Geoderma 265 (2016) 152–163

Contents lists available at ScienceDirect

Geoderma

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

Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference

R.A. Viscarra Rossel [a,*], D.J. Brus [b], C. Lobsey [a], Z. Shi [c], G. McLachlan [d]

[a] CSIRO Land and Water, Bruce E. Butler Laboratory, PO Box 1666, Canberra, ACT 2601, Australia
[b] Alterra, Wageningen University and Research Centre, Environmental Science Group, Wageningen, The Netherlands
[c] Institute of Applied Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China
[d] Agriculture Flagship, CSIRO, Bruce E. Butler Laboratory, PO Box 1666, Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 8 August 2015
Received in revised form 19 October 2015
Accepted 10 November 2015
Available online 3 December 2015

Keywords:
Proximal soil sensing
Soil organic carbon stocks
Visible–near
Infrared spectroscopy
Design-based sampling
Regression estimator
Model-based inference

ABSTRACT

For baselining and to assess changes in soil organic carbon (C) we need efficient soil sampling designs and methods for measuring C stocks. Conventional analytical methods are time-consuming, expensive and impractical, particularly for measuring at depth. Here we demonstrate the use of proximal soil sensors for estimating the total soil organic C stocks and their accuracies in the 0–10 cm, 0–30 cm and 0–100 cm layers, and for mapping the stocks in each of the three depth layers across 2837 ha of grazing land. Sampling locations were selected by probability sampling, which allowed design-based, model-assisted and model-based estimation of the total organic C stock in the study area. We show that spectroscopic and gamma attenuation sensors can produce accurate measures of soil organic C and bulk density at the sampling locations, in this case every 5 cm to a depth of 1 m. Interpolated data from a mobile multisensor platform were used as covariates in Cubist to map soil organic C. The Cubist map was subsequently used as a covariate in the model-assisted and model-based estimation of the total organic C stock. The design-based, model-assisted and model-based estimates of the total organic C stocks in the study area were similar. However, the variances of the model-assisted and model-based estimates were smaller compared to those of the design-based method. The model-based method produced the smallest variances for all three depth layers. Maps helped to assess variability in the C stock of the study area. The contribution of the spectroscopic model prediction error to our uncertainty about the total soil organic C stocks was relatively small. We found that in soil under unimproved pastures, remnant vegetation and forests there is good rationale for measuring soil organic C beyond the commonly recommended depth of 0–30 cm.

1. Introduction

Soil organic carbon (C) helps to maintain soil health and productivity. It provides a primary source of nutrients for plants, helps to aggregate particles and develop soil structure, increases water storage capacity and availability for plants, protects soil from eroding and provides a habitat for soil biota. Capturing and retaining additional C in soil can improve the quality and productivity of the soil to sustain food production and simultaneously also mitigate the emissions of greenhouse gases (GHG).

Thoughtful land use and management practices, such as management-intensive grazing, can help store more soil organic C and offer good potential to improve soil quality, enable profitable food production and reduce net GHG emissions (Machmuller et al., 2015). For baselining and to assess the success of such practices, however, we need to accurately quantify the variability of soil organic C stock in both space and time. Importantly, we should aim to characterize its short range spatial variation, which can be significant, and to monitor over time intervals that enable detection of relatively small changes in C stocks.

Soil sampling protocols and conventional laboratory analyses can be used to directly measure organic C stocks. The protocols typically involve designing a sampling strategy, sampling the 0–30 cm soil layer and measuring the organic C concentration, bulk density and gravel content to derive the organic C stock of the soil in this layer. The methods are time-consuming, expensive, involve much sample handling and preparation and use complex procedures, which can be prone to analytical inaccuracies. The complexity and expense of the conventional approach are greater when there is a need to monitor the organic C stock of deeper soil layers or entire profiles. There is evidence that plants and cultivars with deeper and thicker root systems can input stable forms of organic matter deeper in the soil profile (Jobbágy and Jackson, 2000; Lorenz and Lal, 2005).

Conventional methods for measuring changes in the organic C stocks of soil are therefore impractical. If we are to increase our ability to characterize and monitor changes in soil organic C stocks, we need to

* Corresponding author.
E-mail address: raphael.viscarra-rossel@csiro.au (R.A. Viscarra Rossel).

http://dx.doi.org/10.1016/j.geoderma.2015.11.016

develop rapid, practical, accurate and cheaper methods to measure it (Izaurralde et al., 2013). Proximal soil sensing provides a range of tools that can be used to develop a multi-sensor system to efficiently measure the organic C stock of soil profiles (Viscarra Rossel et al., 2011). For example, electromagnetic induction sensors, gamma radiometers and precise global navigation systems can produce multivariate secondary information to help design sampling strategies and to map soil C (e.g. Simbahan and Dobermann, 2006; Miklos et al., 2010). Soil visible–near infrared (vis–NIR) spectroscopy can be used to measure soil organic C in the laboratory and in situ in the field (Stenberg et al., 2010).

Before we can start measuring with sensors however, we need to know where to sample. Locations can be selected by probability sampling (random sampling with known inclusion probabilities) or by non-probability sampling, giving rise to two widely used philosophies: the design- and the model-based approaches (de Gruijter and ter Braak, 1990; Brus and de Gruijter, 1993; Papritz and Webster, 1995; de Gruijter et al., 2006). In the design-based approach, the source of randomness of an observation is the random selection of the sampling sites. In the model-based approach, randomness originates from a random term in the model of the spatial variation, which is added to the model because our knowledge of the spatial variation is imperfect. Thus, probability sampling is a requirement for the design-based approach, whereas it is not for the model-based.

Choosing the most suitable approach depends, amongst other, on the motivation (Brus and de Gruijter, 1997). For example, the design-based approach might be more suitable if the aim is to obtain estimates of the '*global*' mean or total stock and their accuracies for an area, whose quality is not dependent on the correctness of modelling assumptions. The model-based approach might be preferable if we want to produce a '*local*' map of the soil organic C stock in the area. However, deciding which approach to use is often more complicated because the design-based approach can also be used for estimation of local means, and the model-based approach can be used for global estimation. Further discussion on the merits and disadvantages of each method can be found in de Gruijter and ter Braak (1990), Papritz and Webster (1995), Brus and De Gruijter (1997) and de Gruijter et al. (2006).

The possibility of using a regression model to assist with design-based inference, in a model-assisted approach, was discussed by Särndal et al. (1992) and Brus (2000). The approach uses auxiliary information, captured in a regression model, to improve the accuracy of design-based estimates of means and totals. There are fundamental differences between a model-based and a model-assisted approach. Significantly, the variance of a model-assisted estimate of the mean is a sampling variance, not a model-variance. Unlike the estimates of the model-based variance, the model-assisted estimates of the variance do not rely on the correctness of the model's assumptions. That is, if the assumptions underlying the regression model are violated, the model-assisted approach can still produce an unbiased estimate of the sampling variance (Brus, 2000).

Our aims here are to: (i) demonstrate the use of proximal soil sensors to measure the soil organic C stock of grazing land to a depth of 1 m, (ii) to compare the use of design-based, model-assisted and model-based methods to derive baseline estimates of the mean and total soil organic C stocks and their accuracies in the 0–10 cm, 0–30 cm and 0–100 cm layers, and (iii) to derive maps of soil organic C stocks and their uncertainties for each of the three depth layers.

2. Methods

2.1. Study site

The study area is 2837 ha and is located in the Upper Hunter Valley region, New South Wales, Australia, south of Wollar. It is approximately 300 km northwest of Sydney and 50 km northeast of Mudgee, near the Goulburn River National Park. The region has a temperate climate with an average annual rainfall of approximately 600 mm. Geology consists of shale, sandstone, mudstone conglomerates and coal. Landforms at the site consist of gently sloping colluvium and undulating foothills adjacent to north-flowing tributary creeks that are part of the Goulburn River Catchment. There are steep timbered ridges that surround on the south, west and east. The study area is used mostly for cattle grazing for beef production on rain-fed unimproved pastures, with remnant vegetation and surrounding forests on higher elevations. The soil there belongs to mostly the Dermosol and Kurosol orders in the Australian soil classification (Isbell, 2002), approximately equivalent to Planosols, Phaeozems and Acrisols in the World Reference Base system (IUSS Working Group WRB, 2006).

2.2. Proximal soil multi-sensor survey and data preprocessing

A mobile multi-sensor platform (MMSP) was used to survey the study area. The proximal sensors on the platform were an electromagnetic induction sensor, the EM-38 Mk2 (Geonics, Canada), a gamma radiometer with a 4.2 L NaI crystal detector (Radiation solutions, Canada), and a real-time kinetic global navigation system (RTK-GNS) (Trimble, USA).

The MMSP was driven between 10 to 20 km h^{-1} and the sensor data were recorded at a frequency of 1 Hz on parallel line transects with line spacing between 20 and 60 m. Both the speed and the line spacing depended on the navigability of the terrain. A map of the MMSP tracks is show in Fig. 1a. Using each sensor, the data recorded were: electrical conductivity and magnetic susceptibility recorded from the 0–0.5 m and 0–1 m depths, ($EC_{0.5}$, EC_1, $MS_{0.5}$ and MS_1), respectively; gamma radiometrics total dose, potassium (K), uranium (U) and thorium (Th), recorded from around the top 0.5 m of soil (Cook et al., 1996) and elevation with the RTK-GNS.

We checked the histograms of each sensor's data and checked for outliers using the Mahalanobis distance on their correlations. These and other spurious measurements were removed before proceeding with our analysis.

The gamma U and Th bands possessed significant random noise because of the short integration time that we used for the mobile measurements (i.e. 1 Hz). To improve the signal-to-noise ratio of these data, we aggregated each channel spatially using a moving average and using points within a 20 m radius. Thus, the gamma counts were integrated in space rather than in time (Viscarra Rossel et al., 2007).

We derived variograms for each of the sensor data and interpolated them onto a 5 m grid with ordinary block kriging (Webster and Oliver, 2007). The digital elevation map (DEM), produced by kriging, was used to derive terrain attributes that were thought to help describe the variation in soil organic C across the landscape. To derive the terrain attributes, we used the Geographic Resources Analysis Support System (GRASS) geographic information system (GIS) (GRASS Development Team, 2012). The attributes were slope, aspect, tangential, plan and profile curvatures, flow accumulation and the topographic convergence index (TCI) (GRASS Development Team, 2012).

The maps of the sensor data and the terrain attributes are shown in Fig. 1b–o. Elevation in the study area ranges from 360 m in the north to 485 m on ridges to the south (Fig. 1b). The soil has small electrical conductivity across the site but particularly on the ridges to the west (Fig. 1j). The gamma K counts were generally greater at higher elevations along the ridges (Fig. 1l) and suggest the occurrence of soil derived from parent material that contains K-bearing silicates.

2.3. Soil sampling

We selected sampling locations by probability sampling using a stratified simple random design (de Gruijter et al., 2006). We used the interpolated soil sensor data as the variates in the stratification, but we first reduced dimensionality and eliminated multicollinearity between them, using a principal component analysis (PCA). The PCA was