kurtosis.

Table 2
Descriptive statistics of soil organic carbon content (g kg^{-1}), soil total nitrogen content (g kg^{-1}) and bulk density (BD) (g cm^{-3}) data used in this study.

Parameters	Depths	GR	Elevation	NDVI	Landuse	Slope	MAT	MAP	Prof_cur	TWI
SOC	0–5	0.61**	0.60**	0.26**	−0.47**	0.68**	−0.37**	0.50**	−0.20*	−0.65**
	5–15	0.59**	0.61**	0.27**	−0.46**	0.65**	−0.40**	0.47**	−0.16*	−0.61**
	15–30	0.44**	0.54**	0.33**	−0.34**	0.45**	−0.51**	0.31**	0.07	−0.36**
	30–60	0.43**	0.42**	0.42**	−0.26**	0.46**	−0.52**	0.22**	−0.06	−0.28**
	60–100	0.39**	0.45**	0.40**	−0.22**	0.39**	−0.49**	0.28**	−0.12	−0.25**
STN	0–5	0.54**	0.55**	0.18*	−0.42**	0.59**	−0.37**	0.48**	−0.11	−0.58**
	5–15	0.56**	0.57**	0.21*	−0.40**	0.58**	−0.37**	0.47**	−0.11	−0.53**
	15–30	0.50**	0.53**	0.23**	−0.34**	0.47**	−0.36**	0.39**	−0.13	−0.36**
	30–60	0.50**	0.51**	0.17*	−0.32**	0.49**	−0.31**	0.32**	−0.18*	−0.37**
	60–100	0.16*	0.22**	0.12	−0.04	0.1	−0.12	0.30**	−0.28**	−0.06
BD	0–5	−0.68**	−0.62**	−0.31**	0.38**	−0.68**	0.37**	−0.47**	0.21**	0.54**
	5–15	−0.65**	−0.62**	−0.28**	0.34**	−0.63**	0.31**	−0.39**	0.21*	0.48**
	15–30	−0.50**	−0.56**	−0.08	0.25**	−0.49**	0.12	−0.26**	0.09	0.39**
	30–60	−0.39**	−0.51**	−0.11	0.20*	−0.38**	0.24**	−0.29**	0.04	0.32**
	60–100	−0.59**	−0.37**	−0.15	0.21**	−0.58**	0.27**	−0.42**	0.03	0.41**

Note: GR, ground roughness; NDVI, Normalized Difference Vegetation Index; Landuse, land use; MAT, mean annual temperature; MAP, mean annual precipitation; Prof_cur, profile curvature; TWI, SAGA wetness index.
* $P < 0.05$.
** $P < 0.01$.

Table 3
Average model performance to predict soil organic carbon content [log (SOC), g kg^{-1}], soil total nitrogen [log (STN), g kg^{-1}], and bulk density (g cm^{-3}) based on 100 operations of BRT model.

Parameters	Depth	SD	MAE	RMSE	R^2	LUCC
SOC	0–5	0.55	0.40	0.55	0.65	0.80
	5–15	0.56	0.42	0.56	0.59	0.75
	15–30	0.77	0.52	0.77	0.54	0.69
	30–60	1.09	0.80	1.09	0.50	0.65
	60–100	1.21	0.85	1.21	0.39	0.56
STN	0–5	0.38	0.27	0.38	0.55	0.72
	5–15	0.37	0.28	0.37	0.52	0.70
	15–30	0.42	0.30	0.42	0.49	0.70
	30–60	0.51	0.38	0.51	0.45	0.65
	60–100	0.97	0.65	0.97	0.33	0.48
BD	0–5	0.14	0.11	0.14	0.60	0.85
	5–15	0.12	0.10	0.12	0.57	0.85
	15–30	0.17	0.13	0.17	0.54	0.79
	30–60	0.18	0.14	0.18	0.50	0.79
	60–100	0.22	0.16	0.22	0.45	0.72

Note: SD, standard deviation; MAE, the mean error; RMSE, the root mean squared error; R^2, the coefficient of determination; LUCC, the Lin's concordance correlation coefficient.

3. Results and discussion

3.1. Descriptive statistics

The statistical results of SOC, STN and BD contents are showed in Table 1. Due to the shortage of funds in our research, SOC content measured using wet oxidation method was underestimated compared with the dry combustion method. We revised the measurement results of organic carbon using a correction coefficient (Soon and Abboud, 1991). The corrected data hardly affected our results and conclusion. SOC content was highly variable and ranged from 0 to 235 g kg^{-1} for the topsoil (0–20 cm) and from 0 to 242 g kg^{-1} in the subsoil (Table 1). Mean SOC decreased with soil depth and SOC at 60–100 cm was about seven times lower than the SOC in the 0–5 cm layer, moreover STN at 60–100 cm three times lower than the STN in the 0–5 cm layer. With depth, the standard deviation (SD) of the SOC and STN contents increased. The SD at 0–5 cm was 24 g kg^{-1} and that for the 60–100 cm was approximation 28 g kg^{-1} for SOC. The SOC and STN data were positively skewed at all soil depths with a maximum skewness coefficient at 60–100 cm. The equal-area splines modeled the depth-wise distribution and generated a continuous SOC, STN and BD profile to 1 m depth. The best λ value to fit all soil profiles for SOC, STN and BD data was 0.1. Also mean BD content was found to be increased with soil depth. Up to 30 cm depth of the BD was on average 1.27 g cm^{-3}, whereas it increased to 1.52 g cm^{-3} below 60 cm depth. Bulk density

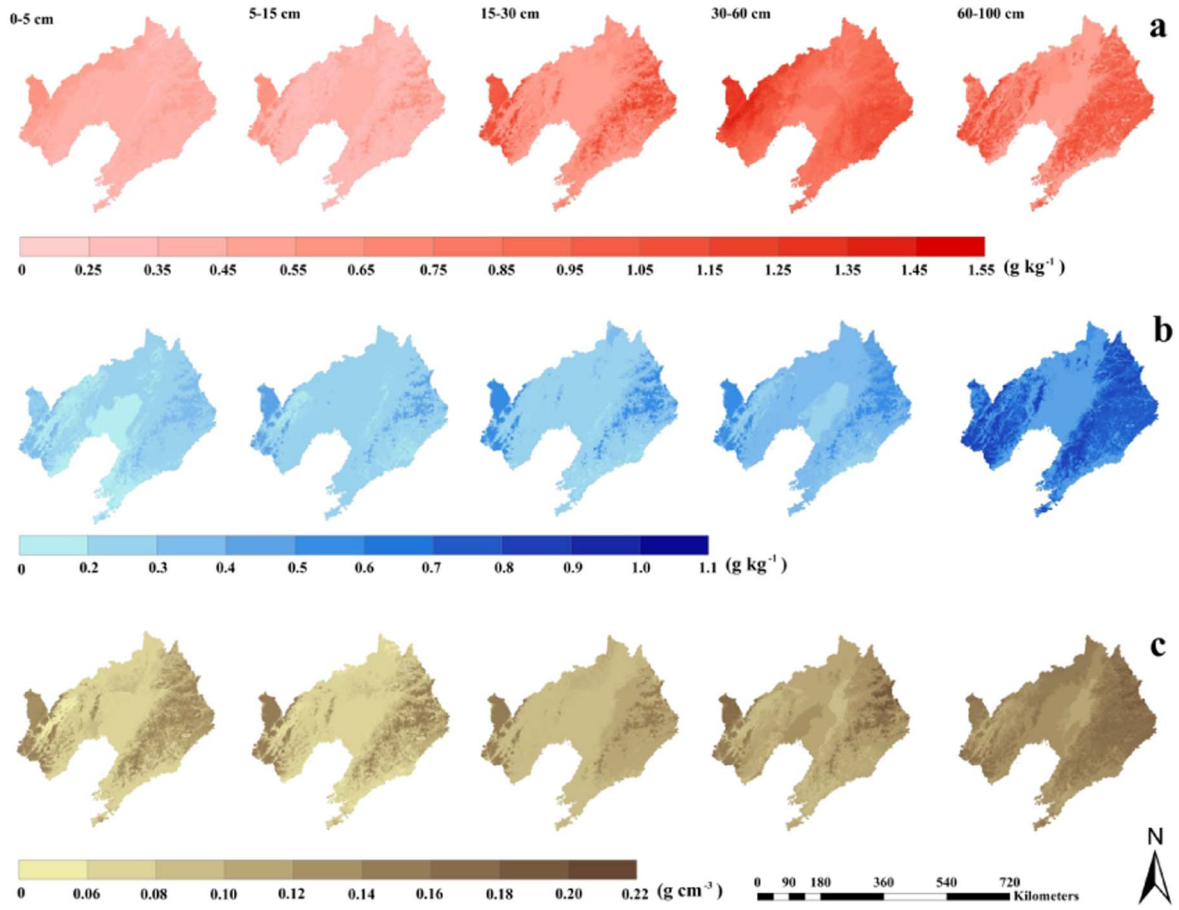

Fig. 2. Standard deviation (SD) of SOC, STN and BD predicted from 100 runs of the boosted regression trees (BRT) models: (a) SD of SOC predicted; (b) SD of STN predicted; and (c) SD of BD predicted.

appeared to be less variable with depth (Table 1).

Linear correlations among SOC, STN and BD with quantitative predictors are showed in Table 2. SOC was positively correlated with ground roughness (GR), elevation, NDVI, landuse, slope, MAT and MAP at all levels. Nevertheless, STN was only positively correlated with GR, elevation and MAP for all five levels. At the same time, the rule of BD was different from STN and SOC on the correlation, which was negatively correlated with GR, elevation slope and MAP. Of more interest for our study, correlations with GR and elevation were all significant. The relation with elevation in this natural area was expected; the slightly higher correlation with the GR was somewhat surprising. Predictors within each group (topography, climate, imagery) and between groups had some collinearity.

3.2. Sampling strategy evaluation, model performance, and uncertainty

Because our study area and high spatial variability of SOC and STN are large, collecting a large number of observations following a sound sampling design is prohibitively impractical (Zhu et al., 2008; Yang et al., 2013). It is very likely that many designed sampling points are not accessible in field. Thus, probability sampling or condition Latin hypercube sampling (cLHS) may be not appropriate. Table 3 showed that BRT model had a steady ability to predict the SOC, STN and BD contents at different layers. BRT model could explain the spatial variation ranging from 39% to 65% for SOC, 33% to 55% for STN, and 45% to 60% for BD at different depths. Those results indicated that sampling strategy can capture the local variation of SOC and STN spatial distribution. Furthermore, this also suggested that the sampling strategy in this study do capture the major pattern of soil-landscape relationships over the area.

For predicting the SOC, STN and BD, 9 environmental variables were used in our BRT model. The model performance was summarized on the basis of R^2, MAE, RMSE and LUCC with 100 model iterations and using 10-fold cross-validation (Table 3). The result showed that the BRT model had excellent predictive capability, with systematically higher R^2 and LUCC, lower MAE and RMSE values in prediction of SOC, STN and BD. In Denmark, Adhikari et al. (2014) used Regression kinging (RK) and Equal-area spline profile function to predict SOC and BD only could explain the change of 23–41% at five layers. Meanwhile, the R^2 values suggested that the BRT model could explain approximately 65%, 56% and 60% (Table 3) within 0–5 cm depth of the total SOC, STN and BD variability in our study. However, the BRT model showed relatively low prediction performance at 60–100 cm depth (R^2, 39%, 33% and 45%), which may be due to the reduction of the explanatory power of the auxiliary environment variables along with the increase of the depth of the soil layer (Adhikari et al., 2014). Although environmental conditions, sampling strategies and validation methods of our study differ from previous studies, the performance of BRT model in these studies is comparable. Martin et al. (2011) developed a BRT model to explain 50%–58% of the variance of SOC in France. A study in the central highlands of Madagascar, Razakamanarivo et al. (2011) explained 61–68% of the total SOC variability using a BRT model. Wang et al. (2013) used a geographically weighted regression (GWR) method to predict mapping of soil total nitrogen (STN) and explain 57% of the total TN variability in Fujian province, China.

To evaluate the modeling uncertainty, we calculated the SDs of validation measurements from 100 runs and found that the BRT model produced low SDs of MAE, RMSE, R^2 and LCCC (Table 3). Although there were different elements of variability and uncertainty in C and N research including the system error, random error and unpremeditated

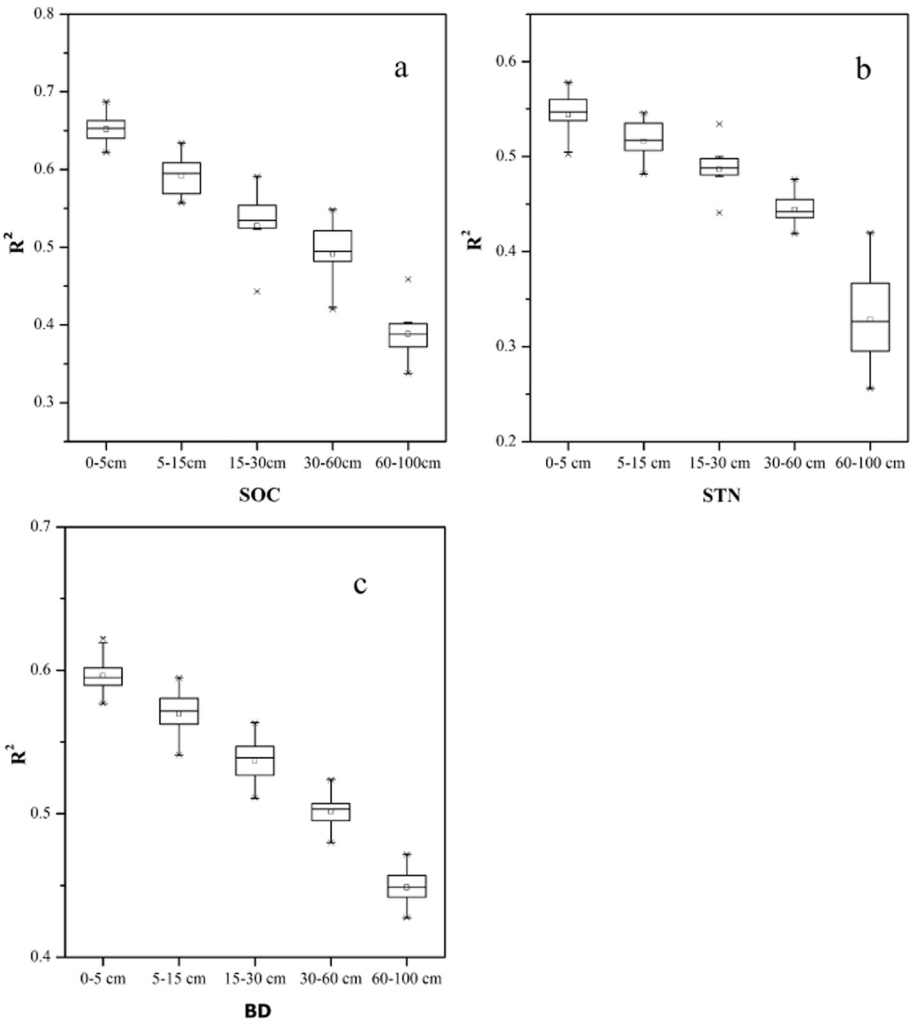

Fig. 3. R^2 distributions from BRT in predicting SOC (a), STN (b) and BD (c) concentrations at different depths based on 100 iterations.

error (Krishnan et al., 2007; L. Yang et al., 2016), the lower SD values of MAE, RMSE and R^2 indicated the BRT model was stable in predicting the SOC, STN and BD contents. Figs. 2, 5, 6 and 7 showed the mean and SD values for these 100 maps. The mean and SD values of the predicted 0–5 cm layer of the 100 runs were 15.2 and 16.0 g kg^{-1} for SOC, 1.0 and 0.5 g kg^{-1} for STN, and 1.2 and 0.2 g cm^{-3} for BD, respectively. Notably, the SOC, STN and BD predicted yielded lower SD values than did the collected samples (Table 1); thus, the predicted SOC, STN and BD variability was smaller than was actually observed. Furthermore, this result indicated that the BRT model could overcome the model uncertainty and had a stable performance. Descriptive statistics of Table 3 also reveals that the BRT model had excellent performance in the spatial prediction of SOC, STN and BD.

To further illustrate the uncertainty of BRT model, we used a boxplot of R^2 distributions at different depths of SOC, STN and BD contents based on 100 iterations (Fig. 3). The uncertainty of the SOC, STN and BD prediction increased with depth. The mean R^2 at different depths ranged from 0.39 to 0.65 for SOC, 0.33 to 0.55 for STN and 0.45 to 0.60 for BD. The high mean R^2 for BRT model was expected because all of the data available were used to train the model benefited from the application of 10-fold cross-validation. Furthermore, adopting a scientific sampling strategy could precisely predict of SOC, TN and BD contents in the complex and heterogeneous landscape research area (Wang et al., 2016). The variation of R^2 for SOC, STN and BD concentrations form BRT model based on 100 iterations was high except 60–100 cm (Fig. 3), indicating instability of the model in its prediction. With the increase of depth, SOC, STN and BD of R^2 had a similar linear decreasing trend. This could be linked to the terrain parameters used as major predictors because most of these parameters explain soil surface phenomena and the uncertainty increases with depth (Adhikari et al., 2014). Due to the different soil types, land use pattern and vegetation in this study area, the BRT model may be overestimated or underestimated the contents of SOC and STN, which may cause an estimation error finally. Furthermore, estimation of the SOC and STN stocks are limited to 100 cm depth in this study which underestimates the stock beyond this depth, since significant amount of stock usually present beyond 100 cm depth.

3.3. Importance of environmental variables

The relative importance (RI) of each environmental variable in BRT model was assessed by iterating 100 times of simulations, normalizing the environmental variables of each model to 100% (Fig. 4). The environmental variables of each variable in the model of SOC, STN and BD showed different RI. There was a large influence of precipitation, landuse, NDVI and some terrain parameters such as elevation, slope gradient, and SAGA wetness index on the spatial distribution of SOC, STN and BD contents.

There were highly significant correlations of SOC, STN and BD contents with all topographic variables (Table 2). Topography is one of five soil formation factors, and it can influence the moisture-temperature conditions and distribution of soil formation substance (Jenny, 1941). Therefore, there was a closely relationship between the spatial variability of soil properties with topographic variables (Charles et al., 2006; L. Yang et al., 2016). Our achievements revealed that the topographic variables had the most importance in predicting SOC, STN and

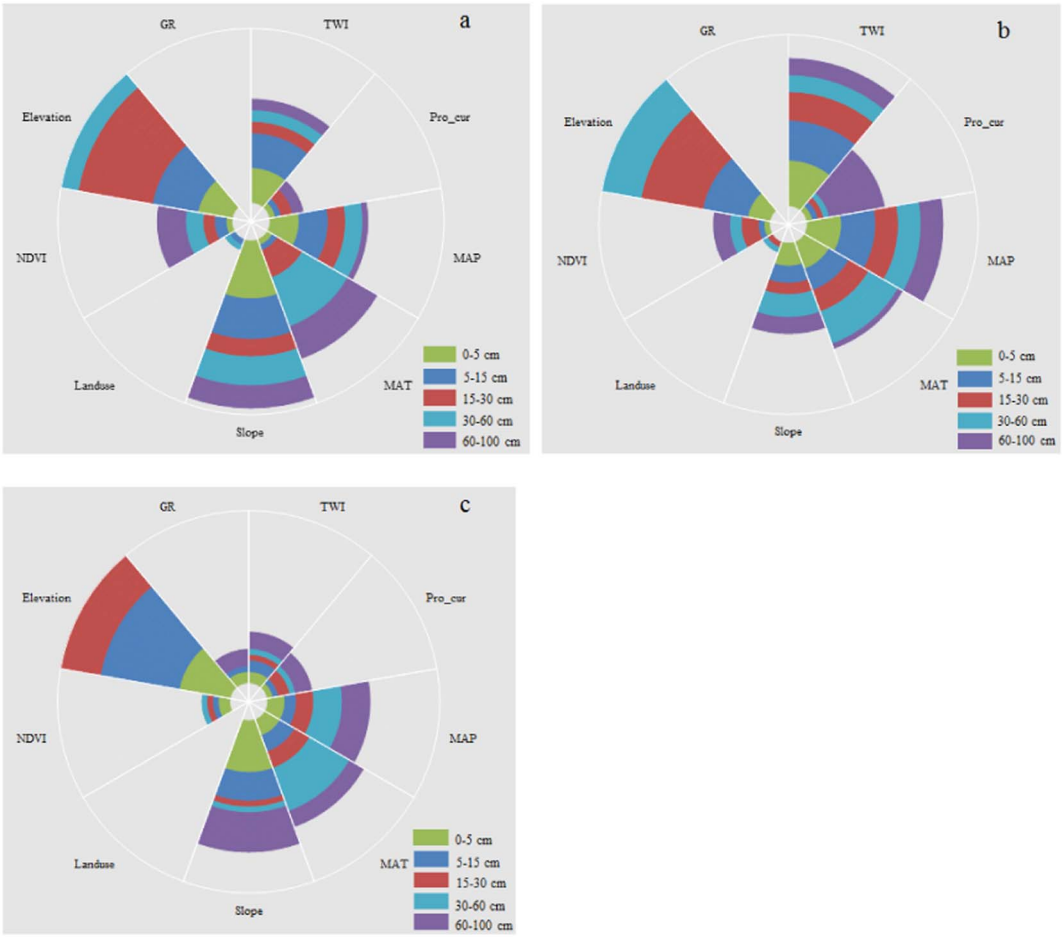

Fig. 4. Relative importance (RI) of each variable as determined from 100 runs of the boosted regression trees model, which are showed in a decreasing order and normalized to 100%. a) RI of soil organic carbon; b) RI of soil total nitrogen; c) RI of bulk density; GR, ground roughness; NDVI, Normalized Difference Vegetation Index; MAT, mean annual temperature; MAP, mean annual precipitation; Prof_cur, profile curvature; TWI, SAGA wetness index.

BD contents (Fig. 3). Among all topographic variables, elevation played the most important role in this process of predicting the distribution of SOC, STN and BD. This could be due to the redistribution under Hydrothermal Condition leading effect of litter decomposition rate, causing the periodic change of carbon and nitrogen, and thus indirectly affecting bulk density change (Tsui et al., 2004; Martin et al., 2014; Adhikari et al., 2014). The SAGA wetness index (TWI) was frequently used to simulate the ideal soil moisture conditions in a watershed quantitatively, and it was the most commonly used variables for predicting the spatial distribution of SOC, STN and BD contents (Adhikari et al., 2014; L. Yang et al., 2016). The contents of SOC, STN and BD showed significantly positive correlations with slope (Table 2). A similar finding was analyzed in China and Spain (Tsui et al., 2004; Wang et al., 2012a; Hontoria et al., 1999; L. Yang et al., 2016). However, a negative effect of slope on SOC distribution was also reported in Hall's (1983). In this study, this contrasting correlation could be explained by the different way of landuse, steeper hills were mainly distributed to woodland and grassland, only a gentler slope can be converted for cultivated land, whereas gentler slopes for cultivated land was relatively lower SOC and STN contents. Furthermore, the relationship between BD with SAGA Wetness Index is just the opposite.

The NDVI and landuse types were the important factors affecting the content of SOC, STN and BD (Jobbágy and Jackson, 2000; Wang et al., 2011; L. Yang et al., 2016; R. Yang et al., 2016a). Wang et al. (2000) reported that NDVI is the main factor to control SOC and STN. The spatial topsoil carbon pattern was highly related to NDVI (Jobbágy and Jackson, 2000). Meanwhile, NDVI was highly predictive of the SOC and STN contents based on the finding of the derived vegetation index reflected vegetation productivity and biomass (Bronson et al., 2004; Liu et al., 2012; Wang et al., 2013). Those findings imply that there was a potential application of remote sensing techniques to mapping the SOC and STN distribution in large regions. However, this finding differs from our conclusions, which suggest that the effect of NDVI may be mediated by using topographic and climatic variables (Wang et al., 2016). Landuse type was another indicator of vegetation, and it appeared to be an important predictor of SOC and STN stocks in the BRT model (Martin et al., 2011; Wang et al., 2011). However, the RI of landuse type was much lower than the RI of NDVI in our models. This difference can be partly explained by the greater predictive ability of quantitative variables (L. Yang et al., 2016) than categorical variables (used by Martin et al., 2011). Furthermore, low sampling density may cause the sample can't really reflect the actual situation in this study area, so lead to certain environment variables in the model prediction less important than practical efficacy (L. Yang et al., 2016).

Precipitation and temperature were the key climatic variables that affect the spatial distribution of SOC, STN and BD in the continental monsoon climate (Follett et al., 2012; Van Wambeke, 2000; Saiz et al., 2012). In this study, elevation has an impact on temperature and precipitation. A highly significant and strongly negative correlation was found between elevation and annual mean maximum temperature, between elevation and annual mean minimum temperature, likewise between elevation and precipitation (Table 3). The presence of highly significant correlations of elevation with temperature, precipitation, suggesting that elevation may be a more significant controlling factor than temperature and precipitation and thereby may serve as a climate proxy and a strong predictor of SOC and STN status (Wang et al.,

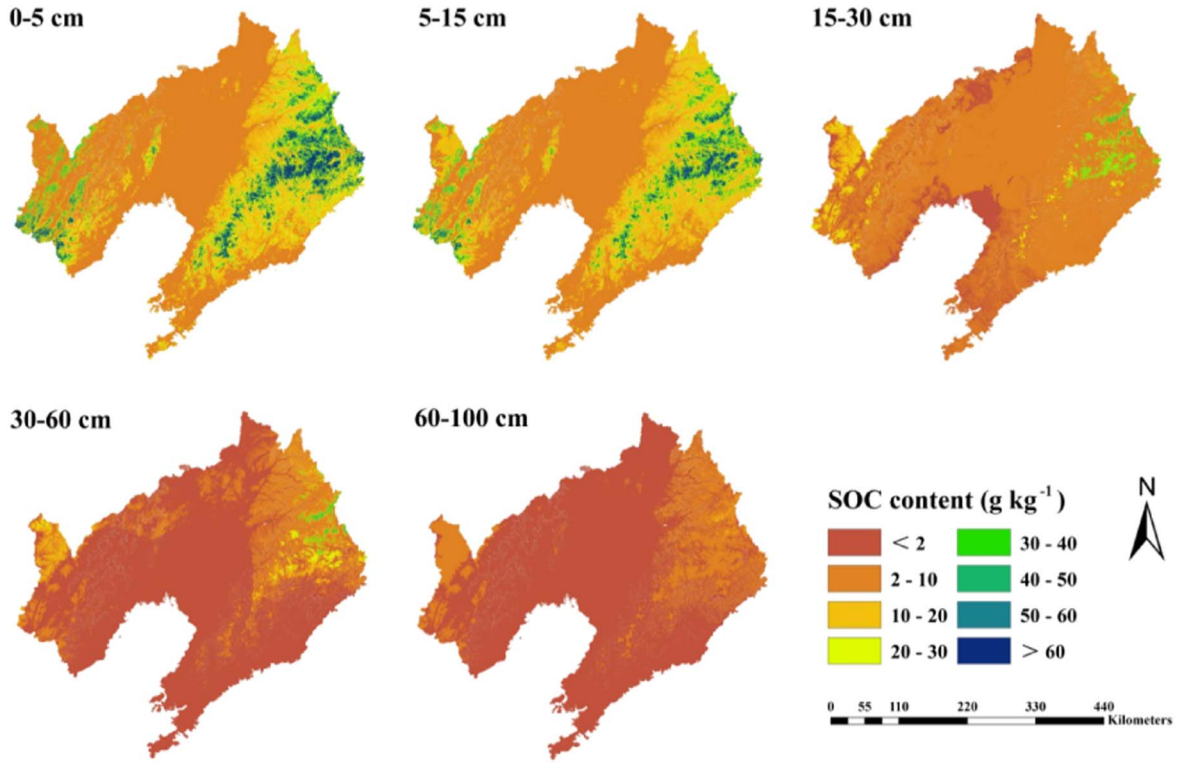

Fig. 5. Spatial distribution of predicted soil organic carbon contents at different depths.

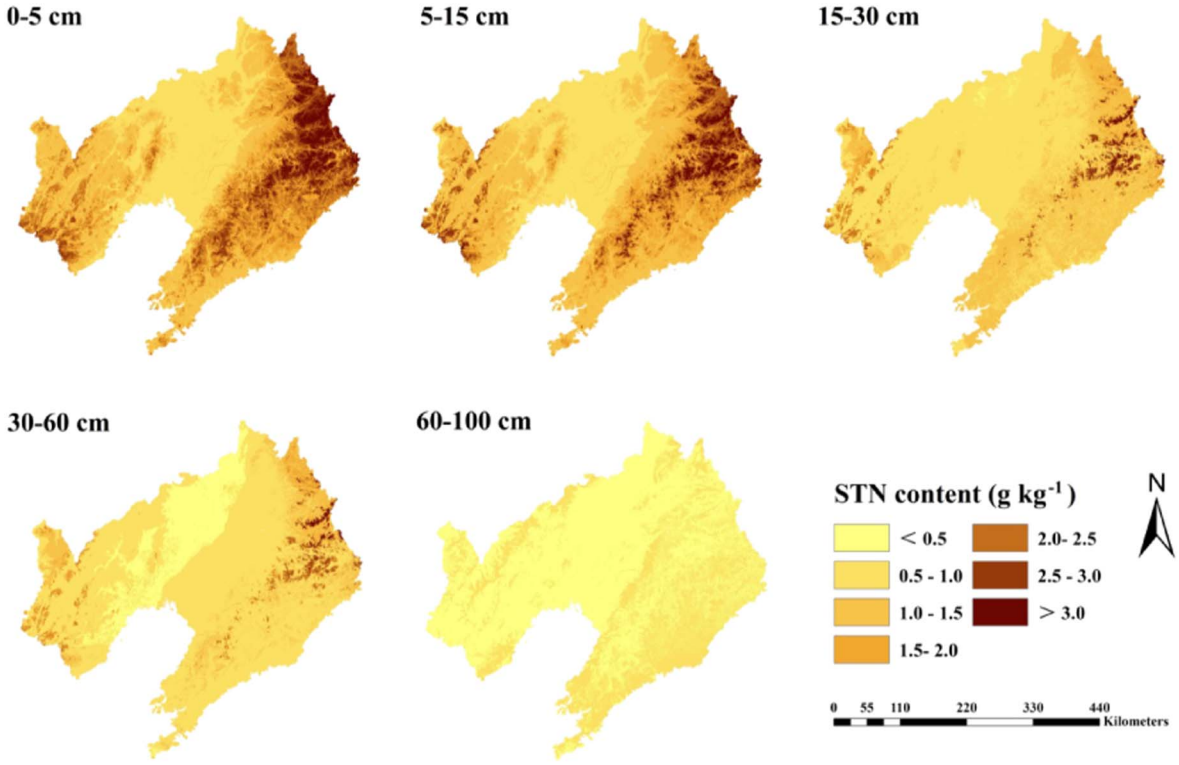

Fig. 6. Spatial distribution of predicted soil total nitrogen contents at different depths.

2012b).

3.4. Modeling of the SOC, STN and BD depth distribution

Predicted maps of SOC (Fig. 5), STN (Fig. 6) and BD (Fig. 7) contents at five soil depths were produced at a resolution of 90 × 90 m. The highest mean SOC and STN contents were in the 0–5 cm layer (mean 15.2 and 1.6 g kg^{-1}). Predicted SOC and STN contents decreased with soil depth and at 60–100 cm, those were on average 1.5 and 0.4 g kg^{-1}, respectively. The soils of northeast and southwest of Liaoning province had relatively higher SOC and STN contents than the rest of the study area. Along the central plains and the coastline,

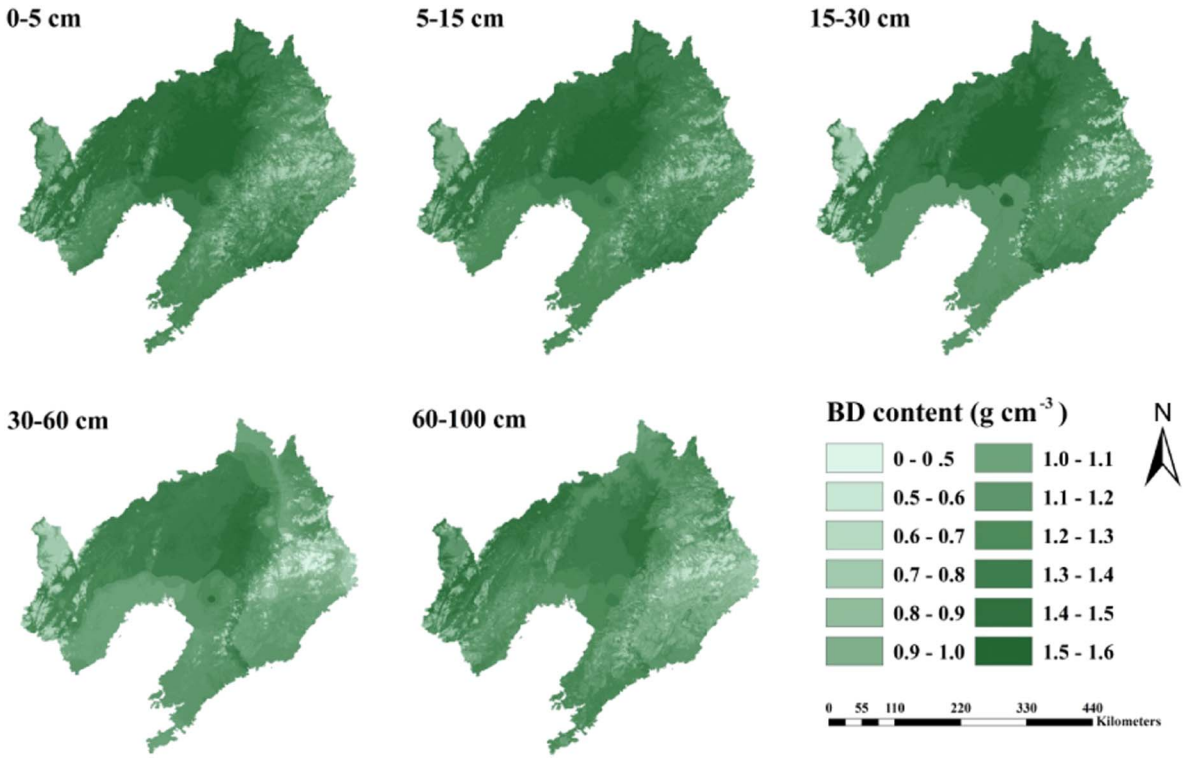

Fig. 7. Spatial distribution of predicted soil bulk density values at different depths.

especially in the south, soils with lower SOC and STN contents were mapped (Figs. 5 and 6). The prediction errors were higher towards the west. The prediction error increased with soil depth. For the 0–5 cm layer, the mean error of SOC was 0.4 g kg^{-1} and it increased to 0.7 g kg^{-1} at 60–100 cm soil depth while the mean error of STN were 0.2 g kg^{-1} (0–5 cm) and 0.6 g kg^{-1} (60–100 cm) (Fig. 2a and b).

The distribution maps of SOC and STN also showed similar spatial distribution patterns (Fig. 8). However, the spatial distribution of BD was contrary to that of SOC and STN contents, and there was a higher BD content than any other areas. High SOC and STN contents are in the northeastern area of Liaoning Province. The spatial distribution pattern of SOC and STN had a strong relationship with the topographic variables, especially the elevation (Fig. 8). The vertical distribution of SOC and STN at the top 1 m depth along latitude 41.6°E was displayed. As expected, SOC and STN contents dominantly concentrated in the topsoil, and decreased with depth. In mountain areas, SOC and STN distributions showed a sharp discontinuity. The effect of elevation on SOC and STN has been demonstrated by recent researches (Powers and Schlesinger, 2002; Tsui et al., 2004; Tian et al., 2007; Podwojewski et al., 2011; Wang et al., 2011). Tsui et al. (2013) reported that the SOC content increased significantly with the increase of elevation. Different elevation gradients affected the input and loss of soil carbon and nitrogen mainly through indirect controls such as precipitation and temperature (Garten and Hanson, 2006; Wang et al., 2011).

3.5. Spatial prediction of SOC and STN stocks

SOC and STN stocks maps were made for two soil depths (0–30 cm and 0–100 cm) (Figs. 9 and 10). For 0–30 cm, average SOC and STN stocks were about 3.1 kg m^{-2} and 0.6 kg m^{-2}, those were about 4.5 kg m^{-2} and 0.9 kg m^{-2} for the top 1 m depth. Most of the northern and southwestern parts of Liaoning Province had > 8.0 kg m^{-2} in the top 30 cm whereas the average stock in the central part of the study area was < 8.0 kg m^{-2}. The stocks of SOC and STN in different soil groups are presented in Table 4. Argosols and Cambosols contained about 91% of the total SOC and STN stocks (Table 4). Other soil groups that contained significant amounts of SOC and STN stocks were Anthrosols (3% vs 3%) and Primosols (5% vs 4%). Although Histosols had SOC and STN stocks of 4.9 and 1.0 kg m^{-2}, its total content was less 0.2 Tg. For all soil groups, > 69% and 49% of the total SOC and STN stocks were in the top 30 cm. This conclusion had been confirmed by Liu et al. (2012), whereby they believed that 43% and 39% of the stocks of SOC and STN in the upper 30 cm. Furthermore, Adhikari et al. (2014) predicted the storage of SOC in Denmark and considered about 59% of SOC stock was in the upper 30 cm.

Of the total storage of SOC and STN were 406 Tg and 63 Tg whereby approximately 69% and 49% distributed in upper 30 cm. Soils under forest had an average SOC stock of 4.1 kg m^{-2} and contained about 245 Tg which was almost 60% of the total estimated SOC stock (Table 5). Another large fraction of SOC and STN stocks were found in the soils of the agriculture, and it had a stock of 120 and 24 Tg in the top 30 cm and about 172 and 52 Tg up to 1 m soil depth. Wetland areas contained large amounts of SOC and STN, and average SOC and STN within 1 m soil depth were about 2.8 kg m^{-2} and 0.9 kg m^{-2} which were nearly 5% and 7% of the total stock. Almost 90% of the total SOC and STN stocks within the top 1 m soil depth were found in the soils under forest, agriculture and wetland.

4. Conclusions

We applied a BRT model to indirectly and digitally map the SOC and STN stocks at different depths for Liaoning Province of China. The results demonstrated that the BRT combination with equal-area spline profile function was robust and easy to produce soil property maps by using less data points of the target variable and several environmental variables. The most important variables affecting the spatial distribution of SOC and STN were elevation, precipitation, temperature, NDVI, landuse, and saga wetness index in our model. Meanwhile, model performance was better for upper soil layers than lower layers and almost all prediction models suffered from lager uncertainties.

In our study area, almost 91% of the total SOC and STN stocks were found in Argosols and Cambosols. About 90% of SOC and STN stocks

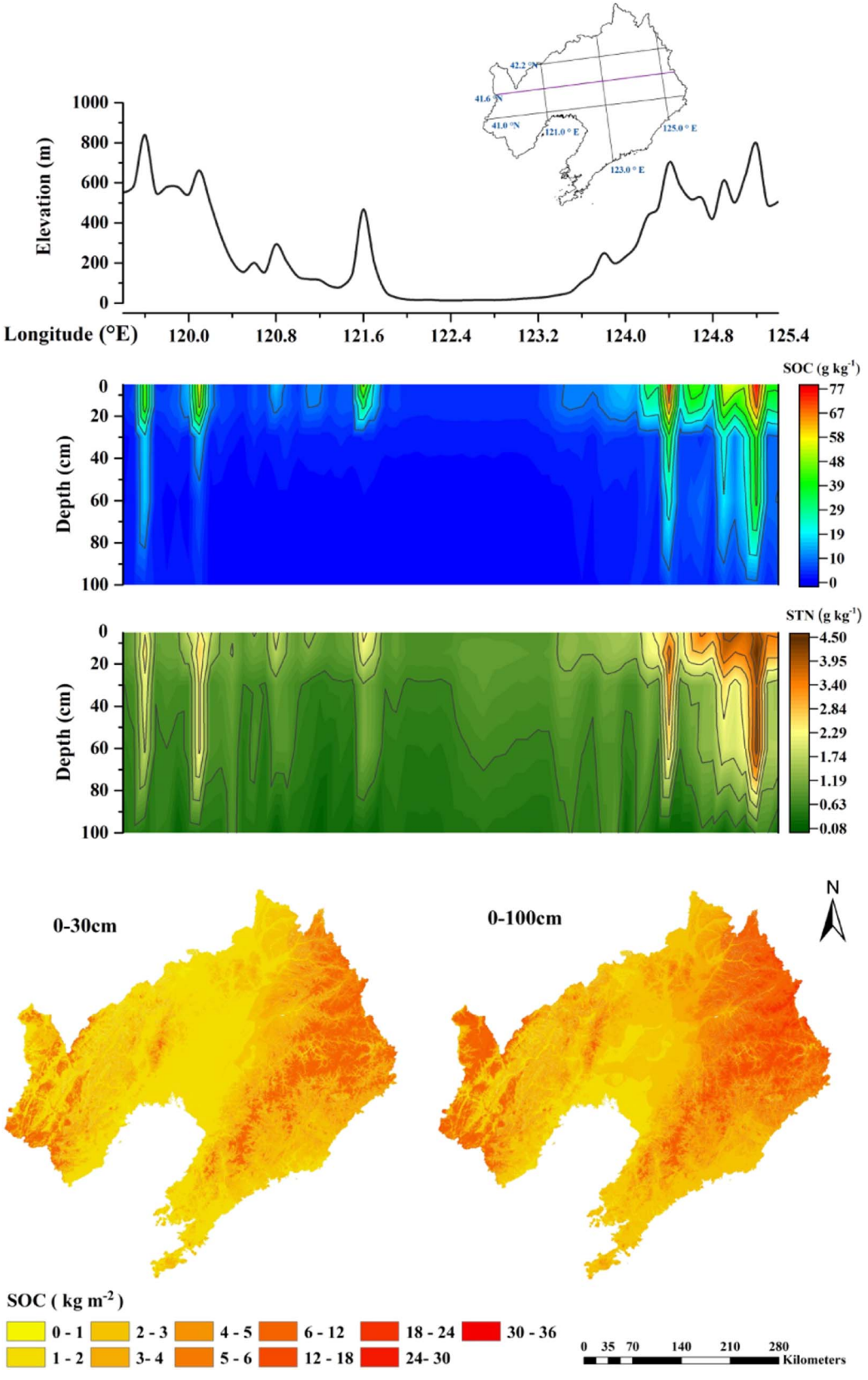

Fig. 8. Vertical distribution of soil organic carbon content and soil total nitrogen at the top 1 m depth along latitude 41.6°E.

Fig. 9. Spatial distribution of soil organic carbon stocks (SOC) at 0–30 and 0–100 cm.

were held in soils under forest, agriculture and wetlands. This had implications for land-use policy making, for instance, the policy on how the current land-use types should be managed to enhance C and N sequestration to mitigate global warming. For soils cultivated with agricultural crops, 70% and 45% of the SOC and STN stocks were found in the top 30 cm, respectively. We conclude that long-term use of fertilizer had a strong enrichment on farmland carbon and nitrogen accumulation.

In addition, this study provides an example for efficiently mapping regional SOC and STN stocks based on GlobalSoilMap procedures.

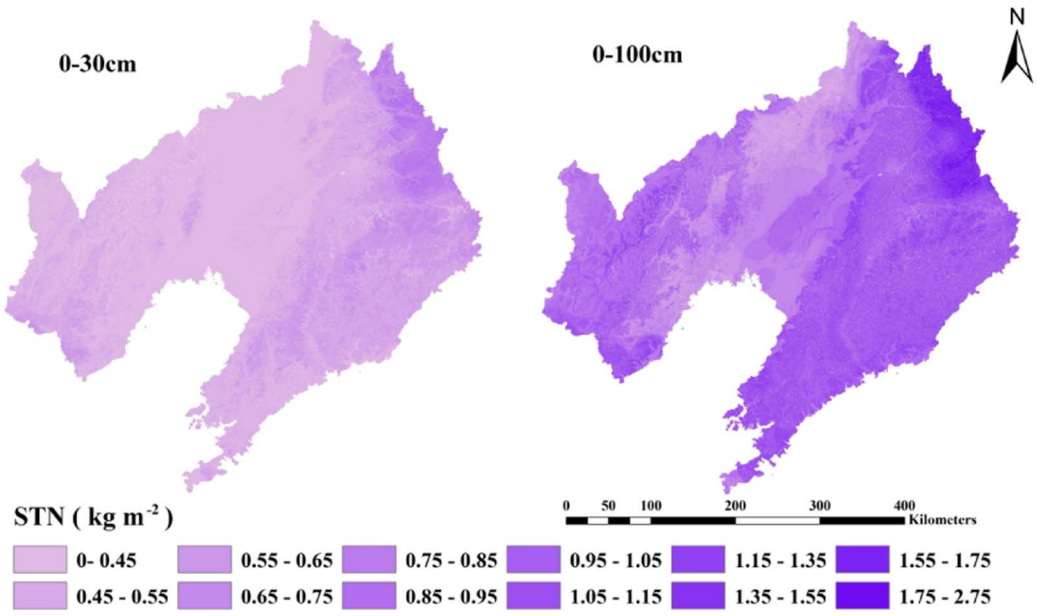

Fig. 10. Spatial distribution of soil total nitrogen stocks (STN) at 0–30 and 0–100 cm.

Table 4
Soil organic carbon and Soil total nitrogen stocks in the top 1 m soil depth according to Chinese soil groups.

Soil groups	Area (km^2)	Average SOC stock ($kg\ m^{-2}$)		Average STN stock ($kg\ m^{-2}$)		SOC stock (Tg)		STN stock (Tg)	
		0–30	0–100	0–30	0–100	0–30	0–100	0–30	0–100
Anthrosols	6064	1.74	2.44	0.39	0.86	10.55	14.80	2.36	5.22
Argosols	45,889	3.27	4.93	0.47	0.92	150.06	226.23	21.57	42.22
Cambosols	73,085	3.04	4.35	0.46	0.93	222.18	317.92	33.62	67.97
Gleyosols	431	1.64	2.42	0.38	0.81	0.71	1.04	0.16	0.35
Halosols	2318	1.62	2.44	0.36	0.79	3.76	5.66	0.83	1.83
Histosols	38	3.75	4.95	0.38	0.95	0.14	0.19	0.01	0.04
Isohumosols	156	3.72	4.76	0.39	0.94	0.58	0.74	0.06	0.15
Primosols	14,187	1.32	1.56	0.32	0.73	18.73	22.13	4.54	10.36
Sum	142,168					406.70	588.71	63.17	128.12

Note: SOC, soil organic carbon; STN, soil total nitrogen.

Table 5
Predicted soil organic carbon and soil total nitrogen stocks in different landuse types derived for two different soil depths.

Major landuse types	Area (km^2)	Soil depth			
		0–30	0–100	0–30	0–100
		SOC stock in Tg		STN stock in Tg	
Forest areas	60,060	244.66	355.93	31.33	58.24
Grassland areas	4561	11.37	17.66	1.83	3.91
Agriculture areas	61,502	120.11	171.57	23.65	52.12
Wetlands areas	10,492	20.52	29.34	4.22	9.29
Artificial surface (urban, industry, roads, etc.)	5388	9.72	13.72	2.08	4.44
Other	165	0.32	0.49	0.06	0.12
Sum	142,168	406.70	588.71	63.17	128.12

Note: SOC, soil organic carbon; STN, soil total nitrogen.

These methods can be tested and used in other parts of the world.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (No. 41371223) and National Scientific-basic Special Fund of China (2008FY110600). Acknowledgements are also extended to Dr. Lin Yang, Chinese Academy of Sciences, Institute of Geographical Sciences and Resources Research, for her technical help during the revision. We kindly acknowledge the anonymous reviewers for their valuable comments on this paper.

References

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.
Arrouays, D., Deslais, W., Badeau, V., 2001. The carbon content of topsoil and its geographical distribution in France. Soil Use Manag. 17 (1), 7–11.
Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B., Hong, S.Y., ... Mendonca-Santos, M., 2014. Chapter three-GlobalSoilMap: toward a fine-resolution global grid of soil properties. Adv. Agron. 125, 93–134.
ASRIS, 2011. ASRIS–Australian Soil Resource Information System. http://www.globalsoilmap.net/specifications (accessed 2007.11.12).
Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261, 110–123.
Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47 (2), 151–163.
Bernoux, M., da Conceição Santana Carvalho, M., Volkoff, B., Cerri, C.C., 2002. Brazil's soil carbon stocks. Soil Sci. Soc. Am. J. 66 (3), 888–896.
Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45.
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 (5), 1695–1704.
Bureau of Statistics Liaoning Province, 2011. Statistical Yearbook of Liaoning 2011. China Statistics Press, Beijing, China (in Chinese).
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.
Carslaw, D.C., Taylor, P.J., 2009. Analysis of air pollution data at a mixed source location using boosted regression trees. Atmos. Environ. 43 (22 – 23), 3563–3570.

Charles, T., Jr, Garten, Hanson, P.J., 2006. Measured forest soil C stocks and estimated turnover times along an elevation gradient. Geoderma 136 (1), 342–352.
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.
Cooperative Research Group on Chinese Soil Taxonomy, 2001. Keys to Chinese Soil Taxonomy. Press of University of Science and Technology of China, Hefei, China (in Chinese).
Dorji, T., Odeh, I.O., Field, D.J., Baillie, I.C., 2014. Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. For. Ecol. Manag. 318, 91–102.
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.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. Ann. Stat. 28 (2), 337–407.
Garten, C.T., Hanson, P.J., 2006. Measured forest soil C stocks and estimated turnover times along an elevation gradient. Geoderma 136 (1), 342–352.
Hall, G.F., 1983. Pedology and Geomorphology. In: Pedogenesis and Soil Taxonomy: I. Concepts and Interactions. In: Wilding, L.P., Smeck, N.E., Hall, G.F. (Eds.), Elsevier Science Publishers Besloten Vennootshap met Beperkte Aansprak-elijkhed, Amsterdam, pp. 117–140 the Netherlands.
Hart, D.M., Humphreys, G.S., 2003. Phytolith depth functions in surface regolith materials. In: Advances in Regolith: Proc. CRC LEME Regional Regolith Symposium, Adelaide, Canberra, and Perth, pp. 13–28. http://www.crcleme.org.au/Pubs/Advancesinregolith/Hart_Humphreys.pdf.
Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77.
Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2013. dismo: species distribution modeling. R package version 0.8-17. http://www.idg.pl/mirrors/CRAN/web/packages/dismo/vignettes/sdm.pdf.
Hontoria, C., Saa, A., Rodríguez-Murillo, J.C., 1999. Relationships between soil organic carbon and site characteristics in peninsular Spain. Soil Sci. Soc. Am. J. 63 (3), 614–621.
Hudson, B.D., 1992. The soil survey as a paradigm-based science. Soil Sci. Soc. Am. J. 56, 836–841.
Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York.
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.
Kalembasa, S.J., Jenkinson, D.S., 1973. A comparative study of titrimetric and gravimetric methods for the determination of organic carbon in soil. J. Sci. Food Agric. 24 (9), 1085–1090.
Karunaratne, S.B., Bishop, T.F.A., Baldock, J.A., Odeh, I.O.A., 2014. Catchment scale mapping of measureable soil organic carbon fractions. Geoderma 219-220, 14–23.
Kempen, B., Brus, D.J., Stoorvogel, J.J., 2011. Three-dimensional mapping of soil organic matter content using soil type–specific depth functions. Geoderma 162 (1), 107–123.
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 C sequestration impacts on Global Climatic Change and Food Security. Science 304 (5677), 1623–1627.
Lehmann, J., Kleber, M., 2015. The contentious nature of soil organic matter. Nature 528 (7580), 60–68.
Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.
Liu, W., Chen, S., Qin, X., Baumann, F., Scholten, T., Zhou, Z., Sun, W., Zhang, T., Ren, J., Qin, D., 2012. Storage, patterns, and control of soil organic carbon and nitrogen in the northeastern margin of the Qinghai–Tibetan Plateau. Environ. Res. Lett. 7 (3), 035401.
Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.
Marr-Lyon, L.R., Gireesh, G.V., Anderson, J.R., 2012. An evaluation of the psychometric properties of the Purdue Pharmacist Directive Guidance Scale using SPSS and R software packages. Social Adm. Pharm. 8 (2), 166–171.
Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., Arrouays, D., 2011. Spatial distribution of soil organic carbon stocks in France. Biogeosciences 8, 1053–1065.
Martin, M.P., Orton, T.G., Lacarce, E., Meersmans, J., Saby, N.P.A., Paroissien, J.B., Joliveta, C., Boulonnea, L., Arrouays, D., 2014. Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma 223, 97–107.
McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
Meersmans, J., Van Wesemael, B., De Ridder, F., Van Molle, M., 2009. Modelling the three-dimensional spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). Geoderma 152 (1), 43–52.
Meersmans, J., Van Wesemael, B., Goidts, E., Van Molle, M., De Baets, S., De Ridder, F., 2011. Spatial analysis of soil organic carbon evolution in Belgian croplands and grasslands, 1960–2006. Glob. Chang. Biol. 17 (1), 466–479.
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.
Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. Adv. Agron. 118 (3), 4.
Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., ... Field, D.J., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86.
Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263, 16–34.
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.
Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. Methods of soil analysis part 3—chemical methods, (methodsofsoilan3) 961–1010.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes - heterotopic cokriging and regression-kriging. Geoderma 67, 215–226.
Olaya, V.F., 2004. A Gentle Introduction to Saga GIS. The SAGA User Group e.V, Göttingen, Germany, pp. 208. http://downloads.sourceforge.net/saga-gis/SagaManual.pdf.
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 (2), 226–239.
Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation—an example from Scotland. Geoderma 232, 284–299.
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 (3), 165–190.
Quilchano, C., Marañón, T., Pérez-Ramos, I.M., Noejovich, L., Valladares, F., Zavala, M.A., 2008. Patterns and ecological consequences of abiotic heterogeneity in managed cork oak forests of Southern Spain. Ecol. Res. 23 (1), 127–139.
R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.
Razakamanarivo, R.H., Grinand, C., Razafindrakoto, M.A., Bernoux, M., Albrecht, A., 2011. Mapping organic carbon stocks in eucalyptus plantations of the central highlands of Madagascar: a multiple regression approach. Geoderma 162 (3–4), 335–346.
Reeves, D.W., 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil Tillage Res. 43 (1–2), 131–167.
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.
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.
Sculla, P., Franklina, J., Chadwickb, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. 27 (2), 171–197.
Soon, Y.K., Abboud, S., 1991. A comparison of some methods for soil organic carbon determination. Commun. Soil Sci. Plant Anal. 22 (9–10), 943–954.
Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., ... Wheeler, I., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agric. Ecosyst. Environ. 164, 80–99.
Tian, G., Justicia, R., Coleman, D.C., Carroll, C.R., 2007. Assessment of soil and plant carbon levels in two ecosystems (woody bamboo and pasture) in Montane Ecuador. Soil Sci. 172, 459–468.
Toutin, T., 2002. Review article: geometric processing of remote sensing images: models, algorithms and methods. Int. J. Remote Sens. 25, 1893–1924.
Tsui, C.C., Chen, Z.S., Hsieh, C.F., 2004. Relationships between soil properties and slope position in a lowland rain forest of southern Taiwan. Geoderma 123 (1–2), 131–142.
Tsui, C.C., Tsai, C.C., Chen, Z.S., 2013. Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. Geoderma 209, 119–127.
Van Wambeke, A.R., 2000. The Newhall Simulation Model for estimating soil moisture & temperature regimes. Conservation Service: Department of Crop and Soil Sciences Cornell University, Ithaca, NY USA. http://www.bfenvironmental.com/pdfs/nsmt.pdf.
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., Huang, M., Shao, X., Mickler, R.A., Li, K., Ji, J., 2004. Vertical distribution of soil organic carbon in China. Environ. Manag. 33 (1), 200–209.
Wang, S., Wang, X., Ouyang, Z., 2011. Effects of land use, climate, topography and soil properties on regional soil organic carbon and total nitrogen in the Upstream Watershed of Miyun Reservoir, North China. J. Environ. Sci. 24 (3), 387–395.
Wang, Z., Liu, G.B., Xu, M.X., Zhang, J., Wang, Y., Tang, L., 2012a. Temporal and spatial variations in soil organic carbon sequestration following revegetation in the hilly

Loess Plateau. China. Catena 99, 26–33.
Wang, J., Pan, X., Liu, Y., Zhang, X., Xiong, Z., 2012b. Effects of biochar amendment in two soils on greenhouse gas emissions and crop production. Plant Soil 360 (1-2), 287–298.
Wang, K., Zhang, C.R., Li, W.D., 2013. Predictive mapping of soil total nitrogen at a regional scale: a comparison between geographically weighted regression and cokriging. Appl. Geogr. 42 (8), 73–85.
Wang, S., Wang, Q., Adhikari, K., Jia, S., Jin, X., Liu, H., 2016. Spatial-temporal changes of soil organic carbon content in Wafangdian, China. Sustainability 8 (11), 1154.
Yang, L., Zhu, A.X., Qi, F., Qin, C.Z., Li, B.L., T. P., 2013. An integrative hierarchical stepwise sampling strategy and its application in digital soil mapping. Int. J. Geogr. Inf. Sci. 27 (1), 1–23.
Yang, L., Qi, F., Zhu, A.X., Shi, J.J., Y.M., 2016a. Evaluation of integrative hierarchical stepwise sampling for digital soil mapping. Soil Sci. Soc. Am. J. 80 (3), 637–651.
Yang, R., Zhang, G., Liu, F., Lu, Y., Yang, F., Yang, F., Yang, M., Zhao, Y., Li, D., 2016b. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecol. Indic. 60, 870–878.
Yang, R., Zhang, G., Yang, F., Zhi, J., Yang, F., Liu, F., Zhao, Y., Li, D., 2016c. Precise estimation of soil organic carbon stocks in the northeast Tibetan Plateau. Sci. Rep. 6.
Zhang, W., Du, T., Wang, J., 2016. Deep learning over multi-field categorical data. In: In European Conference on Information Retrieval. Springer International Publishing, pp. 45–57.
Zhang, Y., Biswas, A., Adamchuk, V.I., 2017. Implementation of a sigmoid depth function to describe change of soil pH with depth. Geoderma 289, 1–10.
Zhu, A.X., Yang, L., Li, B., Qin, C., English, E., Burt, J.E., Zhou, C., 2008. Purposive sampling for digital soil mapping for areas with limited data. In: Hartemink, A.E. (Ed.), Digital Soil Mapping With Limited Data. Springer, pp. 33–245.
Zuo, Y., Serfling, R., 2000. Structural properties and convergence results for contours of sample statistical depth functions. Ann. Stat. 483–499.