Geoderma 263 (2016) 254–263

Contents lists available at ScienceDirect

Geoderma

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

A similarity-based method for three-dimensional prediction of soil organic matter concentration

Feng Liu [a], David G. Rossiter [a,b], Xiao-Dong Song [a], Gan-Lin Zhang [a,*], Ren-Min Yang [a], Yu-Guo Zhao [a], De-Cheng Li [a], Bing Ju [a]

[a] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
[b] Section of Soil & Crop Sciences, School of Integrated Plant Science, Cornell University, Ithaca, NY 14850, USA

ARTICLE INFO

Article history:
Received 15 December 2014
Received in revised form 24 May 2015
Accepted 26 May 2015
Available online 17 June 2015

Keywords:
3D soil mapping
Similarity-based prediction
Depth function model
Soil organic matter concentration

ABSTRACT

This paper presents an approach to predicting three-dimensional (3D) variation of soil organic matter (SOM) concentration by integrating a similarity-based method with depth functions. It was tested in a small hilly landscape. A depth function model was constructed to fit SOM profile distribution using a linear relation in the topsoil and a power function in the subsoil. Then, under the assumption that similar environmental conditions at two sites would lead to the development of similar profile morphologies and thus similar depth function parameters, the similarity-based method was used to spatially interpolate the depth function parameters based on their relationships with environmental variables. With the values of the parameters for every location, a 3D map of SOM distribution was generated. The predicted SOM pattern well reproduced the statistical distribution of the pedon dataset used in this study. The overall mean error (ME) was 0.06 g kg^{-1} and ratio of performance to deviation (RPD) was 2.34. We conclude that the proposed approach is effective and accurate for 3D SOM prediction. It overcomes two drawbacks of the frequently used pseudo 3D soil mapping approach: (1) the neglect of vertical soil pattern when performing horizontal soil predictions, and (2) the repeated applications of depth function fittings in the mapping process, both of which may lead to prediction errors. Moreover, the similarity-based method is a transparent and traceable prediction process, allowing for easy interpretation of its results. This is useful for understanding soil–environmental relationships and processes. The method thus is an attractive alternative to the commonly used non-linear "black-box" techniques such as artificial neural networks.

1. Introduction

Detailed and accurate three-dimensional (3D) soil information is required for estimating soil carbon stocks and modeling hydrological processes (Sanchez et al., 2009). Conventional polygon-based soil mapping techniques usually cannot provide such information. To meet this requirement, the GlobalSoilMap.net project, officially launched by a global consortium of scientists in 2009, intends to develop digital soil mapping methods to produce a fine-resolution, 3D soil data products across the globe (GlobalSoilMap Science Committee, 2013).

Many attempts have been made on 3D soil mapping (Minasny et al., 2013; Arrouays et al., 2014). Most considered it as multiple 2D soil mapping operations at a set of predefined depth intervals (Tekin et al., 2008; Grimm et al., 2008; Vasques et al., 2010a,b; Malone et al., 2009; Adhikari et al., 2013; F. Liu et al., 2013; Odgers et al., 2015). These 2D mapping results are represented as depth averages (for concentrations) or sums (for stocks). These averages can be reconstructed into a full 3D soil property map (Malone et al., 2011; Lacoste et al., 2014). Although multiple 2D mapping is simple to implement, F. Liu et al. (2013) argued that it is a pseudo 3D mapping approach and has two drawbacks. One is that soil variation pattern in the vertical dimension is neglected when performing separate horizontal soil predictions for each depth interval. The other is that depth function fitting is often applied twice in the mapping process. The first is to standardize genetic or field-recorded horizon-based soil observations to pseudo-observations with a set of consistent depth intervals as input to the horizontal prediction models. The second is to fit soil predictions of the depth intervals to produce vertically continuous prediction for each location in the area of interest. Any errors in the fitting are thus repeated and may be magnified.

To overcome these drawbacks, some attempts have been made to develop true 3D soil mapping approaches (Minasny et al., 2006; Meersmans et al., 2009; Mishra et al., 2009; Kempen et al., 2011; Veronesi et al., 2012, 2014). These methods perform horizontal spatial prediction directly on vertical soil variation pattern, represented by the parameters of depth functions. The 1D vertical and 2D horizontal predictions are thus tightly integrated in the mapping process. Moreover, depth function fitting is applied only once in the mapping process. However, in these attempts the depth functions were combined mainly with statistical and geostatistical techniques (i.e., multiple regression

* Corresponding author.
E-mail address: glzhang@issas.ac.cn (G.-L. Zhang).

http://dx.doi.org/10.1016/j.geoderma.2015.05.013
0016-7061/

and ordinary kriging) or data mining techniques (i.e., artificial neural networks and random forests). The former generally requires fairly dense soil observations to ensure reliable model calibration, especially for variogram model estimation (Park and Vlek, 2002; Lacoste et al., 2014). The latter uses "black-box" models with a non-transparent prediction process, making it difficult to interpret their results. This non-transparency is a major concern to soil scientists because the interpretation of the prediction model is desirable for gaining knowledge of soil–landscape processes, e.g., by which soil organic C is accumulated, transformed and distributed in 3D.

An alternative to these techniques is a similarity-based prediction method. This class of methods predicts soil property values at an unsampled location based on its environmental similarity with observed sites. It has no requirements regarding the number and distribution of soil observations (Zhu et al., 2010). Moreover, its prediction process is transparent and traceable, allowing for easy interpretation of its prediction results. It also has one of the merits of the "black-box" models, i.e., the ability to deal with complex and nonlinear relationships between soil and environment. Such relations are likely for soil organic C, due to the complexity of processes by which it is accumulated, transformed and distributed in 3D.

Therefore, the objective of this study was to examine the effectiveness of integrating a similarity-based prediction method with depth functions to predict 3D distribution of soil organic matter (SOM) concentration.

2. Material and methods

2.1. Study area and data sets

The description of the study area can be found in F. Liu et al. (2013). The major soil types in this area include the Haplic Cambisols (Alumic, Dystric), Haplic Cambisols (Dystric, Ferric, Rhodic), Haplic Cambisols (Dystric, Ferric), Haplic Cambisols (Dystric), Haplic Cambisols, Hyperskeletic Leptosols, and Hydragric Anthrosols (Oxyaquic) according to the World Reference Base for soil resources (WRB) (IUSS Working Group WRB, 2014). Fig. 1 shows the general catenary sequence of the soil types in this area. Table 1 lists their typical environmental conditions.

We used the same set of 79 pedon sites as F. Liu et al. (2013). Fig. 2a shows the imbalanced spatial distribution of these sites. The soil sample collection and SOM measurement of soil horizons were described in the previous paper. We also used the same digital elevation model and Landsat Thematic Mapper data. The environmental variables used in this study include elevation, slope gradient, mean curvature, convergence index, topographic wetness index (TWI), and TM bands 3, 4 and 5. Fig. 2b shows the interpreted map of land uses including forests, shrubs, tea plantations, cultivated uplands, and paddy fields, with an overall accuracy of 80%.

2.2. The approach of 3D SOM prediction

The approach contains three steps. First, a depth function model was constructed to fit SOM profile distribution for each pedon. Second, a similarity-based method was used to interpolate the parameters of the depth function across this area. Third, with the values of the parameters for every location, 3D SOM distribution was generated.

2.2.1. Construction of depth function of SOM concentration

We assumed that SOM concentration varies continuously with depth. To find the most accurate depth function with which to describe SOM variation with depth, we evaluated the power, exponential, and logarithmic functions for their ability to match the observed SOM profile variation. These functions were considered due to their mathematical simplicity (only two parameters). The bulk horizon SOM value represents the average value over the horizon (McBratney et al., 2000). We fitted the three functions through the mid-depth of horizon data for each of the 79 profiles. The power function best described the observed SOM variation with depth. But its fittings tended to overestimate SOM concentration near the soil surface. We thus introduced a linear function to describe SOM vertical variation near the soil surface and used a power function to describe that below. The resultant depth function model is continuous but with discontinuous derivatives at the crossover depth:

$$Y = \begin{cases} k_0(ux_0)^{k_1} + a(X - x_0), & X \le x_0 \\ k_0 X^{k_1}, & X > x_0 \end{cases} \quad (1)$$

where Y is the SOM concentration, X is the soil depth, k_0 and k_1 are the parameters of the power function, a is the slope of the linear function, x_0 is the crossover depth of the node between the linear function and the power function, and u is an adjustment factor to compensate for the different uses of the crossover depth (x_0) when considering the effects of different land uses.

For cultivated lands, the first soil horizon is usually the tillage layer, in which SOM concentration was almost constant to the tillage depth due to frequent mixing of the topsoil. For other lands, the first horizon had a near-linear SOM decrease with depth. Thus, the value of a was defined to be the trivial slope 0 for paddy fields, cultivated uplands and tea plantations, and −1 (i.e., linear decrease) for forests and shrub lands. The value of u was set to be 1/2 for cultivated lands and 1 for other lands. From the land use map (Fig. 2b), the values of a and u were derived for every location in this area. The value of x_0 was set to the lower depth limit (d_{h1}) of the first horizon for the cultivated lands and the mid-depth ($d_{h1}/2$) of the first horizon for other lands. Note that d_{h1} is known from the profile description, and is not fit. We fitted this depth function model for each profile and estimated the values of the parameters k_0 and k_1 at all 79 sites using curve fitting by the 'nls' function of the R environment (R Development Core Team, 2012).

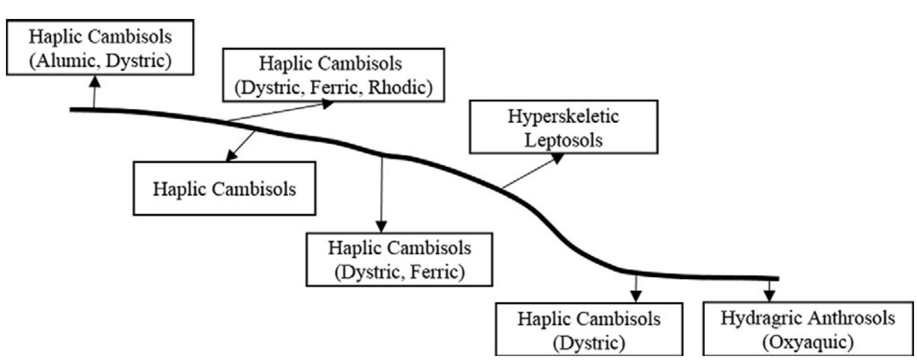

Fig. 1. General catenary sequence of the soil types in this landscape.