Science of the Total Environment 631–632 (2018) 377–389

Contents lists available at ScienceDirect

Science of the Total Environment

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

Accounting for the measurement error of spectroscopically inferred soil carbon data for improved precision of spatial predictions

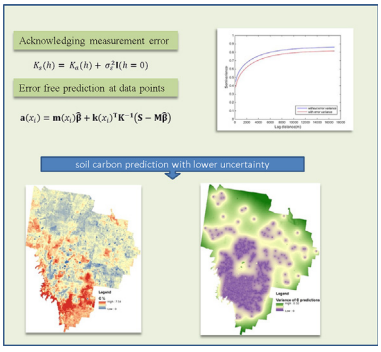
P.D.S.N. Somarathna *, Budiman Minasny, Brendan P. Malone, Uta Stockmann, Alex B. McBratney

Sydney Institute of Agriculture, The University of Sydney, New South Wales, Australia

HIGHLIGHTS

- Measurement errors can be filtered through incorporating in the covariance structure of the spatial model.
- Acknowledging measurement errors in spatial modeling yields a lower uncertainty in spatial predictions.
- MCMC techniques can be used to define the posterior density of measurement error variance.
- Performance of REML-EBLUP approach is comparable to MCMC techniques in terms of bias correction of the spatial model.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 11 September 2017
Received in revised form 11 February 2018
Accepted 25 February 2018
Available online xxxx

Editor: Ouyang Wei

Keywords:
Infrared spectroscopy
Measurement error
Variogram
REML-EBLUP
MCMC
Filtered kriging

ABSTRACT

Spatial modelling of environmental data commonly only considers spatial variability as the single source of uncertainty. In reality however, the measurement errors should also be accounted for. In recent years, infrared spectroscopy has been shown to offer low cost, yet invaluable information needed for digital soil mapping at meaningful spatial scales for land management. However, spectrally inferred soil carbon data are known to be less accurate compared to laboratory analysed measurements. This study establishes a methodology to filter out the measurement error variability by incorporating the measurement error variance in the spatial covariance structure of the model. The study was carried out in the Lower Hunter Valley, New South Wales, Australia where a combination of laboratory measured, and vis-NIR and MIR inferred topsoil and subsoil soil carbon data are available. We investigated the applicability of residual maximum likelihood (REML) and Markov Chain Monte Carlo (MCMC) simulation methods to generate parameters of the Matérn covariance function directly from the data in the presence of measurement error. The results revealed that the measurement error can be effectively filtered-out through the proposed technique. When the measurement error was filtered from the data, the prediction variance almost halved, which ultimately yielded a greater certainty in spatial predictions of soil carbon. Further, the MCMC technique was successfully used to define the posterior distribution of measurement error. This is an important outcome, as the MCMC technique can be used to estimate the measurement error if it is not explicitly quantified. Although this study dealt with soil carbon data, this method is amenable for filtering the measurement error of any kind of continuous spatial environmental data.

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

https://doi.org/10.1016/j.scitotenv.2018.02.302
0048-9697/

1. Introduction

Soil carbon is recognized as a variable central to soil fertility and agricultural productivity. It is also well known for its capacity to serve as a store for atmospheric carbon. Transferring atmospheric CO_2 into long-lived pools and securely storing so that it is not immediately remitted is known as carbon sequestration (Lal, 2004; Yigini and Panagos, 2016). Small increases in soil carbon stocks per unit land area are anticipated to result in significant changes in climate and land use management (Falloon and Betts, 2010). Understanding soil carbon processes for implementing "best practice" for balancing carbon budgets is pivotal for carbon sequestration programs (Dawson and Smith, 2007). These programs need extensive sampling for auditing soil carbon stocks. Similarly, the assessment of soil health would also require conducting extensive measurement of soil carbon.

With the growing need for detailed soil carbon data, existing soil carbon maps and inventories are becoming inadequate, especially for large scale projects (Stevens et al., 2013). Standard techniques of soil carbon measurements such as dry combustion and oxidation analyses can be tedious, time consuming and expensive (Nocita et al., 2014). Conversely, infrared spectroscopy has been demonstrated to be a near comparable measurement technique that has the added advantage of being relatively low cost (Janik et al., 2007; Reeves III, 2010; Rossel and Webster, 2012; Stevens et al., 2013; Viscarra Rossel et al., 2006). The low cost associated with this technique means that mapping studies can afford higher sampling densities, thus enabling a detailed understanding soil carbon spatial variation across landscapes.

The use of infrared spectroscopy for soil analysis has been thriving over the past decade (Bellon-Maurel and McBratney, 2011). These studies have mostly focused on predicting basic soil composition, particularly soil organic carbon (SOC) and texture (Stenberg et al., 2010). Bellon-Maurel and McBratney (2011) provide a detailed review of the studies on the use of NIR and MIR spectroscopic studies for soil carbon inference. The review showed that these soil spectral inference studies are largely dedicated to predicting soil carbon content for point locations. However, it is proposed that these soil spectral inference studies could be further expanded into a spatial context for *optimally* predicting soil carbon content at unsampled locations, and ultimately for soil mapping purposes.

Infrared spectroscopic soil carbon measurement is an indirect mode of measurement. The carbon concentrations are inferred using calibration models based on the characteristics of the absorption spectrum of scanned soil samples. One drawback of using these soil carbon data is the comparatively larger measurement error associated with calibration models compared to the data acquired through standard dry combustion techniques (Bellon-Maurel et al., 2010).

When predicting the soil carbon content spatially, we are interested in the actual value rather than the value distorted by the measurement error. More often than not, measurement error is disregarded. For example, a recent study by Rial et al. (2017) mapped topsoil organic carbon content using Visible-Near Infrared (VNIR) spectroscopic measurements without accommodating within the methodology a procedure for handling the measurement errors in the data.

To achieve an optimal prediction in a spatial modelling exercise, the measurement errors should be filtered out (Cressie, 1991). One way of accounting for the measurement error is to include measurement error variance (σ_ε^2) in the variogram or covariance structure of the spatial model. This is also known as kriging with uncertain data, where the error variance is added to the diagonal of the spatial covariance matrix (Delhomme, 1978; Knotters et al., 1995; Laslett and McBratney, 1990). This filters the measurement error variance from the nugget component of the experimental variogram, ultimately leading to lower uncertainty of spatial predictions.

The accuracy of the spatial predictions can also be influenced by the techniques of model parameter estimation. Conventional techniques using method-of-moments can be biased (Lark et al., 2006), and thus the Residual Maximum Likelihood Method (REML) and Bayesian inference from Markov Chain Monte Carlo (MCMC) analysis are the established techniques for unbiased parameter estimation (Poggio et al., 2016). Lark et al. (2006) used REML for estimating parameters of the covariance function directly from the data, and then the estimated parameters were used for the spatial prediction in what is termed as an empirical best linear unbiased predictor (EBLUP). MCMC simulation can also be applied for estimating the variogram and trend model parameters directly from data. Minasny et al. (2011) advocated the use of MCMC simulation for parameter inference in model-based soil geostatistics including the spatial prediction of soil carbon. The basic advantage of MCMC over REML is that MCMC estimates the underlying uncertainty of the parameters, whereas REML relies on a single realisation of the variogram parameters. However, MCMC estimations are computationally expensive compared to the REML approach due to the slow convergence rates of the former (Mossel and Vigoda, 2006; Poggio et al., 2016).

In this study, we explored the applicability of REML-EBLUP and MCMC simulation for measurement error parameter inference for soil carbon spatial modelling. A combination of laboratory measured (dry combustion), near infrared red (NIR) and mid infrared (MIR) spectra estimated soil carbon data and associated σ_ε^2 were used for predicting soil carbon content across the Hunter Valley region, NSW, Australia. Subsequently, we compared the prediction capability of each model, i.e. incorporating σ_ε^2, and without σ_ε^2.

2. Theoretical context

The stochastic spatial process of soil carbon can be expressed by a linear mixed model.

$$\mathbf{S} = \mathbf{M}\boldsymbol{\beta} + \mathbf{W}\mathbf{u} + \mathbf{e} \tag{1}$$

$\mathbf{S}$ is the vector of n observations, $\mathbf{M}$ is the $n \times p$ design matrix that associates with each value of p fixed effects, and $\boldsymbol{\beta}$ is the vector of p fixed effect coefficients. $\mathbf{u}$ is the vector of q random effects, realisations of variable $\mathbf{u}$, which is associated with the n observations by an $n \times q$ design matrix $\mathbf{W}$. It is assumed that $\mathbf{u}$ is the spatially dependent random variable, while $\mathbf{e}$ independent random errors and $\mathbf{u}$ and $\mathbf{e}$ are independent to each other. Hence, assuming $\mathbf{u}$ and $\mathbf{e}$ are jointly Gaussian,

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma^2\xi\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \sigma^2\mathbf{I}\end{bmatrix}\right) \tag{2}$$

where σ^2 is the variance of the independent error, ξ is the variance ratio between $\mathbf{u}$ and σ^2 and $\mathbf{G}$ is the correlation matrix of $\mathbf{u}$. $\mathbf{e}$ represents both measurement errors and the short scale variations of the spatial process which is geo-statistically known as the nugget effect. Assuming $\mathbf{u}$ is drawn from second order stationary random process, $\mathbf{G}$ can be characterised by a suitable covariance function since it only depends on the relative locations of the observations (Lark et al., 2006).

The Matérn covariance function has been effectively used in soil science (Minasny and McBratney, 2005) to model the covariance structure of the random effects. The Matérn covariance function ($\mathbf{K}$) is given as,

$$K_{ij} = c_0\delta_{ij} + c_1\left[\frac{1}{2^{\upsilon-1}\Gamma(\upsilon)}\left(\frac{h}{r}\right)^{\upsilon}K_\upsilon\left(\frac{h}{r}\right)\right] \tag{3}$$

where K_{ij} is the covariance between observation i and j, h represents the separation distance between i and j, δ_{ij}denotes the Kronecker delta ($\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ when $I \neq j$), $c_0 + c_1$ signifies the sill variance, K_υ is the modified Bessel function of the second kind of order υ. Γ is the gamma function, r denotes the distance or 'range' parameter and υ is the spatial 'smoothness'. The latter parameter allows greater flexibility in modelling the local spatial covariance. The parameters of the covariance function along with σ^2 and ξ can be estimated using REML. This