Geoderma 330 (2018) 9–18

Contents lists available at ScienceDirect

Geoderma

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

Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform

Hongfen Zhu[a,b,c], Wei Hu[d], Yaodong Jing[a,b], Yi Cao[a], Rutian Bi[a,b,*], Wude Yang[c]

[a] College of Resource and Environment, Shanxi Agricultural University, Taigu, Shanxi 030801, China
[b] National Experimental Teaching Demonstration Center for Agricultural Resources and Environment, Shanxi Agricultural University, Taigu, Shanxi 030801, China
[c] College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi 030801, China
[d] New Zealand Institute for Plant & Food Research Limited, Private Bag 4704, Christchurch 8140, New Zealand

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Soil organic carbon (SOC)
Scale
Wavelet transform
Inverse discrete wavelet transform (IDWT)
Covariates

ABSTRACT

The spatial distribution of soil organic carbon (SOC) is vital to agricultural and environmental management, and its variation is influenced by environmental factors operating at different scales and intensities. The objective of this study was to explore the scale-dependent effects of environmental factors on SOC distribution and to predict SOC based on their scale-specific relationships using wavelet transform. The spatial data of SOC, slope, soil water content (SWC), soil bulk density (SBD), sand, silt, and Landsat 8 remote sensing reflectance (Rrs) were extracted at 330 m interval along three transects in the arable land of Taiyuan basin, China. The spatial series of SOC and environmental variables along each of three transects were separated into six detail components and one approximation representing different scales using wavelet decomposition. The specific scale of each detail component was identified by Hilbert transform. The SOC variances over the entire basin were mainly explained (54%–86%) by scales of 12.70 and 21.12 km. Compared with the relationships between SOC and environmental factors at sampling scale, their multiscale correlations were better at larger scales. The SOC estimation using wavelet reconstruction based on predicted SOC at all scale components outperformed its prediction using stepwise multiple linear regression (SMLR) based on the original sampling data. The major contributing scale to SOC prediction was 12.70 km over the entire basin. In the prediction of overall SOC, Rrs was the major predictor in the upstream and downstream portions, whereas soil texture was the major contributor in the midstream portion. In this study, the SOC prediction using wavelet transform based on scale-dependent relationships with environmental variables generated new insights in soil properties estimation, and wavelet transform has potential for determining the multiscale relationships of soil properties with influencing factors.

1. Introduction

Soil organic carbon (SOC) has been recognized as a key factor in soil fertility and environmental management (She et al., 2014). Globally, the SOC pool is approximately three times more than that of the terrestrial vegetation or atmosphere pool (Lal, 2004) and has a significant effect on CO_2 concentration, affecting the rate of climate change and the state of the atmosphere (Eglin et al., 2010). The spatial distribution of SOC is influenced by a range of environmental factors operating at different scales and intensities (Zhou et al., 2016). Therefore, a detailed understanding of the scale-specific relationships between SOC and environmental factors becomes essential for better fertility management and prediction of greenhouse gas emissions.

According to soil formation theory (Jenny, 1941), SOC can be expressed as a function of environmental factors. Traditional statistical methods, including Pearson correlation and linear regression, have been widely used to explore the relationships between SOC and environmental variables at sampling scale (Zhu et al., 2016). However, the spatial distributions of soil variables are scale-dependent (Wiens, 1989), and their characteristics cannot be elucidated only at sampling scales. To reveal the scale-specific relationships between soil variables and their controlling factors, many advanced mathematical methods, including wavelet transform (Biswas, 2018; Huang et al., 2018), wavelet coherency (Biswas and Si, 2011a; Zhu et al., 2017), geostatistical method (Xu and Tao, 2004), empirical mode decomposition (EMD) (Biswas and Si, 2011b), and variations of empirical mode decomposition (multiple empirical mode decomposition [MEMD] and 2-dimentional empirical mode decomposition) (Hu and Si, 2013; Huang et al., 2017; She et al., 2017) have been proposed in soil science. However, the objective of these studies was the scale-dependent analysis rather

* Corresponding author at: College of Resource and Environment, Shanxi Agricultural University, Taigu, Shanxi 030801, China.
E-mail address: sxndbrt@163.com (R. Bi).

https://doi.org/10.1016/j.geoderma.2018.05.029
Received 20 February 2018; Received in revised form 11 May 2018; Accepted 25 May 2018

than the prediction of soil variables.

Wavelet analysis, including wavelet decomposition and wavelet reconstruction (inverse wavelet transform), facilitates the development of multiscale methods (Starck et al., 2007). Wavelet transform is commonly used in signal or spatial series processing by decomposing data series into the transform coefficients, obtaining the variation of the transform coefficients, and reconstructing the data series (Chui, 2016). It has been widely applied to explore the scale-specific spatial heterogeneity of soil variables (Biswas et al., 2013; Zhou et al., 2016) since the introduction of Lark and Webster (1999). Considering the attainability of the original spatial series after scale-dependent analysis using wavelet transform, the objectives of this study were to predict SOC based on its scale-specific relationship with influencing factors along three transects in the Taiyuan basin. Specifically, the objectives included (1) separating the overall variations of spatial series, including SOC and environmental variables, into different scale components using wavelet decomposition; (2) analyzing the relationships between SOC and environmental variables at each scale, and predicting SOC at different scales based on the environmental variables at corresponding scales; and (3) predicting the spatial distribution of SOC at the sampling scale based on predicted SOC at all scales using wavelet reconstruction.

2. Materials and methods

2.1. Study area

The study area is located in the Taiyuan basin in the Chinese Loess Plateau (Fig. 1a) and has an area of 6159 km^2 (37°00′-38°20′ N latitude, 111°30′-113°00′ E longitude). The basin is a typical semiarid area with a mean annual temperature of 9.5 °C, annual precipitation of 425–520 mm, and annual evaporation of 1780 mm. The basin is characterized by a thick loess-covered layer due to dust deposition during the Quaternary; its thickness ranges from 50 to 3000 m, with the loess grain size generally increasing from the center to the margin of the basin (Zhu et al., 2016). The Fen River, which is the second largest tributary of the Yellow River, runs through the basin from northeast to southwest. The major soil types are Calcaric Fluvisols and Calcaric Cambisols under alkaline conditions according to the FAO-90 soil classification (Nachtergaele et al., 2008), and the dominant crops are spring maize and winter wheat.

2.2. Experimental design

Along the Fen River, the Taiyuan basin was divided into upstream, midstream, and downstream portions. Three sampling transects perpendicular to the Fen River were established based on remote sensing images. The transects were proximately 42 km long and sampled at 330 m intervals, producing 121, 128, and 134 sampling points for transect 1, 2, and 3, respectively. If a sampling point was located on non-arable land such as buildings or roads, the nearest point on arable land was used to represent the sampling point. Fig. 1b shows the sampling points and Fig. 1c shows the transect topography, which was bowl-shaped with a depression in the center.

Before maize planting, a total of 383 points were sampled during March 22–31, 2016. The locations of the established soil samples were determined in the field by a GPS. At each sampling location, the undisturbed surface soil was collected using a metallic core cylinder of 100 cm^3 volume (5 cm in height and 5 cm in diameter). Soil bulk density (SBD) and gravimetric soil water content (SWC) was measured by oven-dry method (Hossain et al., 2015). At each location, a composite sample of 0–20 cm soil layer was collected from five sampling points. The samples were air-dried, gently crushed, and passed through a 2 mm sieve for soil particle size (sand and silt content) and SOC content. Sand (0.05–2.0 mm) and silt (0.002–0.05 mm) were measured by the pipette method (Gee and Bauder, 1986). The content of SOC was determined by the dichromate oxidation method (Nelson and Sommers, 1982).

The digital elevation model (DEM) with 30 m resolution was downloaded from an online resource (http://www.gscloud.cn/sources) and used to extract the slope gradient. Landsat-8 (L8) satellite image of March 31, 2016 was acquired from the United States Geological Survey (USGS) Earth Explorer (https://earthexplorer.usgs.gov/), and the sum of its bands except for the pan band and the two Thermal Infrared Sensor (TIRS) bands was used as the environmental factor of the reflectance of remote sensing (Rrs) in this study.

2.3. Discrete wavelet transform

Wavelet transform can be used to analyze the multiscale effects of spatial or time series, which arises from finite spatial or temporal domain (Biswas et al., 2013). For a spatial series Y measured along a location series X, the wavelet transform function is defined as

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} Y_i(X_i)\psi_{a,b}\left(\frac{X_i - b}{a}\right)dX_i \tag{1}$$

where $\psi(X_i)$ is the basic wavelet function and i is the spatial location. The parameter a is the dilation ($a > 1$) or contraction ($0 < a < 1$) factor, and b is the translation or shift of the function (Kumar and Foufoula-Georgiou, 1993). The inverse wavelet transform is defined as

$$Y_i(X_i) = \frac{2}{A+B}\sum_{a,b} W(a,b)\psi_{a,b}\left(\frac{X_i - b}{a}\right) \tag{2}$$

where A and B are upper and lower bounds, respectively. There are two types of wavelet transform, which are the continuous wavelet transform (CWT) and discrete wavelet transform (DWT). If the wavelet coefficients are calculated at continuous scales and locations, the method is defined as CWT. If the coefficients are determined at dyadic scales or scales with a two-fold increment, the method is known as DWT (Lark and Webster, 1999; Zhou et al., 2016). For DWT, a is defined as $a = a_0{}^m$, and b is defined as $b = \mathrm{k}a_0{}^m b_0$, where m and n are integers. Generally, a_0 and b_0 are 1/2 and 1, and the equation for DWT can be expressed as

$$\psi_{m,n}(x) = 2^{\frac{m}{2}}\psi(2^m x - n) \tag{3}$$

A more detailed description of wavelet transform can be found in other publications (Percival and Walden, 2000; Si, 2008).

2.4. Data analysis

The spatial series of SOC and environmental covariates, including slope, SWC, SBD, sand, silt, and Rrs along the three transects, were decomposed into seven scale components, which included six detail components (D1–D6) and one approximation (A6) by the DWT. Meanwhile, the momentary energy and frequency of SOC at the seven scale components were obtained using the Hilbert transform and were converted to period (1/frequency). The specific spatial scales of SOC represented by the detail components and the approximation were calculated after multiplying the period by the sampling interval (0.33 km). The percentage variance contribution of each scale component was calculated as

$$\text{Variance}\ (\%) = \frac{\mathit{Variance\ of\ a\ scale\ component}}{\sum \mathit{variances\ of\ all\ detail\ scales\ and\ approximation}} \times 100 \tag{4}$$

In addition, the correlation coefficients between SOM and the environmental covariates at each scale component were calculated. Finally, the SOC at each scale component were predicted from the environmental factors at the corresponding scale by stepwise multiple linear regression (SMLR), and the predicted SOC contents at the sampling scale were obtained from all predicted values of SOC at all scale components by wavelet reconstruction.

The descriptive statistical analysis for SOC, Pearson correlation