Geoderma 303 (2017) 118–132

Contents lists available at ScienceDirect

Geoderma

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

Digital soil mapping based on wavelet decomposed components of environmental covariates

Xiao-Lin Sun[a,b,c], Hui-Li Wang[d,*], Yu-Guo Zhao[b], Chaosheng Zhang[c,**], Gan-Lin Zhang[b]

[a] Guangdong Provincial Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China
[b] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
[c] International Network for Environment and Health, School of Geography and Archaeology, National University of Ireland, Galway, Ireland
[d] Guangxi Academy of Forestry Sciences, Nanning 530002, China

A R T I C L E I N F O

Keywords:
Digital soil mapping
Wavelet
Multi-scale analysis

A B S T R A C T

Multi-scale soil variations are increasingly employed to improve the accuracy for digital soil mapping (DSM). In this study, we attempted to explore a methodology of wavelet analysis on this topic. The terrain attributes of a study area were decomposed using the wavelet analysis, and the resulted components were applied to map soil organic carbon (SOC) content, pH and clay content using multiple linear regression (MLR) and regression kriging (RK). The results showed that the wavelet components strengthened soil-landscape relationships in terms of correlation coefficients, enhanced soil-landscape modelling in terms of MLR modelling coefficients of determination (R^2). Compared with several standard DSM approaches, i.e., ordinary kriging (OK), MLR and RK with the original terrain attributes, the use of wavelet components improved the prediction accuracy at some scales, but not all the scales. Most of the improvements were at the slight to moderate levels, e.g., 3.66–14.24% increases in the accuracy based on mean error, mean absolute error, root mean square error and R^2. Maps made with wavelet components were relatively smooth and sometimes contained hotspots due to characteristics of wavelet components, which differed a lot from those made by the standard DSM methods. The potential benefits of using wavelet components as predictors in DSM may be further revealed in the future when more predictor selection approaches and mapping methods are considered.

1. Introduction

The main objective of digital soil mapping (DSM) is to provide accurate soil information to meet demands from precise management of natural resources and environment, such as precision agriculture (McBratney et al., 2003; Carré et al., 2007). Though this objective has been achieved to some extent after decades of great efforts devoted to this field, particularly compared with conventional soil mapping, a large gap still exists between the obtained and demanded accuracies, e.g., Yang et al. (2011), Hamiache et al. (2012), Sun et al. (2012c).

To improve mapping accuracy, more and more studies focused on scale- and location-dependent soil variations, e.g., Behrens et al. (2010a,b, 2014), Zhang et al. (2011), Sun et al. (2012a,b), Wang et al. (2013), Miller et al. (2015), Song et al. (2016). This is related to the limitation that standard DSM techniques, such as regression kriging (RK), random forest, ordinary kriging (OK) and multiple linear regression (MLR), are not capable of handling non-stationary soil variation due to the complex non-linear, non-addictive and non-overlapping soil formation processes (Biswas and Si, 2011) and spatially varying relationship between soil and environmental covariates (Song et al., 2016). In general, three approaches were used to deal with the problems: local regression kriging (LRK) (Sun et al., 2012a,b), geographical weighted regression (GWR) (Mishra et al., 2010; Kumar et al., 2012; Wang et al., 2013; Song et al., 2016) and multi-scale terrain analysis (Behrens et al., 2010a,b, 2014; Miller et al., 2015; Roecker and Thompson, 2010).

The main difference of LRK and GWR from their traditional counterparts, i.e., RK and MLR, respectively, is that they build a model for a predefined local area, rather than doing it globally. As a result, these two approaches do not really deal with multi-scale soil variations (Song et al., 2016). In addition, Sun et al. (2012a,b) showed that the performance of LRK was dependent on the inherent local relationships between soil and environmental covariates and was not always the optimal. Although many studies found GWR predicted soil relatively more accurately than MLR, OK and RK, Song et al. (2016) concluded that GWR only outperformed MLR but predicted worse results than RK.

* Correspondence to: H.-L. Wang, Guangxi Academy of Forestry Sciences, 23 Yongwu Road, Nanning 530002, China.
** Correspondence to: C. Zhang, Room 108, Geography, School of Geography and Archaeology, National University of Ireland, Galway, University Road, Galway, Ireland.
E-mail addresses: 397776652@qq.com (H.-L. Wang), chaosheng.zhang@nuigalway.ie (C. Zhang).

http://dx.doi.org/10.1016/j.geoderma.2017.05.017
Received 13 June 2016; Received in revised form 20 April 2017; Accepted 10 May 2017
Available online 18 May 2017
0016-7061/

Multi-scale terrain analysis for soil mapping seems to perform better in handling multi-scale soil variation, as such kind of methods compute terrain attributes within a range of window sizes, resolutions or neighborhoods. According to Behrens et al. (2014), outputs of multi-scale terrain analysis reflected landscape characteristics as driving forces for soil formation at different scales. Behrens et al. (2010a,b, 2014) found that, compared with mapping using standard terrain attributes, mapping using outputs of multi-scale terrain analysis increased prediction accuracies, for example, with a decrease of root mean square error (RMSE) from 16.1% to 11.2% for topsoil silt content and an average 20% increase of F_1-measure (the harmonic mean of precision and recall based on the confusion matrix) for soil classes. Miller et al. (2015) also demonstrated a better performance of multi-scale predictors in soil mapping, i.e., from negligible to 70% increases of the adjusted correlation coefficients (R^2) in soil-landscape modelling.

As a classical approach for multi-scale analysis, wavelet analysis was introduced to investigate scale- and location-dependent soil variation by McBratney (1998) and Lark and Webster (1999). Afterwards, it was applied in a number of studies, e.g., Lark et al. (2003), Si (2003), Lark (2006), Biswas and Si (2011), Biswas et al. (2013). Most of the studies reasonably interpreted multi-scale soil variation to physical soil formation processes. These studies presented the advantages of using wavelet analysis in separating soil variations into different scales by dilating a wavelet using a scale parameter. For example, Lark and Webster (2004) reveals enhanced correlations between soil thickness and slope gradient at different scales and in different directions. Biswas et al. (2013) also found similar results. Thus, Lark and Webster (2004) proposed that wavelet analysis could be used to analyze available dense data such as digital elevation data (DEM) to indicate soil. Later on, Mendonca-Santos et al. (2007) applied wavelet decomposition of environmental covariates into soil mapping and found slightly enhanced prediction accuracies (i.e., a 1% decrease of RMSE for clay content and 2% of misclassification of soil class). Therefore, Mendonca-Santos et al. (2007) recommended decomposition of environmental covariates using wavelet analysis in soil mapping. Recently, Behrens et al. (2010a,b, 2014) also promoted multi-scale analysis on environmental covariates using wavelet. Until now, few studies have discussed about this topic.

To further investigate this topic, this study applied the wavelet decomposed terrain attributes in soil mapping and compared its performances with standard DSM approaches, i.e., OK, MLR and RK. Firstly, the commonly used terrain attributes derived from a DEM were decomposed using wavelet analysis. Secondly, the decomposed components were applied in DSM to predict soil properties using MLR and RK. Thirdly, the resulted prediction accuracies were compared with those obtained from OK, MLR and RK with the original terrain attributes.

2. Materials and methods

2.1. Study area

The study area is located in Nanning, Southwest China, within latitude 22°57′8″–22°58′41″N and longitude 108°20′57″–108°21′54″E (Fig. 1). The total area is 3.03 km^2. The climate is subtropical humid monsoon, with an annual average temperature of 21.6 °C and an annual average rainfall of 1300.6 mm. The landforms are hilly, with elevations between 130 and 300 m, and an average slope of 24.8°. Parent materials of the soil are mud stone, mud shale and sand shale. This area has been used for forestry in history and has been under *Eucalyptus* plantation since a decade ago. The soils are latosolic red soil, referred to Ustisol in the U.S. Soil Taxonomy.

2.2. Soil covariates

A DEM of 10 m cell size for the study area was built based on digitized contour lines, shown in Fig. 1. The DEM was then pre-processed with sinks and pits filled. A number of terrain attributes were then derived from the DEM: elevation, slope, aspect, northernness, easternness, northness, eastness, solar radiation, profile curvature, plan curvature, specific catchment area (SCA), stream power index (SPI), SAGA topographic wetness index (TWI) and topographic position index (TPI). Except solar radiation which was calculated within ArcGIS 10.0 (ESRI, 2012), all the others were computed within SAGA (Conrad et al., 2015). Although the two pairs, i.e., northernness and northness, easternness and eastness, have the same meanings, they were computed in different ways so all of them were used in this study. Northernness were transformed from aspect using abs (180°-aspect) (Samuel-Rosa et al., 2015), and easternness using abs (270°-aspect) for an aspect larger than 90° and 90° + aspect for an aspect smaller than 90°. Northness and eastness were computed using standard cosine and sine transformations from aspect, respectively (Miller et al., 2015). Following Song et al. (2016), SCA was calculated with D8 algorithm and used to compute TWI and SPI with local slope gradient. One year average potential insolation was calculated as solar radiation following Song et al. (2016). Since this area is quite small (i.e., 3.03 km^2) and under the same land use (i.e., *Eucalyptus* plantation), no other environmental covariates like satellite images related to climate, geology and soil type were used. Considering the edge problem in the wavelet analysis (Lark and Webster, 2004), the terrain attributes data on a larger area centering around the study area were input into wavelet analysis. The attributes on sampling sites (depicted in the following) were summarized in Table 1.

2.3. Soil sampling

Samples of this study were taken in three ways. First, conditioned Latin Hypercube Sampling (cLHS) (Minasny and McBratney, 2006; Roudier et al., 2012) was implemented to select 50 samples based on six of the above terrain attributes: elevation, slope, aspect, plan curvature, profile curvature and TWI. This technique was used because it provides a full coverage of the range of each terrain attribute by maximally stratifying the marginal distribution. Only these six terrain attributes were used because they were enough to characterize the main soil formation condition (Sun et al., 2012d). This sampling would be useful for soil mapping with regression. Due to the difficulty of access to some of them, a total of 45 of them were sampled in the field, i.e., red points in Fig. 1. In order to have a full spatial coverage, grid sampling with an interval of 250 m was implemented to select another 50 samples. This would be useful for soil mapping with a spatial random effect (e.g., OK and RK). For the same reason of difficult access, only 45 of them (blue points in Fig. 1) were visited in the field, and some of them were shifted away from originally selected positions. Finally, another 45 random samples were taken. The samples collected using cLHS and grid were used for prediction, while the random samples were used for validation.

At each sampling location, five cores of topsoil 0–20 cm were collected from the four corners and center of a 1 m × 1 m square and were then thoroughly mixed to form one composite sample. Soil organic carbon (SOC) contents were determined by the potassium dichromate oxidation method. Soil pH was determined using a pH meter (1: 2.5 soil to water ratio). Soil clay content was measured using the hydrometer method. Summary statistics of the three soil properties are listed in Table 1. Octile skews (Brys et al., 2003) in the table showed that it was unnecessary to transform the soil data towards normality (Sun et al., 2012c).

2.4. Two-dimensional (2D) discrete wavelet analysis (DWT)

The 2D DWT was conducted in this study to decompose environmental covariates, i.e., terrain attributes mentioned earlier. This technique has been described in detail in many papers such as Lark and Webster (1999, 2004). The wavedec2 function of Matlab 7.8.0 was used in this study to implement the technique. At each scale, i.e., $2^i \times \lambda$