Geoderma 282 (2016) 59–69

Contents lists available at ScienceDirect

Geoderma

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

Modelling soil thickness in the critical zone for Southern British Columbia

Christopher Scarpone [a], Margaret G. Schmidt [a,*], Chuck E. Bulmer [b], Anders Knudby [c]

[a] Soil Science Lab, Department of Geography, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
[b] Department of Geography, University of Ottawa, 75 Laurier Ave E, Ottawa, Ontario K1N 6N5, Canada
[c] British Columbia Ministry of Forests Lands and Natural Resources Operations, Forest Sciences Section, Vernon, BC V1B 2C7, Canada

ARTICLE INFO

Article history:
Received 1 February 2016
Received in revised form 5 July 2016
Accepted 7 July 2016
Available online 15 July 2016

Keywords:
Critical zone
Soil thickness
randomForest
GLM
Kriging
Landscape modelling

ABSTRACT

The Critical Zone (CZ) is defined as the outer layer of the solid Earth, extending from the vegetation canopy through the pedosphere and down to the bottom of the weathered bedrock zone. Most biological, chemical and physical interactions take place in the CZ, and it is here that most terrestrial life is found. Being able to predict the thickness of the pedosphere (i.e. soil thickness, from the mineral soil surface down to bedrock) in the CZ can lead to a better understanding of rates of physical/chemical change, such as carbon sequestration, soil erosion, and water storage. The objective of this study was to accurately map the thickness of the pedosphere on a landscape scale for a ca. 3400 km^2 landscape in Southern British Columbia (BC). The data inputs used were exposed bedrock (EB) points identified from orthophotos, well record (WR) data, and in-situ observations of soil thickness, for which conditioned Latin hypercube sampling (cLHS) was used to define a set of locations where soil thickness was determined. Four methods were then used to model soil thickness as a function of environmental data layers derived from a digital elevation model and satellite imagery: Generalized Linear Model (GLM), Random Forest (RF), GLM Residual Kriging (GLMRK) and RF Residual Kriging (RFRK). An equal weighted random sampling scheme of 100 EB, WR, and in-situ soil thickness points was used with each model. A second sampling scheme was used with the same WR and in-situ soil thickness points and an additional 5000 randomly sampled EB points, used to improve prediction accuracy with limited data sources. Of the modelling methods used, GLMRK proved to be the best method for shallow thicknesses (0 to 2 m) as assessed by Root Mean Square Error (RMSE) values (1.87 m) with the equal weighted sample scheme. The addition of the 5000 EB points substantially improved predictions for shallow soil thickness (RMSE 0.9 m) as well as soil thickness in the 2–5 m range, while having negligible impact on predictions of thicker soils. This demonstrates that an exposed bedrock layer can help constrain shallow soil thickness predictions when used in conjunction with geostatistical approaches such as GLMRK and RFRK for mapping soil thickness.

1. Introduction

The critical zone (CZ) is the environment in which complex interactions between soil, rock, air, water, and living organisms occur to maintain and regulate natural processes. The CZ extends from the top of the vegetation to the solid fresh bedrock below the soil surface (NRC, 2001). The active regions within which these processes take place are the atmosphere, biosphere, hydrosphere, pedosphere, and the lithosphere. The pedosphere, here considered as encompassing all unconsolidated material from the top of the mineral soil to the top of the fresh bedrock, acts as a transition zone between the atmosphere above and the hard rock below (Brady and Weil, 2007) and as the interface for the other four spheres (Lin, 2010).

The pedosphere influences nutrient and chemical exchange rates (Brantley et al., 2007; Chorover et al., 2007), the effects of water storage, water yield and boundary transfer (Field et al., 2015; Yu et al., 2015; Yu et al., 2014), and weathering, erosion and soil production rates (Heimsath et al., 2012; Jin et al., 2010; Ma et al., 2010; Moraetis et al., 2014). These processes are all strongly influenced by the vertical extent of the pedosphere (henceforth: soil thickness). Empirical and quantitative models are needed to predict and describe this aspect of the CZ due to the complexity of directly mapping the thickness of the pedosphere (Lin, 2010). Four categories of such methods have been used to model soil thickness (Li et al., 2011; Pelletier and Rasmussen, 2009):

1. Process based
2. Deterministic
3. Stochastic/geostatistical
4. Combined

* Corresponding author.
E-mail addresses: cscarpon@sfu.ca (C. Scarpone), margaret_schmidt@sfu.ca (M.G. Schmidt), Chuck.Bulmer@gov.bc.ca (C.E. Bulmer), aknudby@uottawa.ca (A. Knudby).

http://dx.doi.org/10.1016/j.geoderma.2016.07.012
0016-7061/

Process based models were some of the first models used to predict soil thickness over larger areas. These models emphasized the relationships of slope angle and slope convexity with soil production rates (Heimsath et al., 1999). These models were very susceptible to stochastic elements in the landscape that could alter erosion rates. An assumed starting soil thickness was required for each model, which introduced errors as an approximate thickness was given. Later, local thickness measurements were introduced to calibrate the models, and higher resolution elevation data were introduced to aid in predictions. This generally improved the model predictions of soil thickness (Pelletier and Rasmussen, 2009). However, these models are typically very complicated to reproduce and are only applicable to small hillslopes which require a large amount of in-situ measurements.

Deterministic methods are also commonly used to predict soil thickness, for example regression models have been used to predict soil thickness from topographic variables (Ziadat, 2010). Quantitative models incorporating the relationship between the distance to bedrock and water well depth have proven to be beneficial in data limited areas of Sweden (Karlsson et al., 2014). However, one issue with these models is the inability to incorporate local variance related to heterogeneous landscapes. These models can therefore only be applied to small areas, where the effect of local trends in the landscapes can be minimized.

Many studies have used geostatistical methods such as kriging and combined methods of residual kriging to predict soil thickness (Kuriakose et al., 2009; Odeh et al., 1994; Sarkar et al., 2013). The combination of both geostatistical and deterministic methods has allowed for the integration of topographic variables to account for spatial uncertainty in the landscape, and thus has led to improvements in soil thickness mapping. Generalized Linear Model (GLM) residual kriging was one of the first combined methods proposed (Odeh et al., 1994). GLM is a weighted linear model that uses a link function to allow distributions other than a normal distribution to be used for predictions (Lane, 2002). Because the uncertainty in the landscape was already incorporated in the residual values, combining the output from the regression model with the kriged residuals was considered a preferred modelling approach compared to kriging the observed soil thicknesses directly (Hengl et al., 2004; Odeh et al., 1995). Other studies have explored different combinations of regression models and kriging procedures to improve upon prediction accuracies. Kuriakose et al. (2009) compared the use of a linear model, block kriging and residual block kriging and found that the residual block kriging performed the best for predicting soil thickness. Residual block kriging was able to account for more of the variability in the landscape when comparing error rates; although validation results were similar to ordinary kriging with residual kriging due to the controlled environment. That study had 259 augered soil points, however, the thickness range was limited from 0 to 4 m and the study site comprised a single watershed.

Random Forest (RF) is a non-parametric decision tree classifier that can predict both discrete and continuous data (Breiman, 2001). The ability of RF to handle large and noisy datasets and to generate variable importance plots has made it an important tool to map soil properties (Heung et al., 2014; Rad et al., 2014; Wiesmeier et al., 2010). Tesfa et al. (2009) used RF in regression mode and compared it to a General Additive Model (GAM) to predict soil thickness at watershed-scale. Their study focused on the creation of new predictors that could help to predict soil thickness. It was found that RF outperformed the GAM model and was able to express more of the spatial variability in the watershed than the GAM model; however they did not further explore the abilities of RF and residual kriging (RFRK).

RFRK has been explored in mapping other soil properties but has not been used to map soil thickness. Guo et al. (2015) explored mapping soil organic matter using RFRK and found that it significantly out-performed a stepwise linear regression and a RF in regression mode. Hengl et al. (2015) used RFRK to map soil nutrient contents for the continent of Africa. Mapping was done at a resolution of 250 m, with minimal data and these authors also found that it outperformed RF. Their results also demonstrate that RFRK can handle a large amount of spatial variation in the landscape in comparison to other methods, even at a continental scale. Even though RFRK has not been explored to map soil thickness, it has proven its ability to map highly variable soil properties, indicating that it could be used for mapping soil thickness at the landscape scale.

Most studies of soil thickness are conducted on a local scale, typically on a single watershed or a hillslope, with very intensive sampling methods. In addition several of the established critical zone observatories are applying detailed information from intensive research plots to interpret processes over larger areas (White et al., 2015). Combination methods for predictions of soil thickness such as residual kriging can prove to be an important tool for mapping soil thickness at the larger landscape scale, especially in data limited areas.

The objectives of this study were to first map the thickness of the pedosphere on a landscape scale; then to compare two deterministic models with residual kriging and assess which model had optimal performance; and finally to assess the predictive power that data on the location of exposed bedrock (EB) adds for mapping soil thickness. The modelling methods GLM, RF, GLMRK and RFRK, were used to predict soil thickness and the methods were compared through variogram variances and prediction errors to determine which method performs the best for mapping soil thickness. The methods were applied to the Tulameen region, Southern British Columbia (BC).

2. Methods

This study compares GLM, RF, GLMRK and RFRK for mapping the thickness of a steady-state soil layer in Southern BC (Fig. 1). Explanatory environmental variables were derived from a 100 m digital elevation model (DEM) and Landsat 7 and ASTER satellite imagery. Calibration and validation soil thickness data points were derived from well records (WR) ($n = 239$), in-situ soil thickness measurements ($n = 174$), and exposed bedrock (EB) locations ($n = 42{,}823$). All models were calibrated with a subset of the soil thickness data ($n = 300$ (100 WR, 100 in-situ, 100 EB)). In addition the models were calibrated with a dataset with 5200 points (100 WR, 100 in-situ and 5000 EB). All models were validated with a random subset of the soil thickness data ($n = 222$ (74 WR, 74 in-situ, 74 EB)) not used for model development. Map accuracy was quantified as the Root-Mean-Squared Error (RMSE) calculated separately for soil thickness ranges 0–2 m, 2–5 m, 5–10 m, and >10 m and with r^2 values for an accuracy metric of the whole model.

2.1. Study area

The Tulameen study area is located in the south central interior of BC Canada (N 49°32′ W 120°45′) (Fig. 2). The area occupies 3435 km^2 of primarily coniferous forest in the Cascade Dry Belt and Thompson Plateau. The Cascade Mountains here are considered to be part of the Coast Mountain range (Holland, 1976) and elevations range from 623 to 2337 m above sea level. Six biogeoclimatic zones are found in this region, including the Interior Douglas Fir, Engelmann Spruce-Subalpine Fir, Ponderosa Pine, Coastal Western Hemlock and Mountain Hemlock, and Interior Mountain-Heather Alpine Monthly average temperatures range from −12 to 27 °C with an average of 550 mm annual precipitation (Lloyd et al., 1990).

The majority of the soils in this region developed after the recession of the Wisconsin glaciation, approximately 12,000 years ago. The dominant parent material in the region is glacial till, with some areas of glacial fluvial and glacial lacustrine deposits. The main soil types found in the forested uplands of this region are Dystric Brunisols and Humo-ferric Podzols, with lesser amounts of Luvisols and Eutric Brunisols. Chernozemic soils are also found on the grasslands at lower elevations (Lord and Green, 1974). Soil pH ranges from 3.6 to 5.2 (Fraser et al., 1989). The thickness of the soil in this region is highly variable, in part because glacial erosion and deposition during the Pleistocene Epoch