Published online November 2, 2017

Pedology

More Data or a Better Model? Figuring Out What Matters Most for the Spatial Prediction of Soil Carbon

P.D.S.N. Somarathna*
Budiman Minasny
Brendan P. Malone
Sydney Institute of Agriculture
School of Life and Environmental Sciences
Univ. of Sydney
New South Wales
Australia

Modeling techniques used in digital soil carbon mapping encompass a variety of algorithms to address spatial prediction problems such as spatial non-stationarity, nonlinearity and multi-colinearity. A given study site can inherit one or more such spatial prediction problems, necessitating the use of a combination of statistical learning algorithms to improve the accuracy of predictions. In addition, the training sample size may affect the accuracy of the model predictions. The effect of varying sample size on model accuracy has not been widely studied in pedometrics. To help fill this gap, we examined the behavior of multiple linear regression (MLR), geographically weighted regression (GWR), linear mixed models (LMMs), Cubist regression trees, quantile regression forests (QRFs), and extreme learning machine regression (ELMR) under varying sample sizes. The results showed that for the study site in the Hunter Valley, Australia, the accuracy of spatial prediction of soil carbon is more sensitive to training sample size compared to the model type used. The prediction accuracy initially increases exponentially with increasing sample size, eventually reaching a plateau. Different models reach their maximum predictive potential at different sample sizes. Furthermore, the uncertainty of model predictions decreases with increasing training sample sizes.

Abbreviations: CCC, concordance correlation coefficient; DEM, digital elevation model; DSM, digital soil mapping; ELMR, extreme learning machine regression; GWR, geographically weighted regression; LMM, linear mixed model; MIR, mid-infrared; ML, maximum likelihood; MLR, multiple linear regression; NDVI, normalized difference vegetation index; NIR, near-infrared; QRF, quantile regression forest; REML, residual maximum likelihood; SSD, standardized squared deviation; TWI, topographic wetness index.

Core Ideas

- **Sample size is the major driving factor of prediction accuracy of soil carbon.**
- **The prediction accuracy increases at a decreasing rate with increasing sample sizes.**
- **Larger sample sizes deliver equally good prediction accuracy despite the model type.**
- **Model type affects the reproducibility (precision) of the predictions.**
- **Uncertainty of model predictions decreases with increasing sample sizes.**

Soil carbon is a key property which controls soil quality, as it is closely related to the structural stability of soil aggregates, soil fertility and plant growth (Blanco-Canqui et al., 2013; McBratney et al., 2014). In addition, soil carbon has the potential to mitigate climate change (Minasny et al., 2017). Consequently it has been the focus of much digital soil mapping (DSM) research in recent times. The number of publications on mapping soil carbon has dramatically risen over the past decade (Grunwald, 2009). A review by Minasny et al. (2013) reveals that the digital mapping of soil carbon employs a diverse range of spatial modeling techniques. In terms of complexity, these methods range from simple linear regression to complex machine learning techniques (Minasny and McBratney, 2016).

Modeling techniques used in soil carbon mapping encompass a variety of statistical methods to address spatial prediction problems such as spatial non-stationarity, nonlinearity and multi-colinearity. Spatial non-stationarity describes a condition where by a general 'global' model fails to explain target variable variation in an area as the soil-landscape relationships are not constant in space (Brunsdon et al., 1998). Nonlinearity describes a condition where the relationship between primary and secondary variables of the model is not directly proportional, but instead varies

Soil Sci. Soc. Am. J.
doi:10.2136/sssaj2016.11.0376
Received 16 Nov. 2016.
Accepted 23 June 2017.
*Corresponding author: (sanjeewani.pallegedaradewage@sydney.edu.au).
© Soil Science Society of America, 5585 Guilford Rd., Madison WI 53711 USA. All Rights reserved.

in a nonlinear fashion. In such situations, the relationships are affected by various factors such as the configuration or relative location other than the values of the variables. This is also termed as 'nonlinear spatial dependence' (Vann and Guibal, 2001). Multi co-linearity occurs when two or more predictor variables in the spatial model are highly correlated. When multico-linearity exists, small changes in the data can cause significant changes in regression coefficients, which eventually lead to higher standard errors of regression coefficients. Furthermore, the regression coefficients can have the incorrect sign and unreasonable magnitude under such circumstances (Kozak, 1997). Ultimately, soil carbon prediction problems are site specific as the soil carbon content is directly related to the spatial properties of the study site. A given study site can inherit one or more aforementioned spatial prediction problems. Dealing adequately with these problems requires the use of a combination of modeling algorithms.

Similarly, the accuracy of a model also depends on the training sample size. A study focusing on decision tree-based modeling algorithms by Morgan et al. (2003) revealed that the prediction accuracy increases at a decreasing rate with sequentially increasing sample size. John and Langley (1996), Frey and Fisher (1999), and Provost et al. (1999) have conducted separate studies using spatial models other than data-mining based models, and came to similar conclusions as Morgan et al. (2003). Taking the analysis a step further, studies have been conducted by Kelley (2007), Kelley and Maxwell (2003) and Maxwell et al. (2008) to establish an optimal sample size for deriving the highest achievable accuracy in estimating multiple linear regression parameters.

This study primarily focuses on identifying how diverse spatial modeling techniques perform under varying training sample sizes, in terms of soil carbon predictions. We selected a range of spatial models commonly used in DSM and also emerging techniques in machine learning literature to include multiple linear regression (MLR), geographically weighted Regression (GWR), linear mixed models (LMM), Cubist models, quantile regression forests (QRF), and extreme learning machine regression (ELMR).

We trained and tested the foregoing models using data collected across the Hunter Valley region, NSW, Australia. We compared the prediction accuracy of these models under varying sampling sizes. We also evaluated the prediction uncertainties of the models and the potential of model ensembles in lowering such uncertainties. The remainder of this paper describes the methods used in detail, discusses the results, and draws conclusions on the performance of model types with respect to the training sample size.

METHODOLOGY

Study Area

The study site is situated in the Lower Hunter Valley, NSW, Australia, in an area known as the Hunter Wine Country Private Irrigation District. This site is located in the southwest portion of the District and has an undulating topography with hills ascending to the south and west (Fig. 1). The area experiences a temperate climate, with warm humid summers and relatively cool winters. Rainfall is uniformly distributed with an annual average of 740 mm. The underlying geology is comprised of predominantly Early Permian siltstones, marl and some minor sandstone and Late Permian siltstones, Middle Permian conglomerates, sandstones and siltstones in minor amounts (Hawley et al., 1995).

Soil Carbon Data

The term 'soil carbon' in this research refers to the total carbon content in the soil. The soil carbon data comes from two

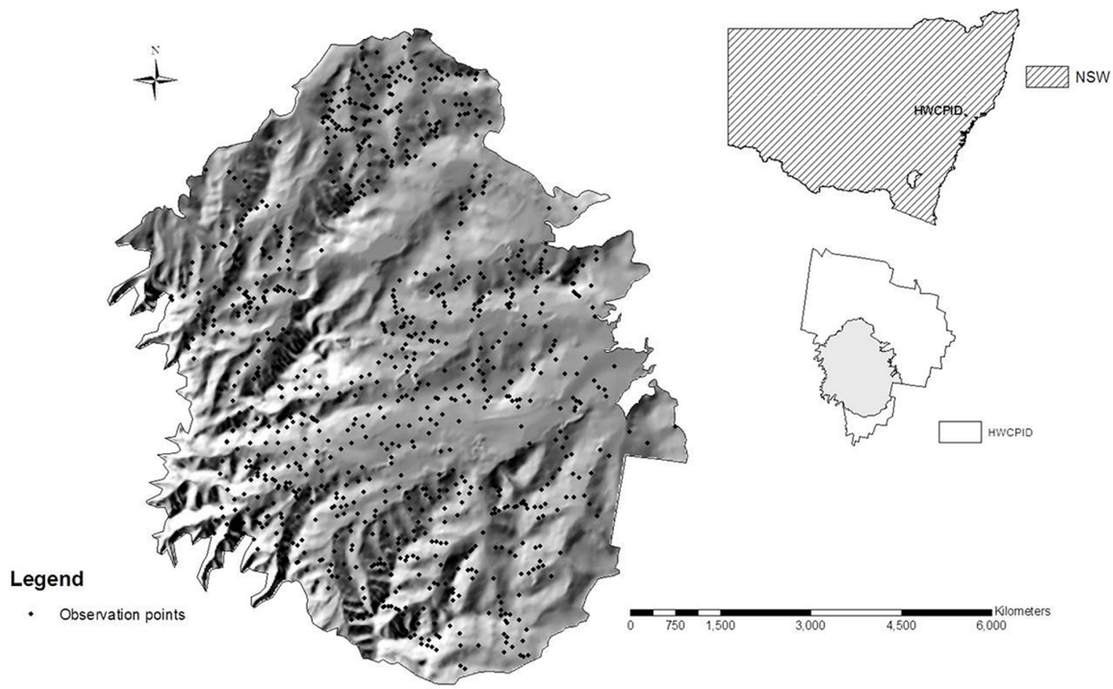

Fig. 1. Spatial distribution of observation points across the Hunter Valley study area in New South Wales, Australia. The layer is a hillshaded surface derived from the available digital elevation model.

sources: (i) data collected between the years of 2001 and 2015 during annual soil surveys performed by students in the Faculty of Agriculture and Environment, University of Sydney, and (ii) data from two PhD research projects (Malone et al., 2011) and (Odgers et al., 2011). Soil samples were collected from depths of 0 to 10 cm and 40 to 50 cm. These samples were taken from 100-cm soil cores from each sampling location; topsoil samples were obtained from 0 to 10 cm and subsoil from 40 to 50 cm. Therefore, the 0- to 10-cm and 40- to 50-cm soil layers will henceforth be referred to as 'topsoil' and 'subsoil', respectively. As the soil carbon data have been collected over several years for different purposes, a consistent method has not been used for soil carbon measurement. Methods used include dry combustion, and spectral inference from near-infrared (NIR) and mid-infrared (MIR) diffuse reflectance measurements.

Dry combustion of the soil samples was conducted using an ElementarVario Max CNS macro elemental analyzer (Elementar Analyses System GmbH, Hanau, Germany) where the carbon content is determined by the loss on ignition at 400°C (Zobeck et el., 2013). The standard deviation of the soil carbon measurement of the ElementarVario Max CNS analyzer is 0.001 to 0.004 g 100 g^{-1} based on standard soil samples.

Soil carbon content was spectrally inferred using the absorption spectrum produced after scanning the soil sample with an analytical spectral instrument (i.e., NIR or MIR). The absorption spectrum has a characteristic shape based on the constituents of the soil, and it is then used to infer the soil carbon content via calibration models. NIR spectroscopic measurements were obtained using an Agrispec portable spectrophotometer with a contact probe attachment (Analytical Spectral Devices, Boulder, CO) and Bruker TENSOR 37 Fourier Transform MIR spectrometer was used to measure the MIR spectral reflectance of soil samples. The collected NIR/MIR spectra were pre-processed to remove the noise, followed by normalizing before using them for the calibrations. Calibration models were derived using a regression tree model called Cubist (Quinlan, 1992; Minasny et al., 2008), where spectral data is linked with soil carbon content measured via the dry combustion method. The calibration data are from a library of 316 soil profile samples from the wheatbelt of southern New South Wales and northern Victoria (Geeves et al., 1995). See Minasny et al. (2008) for a description of the MIR spectral calibration model.

The final database was a pool of dry combusted, and spectroscopically inferred soil carbon data. There were 1435 and 1027 samples in total for the top- and subsoils, respectively. Figure 2 shows the experimental exponential variogram of square-root transformed soil carbon for the two soil layers. The topsoil samples had less variability in carbon content compared to the subsoil samples. Also the range for topsoil carbon is about the one-fifth of the subsoil carbon.

Environmental Covariates

The amount and the spatial distribution of soil carbon and other soil properties are driven by environmental factors such as climate, lithology, topography, flora and fauna, space, and time. This relationship between soil and other spatially referenced factors is described by the SCORPAN spatial prediction function (McBratney et al., 2003). We can improve on this general model by selecting the most important auxiliary variables to build a more parsimonious spatial model. Model types such as Random Forest (Wheeler and Tiefelsdorf, 2005) and Cubist are embedded with methods to tune parameters—for example the number of trees to be constructed or the number of environmental variables to consider in each model fitting procedure in the case of Random Forest models—optimally using inbuilt cross-validation options (Genuer et al., 2010), while some (e.g., MLR, GWR and LMM) are not. As this study is designed to compare several modeling techniques, as a generalization and to improve calibration efficiency, auxiliary variables (environmental covariates) were selected prior to model building. The covariate pool consisted of 22 covariates. Table 1 provides a description of data sources, raster resolution and definition of all covariates that were considered in this study.

Correlation coefficients among covariates and stepwise regression were used to select the most parsimonious model for the study. Highly correlated covariates were removed before proceeding into stepwise regression to avoid issues of multi-colinearity. Stepwise regression can be implemented as a combination of forward and backward elimination of predictor variables where the predictors with higher probabilities than the critical value are removed in a step wise fashion. Therefore, the final model of the stepwise regression contains the most statistically significant variables that best describes the soil carbon variability of the study site. Accordingly, altitude above channel network, analytical hillshade, Landsat band 5, elevation, land cover, normalized difference vegetation index (NDVI), plan curvature, topographic wetness index (TWI), slope direction and terrain roughness

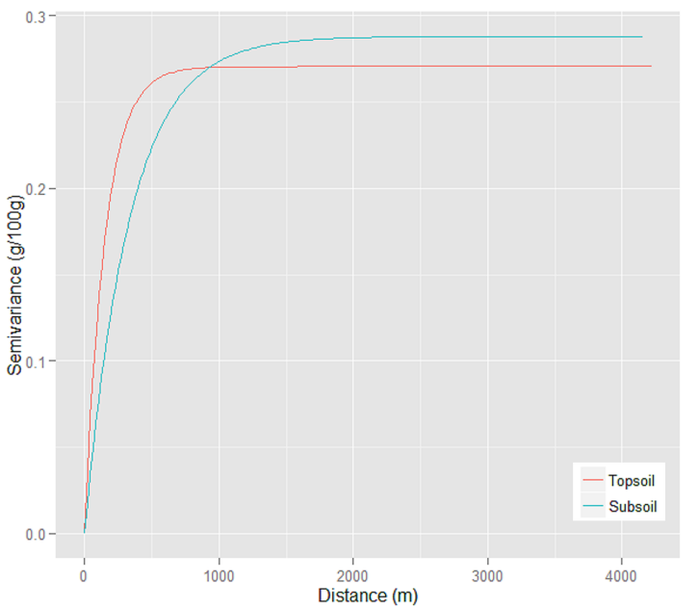

Fig. 2. Experimental variogram of squareroot transformed soil carbon content of topsoil and subsoil.

index were selected as most significant covariates to be used for all considered model types in this study.

Spatial Models

A range of spatial models were chosen to model the distribution of soil carbon. The MLR is the simplest form of spatial model which attempts to minimize the sum of squares to achieve the maximum prediction accuracy. It operates under the assumptions of normality, spatial linearity and spatial stationarity of predictor variables (Hastie et al., 2001).

The GWR is an extended form of MLR which addresses the spatial non-stationarity of predictor variables by allowing the coefficients to vary geographically instead of using global values as in the case of MLR (Brunsdon et al., 1998).

Unlike GWR and MLR, LMMs allow the response variable to have different distributions other than the normal distribution. The LMMs incorporate the spatial stochastic process unexplained by the deterministic trend of the predictor variables. Usually maximum likelihood (ML) or residual maximum likelihood (REML) techniques are used to infer the parameters of the spatial stochastic process (Lark and Cullis, 2004; Lark et al., 2006). The ML and REML approaches give unbiased robust estimates directly from the data unlike other methods, such as residual kriging, which separately model the stochastic process through the residuals (Minasny and McBratney, 2007).

Cubist handles nonlinearity between target and predictor variables via recursive partitioning of the spatial model into local models, which capture local linearity of predictor variables in different regions of the geographical space. Cubist is an ensemble of local models designed to deliver more accurate outputs with relatively lower uncertainty (Holmes et al., 1999; Wang and Witten, 1997).

Quantile regression forest (QRF) is a non-parametric technique used to estimate the conditional quantiles of multi-dimensional predictor variables; hence, QRF is able to estimate more accurate summaries of the conditional distribution of the response variable (Meinshausen, 2006).

Extreme learning machine regression (ELMR) (Huang et al., 2006) is a recently developed machine learning algorithm, which is popular due to its comparatively fast learning speed. It minimizes the training error through improved generalization which avoids perturbation and multi-linearity problems (Ding et al., 2014; Huang et al., 2006).

The following sections describe additional theoretical details about each of the model types considered in this study.

Multiple Linear Regression

The MLR is a relatively simple and frequently used model. It is assumed that the regression function E (Y\X) is linear, or the linear model is a reasonable approximation. The linear regression model can be expressed as:

$$S_{(x)} = \beta_0 + \sum pj = X_i \beta_j \quad [1]$$

Table 1. Environmental covariates with their sources, resolution, and definitions.

Covariate†	Source	Spatial scale	Definition
Landsat Bands 1,2,3,4,5, 7	Landsat 7-(2012)	30 m	Visible (reflected light) bands in the spectrum of blue, green, red, near-infrared (NIR), and mid-infrared (MIR)
Normalized difference vegetation index (NDVI)	Landsat 7	30 m	Ratio, (NIR–Red)/(NIR+Red) [(Band4–Band 3)/(Band4+Band 3)]
Landcover	NSW Dep. of Land and Property Information, Australia	25 m	Physical cover of the study site
Digital elevation model (DEM)	NSW Dep. of Land and Property Information, Australia	25 m	Digital elevation model of the area- representation of the terrain's surface
Plan curvature	DEM	25 m	Curvature types highlight different aspects of the shape or curvature of the slope. Plan curvature is perpendicular to the direction of the maximum slope & relates to the convergence and divergence of flow across a surface
Profile curvature	DEM	25 m	Parallel to the slope and indicates the direction of maximum slope. Affects the acceleration and deceleration of flow across the surface.
Topographic wetness index (TWI)	DEM	25 m	ln(a/tanβ), a = local upslope area draining through a certain point per unit contour length, tanβ = local slope (Sørensen et al., 2006). Known as the tendency of a grid cell in the DEM to accumulate water.
Altitude above channel network	DEM	25 m	Altitude for each grid cell of the DEM above the nearest streamline channel
Analytical hill-shading	DEM	25 m	Hypothetical illumination surface derived using azimuth and altitude of the sun. A relative measure of incident light for analysis.
Light insolation	NSW Dep. of Planning and Environment, 2016	25 m	Amount of solar radiation energy received by a surface area in a particular time period
Mid slope	DEM	25 m	Any position between the top and the bottom of the slope.
Slope direction	DEM	25 m	Direction of the slope
Aspect diretion	DEM	25 m	Slope facing direction
Terrain roughness index (TRI)	DEM	25 m	Amount of elevation difference between adjacent cells of a digital elevation grid
Catchment area	DEM	25 m	Area from which rainfall flows into a river

where β_0 is the intercept of the linear model, X_i represents the auxiliary variables or covariates, β_j represents the unknown coefficients for the auxiliary variables, and p is the number of auxiliary variables (Hastie et al., 2001). Regression methods explore a possible functional relationship between the target variable (soil carbon content) and explanatory variables (environmental covariates).

Geographically Weighted Regression

The GWR (Brunsdon et al., 1998) is an extended form of traditional regression and accounts for spatial non-stationarity by allowing model coefficients to vary spatially. The regression equation can be given as:

$$S_i = a_{i0} + \sum_{k=1}^{m} a_{ik} x_{ik} + \varepsilon_i \quad [2]$$

where S_i is the observation of the dependent variable at location i, a_{ik} is the value of the kth parameter at location i, and the error term, ε_i, is normally distributed with mean zero. For each location, independent local models are calibrated. The observations which are closer to location i have a greater impact in determining the parameter values. The impact is estimated via a weighting scheme using a kernel function. In this study the more commonly used Gaussian kernel function was used:

$$w_{ij} = \begin{cases} \exp(-\frac{d_{ij}}{b})^2, & \text{if } d_{ij} < b \\ 0, & \text{otherwise} \end{cases} \quad [3]$$

where d_{ij} is the distance between observations i and j, and the bandwidth b is used to exclude observations that exceed the distance threshold. The weighting of the data will gradually decrease exponentially as the distance between i and j increases; this is similar to an exponential variogram model. The GWR model will gradually reduce to an ordinary least squares (Hwang et al., 2011) model as the bandwidth increases, and will suffer from over-fitting if the bandwidth decreases to zero. The optimum, b, is computed using cross validation by minimizing the following:

$$\mathrm{CVSS}(b) = \sum_i \left\{ S_i - \hat{S}_i(b) \right\}^2 \quad [4]$$

where $\hat{S}_i(b)$ is the predicted value respective to the optimum band with b.

The GWR has issues of multi-colinearity caused by correlations among local regression coefficients. These correlations can occur between pairs of local regression coefficients at one location, or correlations among two sets of local regression coefficients (Wheeler and Tiefelsdorf, 2005).

The GWR has been widely used in DSM literature and the studies by Mishra et al. (2010); Song et al. (2016); Zeng et al. (2016) are examples for the use of GWR for soil carbon mapping.

Cubist Models

This is a variation of the regression tree model, where the prediction is based on linear regression models instead of discrete values at the terminal nodes. Cubist models spatial non-stationarity indirectly by producing a set of 'if–then' rules, where each rule has an associated multivariate linear model. Whenever a set of covariates matches a rule's conditions, the associated model is used to calculate the predicted value. The algorithm was first described by Quinlan (1992), then followed by extended descriptions from Wang and Witten (1997), and Holmes et al. (1999). Briefly, the Cubist algorithm builds a 'tree' by splitting the data based on the predictors so that it minimizes the intra-subset variation in the class (Holmes et al., 1999). The model associated with each rule is computed using linear least-squares regression. Finally, the linear model is adjusted and simplified to reduce absolute error. Cubist has been used effectively in various soil prediction and digital mapping procedures (Bui et al., 2009; Henderson et al., 2005; Kidd et al., 2014; Minasny et al., 2008; Viscarra Rossel et al., 2014).

Quantile Regression Forests

While the prediction of parametric models is an estimate of the conditional mean of the response variable, the QRF predicts the quantiles which form a more complete summary of the conditional distribution of the response variable. Thus, the QRF (Meinshausen, 2006) is a non-parametric multivariate regression method which builds on the Random Forest decision tree ensembles (Breiman, 2001). Similar to Cubist, the regression trees are constructed by recursive partitioning. Nevertheless, Random Forest is a modified version of bootstrapped trees where the predictions are the results of tree ensembles. Each tree is grown on a random subset of training data. The main difference between QRF and Random Forest is, for each node in a tree, only the mean value of the observations is preserved in Random Forest while QRF keeps all values that fall into the node to assess the conditional distribution. The model prediction, i.e., the conditional distribution, is estimated by the weighted distribution of observed response variables. This method has been used by Rudiyanto et al. (2016) to map peat thickness in Indonesia.

Linear Mixed Models

The spatial model is composed of fixed and random components. The fixed effects, or deterministic trend, u(s), describes the spatial variation of the soil carbon explained by the input covariates. The random component (stochastic residuals), u(s) + e(s), is the unexplained spatial variation (Cressie, 1991).

$$S(x) = \sum_{j=0}^{p} \beta_j X_j + \mathrm{u} + \varepsilon(x) \quad [5]$$

The term e represents both independent measurement errors and microscale variation. This is geostatistically termed as the nugget effect of the spatial variogram. Cressie (1991) explained that the nugget effect of the spatial variogram is made up of two non-negative components, σ_0^2 (microscale variance of the actual value of a variable) and σ_ε^2 of the observed value, u is a spatially dependent second order random process with the variance of $\xi\sigma^2$ which can be estimated via a suitable covariance function.

The current study used residual maximum likelihood (REML), which is an optimization algorithm, to derive unbiased

model parameters directly from the data. Linear mixed models have been used in the DSM studies of Rawlins et al. (2009) and Karunaratne et al. (2014) to analyze the sampling error of soil properties and to map soil organic carbon fractions, respectively.

Extreme Learning Machine Regression

Extreme learning machine methods are feed forward neural networks designed for classification or regression with single layer of hidden nodes. The ELMR adopts a tuning-free strategy for feed forward neural networks. The ELMR is flexible with hidden activation functions and allows the use of nonlinear piecewise continuous functions and their linear combinations. Therefore, ELMR has superior fast learning speed and better generalization than other comparable algorithms such as support vector machines (or SVM) and its variants (Huang et al., 2006).

The relationship between input and output of single hidden layer feed-forward neural network (or SLFN) systems is given by:

$$\sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g\left(\mathbf{w}_i \cdot \mathbf{x}_i + b_i\right) = t_j, \text{ for } j = 1, \ldots, N \qquad [6]$$

where $\mathbf{w}_i$ is the weight vector between the ith neuron in the hidden layer and the input layer; b_i, is the bias of the ith neuron in the hidden layer; $\mathbf{x}_i$ is the jth input data vector; $g(.)$ is an active function of the hidden neuron; $\boldsymbol{\beta}_i$ is the weight vector between the ith hidden neuron and the output layer, $\tilde{N}$; $\tilde{N}$ is the number of hidden nodes; and N is number of training samples. The above equation can also be written as:

$$\mathbf{H}\boldsymbol{\beta} = T \qquad [7]$$

where $\mathbf{H}$ is the hidden layer output matrix of the network.

The ELMR approach has two operational phases: an initialization phase and the sequential learning phase. In the initialization phase, the values of $\mathbf{w}_i$ and b_i are not tuned during training. Random values are assigned for $\mathbf{w}_i$ and b_i according to any continuous sampling distribution. This information is then passed to the learning phase where the weight matrix of hidden-to-output ($\boldsymbol{\beta}_i$) is estimated using following equation:

$$\boldsymbol{\beta} = \mathbf{H}^{-1} T \qquad [8]$$

where $\mathbf{H}^{-1}$ is the Moore–Penrose generalized inverse of the hidden layer output matrix $\mathbf{H}$. Huang et al. (2006) and Huang et al. (2015) provide more detailed descriptions of ELMR.

Neural networks have been used routinely in DSM (e.g., Rudiyanto et al., 2016), however only a few studies have used ELMR for soil carbon prediction with a recent study from Masri et al. (2015).

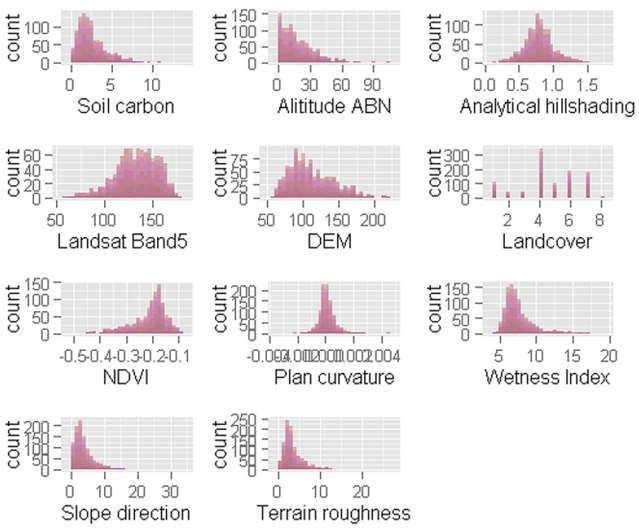

Fig. 3. Histograms of covariates plotted for different training sample sizes.

Training and Validation of Spatial Models

Model Training and Testing

To test how the different spatial modeling techniques performed under varying training sample sizes, the dataset for each soil layer was randomly split into training and testing sets. In this study we used a 70:30 training/testing data split. The training set was used to model the site specific relationship between the soil carbon and the environmental predictors. The held-out test set was used to assess the goodness of fit of the models using accuracy indicators. The training set was further subdivided into different sample sizes. Each sample size had 10 realizations of repeated sampling. Each realization was tested with the same test set using accuracy indicators. The training sample sizes were 100, 200, 300, 400, 500, 700, 800, 900, and 1000 for the topsoil, and 100, 200, 300, 400, 500, and 700 for the subsoil. The sample sizes for soil layers differed as there were fewer subsoil samples available compared to the number topsoil samples. Accordingly, each modeling algorithm MLR, LMM, Cubist, QRF, GWR, and ELMR were trained and tested for all sample sizes across the two soil layers. Figure 3 presents the distributions of model covariates for all sample sizes for the topsoil. The distributions of a covariate for all sample sizes have been plotted on the same plot space. According to these histograms, there is no clear deviation between the distributions of a covariate between the sample sizes for both soil layers.

Prediction Accuracy of Models under Progressive Sample Sizes

Goodness-of-fit statistics used in this study included the root mean squared error (RMSE), Lin's concordance correlation coefficient (CCC) (Lin, 1989), standardized squared deviation (SSD), and prediction variance. These were calculated for each simulation of the spatial models with respect to the different sample sizes and model type.

The RMSE is a measure of prediction accuracy of the model. Lower RMSE values indicate higher prediction accuracy. The CCC evaluates the fidelity to which observed and predicted pairs fall on the 45° line when they are plotted against each other. It is a determinant of both accuracy and precision of the predictions. The CCC is calculated using following equation:

$$\text{CCC} = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + \left(\mu_x - \mu_y\right)^2} \qquad [9]$$

where r is the correlation coefficient, σ_x and σ_y are the variances of observed (x) and predicted (y) values, and μ_x and μ_y are the respective means. The CCC is scaled between -1 and 1, with the latter implying perfect agreement and the former implying perfect reverse agreement.

The SSD(x) measures the prediction model goodness of fit:

$$\mathrm{SSD}(x)=\frac{\left\{S(x)-\hat{S}(x)\right\}^2}{\sigma_x^2} \qquad [10]$$

where $S(x)$ is the measured value, $\hat{S}(x)$ denotes the predicted value with prediction variance σ_x^2. A value closer to 1 for mean SSD(x) indicates a good estimate (Voltz and Webster, 1990) and a median value closer to 0.455 (Lark, 2000) symbolizes kriging with a correct variogram.

Estimations of Prediction Uncertainty

Prediction uncertainty is defined as the variability of model predictions. The uncertainty of model predictions are caused by uncertain model parameters and inputs or approximate and/or incomplete treatment of the spatial relationship of the process being modeled (McKay, 1995). For example this study assumed a linear relationship between soil carbon and the environmental covariates in MLR, however, it can be nonlinear. The stochastic variability of the processes can also contribute to the uncertainty of soil carbon predictions. Hence the uncertainty of spatial model predictions is comprised of three major types of uncertainties; input uncertainty, structural uncertainty and parameter uncertainty (McBratney et al., 2002).

Prediction variance is considered as a measure of uncertainty of model predictions caused by the uncertainty of input parameters of the model. We calculated the variance of the model predictions to compare the reliability of the model predictions in terms of the type of model and training sample sizes. The prediction variance is given by:

$$\sigma_x^2=\left(\frac{1}{n}\right)\left(\sum_{i=1}^{n}\left\{\hat{S}(x)-\hat{\mu}(x)\right\}^2\right) \qquad [11]$$

where $\hat{\mu}(x)$ is the mean of predicted values.

The GWR and LMM prediction functions (in relevant R statistical software packages) are embedded with options to calculate the prediction variance. The MLR and QRF prediction functions allow calculating standard error and/or standard deviation of predictions where the prediction variance is the squared error of the standard error. For Cubist, ELMR and MLR models, the prediction variance was calculated using the repeated sampling for each size of the training samples. For Cubist and ELMR models the prediction variance was calculated using Eq. [11] for repeated samples where n is the number of repeats. For GWR, LMM, MLR, and QRF, the prediction variance is the average of prediction variances of the repeats.

Model Ensembles

Model ensembling is the combining of the predictions of multiple learning algorithms. The resulting ensemble is generally more accurate than any of the individual algorithms within the ensemble (Opitz and Maclin, 1999). Therefore, model ensembles could be a useful technique to enhance the accuracy of soil carbon predictions by combining spatial predictions of several models.

Bagging, boosting, and stacking are commonly used methods for ensembling (Opitz and Maclin, 1999). In this study we used stacking, where the predictions from the MLR, GWR, LMM, Cubist, QRF and ELMR models were combined using different weighting for each algorithm. Weighting was based on the validation R^2 values where models with higher R^2 were given a higher relative weight than the models with lower R^2 when combining model predictions to form the stack. Accordingly, we tested the performance of all possible combinations of the above algorithms using the accuracy indices described previously with regard to prediction accuracy under progressive sample sizes. The stacks consisted of 6, 5, 4, 3, and 2 layers of model predictions in all possible combination of the models being tested.

Spatial Prediction of Soil Carbon

The ultimate objective of the model training was to use the trained model to predict the carbon content at un-sampled location within the study area. Therefore, to compare the effect of sample sizes and models, the spatial distribution of soil carbon in the study area was predicted using LMM and GWR models for three selected sample sizes: 300, 500, and 1000 for the topsoil. The LMM and GWR models were selected for mapping as they gave the highest prediction accuracy. The trained models for each sample size were used to predict carbon content to the same spatial resolution of the environmental covariates (25- by 25-m cell resolution). Additionally, we also mapped the prediction variance of each model to assess uncertainty of the model predictions with respect to the training sample sizes and model type.

RESULTS

Covariates Selection for the Spatial Model

Pearson correlation analysis showed that altitude above channel network terrain roughness index, and elevation covariates had positive correlation whereas NDVI had negative correlation with soil carbon for the topsoil (Fig. 4). Although the other selected covariates had a relatively smaller correlation with soil carbon, they were still influential covariates in the regression analysis as the probability values for these covariates were less than the critical value (0.05). Therefore, they were included in the spatial model. Plan curvature, slope direction, TWI, altitude above channel network and digital elevation model (DEM) exhibited a strong positive correlation with soil carbon in the subsoil layer whereas TWI displayed a negative correlation.

Comparison of Performance Indicators across Models and Training Sample Sizes

Root Mean Squared Error

Each model for specified soil layers had ten replicates of validation statistics for each training sample size. We have described the variations in RMSE values of replicates for validation

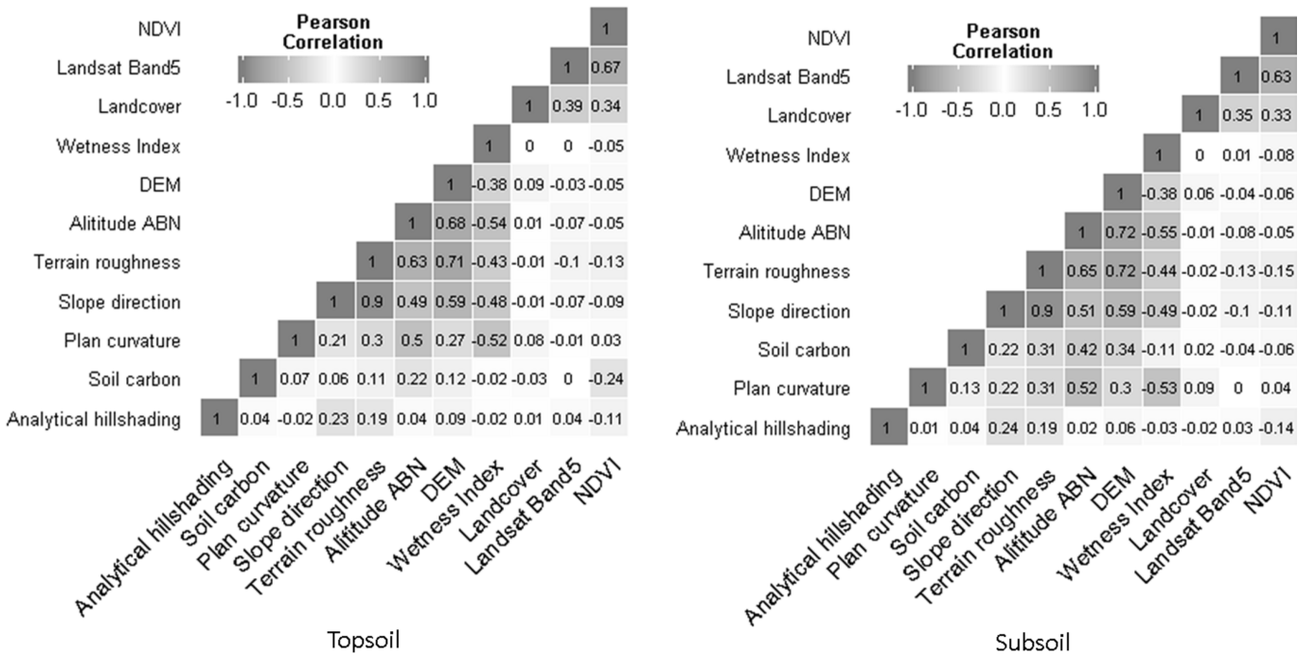

Fig. 4. Correlation among the soil carbon and environmental covariates for topsoil and subsoil layers.

using the boxplots (Fig. 5). For most of the models the variation in RMSE between replicates was high at small training sample sizes. At small sample sizes (<300) this variation was around 5 to 10% while the difference for higher training sample sizes was negligible for both soil layers.

For topsoil, average RMSE values of MLR ranged between 0.6 and 0.46 between the 100 and 1000 sample sizes. Accordingly, there was a 14% improvement in the accuracy of predictions for MLR with increasing training sample size. The respective RMSE values for GWR, LMM, QRF, Cubist and ELMR were 0.5–0.48, 0.50–0.46, 0.49–0.44, 0.50–0.46 and 0.41–0.37. The percentage accuracy improvements for the respective models were 2, 4, 5, 4, and 3%. In addition, the RMSE values for each sample size for all models except ELMR had very

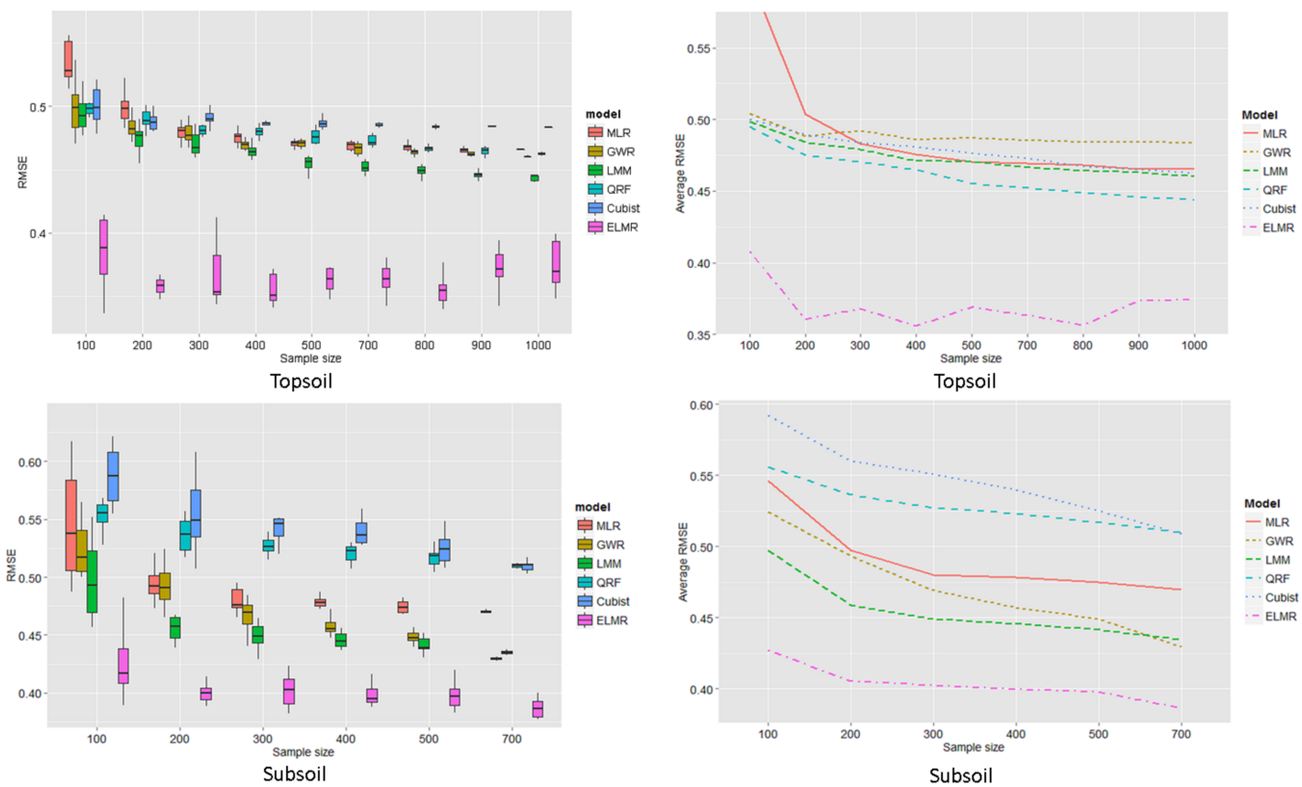

Fig. 5. Boxplot of RMSE values on the validation data (left), and learning curves (right).

similar values. ELMR had approximately 10% lower RMSE values for each sample size compared to the other models.

For the subsoil, average RMSE values for MLR improved from 0.55 to 0.47 when the training sample size increased from 100 to 700. This indicates an 8% improvement in prediction accuracy for MLR. The respective values for GWR, LMM, QRF, Cubist, and ELMR were 0.52-0.43, 0.50–0.43, 0.56–0.51, 0.59–0.51 and 0.43–0.39. Therefore, the accuracy enhancement between the lowest and highest sample sizes for these models are 9, 7, 5, 8 and 4%, respectively. The RMSE values between models for each sample size were very similar with the exception of ELMR. In the topsoil layer ELMR had an approximately 10% lower RMSE value for all sample sizes compared to the other models.

Learning Curves

Learning curves can be defined as a measure of predictive performance on a given domain as a function of some measure of varying amounts of learning effort. The most common form of learning curves shows predictive accuracy on the test data set as a function of the number of training samples (Perlich, 2010).

The learning curves for each model for both soil layers were computed using the average RMSE values of replicates of each training sample size (Fig. 5). According to this figure, the learning curves had two clusters. One cluster was comprised of ELMR and all other models were in the other cluster. This shows that ELMR performed very differently to the other models. All other models converged into steady state after a certain sample size while ELMR convergence still fluctuated with samples sizes nearing and equal to the maximum sample size.

Concordance Correlation Coefficient

Figure 6 presents the average validation CCC and associated standard error of replicates for each sample size for all described modeling scenarios for both soil layers. For the topsoil, the average CCC of MLR improved from 0.24 to 0.36 when the training sample sizes increased from 100 to 1000. The respective statistics for the other models, Cubist, GWR, LMM, QRF and ELMR were 0.22–0.27, 0.28–0.41, 0.31–0.43, 0.18–0.35, and 0.17–0.19. Therefore, prediction accuracy and precision gains of the models for the increasing training sample sizes were 12, 5, 13, 12, 17, and 2% for MLR, Cubist, GWR, LMM, QRF and ELMR, respectively. For each sample size, the average CCC values increased in an ascending order according to the manner; ELMR < QRF < Cubist < MLR < GWR < LMM. Accordingly, ELMR had the lowest average CCC while LMM had the highest for each training sample size. The differences between the lowest and highest CCC values between these two models were 14, 18, 17, 15, 19, 22, 21, 26, and 22% for the sample sizes 100, 200, 300, 400, 500, 700, 800, 900, and 1000, respectively.

For subsoil the average CCC values for all models and for all sample sizes were higher than the topsoil. Nevertheless, the patterns between the sample sizes and the models were similar to the topsoil. ELMR had the lowest and LMM has the highest CCC values while the order of accuracy gain of the models followed that of the topsoil. The differences between the lowest and highest CCC values between these two models were 34, 35, 38, 35, 35, and 31%, respectively, for the sample sizes 100, 200, 300, 400, 500, and 700.

Standardized Squared Deviation

Bias, mean SSD, median SSD and the prediction variances of the model validations for both soil layers respective to the model type and sample sizes are given in Table 2. The ELMR had the highest biased predictions while LMM had the lowest biased predictions for both soil layers. There was no clear difference between the bias values of different sample sizes of each model except for Cubist and ELMR models of the subsoil.

A mean SSD of 1 indicates an accurate prediction. The mean SSD values were closer to 1 for GWR, LMM and QRF for both soil layers (ranges 0.85–1.13 for the topsoil, and 0.0.85–1.17 for the subsoil). The MLR, Cubist and ELMR had values further from 1 for both soil layers. The values ranged between 0.63 to 1.33 for the topsoil and from 0.37 to 0.75 for the subsoil. Median SSD values followed a similar pattern for soil layers, models and sample sizes. Accordingly, GWR, LMM and QRF had values closer to best estimate of 0.455. It is important to note that there were no clear differences of mean and median SSD values for varying training sample sizes of each model.

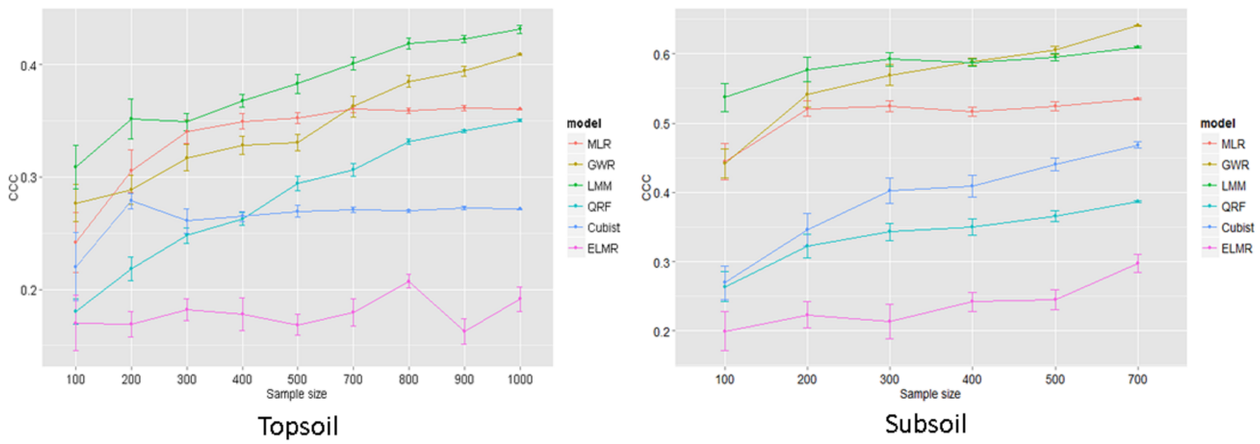

Fig. 6. The concordance correlation coefficient values and their standard errors of the models for topsoil and subsoil layers.

Prediction Variance

The prediction variances were quite similar between sample sizes of each model for both soil layers. The ELMR had the lowest prediction uncertainty while Cubist had the highest prediction variance for topsoil for all sample sizes. For the subsoil, MLR had the highest prediction uncertainty while LMM and GWR displayed a similar low level of prediction uncertainty. There were no noticeable differences between the prediction variances among the sample sizes for all model types.

Model Ensembles

Among the tested ensembles, neither model stacks gave a significant improvement in the accuracy. The highest accuracy gain was 2% delivered by an ensemble of LMM and GWR modeling algorithms.

Spatial Prediction of Soil Carbon

Figures 7 and 8 demonstrate the LMM and GWR predicted spatial distribution of topsoil carbon content over the study area. Overall the spatial patterns of the soil carbon content were consistent across the models and sample sizes. The northern part of the study area had lower carbon content and it gradually increased toward the south. The prediction variance decreased as the training sample size increased for both modeling scenarios.

DISCUSSION

Model Type and Training Sample Size

This study examined how the accuracy of soil carbon predictions depends on the model type and training sample size. The RMSE values across the models for any given sample size were very similar. For the chosen study site it appears that the models have a similar prediction accuracy. Prediction accuracy increases with the increasing sampling size for all models, which implies that, for this study site, soil carbon prediction accuracy is more sensitive to training sample size than the type of spatial model. ELMR is an exception here as it scored the lowest prediction error across all sample sizes, while its prediction accuracy is less sensitive to the training sample size.

When the RMSE values between models are compared, LMM predictions were comparatively more accurate than MLR, GWR, QRF and Cubist model predictions. The LMM is the only model that considers the spatial auto-correlation of the

Table 2. Bias, mean standardized squared deviation (SSD), median SSD, and prediction percentage, falling within 95% confidence interval limits of each model for the topsoil (0-10 cm) and subsoil (40-50 cm) layers.

	0–10 cm									40–50 cm					
	Sample size									Sample size					
Model†	100	200	300	400	500	700	800	900	1000	100	200	300	400	500	700
							Bias								
MLR	-0.03	-0.01	0.01	0	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.03
GWR	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.04	0.02	0.01	0.01	0.01	0
LMM	0.01	0.01	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.03
QRF	-0.01	0.01	0.01	0.01	0	0	0	0	0	0.04	0.03	0.02	0.03	0.02	0.02
Cubist	0.03	0.04	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.08	0.08	0.08	0.01	0.01	0.01
ELMR	0.15	0.19	0.14	0.14	0.15	0.17	0.16	0.18	0.20	0.11	0.05	0.03	0.02	0.01	0.01
							Mean SSD								
MLR	1.33	0.84	0.72	0.74	0.75	0.69	0.69	0.7	0.7	0.59	0.38	0.37	0.39	0.42	0.37
GWR	0.97	0.93	0.89	0.86	0.86	0.84	0.87	0.87	0.86	1.02	0.95	0.92	0.91	0.92	0.85
LMM	1.13	0.91	0.93	0.87	0.86	0.87	0.88	0.85	0.85	1.17	0.92	0.98	0.99	1.02	1
QRF	0.97	1.07	0.96	0.96	0.94	0.95	0.94	0.94	0.94	1.17	1.09	0.98	1	0.95	0.92
Cubist	0.82	0.69	0.67	0.68	0.67	0.66	0.67	0.66	0.67	0.69	0.64	0.63	0.66	0.68	0.56
ELMR	0.84	0.66	0.71	0.69	0.65	0.63	0.66	0.66	0.64	0.75	0.71	0.66	0.58	0.6	0.46
							Median SSD								
MLR	0.27	0.25	0.24	0.24	0.24	0.23	0.23	0.23	0.23	0.22	0.14	0.14	0.15	0.15	0.13
GWR	0.32	0.32	0.31	0.3	0.3	0.3	0.3	0.31	0.32	0.41	0.38	0.36	0.38	0.35	0.35
LMM	0.41	0.34	0.35	0.31	0.3	0.32	0.31	0.3	0.29	0.44	0.36	0.37	0.38	0.36	0.36
QRF	0.34	0.34	0.34	0.33	0.33	0.33	0.33	0.33	0.33	0.51	0.53	0.48	0.49	0.49	0.48
Cubist	0.25	0.24	0.24	0.24	0.23	0.23	0.23	0.23	0.23	0.22	0.2	0.21	0.21	0.22	0.18
ELMR	0.27	0.25	0.25	0.25	0.24	0.25	0.25	0.25	0.25	0.26	0.27	0.27	0.23	0.25	0.2
							Prediction Variance								
MLR	0.29	0.31	0.33	0.31	0.3	0.32	0.32	0.31	0.31	0.62	0.79	0.65	0.6	0.54	0.6
GWR	0.26	0.26	0.26	0.26	0.26	0.26	0.25	0.25	0.24	0.49	0.48	0.51	0.5	0.49	0.49
LMM	0.23	0.25	0.24	0.25	0.24	0.23	0.23	0.23	0.23	0.22	0.23	0.20	0.20	0.19	0.19
QRF	0.27	0.27	0.26	0.26	0.26	0.26	0.25	0.25	0.25	0.28	0.26	0.28	0.27	0.27	0.27
Cubist	0.31	0.34	0.36	0.35	0.35	0.36	0.35	0.35	0.35	0.49	0.48	0.51	0.5	0.49	0.49
ELMR	0.2	0.2	0.17	0.18	0.17	0.18	0.18	0.19	0.19	0.23	0.21	0.21	0.21	0.21	0.21

† MLR, multiple linear regression; GWR, geographically weighted regression; LMM, linear mixed model; QRF, quantile regression forest; ELMR, extreme learning machine regression.

predictor variable while other models only account for the deterministic component. Spatial autocorrelation appears to be an essential component in predicting soil carbon content for this particular study site.

Learning curves provide a clear picture of a model's behavior with increasing sample sizes. When the sample size is increased, the model prediction accuracy increased at a decreasing rate. Learning curve of each modeling algorithm fully converged at different sample sizes. This indicates that the number of samples required for the optimum performance depends on the model type. For example, MLR and GWR reached their steady state of convergence very early while other models were still converging. In general, this particular study site required more than 15 samples for a square kilometer to achieve an optimum predictive performance. It is noteworthy that the same model fully converges at a different sample size for the top and subsoil layers. For example, GWR fully converged at 700 data points for the topsoil but for the subsoil, the full convergence was not reached with this sample size.

According to the accuracy and reproducibility of predictions given by CCC values, LMM, GWR, and MLR delivered more accurate and precise predictions than the QRF, Cubist, ELMR models. These statistics suggest that accuracy and repeatability of soil carbon predictions depend on the model type (Somarathna et al., 2016). Although ELMR had the lowest CCC, it also had the lowest RMSE, most likely due to the fact that ELMR predictions were heavily biased compared to the other models (Table 2). The CCC values also depend on the training sample size; however, the differences of CCC between the models were greater than the gain of CCC between progressive sample sizes. Therefore, reproducibility or precision of soil carbon predictions for the study site mostly depended on the model type than the training sample size.

The SSD values are also accuracy indicators of predictions. Mean SSD values (Table 2) revealed that for both soil layers, GWR, LMM and QRF generated equally good estimates for all sample sizes and MLR, Cubist, and ELMR resulted in comparatively poorer estimates. The SSD values decreased with increasing training sample size. These values suggest that the prediction variance increases with the increasing number of training points. This may have been caused by increasing number of outliers (ob-

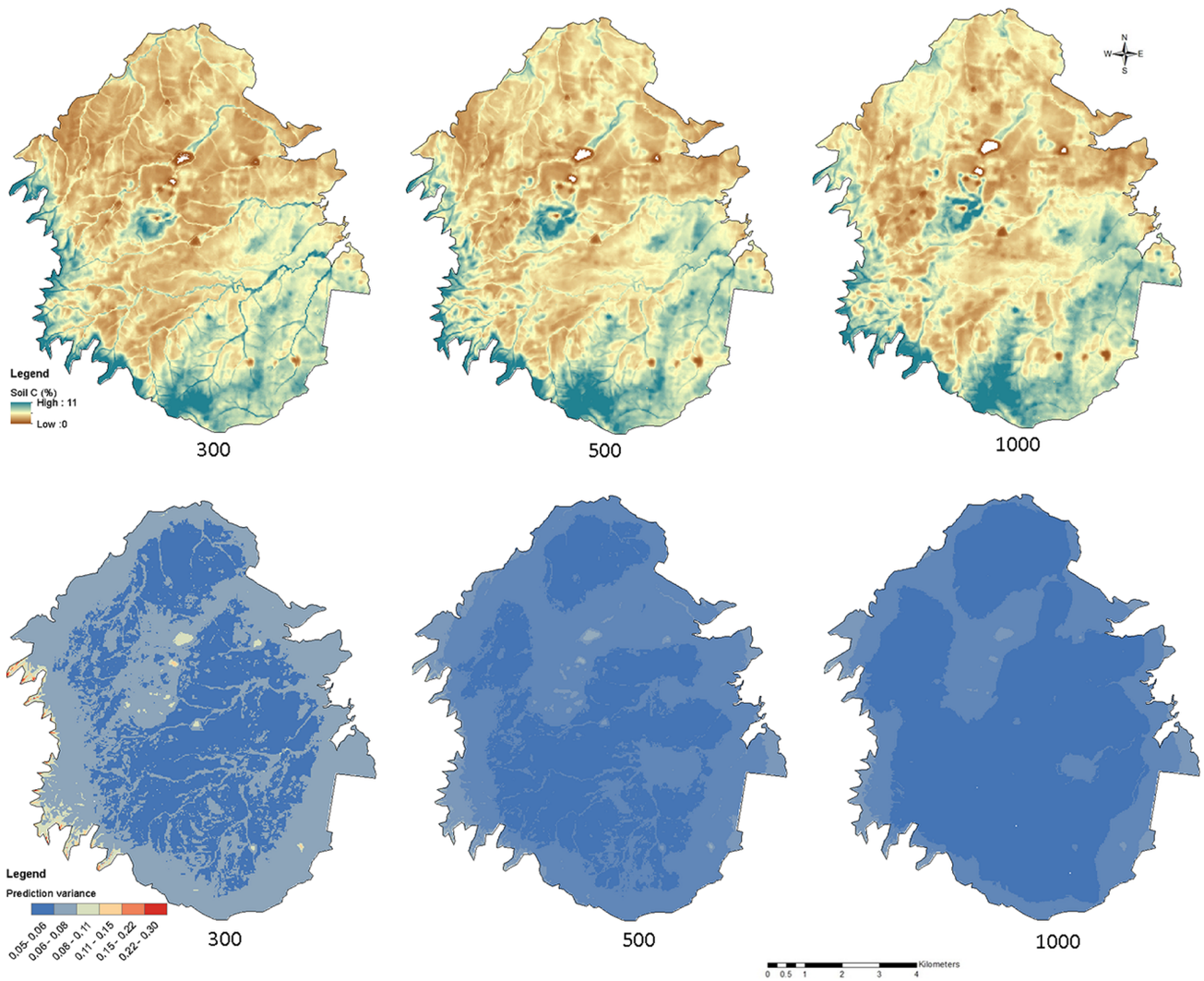

Fig. 7. LMM predicted carbon content (%) and the associated prediction variance (bottom) for the sample sizes 300, 500, and 1000 for topsoil.

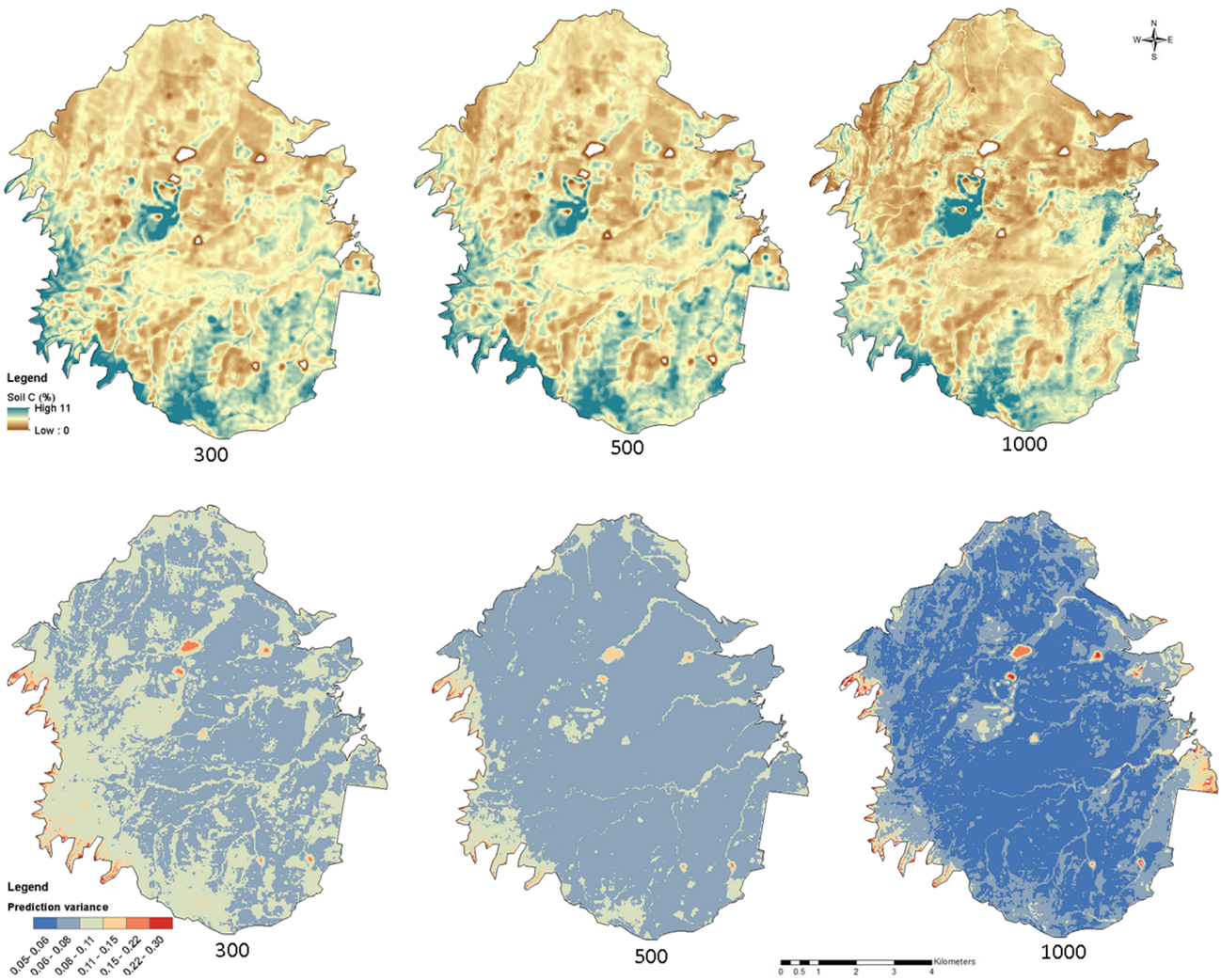

Fig. 8. GWR predicted carbon content (%) and the associated prediction variance (bottom) for the sample sizes 300, 500, and 1000 for topsoil.

servations that lies an abnormal distance from other values in a random sample from a population) as the number of samples grew. The median values of the SSD were much lower than 0.455 for all models and sample sizes. This indicates all models tended to overestimate the predictions. The calculated bias values also suggest that all tested models in this study had a propensity to overestimate. The comparison of SSD values between the models and the training sample sizes revealed that the differences between the models were relatively higher than the differences between sample sizes. This emphasizes that some models (GWR, LMM, QRF) generated more accurate estimates of soil carbon than other models (MLR, ELMR).

Prediction variance is an indication of uncertainty of predictions. According to the results, the degree of uncertainty of model predictions was more or less similar across the training sample sizes, however there were more noticeable differences of prediction variances between the models, especially for the subsoil. For example, LMM, QRF and ELMR delivered more certainty in predictions than MLR, Cubist and GWR. Therefore, the uncertainty of prediction most likely depends on the model type rather than the sample size.

Despite the use of a variety of techniques, the accuracy of spatial soil carbon predictions always remains low (the maximum CCC of model prediction is 0.6). This confirms the observations made by (Heuvelink and Webster, 2001) that it is impossible to completely capture the local variability of soil carbon through a deterministic model. LMM and GWR delivered more accurate predictions. LMM capture the spatial autocorrelation of soil carbon through estimating the experimental variogram directly from the data. Hence, spatial auto-correlation is also an important factor to be considered in predicting soil carbon content. This correlation was more significant for subsoil carbon of the study site. This may have caused by the abundance of marl in the subsoil layer, which occurs through the study area and HWCPID in general. Similarly, GWR treats the covariates as spatially non-stationary and so it is capable of capturing most of the non-stationarity of the deterministic trend of the spatial model.

The current study suggests that the training sample size had a substantial effect on prediction accuracy of the model. Regardless of model type, a comparatively higher degree of prediction accuracy can be achieved with large sample sizes using any spatial model. Learning curves suggested that all models require at least

300 training samples to achieve a state closer to the optimum accuracy. Overall, the combination of sufficient sample size and the right spatial modeling techniques designed for capturing random non-stationarity will deliver more accurate soil carbon predictions.

Do Model Ensembles Improve Accuracy?

The maximum 2% accuracy gain of model ensembles demonstrated that even combining algorithms could not deliver a clear improvement in soil carbon predictions for this study site. It suggests the limitations of capturing the local variability of soil carbon through deterministic models; thus performing ensembles of different machine learning models is not recommended. Ensembles can only prove beneficial if different methods of mapping or datasets were combined (e.g., Dobarco et al., 2017; Malone et al., 2014).

Comparing the Spatial Predictions of Soil carbon

According to Fig. 7, it is evident that the LMM model had the tendency to overestimate the topsoil carbon content with lower training sample sizes. Accordingly, the uncertainty of the model predictions decreased with increasing training sample size. Although we did not observe a distinct variation of prediction variance across sample sizes in the validation procedure, it is clearly visible in the maps. This may be due to validation test points and training points that are closely located, whereas in the maps the predictions are mostly on un-sampled locations.

The GWR model tended to overestimate when the model was trained with fewer samples (Fig. 8) and in turn the uncertainties of the model predictions decreased. When the predictions of GWR and LMM were compared, LMM predictions were found to be more reliable as LMM had a lower prediction variance than GWR.

CONCLUSIONS

For this study site, the accuracy of spatial prediction of soil carbon is less likely to depend on the model type used, yet the training sample size has a clear effect on model prediction accuracy. The difference between model realizations become insignificant when the models are trained with comparatively larger sample sizes. The models such as MLR, LMM and GWR seem to deliver more precise predictions than the other considered models. The prediction accuracy increased at a decreasing rate with increasing sample sizes. Most of the models require a minimum of 15 samples per square kilometer to reach their maximum predictive capability for this particular study site.

ACKNOWLEDGMENTS

The authors are grateful to the Australian Dep. of Agriculture, Round 2, Filling the Research Gap Program for supporting this research (Grant No. 1194105-66). A special thank is extended to the two anonymous reviewers and associate editor of Soil Science Society of America Journal whose constructive comments helped to improve the manuscript.

REFERENCES

Blanco-Canqui, H., C.A. Shapiro, C.S. Wortmann, R.A. Drijber, M. Mamo, T.M. Shaver, and R.B. Ferguson. 2013. Soil organic carbon: The value to soil properties. J. Soil Water Conserv. 68(5):129A–134A. doi:10.2489/jswc.68.5.129A

Breiman, L. 2001. Random forests. Mach. Learn. 45(1):5–32. doi:10.1023/A:1010933404324

Brunsdon, C., S. Fotheringham, and M. Charlton. 1998. Geographically weighted regression- modelling spatial non-stationarity. Statistician 47:431–443.

Bui, E., B. Henderson, and K. Viergever. 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. Global Biogeochem. Cycles 23(4):GB4033.

Cressie, N.A.C. 1991. Statistics for spatial data. John Wiley & Sons, New York.

Ding, S., X. Xu, and R. Nie. 2014. Extreme learning machine and its applications. Neural Comput. Appl. 25(3-4):549–556. doi:10.1007/s00521-013-1522-8

Dobarco, M.R., D. Arrouays, P. Lagacherie, R. Ciampalini, and N.P. Saby. 2017. Prediction of topsoil texture for Region Centre (France) applying model ensemble methods. Geoderma 298:67–77.

Frey, L.J., and D.H. Fisher. 1999. Modeling decision tree performance with the power law. Artificial Intelligence and Statistics 99, Proceedings of the International Conference on Artificial Intelligence and Statistics. January 1999. Fort Lauderdale, FL.

Geeves, G.W., H.P. Cresswell, B.W. Murphy, P.E. Gessler, C.J. Chartres, I.P. Little, and G.M. Bowman. 1995. The physical, chemical and morphological properties of soils in the wheat-belt of southern NSW and northern Victoria. NSW Dep. of Conservation and Land Management, and CSIRO Division of Soils Occasional Report. CSIRO Division of Soils, Glen Osmond, S. Aust.

Genuer, R., J.-M. Poggi, and C. Tuleau-Malot. 2010. Variable selection using random forests. Pattern Recognit. Lett. 31(14):2225–2236. doi:10.1016/j.patrec.2010.03.014

Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152(3-4):195–207. doi:10.1016/j.geoderma.2009.06.003

Hastie, T., R. Tibshirani, and J.H. Friedman. 2001. The elements of statistical learning: Data mining, inference, and prediction. Springer, New York. doi:10.1007/978-0-387-21606-5

Hawley, S., R. Glen, and C. Baker. 1995. Newcastle coalfield regional geology sheet 1:100000. Geological Survey of New South Wales. New South Wales Dep. of Mineral Resources, Sydney.

Henderson, B.L., E.N. Bui, C.J. Moran, and D.A.P. Simon. 2005. Australia-wide predictions of soil properties using, decision trees. Geoderma 124(3-4):383–398. doi:10.1016/j.geoderma.2004.06.007

Heuvelink, G.B.M., and R. Webster. 2001. Modelling soil variation: Past, present, and future. Geoderma 100(3–4):269–301. doi:10.1016/S0016-7061(01)00025-8

Holmes, G., M. Hall, and E. Frank. 1999. Generating rule sets from model trees. In: N. Foo, editor, Advanced topics in artificial intelligence. AI 1999. Lecture Notes in Computer Science, vol 1747. Springer-Verlag, Berlin. p. 1–12. doi:10.1007/3-540-46695-9_1

Huang, G.-B., Q.-Y. Zhu, and C.-K. Siew. 2006. Extreme learning machine: Theory and applications. Neurocomputing 70(1-3):489–501. doi:10.1016/j.neucom.2005.12.126

Huang, G., G.-B. Huang, S. Song, and K. You. 2015. Trends in extreme learning machines: A review. Neural Netw. 61:32–48. doi:10.1016/j.neunet.2014.10.001

Hwang, T., C.H. Song, P.V. Bolstad, and L.E. Band. 2011. Downscaling real-time vegetation dynamics by fusing multi-temporal MODIS and Landsat NDVI in topographically complex terrain. Remote Sens. Environ. 115(10):2499–2512. doi:10.1016/j.rse.2011.05.010

John, G.H., and P. Langley. 1996. Static versus dynamic sampling for data mining. IN: KDD-96 Proceedings, Association for the Advancement of Artificial Intelligence, Palo Alto, CA. p. 367–370.

Karunaratne, S.B., T.F.A. Bishop, J.A. Baldock, and I.O.A. Odeh. 2014. Catchment scale mapping of measureable soil organic carbon fractions. Geoderma 219-220:14–23. doi:10.1016/j.geoderma.2013.12.005

Kelley, K. 2007. Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. Behav. Res. Methods 39(4):755–766. doi:10.3758/BF03192966

Kelley, K., and S.E. Maxwell. 2003. Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. Psychol. Methods 8(3):305–321. doi:10.1037/1082-989X.8.3.305

Kidd, D.B., B.P. Malone, A.B. McBratney, B. Minasny, and M.A. Webb. 2014. Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. Soil Res. 52(2):107–119. doi:10.1071/SR13100

Kozak, A. 1997. Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. Can. J. Forest Res. 27(5):619–629. doi:10.1139/x97-011

Lark, R.M. 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. Eur. J. Soil Sci. 51(4):717–728. doi:10.1046/j.1365-2389.2000.00345.x

Lark, R.M., and B.R. Cullis. 2004. Model-based analysis using REML for inference from systematically sampled data on soil. Eur. J. Soil Sci. 55(4):799–813. doi:10.1111/j.1365-2389.2004.00637.x

Lark, R.M., B.R. Cullis, and S.J. Welham. 2006. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. Eur. J. Soil Sci. 57(6):787–799.

Lin, L.I. 1989. A concordance correlation-coefficient to evaluate reproducibility. Biometrics 45(1):255–268. doi:10.2307/2532051

Malone, B.P., J.J. de Gruijter, A.B. McBratney, B. Minasny, and D.J. Brus. 2011. Using additional criteria for measuring the quality of predictions and their uncertainties in a digital soil mapping framework. Soil Sci. Soc. Am. J. 75(3):1032–1043. doi:10.2136/sssaj2010.0280

Malone, B.P., B. Minasny, N.P. Odgers, and A.B. McBratney. 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232:34–44.

Masri, D., W.L. Woon, and Z. Aung. 2015. Soil property prediction: An extreme learning machine approach. In: S. Arik, T. Huang, W.K. Lai, and Q. Liu, editors, Neural information processing, Part 2. Lecture Notes in Computer Science. Springer. p. 18–27.

Maxwell, S.E., K. Kelley, and J.R. Rausch. 2008. Sample size planning for statistical power and accuracy in parameter estimation. Annu. Rev. Psychol. 59:537–563. doi:10.1146/annurev.psych.59.103006.093735

McBratney, A.B., B. Minasny, S.R. Cattle, and R.W. Vervoort. 2002. From pedotransfer functions to soil inference systems. Geoderma 109(1-2):41–73. doi:10.1016/S0016-7061(02)00139-8

McBratney, A.B., M.L.M. Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117(1-2):3–52. doi:10.1016/S0016-7061(03)00223-4

McBratney, A.B., U. Stockmann, D.A. Angers, B. Minasny, and D.J. Field. 2014. Challenges for soil organic carbon research. In: A.E. Hartemink and K. McSweeney, editors, Soil Carbon. Spring, New York. p. 3–16. doi:10.1007/978-3-319-04084-4_1

McKay, M.D. 1995. Evaluating prediction uncertainty. US Nuclear Regulatory Commission, Washington, DC. doi:10.2172/29432

Meinshausen, N. 2006. Quantile regression forests. J. Mach. Learn. Res. 7:983–999.

Minasny, B., B.P. Malone, A.B. McBratney, D.A. Angers, D. Arrouays, A. Chambers, et al. 2017. Soil carbon 4 per mille. Geoderma 292(Supplement C):59–86.

Minasny, B., and A.B. McBratney. 2007. Corrigendum to "Spatial prediction of soil properties using EBLUP with the Matern covariance function" [Geoderma 140 (2007) 324–336]. Geoderma 142(3-4):357–358. doi:10.1016/j.geoderma.2007.09.003

Minasny, B., and A.B. McBratney. 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264:301–311. doi:10.1016/j.geoderma.2015.07.017

Minasny, B., A.B. McBratney, B.P. Malone, and I. Wheeler. 2013. Digital mapping of soil carbon. In: D.L. Sparks, editor, Advances in Agronomy, Vol. 118. Elsevier Academic Press Inc, San Diego, CA. p. 1–47. doi:10.1016/B978-0-12-405942-9.00001-]3

Minasny, B., A.B. McBratney, and S. Salvador-Blanes. 2008. Quantitative models for pedogenesis- A review. Geoderma 144(1-2):140–157. doi:10.1016/j.geoderma.2007.12.013

Mishra, U., R. Lal, D.S. Liu, and M. Van Meirvenne. 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil Sci. Soc. Am. J. 74(3):906–914. doi:10.2136/sssaj2009.0158

Morgan, T.M., C.S. Coffey, and H.M. Krumholz. 2003. Overestimation of genetic risks owing to small sample sizes in cardiovascular studies. Clin. Genet. 64(1):7–17. doi:10.1034/j.1399-0004.2003.00088.x

NSW Dep. of Planning and Environment. 2016. Standard instrument local environmental plan, Land use zoning. New South Wales Government, Sydney, Australia. https://www.planningportal.nsw.gov.au/planning-tools/open-data (verified 10 Sept. 2017).

Odgers, N.P., A.B. McBratney, and B. Minasny. 2011. Bottom-up digital soil mapping. I. Soil layer classes. Geoderma 163(1–2):38–44. doi:10.1016/j.geoderma.2011.03.014

Opitz, D., and R. Maclin. 1999. Popular ensemble methods: An empirical study. J. Artif. Intell. Res. 11:169–198.

Perlich, C. 2010. Learning Curves in Machine Learning. In: C. Sammut and G.I. Webb, editors, Encyclopedia of Machine Learning. Springer US, Boston, MA. p. 577–580.

Provost, F., D. Jensen, and T. Oates. 1999. Efficient progressive sampling, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, p. 23-32. doi:10.1145/312129.312188

Quinlan, J.R. 1992. Learning with continuous classes. In: Proc. of the Fifth Australian Joint Conference on Artificial Intelligence World Scientific, Singapore. p. 343–348.

Rawlins, B.G., A.J. Scheib, R.M. Lark, and T.R. Lister. 2009. Sampling and analytical plus subsampling variance components for five soil indicators observed at regional scale. Eur. J. Soil Sci. 60(5):740–747. doi:10.1111/j.1365-2389.2009.01159.x

Rudiyanto, B. Minasny, B.I. Setiawan, C. Arif, S.K. Saptomo, and Y. Chadirin. 2016. Digital mapping for cost-effective and accurate prediction of the depth and carbon stocks in Indonesian peatlands. Geoderma 272:20–31. doi:10.1016/j.geoderma.2016.02.026

Somarathna, P.D.S.N., B.P. Malone, and B. Minasny. 2016. Mapping soil organic carbon content over New South Wales, Australia using local regression kriging. Geoderma Regional 7(1):38–48. doi:10.1016/j.geodrs.2015.12.002

Song, X.D., D.J. Brus, F. Liu, D.C. Li, Y.G. Zhao, J.L. Yang, and G.L. Zhang. 2016. Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China. Geoderma 261:11–22. doi:10.1016/j.geoderma.2015.06.024

Sørensen, R., U. Zinko, and J. Seibert. 2006. On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. Hydrol. Earth Syst. Sci. Discuss. 10(1):101–112. doi:10.5194/hess-10-101-2006

Vann, J., and D. Guibal. 2001. Beyond ordinary kriging: An overview of non-linear estimation. Mineral Resource and Ore Reserve Estimation, The AusIMM Guide to Good Practice (Monograph 23). Australian Institute of Mining and Metallurgy, Carlton, Victoria, Australia.

Viscarra Rossel, R.A., R. Webster, E.N. Bui, and J.A. Baldock. 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. Glob. Change Biol. 20(9):2953–2970. doi:10.1111/gcb.12569

Voltz, M., and R. Webster. 1990. A comparison of kriging, cubic-splines and classification for predicting soil properties from sample information. J. Soil Sci. 41(3):473–490. doi:10.1111/j.1365-2389.1990.tb00080.x

Wang, Y., and I.H. Witten. 1997. Induction of model trees for predicting continuous classes. Working paper series, ISSN 1170-487X. Dep. of Computer Science, Univ. of Waikato, Hamilton, New Zealand.

Wheeler, D., and M. Tiefelsdorf. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7(2):161–187. doi:10.1007/s10109-005-0155-6

Zeng, C.Y., L. Yang, A.X. Zhu, D.G. Rossiter, J. Liu, J.Z. Liu, C.Z. Qin, and D.S. Wang. 2016. Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method. Geoderma 281:69–82. doi:10.1016/j.geoderma.2016.06.033

Zobeck, T.M., M. Baddock, R.S. Van Pelt, J. Tatarko, and V. Acosta-Martinez,. 2013. Soil property effects on wind erosion of organic soils. Aeolian Res. 10:43-51.