Geoderma Regional 7 (2016) 38–48

Contents lists available at ScienceDirect

Geoderma Regional

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

Mapping soil organic carbon content over New South Wales, Australia using local regression kriging

P.D.S.N. Somarathna *, B.P. Malone, B. Minasny

Faculty of Agriculture and Environment, Department of Environmental Sciences, The University of Sydney, New South Wales, Australia

ARTICLE INFO

Article history:
Received 26 September 2015
Received in revised form 16 December 2015
Accepted 16 December 2015
Available online 24 December 2015

Keywords:
Carbon sequestration
Soil organic carbon
Multiple linear regression kriging
Digital soil mapping
Luvisols
Vertisols
Plinthosols

ABSTRACT

Various digital soil mapping techniques ranging from simple linear models to complex machine learning techniques have been employed for soil organic carbon (SOC) mapping. When SOC mapping over a large region is required, the usual approach has to employ a model calibrated for the whole area. An alternative is to use a series of locally calibrated models to map smaller areas that collectively cover the large region of interest. The accuracy of the SOC products generated by these two approaches can potentially vary. However, performance of whole-area calibrated models versus locally calibrated models in mapping SOC of large extents has seldom been explored in detail, particularly with respect to the type of model being employed. Our study aims to fill this gap by evaluating the SOC prediction performance of three common models, multiple linear regression (MLR), Regression tree model; Cubist and Support Vector Regression (SVR) that are calibrated locally and for the whole study area. This study was carried out using eight identified local areas in New South Wales (NSW), Australia and across the whole state entirely. Every model was calibrated separately for each local area and for the entire state. The local and whole-area models were validated using the same test data set over 50 realizations. In particular, local prediction accuracy of whole-area calibrated models was compared to that of locally calibrated models. The models were tested separately for the standard soil depth layers including 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm; 60–100 cm. The results show that SVR models have a superior performance out of three tested models for all standardized depth layers. In general the local models outperform the whole-area models for all three tested models with respect to the accuracy of predictions. All models displayed area specific performances proving the importance of inclusion of prevailing local conditions in SOC modelling and mapping. Therefore, we introduce a moving window approach where a hybrid series of locally calibrated models and a whole-area calibrated model can be used against using one calibrated model for the modelling very large mapping extents. Moving window approach provides more accurate results having the lowest error compared to the whole-area model. Also it provides the least biased predictions. Therefore, this novel approach provides a promising way of increasing the efficiency and accuracy of digital soil mapping.

1. Introduction

Soil organic carbon (SOC) is one of the most researched soil properties due to its importance in agronomic sustainability (Reeves, 1997) and carbon sequestration potential. Carbon sequestration is seen as the best solution to reduce atmospheric carbon where both agriculture and the environment are benefited. Consequently, several global and national policy initiatives that revolve around the carbon sequestration potential of SOC have come to the forefront (O'Rourke et al., 2015). A carbon offset scheme known as the Carbon Farming Initiative (CFI) instigated in Australia is a perfect example. Such programs rely on accurate estimates of SOC content over the spatial extent of interest which can be represented by a baseline SOC map. SOC mapping has been greatly benefited by Digital soil mapping (DSM). During the last decade, various DSM techniques ranging from simple linear models to complex machine learning techniques have been employed for SOC mapping (Minasny et al., 2013).

These techniques include, kriging (Cambule et al., 2014; Dai et al., 2014), co-kriging (Odeh et al., 1995; Phachomphon et al., 2010), regression kriging (Mora-Vallejo et al., 2008; de Brogniez et al., 2014; Dorji et al., 2014; Piccini et al., 2014), Linear mixed models (Rawlins et al., 2009; Karunaratne et al., 2014), machine learning techniques such as Artificial neural networks (Minasny and McBratney, 2002; Malone et al., 2009; Zhao et al., 2010), Support Vector Regression (Ballabio, 2009), Regression tree models, such as Cubist (Adhikari et al., 2014; Miklos et al., 2010; Rossel et al., 2014) and Random forests (Wiesmeier et al., 2011; Subburayalu and Slater, 2013; Hengl et al., 2015) and Generalised Additive Models (GAM) (Poggio et al., 2013; de Brogniez et al., 2014).

* Corresponding author.
E-mail address: sanjeewani.pallegedaradewage@sydney.edu.au (P.D.S.N. Somarathna).

http://dx.doi.org/10.1016/j.geodrs.2015.12.002

Foregoing techniques and models can be seen employed at various scales ranging from small farm areas to larger regional and continental extents for SOC mapping. When the requirement is to map SOC of a larger area, the preferred approach has to use a single calibrated model to map the entire area. Alternatively, a series of locally calibrated models can be used to map small areas that collectively cover the large region of interest where there is a fairly reasonable sampling density or the usage of hybrid series of local and whole-area calibrated models for the areas with dense and sparse observation points respectively. The latter approach is very uncommon in DSM literature to the best of our knowledge. These two approaches coupled with different model types such as multiple linear regression (MLR) Cubist and Support Vector Regression (SVR) could produce results that are of varying accuracies. Performance of such models in predicting SOC over large spatial extents has seldom been compared with respect to whole-area and locally calibrated models. Therefore, this study aims to examine the SOC prediction capability of MLR, Cubist and SVR with respect to local versus whole-area model training and application. The study is carried out using eight identified local areas for the localized studies in the state of NSW and a whole area study covering all of NSW. Based on the results, we make recommendations on the best combinations of model type and spatial extent used for calibration.

2. Methods

2.1. Study area

The study area is the state of New South Wales (NSW), Australia which covers approximately 810,000 km^2. The Great Dividing Range which runs approximately north to south in the east has a major impact on the State's distribution of rainfall that results in four distinct climatic zones. The area to the west of the Great Dividing Range which represents majority of NSW has an arid to semi-arid climate. The average annual rainfall for this area ranges from 150 mm to 500 mm. The climate along the flat, coastal plain east of the dividing range varies from cool oceanic to humid subtropical from south to far north of the state. The area has a higher annual rainfall ranges from 800 mm to 2000 mm. (Stormy Weather, Bureau of Meteorology). About 65% of the area is occupied by grazing lands which comprises of both native and modified pastures. The nature conservation areas which accounts for around 7.6% of the of total land use are mostly located in the eastern coastal areas. Dry land crops occupy about 9% of the area, while about 7% of the land is minimally used. (Catchment Scale Land use data, Department of Agriculture, Australia).

2.2. Data sets and data processing

2.2.1. Soil data

SOC data consists of the University of Sydney research data and the Terrestrial Ecosystem Research Network (TERN) data that are collected by different institutions for various purposes. There were 5386 observation sites in total. The data were clustered as they came from different survey projects from 1995 to 2014. The observed SOC content is given by the g/100 g. Since the distribution was positively skewed, the data was log-transformed for modelling procedures. The spatial distribution of those sampling points within the study area is shown in Fig. 1.

2.2.1.1. Harmonizing observed soil profile data. The sampling depths of the soil profiles were different to each other. For further analysis of data, it is imperative to have a common depth interval range across all sampling points. Malone et al. (2009) generalized and extended the quadratic spline model of Bishop et al. (1999) and formulated a smoothing spline function for vertical prediction of soil properties into specified common depth interval range.

The smoothing parameter (λ) of the quadratic spline function (Malone et al., 2009), is a determinant of the accuracy of prediction. It is crucial to find out the best λ value that minimizes the prediction error. Therefore, 506 sampling points which have more than 4 layers of measurements were selected to find out the best fitting λ value. The weighted average of the first two layers and the third and fourth layers were calculated to form two layers for each profile. Those values were then used to predict SOC values for the original sampling depths with respect to a series of λ values (0.00001, 0.0001, 0.001, 0.01, 0.1, 0.5, 1, 2, and 5). The λ value which gave the minimum mean squared error (MSE) value was selected as the best smoothing parameter. Then, the depth intervals (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm) corresponding to the digital soil mapping specifications of the GlobalSoilMap project (Arrouays et al., 2014) were used as the harmonized depth intervals for spatial prediction models of SOC.

2.2.2. Environmental covariates

The content and the spatial distribution of SOC in an ecosystem are driven by the environmental factors such as climate, underlying lithology, topography, fauna and flora. Introduced by Jenny (1941), this concept was generalized and formalized by McBratney et al. (2003) as the

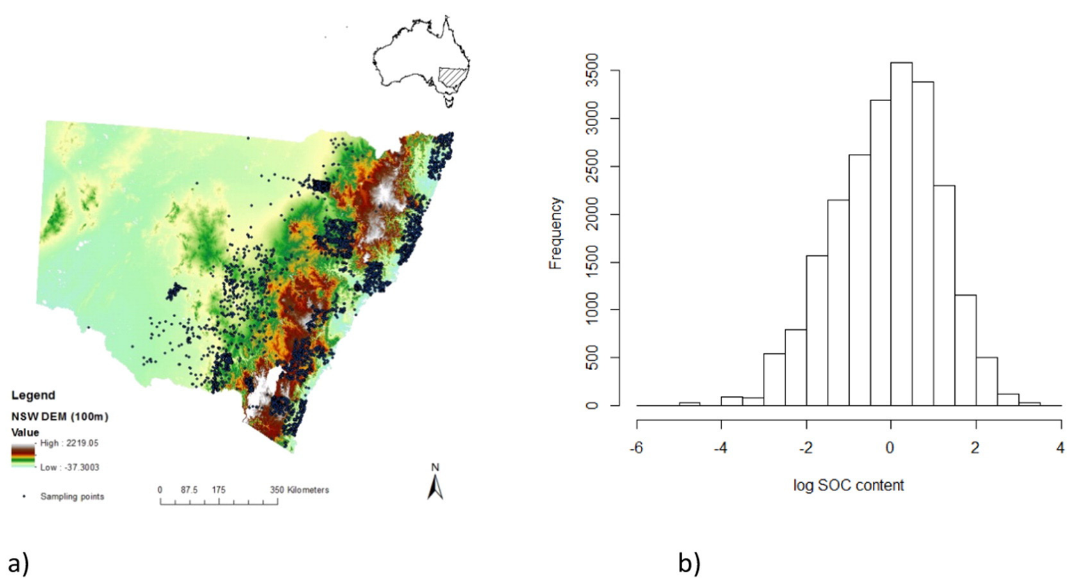

Fig. 1. (a) Spatial distribution of the sampling points in NSW, Australia, (b) histogram of observed SOC in log scale.