Catena 163 (2018) 257–275

Contents lists available at ScienceDirect

Catena

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

# Selection of optimal scales for soil depth prediction on headwater hillslopes: A modeling approach

Xiaole Han[a,b,c], Jintao Liu[a,b,*], Subhasis Mitra[c], Xiaopeng Li[d], Puneet Srivastava[c], Sandra M. Guzman[c], Xi Chen[a,b]

[a] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, People's Republic of China
[b] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
[c] Biosystems Engineering, Auburn University, 200 Corley Building, Auburn, AL 36849-5417, USA
[d] Institute of Soil Science, Chinese Academy of Sciences, 210008 Nanjing, People's Republic of China

ARTICLE INFO

*Keywords:*
Soil depth
Digital elevation model
Terrain attributes
Ground penetrating radar
Artificial neural networks

ABSTRACT

Landform attributes derived from digital elevation models (DEMs) are the most commonly used factors to predict soil depths on hillslopes. However, the selection of an appropriate algorithm to calculate terrain attributes and identify an optimal DEM resolution remains ambiguous. In this study, we propose a method to use high-resolution DEM and spatial soil depth data to obtain terrain attributes and identify the optimal DEM resolution to predict soil depths at hillslope scales. A geophysical method (ground penetrating radar, GPR) was used to investigate the reasons for the optimal DEM resolution findings. Point-scale soil depth from 116 sites and elevation data were collected from two adjacent headwater hillslopes (H1: 0.42 ha and H2: 0.31 ha) at the Hemuqiao hydrological experimental station in Southeast China. The elevation datasets were collected using a total station at a variable spacing level and were then used to derived DEMs at nine spatial resolutions: 0.25, 0.50, 0.75, 1.00, 2.00, 3.50, 5.00, 7.50 and 10.00 m. Nine primary and secondary topographic attributes using the nine spatial resolution DEMs were then derived. Two different algorithms (D8 and D$\infty$) for calculating contributing areas and related secondary topographic attributes were compared. We used and compared both linear (multiple linear regression, MLR) and non-linear (artificial neural network, ANN) models for soil depth prediction. Results demonstrated that the two models performed well for predicting soil depth. Specifically, MLR performed better than the non-linear model of ANNs. Additionally, we found that the multiple-direction algorithm (D$\infty$) allowed flow divergence and avoided abrupt changes in soil depth predictions (orphan cells) and should be adopted for model construction. The D$\infty$ algorithm performed better in divergent areas, such as ridges and side slopes, and the D$\infty$ algorithm also worked well in convergent areas, such as valleys. Moreover, our results demonstrated that moderate (e.g., 2.00 m) resolution topographic attributes, instead of the finest resolution, achieved the best prediction with the lowest root mean square error (RMSE) and mean absolute error (MAE) and the highest values of the coefficient of determination ($R^2$). Moreover, the GPR results indicated that the valley accumulated more soils than side-slope areas, and a sharp increase in soil depth was found in areas adjacent to the valley. Comparing the optimal DEM resolution and valley width obtained by GPR, we found that average valley width (*AVW*) should be considered a good measure for choosing the optimal DEM resolution for soil depth prediction.

## 1. Introduction

Soil depth, which here is defined as the vertical distance from the soil surface to the underlying weathered bedrock and lacking relict rock structure (Dietrich et al., 1995; Heimsath et al., 1997; Han et al., 2016), is a major factor controlling soil water storage, evapotranspiration, and atmosphere-soil-plant interaction. The pattern of soil depth distribution over hillslopes exerts significant influence on runoff volume and the runoff coefficient. (Hoover and Hursh, 1943; Pelletier and Rasmussen, 2009; Fu et al., 2011; Liu et al., 2013). However, the soil depth distribution on hillslopes is usually unknown. Field measurements are costly and time-consuming. Therefore, there is an urgent need for high-quality soil depth survey maps (Dahlke et al., 2009; Tesfa et al., 2009; Liu et al., 2013).

Various models (or methods) have been developed in the past aimed at generating high-resolution soil depth maps. The process-based

* Corresponding author at: State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, People's Republic of China.
*E-mail address:* jtliu@hhu.edu.cn (J. Liu).

https://doi.org/10.1016/j.catena.2017.12.026
Received 11 April 2017; Received in revised form 5 December 2017; Accepted 22 December 2017
0341-8162/ 

geomorphic modeling approach was developed from the perspective of soil evolution and applies mass balance equations to delineate processes of soil weathering and transformation and the transport of soil through the landscape over geomorphological timescales (Dietrich et al., 1995; Heimsath et al., 1997; Pelletier and Rasmussen, 2009; Liu et al., 2013). For example, Dietrich et al. (1995) proposed a method for predicting colluvial soil depth by assuming that soil production rates are depth-dependent and slope transport rates are slope-dependent. This method was successfully validated by measuring the in-situ production of cosmogenic $^{10}$Be and $^{26}$Al concentrations in the bedrock under different depths and elucidated an exponential decline of soil production rates with increasing soil depth (Heimsath et al., 1997). However, these models are limited in application because of the complexity of measurements, complicated in solution techniques, and not adaptable to anthropogenic land disturbances (Kuriakose et al., 2009), which limits their application for soil depth prediction.

In contrast, stochastically-based models avoid the mechanistic explanation of soil evolution and focus on directly establishing relationships between soil depth and a set of influencing factors that can be easily estimated (Moore et al., 1993; Zhu, 2000; Ziadat, 2005; Penížek and Borůvka, 2006; Kuriakose et al., 2009; Tesfa et al., 2009; Mehnatkesh et al., 2013; Yang et al., 2014; Bagheri Bodaghabadi et al., 2015). Various stochastic models have been used in the past such as multiple linear regression and maximum likelihood classification (Ziadat, 2005), expert knowledge and fuzzy logic (Zhu et al., 2001), a generalized linear model (Shary et al., 2017) and random forests (Tesfa et al., 2009; Behrens et al., 2010; Möller and Volk, 2015). Among the models, the regression model is the most common method used to predict soil depth because of its ability to be applied at a regional scale, ease of calculation, and ability to estimate error (Ziadat, 2005). Various studies in the past have used linear regression models to correlate and predict soil depth using various terrain attributes (Moore et al., 1993; Ziadat, 2005; Mehnatkesh et al., 2013; Möller and Volk, 2015; Shary et al., 2017). For example, Möller and Volk (2015) used the mass balance index (MBI) to characterize the process domains of soil loss and accumulation. Negative MBI values represent areas of net deposition, such as depressions, and positive values indicate net erosion, such as convex hillslopes. Shary et al. (2017) provided examples for the selection and derivation of terrain attributes. They pointed out the importance of evaluating the significance of predictors in a model. In these studies, digital elevation models (DEMs) and their derived surface attributes are used as explanatory variables to construct linear regression models for soil depth modeling. However, relationships between soil characteristics and topographic variables are usually non-linear in nature, which renders the conventional regression methods unreliable (Zhao et al., 2009). These non-linear relationships can be managed by models such as artificial neural networks (ANNs) and generalized additive models (GAMs) (Shary et al., 2017). ANNs have been widely used in soil science research since they are data-driven, self-adapted and can be used to describe complex non-linear relationships. ANNs have been extensively used for predicting soil properties (both soil physical and soil chemical properties) and the spatial distribution of taxonomic classes (Zhu, 2000; Behrens et al., 2005; Zhao et al., 2009; Bagheri Bodaghabadi et al., 2015). For example, Zhu (2000) developed a neural network approach to populate a soil similarity model that was used to provide information on the detailed spatial variation of soil properties for hydro-ecological modeling. They concluded that the soil map by the ANN approach revealed much higher quality than conventional soil maps. Therefore, the implementation of models such as ANN provides an alternative method to refine the quality of soil mapping, especially when the relationship between soil characteristics and topographic attributes is unknown.

Success in predicting regional soil depth is highly related to the appropriate ways to obtain these unknown attributes, which is affected by the spatial resolution of the DEMs used (Yang et al., 2014), one reason being that landscape-related processes occur at different scales. For example, soil erosion processes show an obvious dependence on space and time and are embedded in a multi-hierarchical system ranging from field to landscape scales. Each scale requires the provision of data at a scale-adequate resolution and with scale appropriate methods (Volk et al., 2010). Therefore, the choice of DEM resolution should be adapted to the context of the implemented analysis (Behrens et al., 2010; Kim and Zheng, 2011). In these studies, choosing the optimal DEM is critical for determining the terrain attributes.

Studies in the past have attempted to determine optimal DEM resolutions for soil property predictions (e.g., potassium, pH, total phosphorous, nitrate and topsoil silt content) (Behrens et al., 2010; Kim and Zheng, 2011; Yang et al., 2014). The gradual emergence of very high-resolution elevation data, such as from the LiDAR technique (providing sub-meter level DEM resolution), has offered greater details for landscape characterization (Leempoel et al., 2015). However, former studies have presented contrasting views regarding the optimal DEM resolution for model construction. For example, Vaze et al. (2010) suggested using high-resolution DEMs instead of contour-derived low-resolution DEMs for the derivation of hydrological features. However, Zhang and Montgomery (1994), Smith et al. (2006), Behrens et al. (2010), Kim and Zheng (2011), and Möller and Volk (2015) suggested that it was not always the highest resolution DEM that made the best prediction. They proposed that high-resolution DEMs should be tested against the coarser resolutions for predicting hydrologic attributes.

Very few studies have been conducted in the past to determine the optimal resolution for soil depth modeling. Yang et al. (2014) used grey relational analysis (GRA) to determine the relationship between soil depth and terrain attributes under multiple resolutions before selecting the best resolution for soil depth prediction. Note that the GRA results indicated that a moderate resolution DEM (10 m) should be adopted instead of the finest resolution (5 m). Other studies, for example Möller and Volk (2015), were based on the effective map scale (EMS) approach to detect operation scales and the statistical and spatial visualization of scale-specific inaccuracies. Möller and Volk (2015) also concluded that the original high-resolution digital elevation model (hrDEM) and the smallest scale level are characterized by poorer prediction results. Therefore, in this study we aimed to determine the optimal DEM resolution for predicting soil depth by implementing both linear (multiple linear regression) and non-linear models (artificial neural networks) often used in the soil sciences and to propose a simplified methodology for determining an optimal DEM resolution using average valley width.

To achieve these goals, we used accuracy metrics of model performance, such as the root mean square error (RMSE), the mean absolute error (MAE) and the coefficient of determination ($R^2$), along with continuous soil profile imaging using ground-penetrating radar (GPR), to estimate the efficiency of our proposed optimal resolution soil depth procedure. Additionally, we used the single direction (D8; O'Callaghan and Mark, 1984) and multiple directions ($D\infty$; Tarboton, 1997) algorithms to calculate the contributing area. The development of optimization procedures for soil depth prediction, such the one evaluated in this study, will provide better information for soil scientists to make improved assessments on hillslope processes.

## 2. Materials and methods

### 2.1. Study area

Two steep headwater hillslopes with areas of 0.42 ha (H1) and 0.31 ha (H2) were selected for the study (Figs. 1 and 2). The areas are within the Hemuqiao hydrological experimental station (119°47′E, 30°34′N, 135 ha) located upstream of the Taihu Basin in southeastern China. The shapes of H1 and H2 are similar, with a valley in the middle and two-facet hillslopes on each side (Fig. 1c and d). Elevation in H1 ranges from 312 to 367 m and in H2 from 304 to 364 m above the mean