Catena 166 (2018) 134–146

Contents lists available at ScienceDirect

Catena

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

Mapping total soil nitrogen from a site in northeastern China

Shuai Wang[a,1], Xinxin Jin[a,1], Kabindra Adhikari[b], Wenwen Li[c], Miao Yu[d], Zhenxing Bian[a,*], Qiubing Wang[a,*]

[a] College of Land and Environment, Shenyang Agricultural University, Shenyang, Liaoning Province 110866, China
[b] University of Arkansas, Department of Crop, Soil, and Environmental Sciences, Fayetteville, AR 72701, United States
[c] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA
[d] College of Sciences, Shenyang Agricultural University, Shenyang, Liaoning Province 110866, China

ARTICLE INFO

Keywords:
Digital soil mapping
Remote sensing
Random forest

ABSTRACT

Precise mapping of the spatial distribution of total soil nitrogen (TSN) is essential for soil resource management, agronomic sustainability, and nitrogen sequestration potential. Present study used three random forest (RF) and three multiple linear stepwise regressions (MLSR) models to map the distribution of topsoil (0–20 cm) TSN content in Lushun City in China's northeastern Liaoning Province. A total of 12 covariates (including topography, climate, and remote sensing images) and 115 soil samples were employed. An independent validation set of 23 samples was used to verify the performance of the model based on mean absolute prediction error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and Lin's concordance correlation coefficient (LCCC). Accuracy assessments showed that the RF model, combined with all environmental variables, had the best prediction performance. The second best was produced by the use of remote sensing alone. The third was the model that used only topographic and climatic variables alone. Remote sensing was not significantly inferior to the use of the model with all variables and can be used to build a realistic model. In the model with all covariates, the distribution of TSN is mainly explained by remote sensing, followed by topography variables, and climatic variables; their relative importance (RI) was 48%, 36%, and 16%, respectively. The results of remote sensing on the robust dependence of TSN should provide a link to dense natural vegetation in this forested environment. Additionally, remote sensing and its derived environmental variables should be used as the main predictors when mapping TSN in forested areas and other areas with similar vegetation coverage.

1. Introduction

Soil contains one of the most important nitrogen stocks in terrestrial ecosystems (Batjes, 1996). Measurements of total Soil nitrogen (TSN) provide a major index of soil nutrients; thus TSN has a strong relationship with soil resource management, agronomic sustainability, and nitrogen sequestration potential (Reeves, 1997; Kaushal et al., 2006; Yang et al., 2010; Qu et al., 2012; Ruiz et al., 2013; Wang et al., 2017b). Reliable estimates of the spatial distribution of TSN and factors controlling it are essential for understanding regional N cycling and establishing soil N sequestration programs (Nieder and Benbi, 2008; Jelinski and Kucharik, 2009; Zhao et al., 2011).

Natural ecological processes (Chaminade, 2005a; Tu, 2011; Wang et al., 2013) and the production practices of human society (Kaushal et al., 2006; Qu et al., 2012) influence the spatial distribution of TSN. Thus, precisely predicting TSN content at the regional scale is an extremely challenging task. However, it is difficult to obtain the regional details of the distribution TSN by dense sampling on a large scale due to the high cost in sample acquisition and analysis (Yang et al., 2016). A convenient and low-cost technique called digital soil mapping (DSM) has been introduced to estimate the distribution of TSN across large-scale areas based on a reduced sampling dataset and environmental variables (Wang et al., 2016). The basic assumption of DSM is that soil-landscape model is a function of Jenny's (1941) equation, which states that the formation of soil is caused by the interactions between climate, organisms, time, relief, and parent material.

Vegetation intensity is one of the major covariate related to TSN in digital soil mapping, especially in areas with good natural vegetation coverage (Yang et al., 2016). It can be obtained through remote sensing derivatives such as vegetation map, biomass map, and vegetation index, which has been widely applied to TSN prediction based on various DSM methods (McBratney et al., 2003; Wang et al., 2016; Yang et al., 2016).

* Corresponding authors at: College of Land and Environment, Shenyang Agricultural University, No. 120 Dongling Road, Shenhe District, Shenyang, Liaoning 110866, China.
E-mail addresses: zhx-bian@hotmail.com (Z. Bian), wangqbsy@yahoo.com (Q. Wang).

[1] These authors contributed equally to this paper.

https://doi.org/10.1016/j.catena.2018.03.023
Received 12 July 2017; Received in revised form 22 March 2018; Accepted 25 March 2018
0341-8162/

Image data from Landsat TM, SPOT, and IKONOS have successfully been used to study the spatial distribution of TSN in previous studies (Dalal and Henry, 1986; Barnes et al., 2003; Sullivan et al., 2005). These studies frequently used the linear regression equation and the image band value to predict TSN content in uniform soils or the bare soil surface or partial vegetation coverage area, so as to minimize the impact of vegetation on TSN.

Of the remote sensing satellites, Landsat 5 has been the most significant source of earth resource information in the world and data has been widely used in DSM. Remote sensing data provides a direct representation of the surface, and if it can be shown to be closely-related to TSN, DSM of this property could be considerably simplified. Therefore, it was possible to predict TSN content directly by remote sensing data, because TSN variation was influenced by natural vegetation as well as by human activities, especially the surface TSN in natural environments, which has been proved to have a good correlation with the above-ground-biomass (Yang et al., 2015).

A variety of DSM techniques have been used to predict soils distribution including TSN. Multiple linear regression (Selige et al., 2006; Wang et al., 2017a), regression kriging (Sumfleth and Duttmann, 2008; Hengl et al., 2015), ordinary kriging (Wang et al., 2013; Elbasiouny et al., 2014), random forest (RF) model (Hengl et al., 2015), geographically weighted regression (Wang et al., 2013), Cubist models (Bui et al., 2006; Adhikari et al., 2013), and principal component regression (Chang et al., 2001). Although many DSM techniques have been applied to estimate TSN, the tree based models used in previous studies can seldom be retrieved.

Tree based models such as random forest (RF) has been widely used in the prediction of soil properties (Peters et al., 2008; Stoorvogel et al., 2009; Hengl et al., 2015; Wang et al., 2016). The RF model accomplishes the task of improving the performance of tree model by combining many single trees; specifically, it is more stable than traditional single tree models (Breiman, 2001). As a data mining method, the RF model increases the predictive performance by reducing over-learning and over-fitting the RF model can deal with missing and outlier data types, interaction between variables, and evaluate the final fitting of the model (Breiman, 2001; Skurichina and Duin, 2002; Yang et al., 2016).

A number of environmental variables and DSM techniques could be employed to map soil attributes, however, an approach combining RF and remote sensing data to predict TSN in a forested site is still limited and rarely reported in the literature (Wang et al., 2015; Xu et al., 2017, 2018). We evaluated the potential of using RF and remote sensing for mapping topsoil (0–20 cm) TSN in an ecological region in northeastern China. The primary objectives were to (1) construct an RF model based on data collected at 92 sample point and remote sensing data to predict TSN content; (2) explore the important role of remote sensing images in delineating variations in TSN, and (3) validate the performance of this process and analyze its potential application prospects.

2. Materials and methods

2.1. Study area

The study area (38.72°–38.97°N, 121.08°–121.47°E) covers a total area of 512 km^2 in Lushun District, Liaoning Province, China (Fig. 1). The main land use in Lushun District can be classified as forest, cultivated land, urbanized land, rivers and reservoirs, and roads (Fig. 2). Forested regions are mainly distributed in Eastern and Western Mountain Areas, accounting for 54.5% of the study area. Much of the natural vegetation is warm-temperate deciduous broad-leaved forest where the main tree species are Pinaceae, Taxodiaceae, Cupressaceae and Ginkgoaceae. The study area is a national natural reserve, famous scenic spot, and forest park. The elevation of this area increases as one moves from the southwest towards the northeast, with elevations ranging from 0 m to 466 m above sea level. The region has a temperate continental monsoon climate characterized by mild temperatures and sufficient sunshine. Annual mean precipitation ranges between 590 and 780 mm with more than half of the rainfall concentrated in frequent rainstorms during the rainy season (June–August). The annual mean temperature is 10 °C, peaking at 27.5 °C in August with December as coldest month at temperature is 8.2 °C. The main geomorphological types include hilly mountains, littoral karst, and marine erosional granite landforms. Based on the soil map of the second soil survey (OSSLP, 1990) and the soil reference on the bases of Chinese soil taxonomy (Gong et al., 2002), we were able to compile a soil map of the study area according to WRB legend (IUSS Working Group, 2006). Major soil types found in the study area are Anthrosols, Cambisols, Gleysols, Histosols, Leptosols, Luvisols, and Phaeozems. Cambisols is the most widely distributed soil type accounting for 58% of the total area followed by Luvisols (16%), which is mainly distributed in the eastern and western coastal mountains. Anthrosols are common in the north and central areas whereas, Histosols, Gleysols and Phaeozems are found in the eastern and western mountainous area along river valleys, flat or low foothills, and diving seepage areas. Leptosols is mainly distributed in the slope terrain of the eastern and western mountainous areas.

2.2. Soil observations

Rugged terrain, the expense of sampling, and forests that cover a very large area (54.5% of the region) made extensive field sampling impractical in this study. In order to represent the spatial characteristics of TSN content in complex geographical landscapes, this study adopted a purposive sampling strategy, using the method of Zhu et al. (2008). Based on the pedogenesis of TSN of the study area, the main environmental factors considered in sampling design were soil types (Anthrosols, Cambisols, Gleysols, Histosols, Leptosols, Luvisols, and Phaeozems), land use (forest, cultivated land, and urbanized land) and terrain conditions (elevation, slope, and aspect) (Zeng et al., 2016, 2017). The study area was first stratified using the combination of soil type and land use type (Yang et al., 2011). Thirteen soil-land-use units were obtained using a fuzzy C-means classification method based on seven soil subgroups from the Second National Soil Survey (OSSLP, 1990) and three land-use types including forest, cultivated land, and urbanized land. Within each soil-land-use type unit, nine or ten samples were taken at different landform positions by local soil experts. This method can allow researchers to obtain a small number but representative soil samples from an area (An et al., 2017). Finally, a total of 115 soil samples were collected in 2011 from the topsoil depth (0–20 cm) excluding litter layer, if present (Fig. 1). Randomly selected 80% samples were used as model calibration (n = 92), and the remaining 20% for model validation and accuracy assessment (n = 23). The geographical coordinates of sample locations were recorded using a handheld global positioning system. Soil samples were subsequently dried, ground, and sieved at 2 mm and measured for TSN content using the semimicro-Kjeldahl method (Bremner and Mulvaney, 1982) at the Key Laboratory of Agricultural Resources and Environment of Liaoning Province, Shenyang Agricultural University, Shenyang, China.

2.3. Environmental data

Twelve environmental variables representing topography, climate, and remote sensing data were adopted as predicting covariates of TSN. All covariates are generated and converted into raster data (30 × 30 m grid) using ArcGIS 10.1 (Environmental Systems Research Institute (ESRI), Redlands, CA, 2012). In order to accurately predict the spatial distribution of TSN, high-precision of covariates served as one of the essential elements. Although the high accuracy of the covariates can provide relatively detailed information about the attributes, 30-m resolution is sufficient to present the spatial distribution characteristics of TSN in this study area.