Ecological Indicators 45 (2014) 184–194

Contents lists available at ScienceDirect

Ecological Indicators

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

Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau

Fuqiang Dai[a,b], Qigang Zhou[a], Zhiqiang Lv[a], Xuemei Wang[b], Gangcai Liu[b,*]

[a] College of Tourism and Land Resources, Chongqing Technology and Business University, Chongqing 400067, China
[b] Key Laboratory of Mountain Hazards and Earth Surface Processes, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences and Ministry of Water Resources, Chengdu 610041, China

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 21 March 2014
Accepted 1 April 2014

Keywords:
Soil organic matter
Digital soil mapping
Artificial neural network
Ordinary kriging
Accuracy improvement

ABSTRACT

Soil organic matter (SOM) content is considered as an important indicator of soil quality. An accurate spatial prediction of SOM content is so important for estimating soil organic carbon pool and monitoring change in it over time at a regional scale. Due to the unfavourable natural conditions in Tibetan Plateau, soil sampling with high density is time consuming and expensive. As a result, little research has focused on the spatial prediction of SOM content in Tibet because of shortage of data. We used a two-stage process that integrated an artificial neural network (ANN) and the estimation of its residuals by ordinary kriging to produce accurate SOM content maps based on sparsely distributed observations and available auxiliary information. SOM content data were obtained from a soil survey in Tibet and were used to train and validate the ANN-kriging methodology. Available environmental information including elevation, temperature, precipitation, and normalized difference vegetation index were used as auxiliary variables in the ANN training. The prediction accuracy of SOM content was compared with those of ANN, universal kriging, and inverse distance weighting (IDW). A more accurate prediction of SOM content was obtained by ANN-kriging, with lower global prediction errors (root mean square error = 6.02 g kg^{-1}) and higher Lin's concordance correlation coefficient (0.75) for validation sampling sites compared with other methods. Relative improvements of 26.94–37.10% over other methods were observed in the prediction of SOM content. In conclusion, the proposed ANN-kriging methodology is particularly capable of improving the accuracy of SOM content mapping at large scale.

1. Introduction

Soil organic matter (SOM) content is considered as an important indicator of soil quality which greatly contributes to the ecosystem services such as food production and carbon sequestration. SOM has an important effect on many soil properties and major biogeochemical cycles and is a strong indicator of fertility and land degradation (Manlay et al., 2007). With increasing SOM content, the physical and hydraulic properties of soil are improved, and its nutrient-holding capacity is enhanced (Benjamin et al., 2008). A better understanding of SOM content and its spatial variation is essential for using soils efficiently and maintaining soil productivity. Furthermore, soil, being an important sink of atmospheric carbon, contains about three times more organic carbon (an essential part of SOM) than vegetation and about twice as much carbon as the atmosphere (Batjes and Sombroek, 1997). Therefore, an accurate SOM content map is favoured to estimate the baseline of the organic carbon pool in soil and monitor the change over time at various scales so as to understand responses of the terrestrial system to climate change (Milne et al., 2007) and to verify soil carbon sequestration (Smith, 2004).

Digital soil mapping (DSM) in soil science, also referred to as predictive soil mapping or pedometric mapping, is the creation and the population of a geographically referenced soil database, generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships (McBratney et al., 2003; Scull et al., 2003). DSM has paid more attention on mapping SOM content as well as estimating carbon stocks (Karunaratne et al., 2014). But prediction accuracy of SOM content for the typical physiogeographical area such as Tibetan Plateau is too low to help with estimating the soil organic carbon pool and monitoring changes in it over time. Fortunately, a global project has been formed to make a new digital soil map of the world using emerging technologies, in view of a need for accurate and spatially referenced soil information (Sanchez et al., 2009). The need is just consistent with developing a methodology

* Corresponding author. Tel.: +86 28 85231287; fax: +86 28 85238973.
E-mail address: liugc@imde.ac.cn (G. Liu).

http://dx.doi.org/10.1016/j.ecolind.2014.04.003
1470-160X/

in the study that allows for improvements in accurately predicting SOM content at large scale. During the last decades, various DSM methods have been used for mapping SOM content based on soil observations. Generally, there are two main types of interpolation methods: deterministic and stochastic. The former uses deterministic techniques, which calculate the unknown values based on parameters that control either (i) the extent of the similarity of values (e.g., IDW, inverse distance weighting) or (ii) the degree of smoothing of the surface (e.g., RBF, radial basis function). The latter are stochastic techniques, which assume that at least some of the spatial variation of natural phenomena can be modelled by random processes with spatial autocorrelation. Kriging is one of the most widely used methods among the stochastic techniques and is the best linear unbiased estimator in the sense that it minimizes the variance of the estimation error (Webster and Oliver, 2001). Kriging methods have shown considerable advantages in the prediction of soil properties including SOM content, compared with deterministic interpolation methods (Liu et al., 2008; Schloeder et al., 2001; Worsham et al., 2010). Although sampling density could be reduced to a certain degree relative to the deterministic techniques, the accuracy of the SOM content maps with kriging interpolation is strongly dependent on the number of soil samples. Further, a large variability in elevation, together with difference in climate and vegetation, can result in strong local variation of SOM content. In such cases, the reliability of kriging estimates deceases, as site-specific measurements can be influenced by strong local variation (Alsamamra et al., 2009). Alternatively, cokriging has been used to overcome this limitation by incorporating auxiliary variables that may compensate for the lack of data and the small sample size (Odeh et al., 1995). This method is valuable when the auxiliary variable is highly correlated with the SOM content and is sampled intensely (Wu et al., 2009). Previous studies have shown that environmental variables, such as terrain attributes (Mueller and Pierce, 2003), remotely sensed data (Wu et al., 2009), and climate data (Mishra et al., 2010), have had the potential as useful auxiliary variables for improving the accuracy and reliability of SOM prediction. Instead of directly including the auxiliary variables in the kriging process, regression kriging was proposed which considers the auxiliary environmental data as independent variables in multiple linear regression and then incorporates the prediction of soil attributes and the kriging of residuals (Frogbrook and Oliver, 2001; Piccini et al., 2014; Takata et al., 2007; Zhao et al., 2014). The accuracy of these methods is still dependent on the density and size of sampling sites, as these methods are based on interpolation, which requires some data as inputs. Therefore, a more efficient method is required to improve the accuracy of interpolation methods for producing high-resolution SOM content maps.

Essentially, the non-linearity and multicollinearity problems existed in the interpolation process of both cokriging and multiple linear regression (Alvarez et al., 2011; Gautam et al., 2011). To overcome these problems along with the limited soil observations and available auxiliary information, linear mixed model (Karunaratne et al., 2014), support vector regression (SVR) (Ballabio, 2009), and artificial neural network (ANN) (Malone et al., 2009; Zhao et al., 2010) were introduced to construct the relationships between SOM content and other environmental variables and thus to produce more accurate SOM content maps. Further, fewer studies used other machine learning approaches like Cubist (Lacoste et al., 2014), boosted regression tree (Martin et al., 2011), and Random Forest (RF) (Subburayalu and Slater, 2013) to predict SOM in order to improve prediction accuracy. However, a major disadvantage of these methods is that the SOM content at a particular grid node is derived only from auxiliary information at each point, regardless of the spatial autocorrelation of the surrounding measured data (Takata et al., 2007). Integrating ordinary kriging of residuals from ANN can incorporate the spatial autocorrelation of measured values which can lead to better predictions and lower error.

The Tibetan Plateau has been considered the best laboratory for environmental research because of the prominent responses of the plateau's biogeochemical cycle to global changes (Bi, 1997). According to the pattern and change of soil organic carbon storage in China, the SOM content has decreased in southern Tibet due to anthropogenic activities and natural factors (Wang et al., 2003; Dai et al., 2011). To scale down this result to a regional level, it is necessary to develop a method for improving the accuracy of SOM predictions. To our knowledge, however, little research has been conducted on the spatial prediction of SOM content in the south of Tibet (Li et al., 2009). In this study, a two-stage process methodology integrating ANN and kriging was proposed for predicting SOM content. The primary objective is to investigate the improvement in accuracy of SOM content mapping achieved by integrating ANN and residual estimation by kriging over ANN, universal kriging, and IDW.

2. Materials and methods

2.1. Study area

This study was conducted in one of seven physiogeographical units in Tibet classified by the Chinese Academy of Sciences (CAS, 1982) and is located in the south of Tibetan Plateau with a total area more than 220,000 km^2 (Fig. 1). Broad valleys and basins are the main types of topography in the area. It is characterized by plateau cold monsoon and semi-arid climate, with a mean annual temperature of 1–7.5 °C, a mean annual relative humidity of 40–50%, and a mean annual precipitation of 200–500 mm represent the ranges of means for different portions of the study area. According to Chinese soil classification, the main soil types are mountain shrub steppe soils and alpine steppe soils (Inceptisols in the USDA Soil Taxonomy and Cambisols in the FAO World Reference Base for Soil Resources). This area is an important agricultural zone in Tibet where the dominant land use is irrigated cropping systems producing grains and crops.

2.2. Soil observations and environmental variables

A total of 244 soil samples (0–25 cm) were collected over the study area in 2007 (Fig. 1), taking soil type, lithologic characters, vegetation, and topography into consideration. A large sampling grid size was designed because of the extent of the study area and its unfavourable natural conditions. Sampling sites were selected on a 64×64 km^2 grid for forest, pasture, and waste land and on an 8×8 km^2 grid for arable land. The longitude and latitude of each sampling site were recorded with a global positioning system (GPS). A portion of 3 kg of soil at each sampling site, taken from composite samples, was air dried and sieved to pass a 2-mm mesh to determine SOM content with wet oxidation with external heat applied (Liu, 1996; Mebius, 1960). In the procedure, SOM was oxidized by potassium dichromate and heated to about 170–180 °C for five minutes, and then excess potassium dichromate was determined by titration with standard 0.2 mol L^{-1} ferrous sulfate ($FeSO_4$). The SOM content was estimated by multiplying the amount of organic carbon by the conventional carbon factor (1.724) (Read and Ridgell, 1922).

In the study area, a digital elevation model (DEM) with a spatial resolution of 90 m was obtained from the Environmental and Ecological Science Data Center for West China, National Natural Science Foundation of China. Low-elevation areas were located in the valleys of the middle reaches of the Yarlung Zangbo River, whereas high-elevation areas were located in the south of