Geoderma 230–231 (2014) 119–130

Contents lists available at ScienceDirect

Geoderma

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

# Spatial prediction of soil organic carbon stock using a linear model of coregionalisation

T.G. Orton [a,b,*], M.J. Pringle [b], K.L. Page [b], R.C. Dalal [b], T.F.A. Bishop [a]

[a] Faculty of Agriculture and Environment, The University of Sydney, 1 Central Avenue, Australia Technology Park, Eveleigh, NSW 2015, Australia
[b] EcoSciences Precinct, Department of Science, Information Technology, Innovation and the Arts, GPO Box 5078, Brisbane, QLD 4001, Australia

ARTICLE INFO

*Article history:*
Received 17 June 2013
Received in revised form 11 April 2014
Accepted 13 April 2014
Available online 4 May 2014

*Keywords:*
Regional scale
Mapping
Multivariate geostatistics
Soil carbon

ABSTRACT

Spatial prediction of soil organic carbon (SOC) stock from concentration and bulk density measurements typically employs a two-step procedure. First, SOC concentration and bulk density data are used to calculate the SOC stock at each data location. Second, the calculated SOC stock at each data location is interpolated to give predictions at unsampled locations, often by applying a univariate geostatistical method to the stock. We refer to this as a 'calculate-then-model' approach. In this study, we investigate an alternative method to predict SOC stock at unsampled locations, based on a 'model-then-calculate' approach. We first consider the spatial covariation of the SOC concentration and bulk density data. This covariation can be modelled using a multivariate geostatistical approach, namely, the linear model of coregionalisation (LMCR), which deals with data on two (or more) correlated variables that together exhibit spatial correlation. The LMCR can be used to calculate a multivariate prediction distribution at an unsampled location for SOC concentration and bulk density in all depth intervals (for which observations were made). A prediction of the SOC stock can subsequently be calculated from this distribution. We compare the 'calculate-then-model' and 'model-then-calculate' approaches using data on SOC concentrations and bulk densities from croplands across Queensland, Australia. Cross-validation results show that the two approaches give similar prediction accuracies and reasonable uncertainty assessments. Therefore, the more simple 'calculate-then-model' approach might be favoured when both SOC concentration and bulk density measurements are available at all data locations. However, when some locations have just SOC concentration data — a commonly encountered situation due to difficulties in obtaining bulk density measurements under field conditions — then the 'model-then-calculate' approach offers a useful alternative; we show that it still gives good predictions and a fair assessment of uncertainty in this situation. We also show the importance of the LMCR for modelling the correlation between SOC concentration and bulk density data for the different depth intervals within a 'model-then-calculate' approach; if this correlation is not accounted for (i.e. modelling of the SOC concentration and bulk density data for the different depth intervals is performed independently), then very poor uncertainty assessments result.



## 1. Introduction

Spatial prediction of soil organic carbon (SOC) stock has become a key issue over recent years, because of the potential impacts of carbon on climate change and soil health (Bellamy et al., 2005; Lal, 2004; Saby et al., 2008). This issue has received considerable attention, in part because of the large variation displayed by SOC at all scales from national to field, and also due to the expense of obtaining accurate measurements of SOC. As a result, research into approaches to improve spatial prediction of SOC stock is on-going (Minasny et al., 2013).

Many different techniques have been used to predict SOC stock at locations where it has not been measured, and thereby create a map (Grunwald, 2009). Typically, data from each of $N$ sampled sites are first used to calculate the SOC stock at these sites through:

$$c_{stock} = c\,\rho\,(1-r)\,\alpha, \quad (1)$$

where $c$ is the SOC concentration, $\rho$ the bulk density, and $r$ the proportion of rock fragments for the sample. (In this study of SOC in cropping land, we will assume zero rock fragments.) The final term, $\alpha$, is a factor to get to the appropriate units for SOC stock (Mg ha$^{-1}$) down to the relevant reference soil depth (commonly 30 cm, following the IPCC guidelines; IPCC, 1996), or preferably, to a fixed soil mass (Gifford and Roderick, 2003; McBratney and Minasny, 2010). Subsequently, the calculated stock at the $N$ sites is either interpolated (perhaps through a geostatistical approach, e.g. Cambule et al., 2014; Phachomphon et al., 2010), or used to build some model to predict the SOC stock at unsampled locations, for example, boosted regression trees (Martin

* Corresponding author. Tel.: +61 7 3170 5772.
*E-mail address:* Thomas.Orton@science.dsitia.qld.gov.au (T.G. Orton).

http://dx.doi.org/10.1016/j.geoderma.2014.04.016
0016-7061/

et al., 2011). We refer to this general approach as a 'calculate-then-model' approach.

An alternative to this is a 'model-then-calculate' approach. First, the variables for which we have data are modelled and interpolated onto a prediction grid. Typically, the dataset will include measurements of the bulk density and SOC concentration for multiple depth intervals. Subsequently, the interpolated values of these variables are used to calculate estimates of stock, whilst accounting for the uncertainty of the interpolations. Since the relationship between SOC stock and its components is non-linear, we would not expect the model-then-calculate and calculate-then-model approaches to give the same results.

For the 'model' step of this model-then-calculate approach, we could employ any of a number of methods (as referred to above); here we consider geostatistical approaches. A first naïve approach might be to apply univariate geostatistical methods (Webster and Oliver, 2001) to the data for each of the variables independently. This approach, however, ignores any correlation between the SOC concentration and bulk density and between the data for the different depth intervals. Panda et al. (2008) and Goidts et al. (2009) have stressed the importance of accounting for correlations between SOC stock constituents when calculating SOC stock estimates. This 'covariation' can be modelled using a multivariate geostatistical approach, namely, the linear model of coregionalisation (LMCR), which is designed to deal with data on two (or more) correlated variables that together exhibit spatial correlation (Webster and Oliver, 2001). The LMCR can be used to calculate a bivariate prediction distribution for SOC concentration and bulk density at an unsampled location; if data from multiple depth intervals are available, a multivariate prediction of the SOC concentration and bulk density for all intervals can be calculated. A prediction of the SOC stock calculated from this distribution will therefore explicitly account for the spatial correlation for both constituents of the SOC stock, the correlation between them, and the correlation between the data for the different depth intervals. If the LMCR is able to represent the variability of each of the SOC stock components, then a model-then-calculate approach could be useful, particularly when data on the components are incomplete (e.g. some missing bulk density measurements). Allen et al. (2010) described how a LMCR of SOC concentration and bulk density could be advantageous for designing sampling schemes. However, we are unaware of any studies that have applied this approach to the prediction of SOC stock.

In this work, we consider the calculate-then-model and model-then-calculate frameworks for prediction of SOC stock using data on SOC concentrations and bulk densities from croplands across Queensland, Australia. The data are measurements of SOC concentrations and bulk densities from three sampled depth intervals; therefore, the model-then-calculate approach consists of modelling the spatial distribution of six variables. We use cross-validation to compare the methods under different scenarios for the available data. First, we consider a full dataset (i.e. all data locations consist of measurements of all variables). Second, we investigate prediction performance when some data locations do not have measurements of bulk density: this represents a commonly-encountered situation, due to conventional bulk density sampling methods in the field being time-consuming and labour-intensive (Allen et al., 2010; Don et al., 2007; Holmes et al., 2011). We draw conclusions and provide some recommendations as to the most appropriate methodology in each particular scenario.

## 2. Case study data

### 2.1. Study region, soil sampling and analysis

Our focus in this study is on the cropping lands of Queensland. The grey area in Fig. 1 covers three of the five zones that constitute Queensland's Strategic Cropping Land: the Western Cropping, Eastern Darling Downs and Granite Belt zones (The State of Queensland, 2011). When we consider mapping SOC stock across Queensland's cropping lands, we focus on this area. We note that this area does also include other land uses (grazing lands and national parks in particular), for which our predicted SOC stock is not appropriate. An alternative would be to use a more refined outline of the cropping lands as our mapping area. However, this would result in very fragmented maps that were difficult to compare. Therefore, for the purposes of this study, in which the main interest is the comparison of methods for spatial prediction of SOC stock, the outline shown in Fig. 1 is sufficient.

The data we work with were reported in a previous study (Page et al., 2013), and we provide only a brief description here. Soil samples were collected from 179 sites across the cropping lands of Queensland. We use data from 172 of these sites (Fig. 1), for which composite soil samples were formed; the remainder we discarded because no composite was made. Each site was marked out with a 25-m × 25-m grid with 5-m spacings (i.e. 36 locations), and 10 of these locations were selected randomly for sampling. Soil cores were collected from each of these 10 locations down to 30-cm depth using a hydraulic push rig. The cores were divided into 10-cm depth intervals, and then combined for each interval to create composite samples for the 0–10-cm, 10–20-cm and 20–30-cm intervals. SOC concentrations were determined for the composite samples using the dry combustion (LECO) method. An Additional three soil cores were taken to 30 cm from random locations within the 25-m × 25-m area using a 42-mm internal diameter tube, and split into the three depth intervals for bulk density measurements. All sites were sampled during a fallow period. We refer to Page et al. (2013) for a more detailed description of the analysis.

Fig. 2 shows histograms and scatterplots of the three bulk density variables (denoted as $\rho_1$, $\rho_2$ and $\rho_3$), and SOC-concentration variables (denoted as $c_1$, $c_2$ and $c_3$), for the three sampled depth intervals. The plots show the negative correlation of bulk density with SOC concentration, and the positive correlation between data for the three sampled depth intervals. This suggests that a statistical modelling approach that makes use of the joint variability of these six variables, such as the LMCR, could be advantageous. The histograms suggest some skewness in the SOC concentration data.

### 2.2. Calculation of SOC stock from soil-profile bulk density and SOC concentration

In this paper, our primary variable of interest is the SOC stock, calculated according to the fixed soil-mass system (Gifford and Roderick, 2003) to a nominal reference depth of 30 cm (IPCC, 1996). Given the concentrations and bulk densities of the three depth intervals at any particular location, we calculated the stock as follows. The approach makes use of spline functions, which can be fitted to soil profile data to represent the variation of bulk density and SOC concentration with depth (Bishop et al., 1999). Since interest is in the SOC stock down to a fixed soil mass (which corresponds to different depths at different locations), such spline functions prove useful for calculating the stock (Malone et al., 2009; Pringle et al., 2011). We note that for the 'model-then-calculate' approach, the calculation was embedded within a Monte Carlo routine to account for the uncertainty for the predicted SOC concentration and bulk density in the three depth intervals, as detailed in Section 3.1.

First, the soil mass (megagram per square metre of ground area) to a depth of 30 cm was calculated for each sample by summing the soil masses of the three depth intervals (the soil mass of each interval is given by the bulk density, in Mg $m^{-3}$, multiplied by the length of each interval, in m). The minimum mass per unit area to 30 cm over all 172 samples was 0.30 Mg $m^{-2}$, which we refer to as the reference mass, $m_{ref}$. An equal-area spline (Bishop et al., 1999; Malone et al., 2009) was then fitted (with smoothing parameter, $\lambda = 0.01$), to model the variation of bulk density through each soil profile. For each profile, with bulk density profile data $\boldsymbol{\rho}_i$, the spline was used to calculate cumulative soil masses (per unit ground area) for small increments of the depth (we use 1-cm increments), so that the depth in m, $d_{ref}(\boldsymbol{\rho}_i, m_{ref})$,