Catena 165 (2018) 576–589

Contents lists available at ScienceDirect

Catena

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

Examining soil organic carbon distribution and dynamic change in a hickory plantation region with Landsat and ancillary data

Wei Lu[a], Dengsheng Lu[a,*], Guangxing Wang[b], Jiasen Wu[a], Jianqin Huang[a], Guiying Li[a]

[a] State Key Laboratory of Subtropical Silviculture, School of Environmental & Resource Sciences, Zhejiang Agriculture and Forestry University, China
[b] Department of Geography, Southern Illinois University at Carbondale, IL 62901, USA

ARTICLE INFO

Keywords:
Soil organic carbon
Dynamic change
Driving factors
Random forest
Geographically weighted regression
Hickory plantation

ABSTRACT

Soil organic carbon (SOC) is an important soil property relating to soil formation, structure, and water-holding capacity. Remote sensing has been used for predicting SOC but remains a challenge due to the complex and indirect relationship between SOC and remote sensing variables. In this research we explored the approaches of predicting SOC distribution in a hickory plantation region using random forest (RF) and geographically weighted regression (GWR) based on Landsat and ancillary data, analyzed SOC spatial distribution and dynamic change between 2008 and 2013 through a thresholding approach, and examined major factors that resulted in SOC degradation in the young and mature hickory plantations using a logistic regression. The results showed that RF outperformed GWR in the prediction of SOC and provided stable and reliable SOC predictions with root mean squared errors of 4.6 $g\,kg^{-1}$ in 2008 and 4.4 $g\,kg^{-1}$ in 2013. A large area of hickory plantation was experiencing SOC decrease. The analysis of major factors causing SOC degradation indicated that steep slope and high proportion of silt component in soil resulted in SOC decrease in the young hickory plantations, and high elevation, high proportion of silt component in the soil, and the increase of soil fraction in the ground cover led to SOC decrease in the mature hickory plantations. This research provides valuable approaches to spatially predict SOC and identify major factors driving SOC degradation, which will be useful for adopting better measures to improve management of hickory plantations.

1. Introduction

Soil organic carbon (SOC) is one of the important soil properties that relate to soil formation and structure and water-holding capacity (Reeves, 1997; Y. Wang et al., 2016a). The SOC's spatial distribution and dynamic balance are important indicators in evaluating soil quality and sustainable use. SOC quantity and spatial patterns are thus related to local ecological environments, the characteristics of plant communities, and human activities (Su et al., 2005; C. Wu et al., 2016; J. Wu et al., 2013). It is reported that carbon in soil is three times greater than in terrestrial vegetation, and has twice the amount of the atmospheric carbon (Davidson et al., 2000; Eswaran et al., 1993; K. Wang et al., 2012). Therefore, the spatial distribution and dynamic change of SOC are considerably significant in the control of global environmental and climate change, and sustainability of natural resources (K. Wang et al., 2012). Since SOC is dynamic and different factors such as topographic features and human activities can affect its spatial patterns and dynamic change, timely updating of SOC distribution is necessary.

Traditional methods of obtaining SOC content are based on field surveys at typical sites, but cannot provide SOC's spatial distribution at larger scales. One option to produce SOC spatial distribution is to develop a prediction model through establishing the relationships between SOC and ancillary variables such as vegetation, terrain, and soil properties (Mishra et al., 2010; C. Wu et al., 2009). However, different data quality and spatial resolutions of the ancillary variables often lead to large uncertainties of SOC predictions. An alternative is to use remote sensing technologies to develop an SOC prediction model because remote sensing data can repeatedly capture land surface features of large areas (C. Wu et al., 2009), but remotely sensed images cannot directly measure SOC under forest canopies. A common approach is to

Abbreviations: CC, correlation coefficient; CO, contrast; DEM, digital elevation model; DI, dissimilarity; DVI, difference vegetation index; EN, entropy; EVI, enhanced vegetation index; GLCM, grey-level co-occurrence matrix; GVSH, vegetation-shade index; GVSO, vegetation-soil index; GWR, geographically weighted regression; HO, homogeneity; LS-factor, slope length and steepness factor; ME, mean; MFD, multiple flow direction; MLR, multiple linear regression; MNDWI, modified normalized difference water index; NMI, normalized multi-component index; RF, random forest; RMSE, root mean squared error; SCA, specific catchment area; SFD, single flow direction; SM, second moment; SOC, soil organic carbon; TVI, transformed vegetation index; TWI, topographic wetness index; VA, variance

* Corresponding author at: Zhejiang Agriculture and Forestry University, Hangzhou 311300, Zhejiang, China,
E-mail addresses: harrylue@sina.com (W. Lu), luds@zafu.edu.cn (D. Lu), gxwang@siu.edu (G. Wang), jswu@zafu.edu.cn (J. Wu), huangjq@zafu.edu.cn (J. Huang), gyli@zafu.edu.cn (G. Li).

https://doi.org/10.1016/j.catena.2018.03.007
Received 8 August 2017; Received in revised form 21 February 2018; Accepted 7 March 2018
0341-8162/

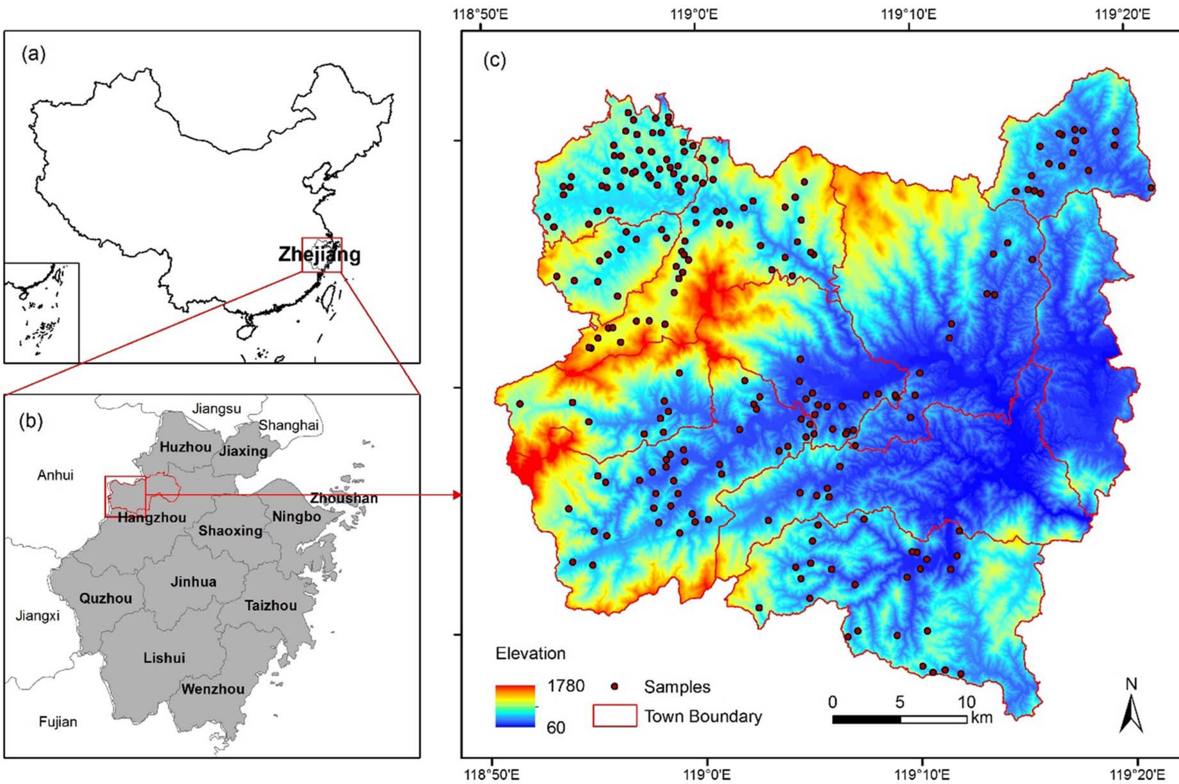

Fig. 1. Study area in western Lin-An District, Zhejiang Province, China, and spatial distribution of soil samples over hickory plantations: (a) location of Zhejiang Province; (b) location of Lin-An District in Zhejiang Province; and (c) spatial distribution of soil samples over elevation in the study area.

combine remote sensing and ancillary data to develop SOC models to indirectly predict SOC distribution (Kumar et al., 2012; Liu et al., 2015; Mishra et al., 2010; K. Wang et al., 2012, 2013; C. Wu et al., 2009). The widely used remote sensing variables include spectral bands, vegetation indices, and textural images, while ancillary attributes mainly consist of topographic factors (e.g., slope, aspect, elevation) from digital elevation models (DEMs) and SOC-relevant soil attributes (Mishra et al., 2010; K. Wang et al., 2012, 2013). Previous studies used spatial interpolation algorithms such as multiple linear regression (MLR), ordinary kriging, regression kriging, geographically weighted regression (GWR), and geographically weighted regression kriging to produce the spatial patterns of SOC (Gollini et al., 2015; Harris et al., 2010; Huang et al., 2012; Kumar et al., 2012; Liu et al., 2015; Mishra et al., 2009; Song et al., 2016; K. Wang et al., 2012; C. Wu et al., 2009; Zhang et al., 2011).

The algorithms have their own merits and shortcomings. For example, MLR is a global method and often leads to unbiased SOC prediction of a study area but ignores the heterogeneity (local spatial variability) (Hastie et al., 2009), resulting in large uncertainties in various locations. Ordinary kriging is a common algorithm for spatial interpolation by taking spatial autocorrelation and heterogeneity into account (Mishra et al., 2009). Some studies have indicated that MLR and regression kriging can predict SOC more accurately than ordinary kriging (Kumar and Lal, 2011; Kumar et al., 2012; Song et al., 2016; K. Wang et al., 2012, 2013; Zhang et al., 2011). A comparative analysis of GWR and regression kriging for SOC prediction has shown that GWR had a better performance than regression kriging (K. Wang et al., 2012) and a similar conclusion was also obtained by Mishra et al. (2010).

In recent years, machine learning algorithms such as random forest (RF), artificial neural network (ANN), and support vector regression (SVR) have been utilized to develop the relationships between SOC and the predictors (Aitkenhead and Coull, 2016; Y. Wang et al., 2016a; Were et al., 2015). RF is an ensemble-learning algorithm that can deal with noise and large datasets (Breiman, 2001). Essentially, RF consists of many regression trees, each tree creates a prediction, and then the most accurate prediction can be output. For each node in each regression tree, a small set of input variables are randomly selected for splitting each node, and the criterion for choosing the best split variables is based on the lowest GINI index and error measure (Breiman, 2001; L. Wang et al., 2016b). Thus, RF has advantages in dealing with multiple variables and nonlinear data, preventing overfitting, and keeping model stability compared with the methods based on the multivariate statistics (Y. Wang et al., 2016a). The examination of using RF for SOC prediction found that RF was effective even in complex topographic regions (L. Wang et al., 2016b).

Although remote sensing–based approaches have been employed for SOC prediction, rarely has research been conducted on SOC prediction in plantations. In particular, the research on SOC dynamic change has not been explored because of high uncertainty in SOC prediction, the serious impacts of human-induced factors on SOC, and lack of field survey data. Based on the sample plot data collected in 2008 and 2013 at the same locations, we explored the use of RF and GWR to predict SOC spatial distribution in a hickory (*Carya cathayensis*) plantation region. The objective of this research was to identify a suitable approach to generate the SOC spatial distributions through the combination of Landsat and ancillary data in 2008 and 2013, respectively, and to enhance the understanding of the SOC spatial patterns and their temporal trends through examining the SOC dynamic change. Another objective is to identify the major factors resulting in SOC degradation using a logistic regression model so that better measures can be adopted to improve management of hickory plantations.

2. Material and methods

2.1. Study area

Hickory products (nuts) have become one of the most important economic sources for local farmers in the past three decades, thus hickory plantations have experienced rapid expansion. However, the