Science of the Total Environment 630 (2018) 367–378

Contents lists available at ScienceDirect

Science of the Total Environment

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

High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia

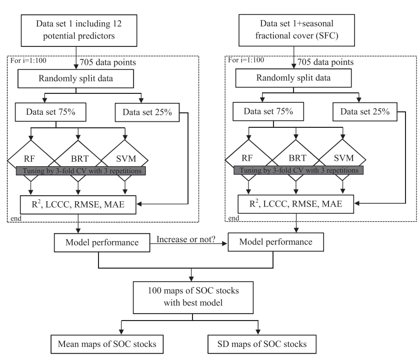
Bin Wang [a,*], Cathy Waters [b], Susan Orgill [a], Jonathan Gray [c], Annette Cowie [d], Anthony Clark [b], De Li Liu [a,e]

[a] NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, NSW 2650, Australia
[b] NSW Department of Primary Industries, Orange Agricultural Institute, NSW 2800, Australia
[c] Science Division, NSW Office of Environment and Heritage, PO Box 644, Parramatta, NSW 2124, Australia
[d] NSW Department of Primary Industries, Trevenna Rd, Armidale, NSW 2351, Australia
[e] Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, NSW 2052, Australia

HIGHLIGHTS

- Remote sensing covariates improved the estimation of SOC stocks.
- Prediction accuracy of tree-based models was superior to support vector machine.
- Digital soil mapping for SOC was practical and cost-effective in semi-arid rangelands.
- Fractional cover data influenced SOC stock at the soil surface.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 December 2017
Received in revised form 29 January 2018
Accepted 17 February 2018
Available online xxxx

Editor: Ouyang Wei

Keywords:
Soil organic carbon stocks
Seasonal fractional cover
Remote sensing
Machine learning
Digital soil mapping

ABSTRACT

Efficient and effective modelling methods to assess soil organic carbon (SOC) stock are central in understanding the global carbon cycle and informing related land management decisions. However, mapping SOC stocks in semi-arid rangelands is challenging due to the lack of data and poor spatial coverage. The use of remote sensing data to provide an indirect measurement of SOC to inform digital soil mapping has the potential to provide more reliable and cost-effective estimates of SOC compared with field-based, direct measurement. Despite this potential, the role of remote sensing data in improving the knowledge of soil information in semi-arid rangelands has not been fully explored. This study firstly investigated the use of high spatial resolution satellite data (seasonal fractional cover data; SFC) together with elevation, lithology, climatic data and observed soil data to map the spatial distribution of SOC at two soil depths (0–5 cm and 0–30 cm) in semi-arid rangelands of eastern Australia. Overall, model performance statistics showed that random forest (RF) and boosted regression trees (BRT) models performed better than support vector machine (SVM). The models obtained moderate results with R^2 of 0.32 for SOC stock at 0–5 cm and 0.44 at 0–30 cm, RMSE of 3.51 Mg C ha^{-1} at 0–5 cm and 9.16 Mg C ha^{-1} at 0–30 cm without considering SFC covariates. In contrast, by including SFC, the model accuracy for predicting SOC stock improved by 7.4–12.7% at 0–5 cm, and by 2.8–5.9% at 0–30 cm, highlighting the importance of including SFC to enhance the performance of the three modelling techniques. Furthermore, our models produced a more accurate

Abbreviations: SOC, soil organic carbon; SFC, seasonal fractional cover data; RF, random forest; BRT, boosted regression trees; SVM, support vector machine; RMSE, root-mean-square error; C, carbon; DSM, digital soil mapping; DEM, digital elevation model; SLGA, Soil and Landscape Grid of Australia; OOB, out-of-bag; NSW, New South Wales; R^2, regression coefficient of determination; MAE, mean absolute error; LCCC, Lin's Concordance Correlation Coefficient; RI, the relative improvement; OEH, Office of Environment and Heritage.

* Corresponding author.
E-mail address: bin.a.wang@dpi.nsw.gov.au (B. Wang).

https://doi.org/10.1016/j.scitotenv.2018.02.204

and higher resolution digital SOC stock map compared with other available mapping products for the region. The data and high-resolution maps from this study can be used for future soil carbon assessment and monitoring.

1. Introduction

Globally, rangelands account for approximately half of the world's land mass, providing a key role in the mitigation of climate change. The extensive areas occupied by rangelands can potentially store huge amounts of carbon both in biomass and soil organic matter (Bikila et al., 2016). Australian rangelands extend across low rainfall environments accounting for approximately 81% of national land area (http://www.environment.gov.au/land/rangelands). It is estimated that Australia's rangeland soils store between 34 and 48 Gt of carbon, representing a sequestration potential of 78 Mt C per year (Keating et al., 2009). Soil organic carbon (SOC) is also recognized as the most important indicator of soil fertility and playing a vital role in a range of soil processes (Schillaci et al., 2017a). While the accurate assessment of SOC stock is essential to enhance this resource, quantifying and mapping SOC stocks in the rangelands is challenging due to low levels of SOC and the inherently patchy spatial and temporal patterns of vegetation and soil resources (Waters et al., 2015). Using direct measurement (field survey including soil sampling and laboratory analyses) to determine SOC stocks is both time consuming and costly (Bartholomeus et al., 2011), and prohibitive at large scales (regional, national or global).

Digital soil mapping (DSM) techniques are a useful tool to reduce sampling and analytical costs while still obtaining reliable results (Jeong et al., 2017). DSM is the procedure of creating spatial soil information based on mathematical or statistical relationships between field soil observations and related environmental covariates or factors (e.g. climate, vegetation, relief, parent material and time) to understand spatial and temporal variation in soil type and other soil properties in the form of rasters of prediction (Camera et al., 2017; Jeong et al., 2017; Lagacherie et al., 2006; Malone et al., 2016; Minasny and McBratney, 2016). In DSM, these environmental variables can be retrieved from available digital elevation model (DEM), readily accessible remote sensing data and other data sources (such as climate data). The past few decades have seen the growth of DSM as a sub-discipline of soil science, experiencing a continuous expansion mainly due to its increased efficiency (Kempen et al., 2012) and accuracy (Lorenzetti et al., 2015) compared to conventional field soil mapping techniques. With continual growth in computational capacities, the great explosion of 'Big Data' involved with the development of data-mining algorithms, geographic information systems, and the increased availability of spatial data (DEM and satellite imagery) (Minasny and McBratney, 2016), DSM is likely to play an increasingly important role in the future monitoring of changes in soil properties and characteristics.

Recently, DSM has been successfully applied to map SOC stocks under a range of environments (Bonfatti et al., 2016; Gray et al., 2015; Ottoy et al., 2017; Schillaci et al., 2017a; Wang et al., 2017; Were et al., 2015; Yang et al., 2016). These advances in DSM of SOC mainly result from the development of machine learning techniques and the availability of high-quality covariates. The success of machine learning in DSM is related to several advantages over traditional soil survey. These advantages have been summarized as: 1) DSM is easy to update because predicting models can be stored and rerun when new data become available; 2) Different models of spatial variation can be chosen due to the availability of computing power to process large data sets; 3) The proper use of data mining tools and progress in geographic information systems results in predictions with quantified uncertainty (Kempen et al., 2012; Minasny and McBratney, 2016).

In Australia, a recent project has produced the Soil and Landscape Grid of Australia (SLGA) (http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html) (Grundy et al., 2015; Viscarra Rossel et al., 2015) which is based on recent digital soil mapping methods and integrates historical soil information and novel spatial modelling to generate nationwide digital maps of soil attributes including SOC (3 arc sec, approximately 90 m, resolution) (Grundy et al., 2015; Viscarra Rossel et al., 2014). However, the SLGA's accuracy varies between soil depths and soil textures across Australia. For example, the accuracy of SLGA products was higher in clay soil ($R^2 = 0.53$) than that in silt soil ($R^2 = 0.46$) at 0–5 cm. In addition, the range in time since the surveyed soil data were collected may result in poor estimates of the current status of attributes that are dynamic and responsive to land management practices, such as SOC (Grundy et al., 2015). Similarly, DSMs (100 m resolution) have been produced for key soil properties over New South Wales in eastern Australia (OEH, 2017) derived through quantitative modelling techniques (mainly multiple linear regressions) that are based on relationships between soil attributes and different environmental variables. These existing map products were produced at a national or state level, so they may not provide reliable information on SOC stocks down to the local or farm levels. This information is fundamental to monitor changes in the SOC stocks as a consequence of land management and is not available for the semi-arid rangelands of eastern Australia.

Accurate predictions of SOC stocks at smaller spatial scales are central in assessing the carbon sink capacity of soils, temporal changes due to seasonal conditions as well as the influence of management (Wang et al., 2017). Remote sensing data have gained attention in the past few decades as a promising secondary data source for improving DSM due to their high accessibility, resolution and availability at a range of scales. Forkuor et al. (2017) summarized the advantages of soil data sources derived from remote sensing as (1) contain extractable soil information, e.g. spectral reflectance, (2) have large spatial coverage and therefore permit mapping of inaccessible areas, (3) produce consistent and comprehensive data both in time and space and (4) provide possibilities of supplementing or at least reducing traditional labour-intensive soil sampling in soil surveys. Based on these advantages, numerous studies have explored the use of remote sensing data with varying spatial, temporal and spectral characteristics in digital soil mapping (Forkuor et al., 2017; Rudiyanto et al., 2016; Schillaci et al., 2017a,b; Wang et al., 2017; Yang et al., 2015). For example, Schillaci et al. (2017a) found that the integration of remote sensing with other environmental predictors increased the predictive ability compared to models built without remote sensing covariates. Previous studies in the semi-arid rangelands have shown clear relationships between ground cover (perennial grass and litter cover) and SOC stock (Orgill et al., 2017b; Waters et al., 2015, 2016). These relationships indicate that suitable satellite-derived covariates such as seasonal fractional cover (SFC) data may be useful in the estimation of SOC stocks in these semi-arid environments. However, the efficacy of SFC is improving prediction of SOC in semi-arid rangelands has not been tested.

The aim of the present study was to determine a reliable method for mapping the SOC stocks in the semi-arid rangelands of eastern Australia through different machine learning techniques using a set of environmental covariates obtained from remote sensing, and specifically to investigate whether inclusion of SFC improves prediction of SOC. We compared the influence of two groups of predictor variables on machine learning model performance in the study area; (1) 12 covariates (referred to as data set 1) including parent material, relief, climate and radiometric variables that represented a large suite of potentially useful covariates, (2) the covariates in data set 1 plus 16 additional covariates (annual seasonal fractional ground cover; mean value of Band 1 (bare