Geoderma Regional 7 (2016) 38–48

Contents lists available at ScienceDirect

Geoderma Regional

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

Mapping soil organic carbon content over New South Wales, Australia using local regression kriging

P.D.S.N. Somarathna *, B.P. Malone, B. Minasny

Faculty of Agriculture and Environment, Department of Environmental Sciences, The University of Sydney, New South Wales, Australia

ARTICLE INFO

Article history:
Received 26 September 2015
Received in revised form 16 December 2015
Accepted 16 December 2015
Available online 24 December 2015

Keywords:
Carbon sequestration
Soil organic carbon
Multiple linear regression kriging
Digital soil mapping
Luvisols
Vertisols
Plinthosols

ABSTRACT

Various digital soil mapping techniques ranging from simple linear models to complex machine learning techniques have been employed for soil organic carbon (SOC) mapping. When SOC mapping over a large region is required, the usual approach has to employ a model calibrated for the whole area. An alternative is to use a series of locally calibrated models to map smaller areas that collectively cover the large region of interest. The accuracy of the SOC products generated by these two approaches can potentially vary. However, performance of whole-area calibrated models versus locally calibrated models in mapping SOC of large extents has seldom been explored in detail, particularly with respect to the type of model being employed. Our study aims to fill this gap by evaluating the SOC prediction performance of three common models, multiple linear regression (MLR), Regression tree model; Cubist and Support Vector Regression (SVR) that are calibrated locally and for the whole study area. This study was carried out using eight identified local areas in New South Wales (NSW), Australia and across the whole state entirely. Every model was calibrated separately for each local area and for the entire state. The local and whole-area models were validated using the same test data set over 50 realizations. In particular, local prediction accuracy of whole-area calibrated models was compared to that of locally calibrated models. The models were tested separately for the standard soil depth layers including 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm; 60–100 cm. The results show that SVR models have a superior performance out of three tested models for all standardized depth layers. In general the local models outperform the whole-area models for all three tested models with respect to the accuracy of predictions. All models displayed area specific performances proving the importance of inclusion of prevailing local conditions in SOC modelling and mapping. Therefore, we introduce a moving window approach where a hybrid series of locally calibrated models and a whole-area calibrated model can be used against using one calibrated model for the modelling very large mapping extents. Moving window approach provides more accurate results having the lowest error compared to the whole-area model. Also it provides the least biased predictions. Therefore, this novel approach provides a promising way of increasing the efficiency and accuracy of digital soil mapping.

1. Introduction

Soil organic carbon (SOC) is one of the most researched soil properties due to its importance in agronomic sustainability (Reeves, 1997) and carbon sequestration potential. Carbon sequestration is seen as the best solution to reduce atmospheric carbon where both agriculture and the environment are benefited. Consequently, several global and national policy initiatives that revolve around the carbon sequestration potential of SOC have come to the forefront (O'Rourke et al., 2015). A carbon offset scheme known as the Carbon Farming Initiative (CFI) instigated in Australia is a perfect example. Such programs rely on accurate estimates of SOC content over the spatial extent of interest which can be represented by a baseline SOC map. SOC mapping has been greatly benefited by Digital soil mapping (DSM). During the last decade, various DSM techniques ranging from simple linear models to complex machine learning techniques have been employed for SOC mapping (Minasny et al., 2013).

These techniques include, kriging (Cambule et al., 2014; Dai et al., 2014), co-kriging (Odeh et al., 1995; Phachomphon et al., 2010), regression kriging (Mora-Vallejo et al., 2008; de Brogniez et al., 2014; Dorji et al., 2014; Piccini et al., 2014), Linear mixed models (Rawlins et al., 2009; Karunaratne et al., 2014), machine learning techniques such as Artificial neural networks (Minasny and McBratney, 2002; Malone et al., 2009; Zhao et al., 2010), Support Vector Regression (Ballabio, 2009), Regression tree models, such as Cubist (Adhikari et al., 2014; Miklos et al., 2010; Rossel et al., 2014) and Random forests (Wiesmeier et al., 2011; Subburayalu and Slater, 2013; Hengl et al., 2015) and Generalised Additive Models (GAM) (Poggio et al., 2013; de Brogniez et al., 2014).

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

http://dx.doi.org/10.1016/j.geodrs.2015.12.002
2352-0094/

Foregoing techniques and models can be seen employed at various scales ranging from small farm areas to larger regional and continental extents for SOC mapping. When the requirement is to map SOC of a larger area, the preferred approach has to use a single calibrated model to map the entire area. Alternatively, a series of locally calibrated models can be used to map small areas that collectively cover the large region of interest where there is a fairly reasonable sampling density or the usage of hybrid series of local and whole-area calibrated models for the areas with dense and sparse observation points respectively. The latter approach is very uncommon in DSM literature to the best of our knowledge. These two approaches coupled with different model types such as multiple linear regression (MLR) Cubist and Support Vector Regression (SVR) could produce results that are of varying accuracies. Performance of such models in predicting SOC over large spatial extents has seldom been compared with respect to whole-area and locally calibrated models. Therefore, this study aims to examine the SOC prediction capability of MLR, Cubist and SVR with respect to local versus whole-area model training and application. The study is carried out using eight identified local areas for the localized studies in the state of NSW and a whole area study covering all of NSW. Based on the results, we make recommendations on the best combinations of model type and spatial extent used for calibration.

2. Methods

2.1. Study area

The study area is the state of New South Wales (NSW), Australia which covers approximately 810,000 km^2. The Great Dividing Range which runs approximately north to south in the east has a major impact on the State's distribution of rainfall that results in four distinct climatic zones. The area to the west of the Great Dividing Range which represents majority of NSW has an arid to semi-arid climate. The average annual rainfall for this area ranges from 150 mm to 500 mm. The climate along the flat, coastal plain east of the dividing range varies from cool oceanic to humid subtropical from south to far north of the state. The area has a higher annual rainfall ranges from 800 mm to 2000 mm. (Stormy Weather, Bureau of Meteorology). About 65% of the area is occupied by grazing lands which comprises of both native and modified pastures. The nature conservation areas which accounts for around 7.6% of the of total land use are mostly located in the eastern coastal areas. Dry land crops occupy about 9% of the area, while about 7% of the land is minimally used. (Catchment Scale Land use data, Department of Agriculture, Australia).

2.2. Data sets and data processing

2.2.1. Soil data

SOC data consists of the University of Sydney research data and the Terrestrial Ecosystem Research Network (TERN) data that are collected by different institutions for various purposes. There were 5386 observation sites in total. The data were clustered as they came from different survey projects from 1995 to 2014. The observed SOC content is given by the g/100 g. Since the distribution was positively skewed, the data was log-transformed for modelling procedures. The spatial distribution of those sampling points within the study area is shown in Fig. 1.

2.2.1.1. Harmonizing observed soil profile data. The sampling depths of the soil profiles were different to each other. For further analysis of data, it is imperative to have a common depth interval range across all sampling points. Malone et al. (2009) generalized and extended the quadratic spline model of Bishop et al. (1999) and formulated a smoothing spline function for vertical prediction of soil properties into specified common depth interval range.

The smoothing parameter (λ) of the quadratic spline function (Malone et al., 2009), is a determinant of the accuracy of prediction. It is crucial to find out the best λ value that minimizes the prediction error. Therefore, 506 sampling points which have more than 4 layers of measurements were selected to find out the best fitting λ value. The weighted average of the first two layers and the third and fourth layers were calculated to form two layers for each profile. Those values were then used to predict SOC values for the original sampling depths with respect to a series of λ values (0.00001, 0.0001, 0.001, 0.01, 0.1, 0.5, 1, 2, and 5). The λ value which gave the minimum mean squared error (MSE) value was selected as the best smoothing parameter. Then, the depth intervals (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm) corresponding to the digital soil mapping specifications of the GlobalSoilMap project (Arrouays et al., 2014) were used as the harmonized depth intervals for spatial prediction models of SOC.

2.2.2. Environmental covariates

The content and the spatial distribution of SOC in an ecosystem are driven by the environmental factors such as climate, underlying lithology, topography, fauna and flora. Introduced by Jenny (1941), this concept was generalized and formalized by McBratney et al. (2003) as the

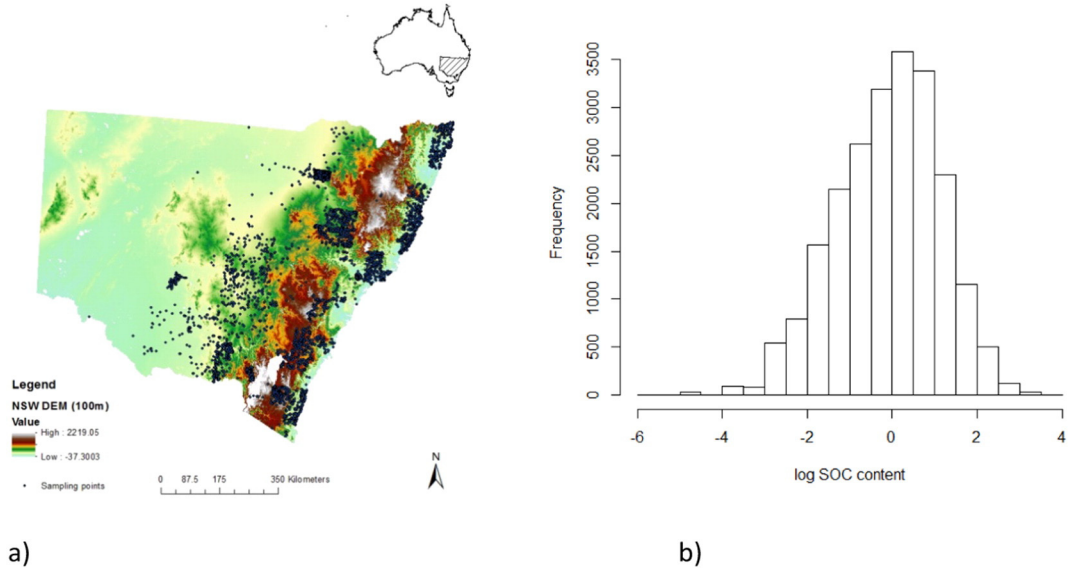

Fig. 1. (a) Spatial distribution of the sampling points in NSW, Australia, (b) histogram of observed SOC in log scale.

SCORPAN model, which is the basis for contemporary digital soil mapping (DSM). Since the introduction of the SCORPAN spatial soil prediction function (SSPF) concept, there has been a significant interest in using the model (sorpan-SSPF) all over the world for generating soil information from sparse data sets (Lagacherie, 2008). Following SCORPAN a number of environmental covariates were selected as the auxiliary variables along with point data for the prediction of SOC at unsampled locations. These environmental covariates included Landsat 7 satellite information, average annual rainfall, average annual evapotranspiration, average annual temperature, digital elevation model (DEM) and gamma radiometric data.

2.2.2.1. Landsat 7 — 2012 image (30 m). The Landsat 7 satellite is equipped with the Enhanced Thematic Mapper Plus (ETM+) which is a multispectral scanning radiometer that detects 7 spectral bands in the visible and near-infrared wavelengths. The band ratios b3/b2, b3/b7 and b5/b7, were derived. These are more commonly referred as soil enhancement ratios (Malone et al., 2009). Further, Normalised Difference Vegetation Index (NDVI) which is defined as the ratio between (b4 − b3) and (b4 + b3), was also derived as an approximation to the vegetation cover. Landsat 7 ETM+ image was taken in the year 2012.

2.2.2.2. Climatic data (90 m). The raster grids of average annual rainfall, average annual evapotranspiration, and average annual temperature were obtained from the Bureau of Meteorology Australia. The average values are referred to the ten year average from year 2000–2010 for all aforementioned climatic raster grids.

2.2.2.3. Gamma-radiometric data (90 m). The measurement of naturally emitted gamma radiation from the ground surface is referred to gamma radiometric data. "The most commonly detected gamma-emitters for geophysical purposes are potassium (^{40}K) and Thorium (^{232}Th) and Uranium (^{232}U). The formation on their relative abundance provides indirect evidence about the distribution of soil-forming minerals in the landscape (Cook et al., 1996)". The gamma radiometric data collected by Geosciences Australia was used in the study.

2.2.2.4. The digital elevation model (DEM) (30 m). From the DEM, first and second derivatives; slope, Topographic Wetness Index (TWI), (Beven & Kirby, 1979) covariates were derived. Further, Multi Resolution Valley Bottom Flatness Index (MrVBF) (Gallant and Dowling, 2003), was also used as a DEM derived topographic index in this study. Moore et al. (1993) provides a detailed description of how some or all of these derivatives have been used to derive relationships with the spatial distribution of various soil properties.

2.2.3. Covariate layers processing and database compilation

The original raster layers from various sources were resampled to 100 × 100 m grid cell resolution using nearest neighbour resampling method, and all layers were re-projected to a common coordinate reference system. The pixel values of the raster layers (environmental covariates) that correspond to the coordinates of sampling points were extracted and compiled in a database.

2.3. The SOC prediction models

Multiple linear regression (MLR) is a simple and frequently used method in SOC mapping and other DSM activities (Minasny and McBratney, 2002; Stevens et al., 2014; Phachomphon et al., 2010; Mishra et al., 2010). Regression tree such as Cubist is also a popular technique among the DSM community used for SOC and other soil properties mapping (Lacoste et al., 2014; Rossel et al., 2014; Bui et al., 2009; Kidd et al., 2014). The use of Support Vector Machines (SVM) models for SOC mapping has also a growing attention (Ballabio, 2009; Kanevski et al., 2002). Therefore, these three models were selected for testing whether the SOC prediction capabilities of models depend on the extent of mapping. An overview of the method of models selection is given by Fig. 2.

2.3.1. MLR

The method used for the spatial interpolation is essentially a linear model (LM). It is assumed that the regression function E (YIX) is linear, or the linear model is a reasonable approximation. The linear regression model can be expressed as,

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \tag{2.2.6}$$

where β_0 is the interception of the linear model, X_j represents the auxiliary or secondary variables or covariates and β_j are 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 primary variable (soil carbon content) and explanatory variables (SCORPAN factors).

2.3.2. Cubist model

This is a variation of a Regression tree model, where the prediction is based on linear regression models instead of discrete values. The Cubist model produces 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 cryptically by Quinlan (1992), and further clarified by Wang and Witten (1996), and Holmes et al. (1999). Briefly, Cubist 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). Then the model associated with each rule was computed using the conventional 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 mapping procedures (e.g., Henderson et al., 2005; Minasny et al., 2008; Bui et al., 2009; Rossel et al., 2014; Kidd et al., 2014).

2.3.3. Support Vector Regression (SVR)

The present form of Support Vector Machines (SVM) originated from the work by Vapnik and co-workers developed for the classification purposes where two classes are separated by the optimal separating hyperplane which is obtained by maximizing the margin between classes' closest points. Support Vector Regression (SVR) is a generalization of SVM, and used as a technique for nonlinear classification and regression. The ε-SVR differs from the classic regression due to the use of loss function to define the borders (hyperplane) of the regression function. Hence the regression function lies between ± ε (maximum error). Therefore, the loss is equal to 0 if the difference between the predicted and measured values is less than ε. A detailed explanation of SVM and SVR can be found in Smola and Scholkopf (2004); Wang et al. (2009)and Ballabio (2009). It has been used in soil mapping by Ballabio (2009) and Padarian et al. (2014).

2.4. Model training and testing

Prior to mapping, the models' performances were tested for both whole-area and local scenarios. Firstly, the MLR, Cubist and SVM models were trained and validated for the entire study area. Then, the models were locally trained and tested for selected local regions in order to compare the whole-area trained model performances against the locally trained models. Model performance was evaluated using root mean squared error (RMSE), correlation coefficient (R^2) and Lin's (1989) concordance correlation coefficient (CCC). Fig. 2 provides schematic representation of the model selection procedure for mapping the entire study area using a single model.

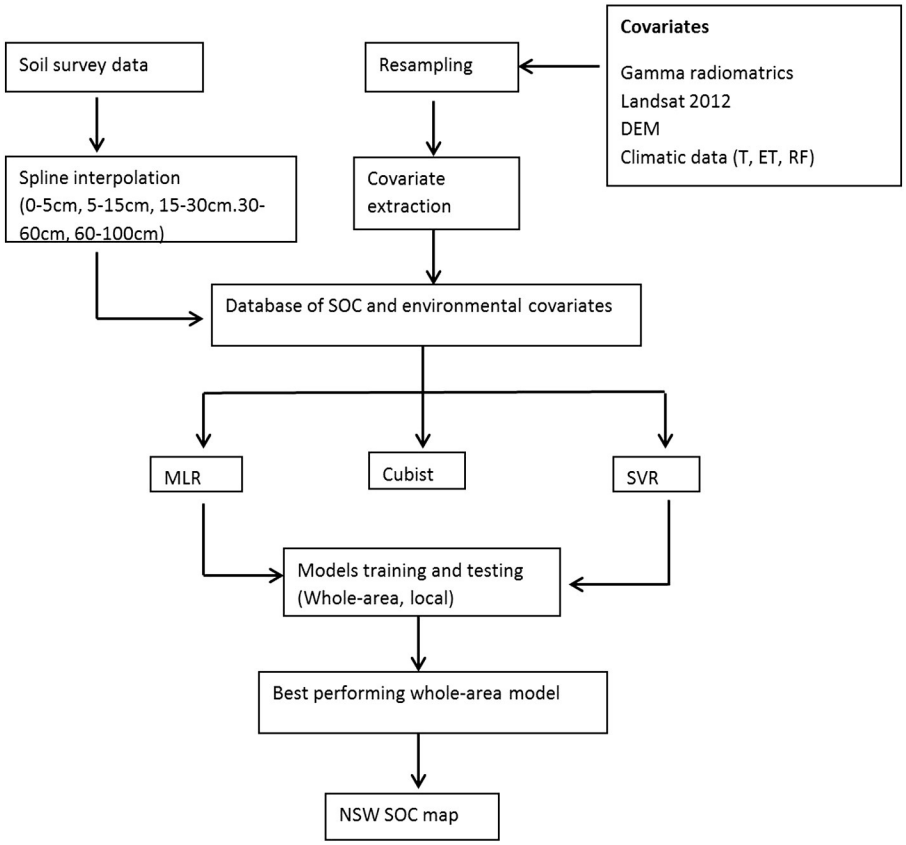

Fig. 2. Flowchart — assessing the model performances and generation of SOC maps for different soil depth layers for NSW.

2.4.1. Whole-area-model training and validation

Three models: MLR, Cubist, SVR were spatially trained for the whole area. The performances of these models were tested using 50 realizations of random, 70:30 calibration: testing data splits. Accordingly, each whole-area model was trained and validated for all standard depth layers independently (Fig. 2). The performance indices were calculated for each model and for each simulation for all standard depths.

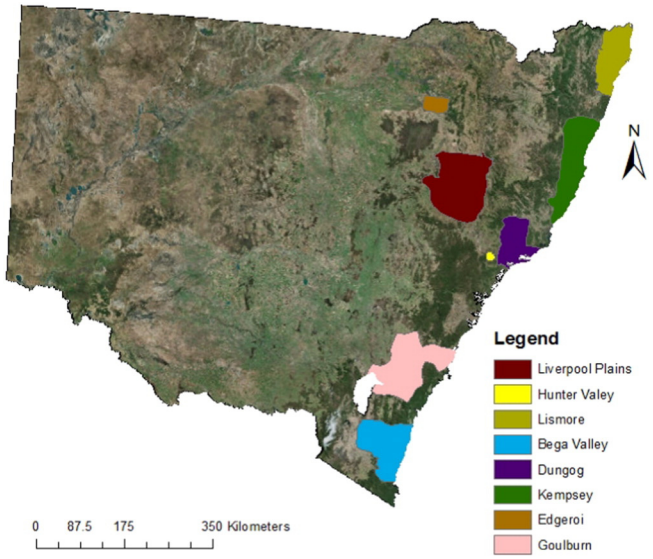

Fig. 3. Eight local regions in NSW.

2.4.2. Testing whole-area models against the local models

Most of the data points of the study area were concentrated in certain areas corresponding to the survey areas (Fig. 3). Eight such regions were identified as the local areas, to test the whole-area calibrated model performances against the locally calibrated models. Each local region was given a name based on the location for identification purposes. A brief description of the eight local regions is given in Table 1. The area of the local regions ranges from 261 km^2 to 13,694 km^2 with data density of 3.02 to 0.03 km^{-2}.

Local models (MLR, Cubist, and SVR) for each local region were also trained in a similar fashion, with 70% of data used in training and the held back 30% was used for testing under 50 realizations. The spatial training and 50 realizations were applied for each model and for each local area and for each specified depth interval. Also, the performances of all whole-area models were tested across the 50 realizations where the same validation data set used as per in the local models validation.

2.5. Spatial prediction of SOC for NSW

We used two approaches for predicting the SOC content across NSW. The first approach was to use a single model calibrated for the entire study area (New South Wales) to predict the SOC content onto a 100 m grid. The second method is a novel approach where a hybrid series of local and whole-area models are used to predict the SOC content on to 100 m grid of the entire study area using moving windows, and is hereafter referred to as the moving window (MW) approach.

2.5.1. Mapping NSW SOC content using a whole-area model

The best performing whole-area model between MLR, Cubist and SVR was selected for the mapping of SOC. Then spatially trained whole-area model with 70% of data was used to predict SOC content

Table 1
Local area.

Regions' given name	Localities	Area (km^2)	Number of samples	Sampling density ($n\ km^{-2}$)
Lismore	Clarence Valley, Richmond Valley, Lismore, Ballina, Byron and Tweed	7096	305	0.04
Kempsey	Greater Taree, Port Macquarie-Hastings, Kempsey, Nambucca, Bellingen and Coffs Harbour	10,197	465	0.05
Liverpool Plains	Hunter Sire, Liverpool Plains Tamworth, Gunnedah	13,694	686	0.05
Dungog	Newcastle, Port Stephens, Maitland, Dungog, Great Lakes and Gloucester	5619	314	0.05
Bega Valley	Bega Valley, Snowy River, Cooma-Monaro	9022	386	0.04
Goulburn	Kiama, Shellharbour and inland Goulburn, Upper Lachlan Shire, Queanbeyan, and Palerang	12,425	412	0.03
Edgeroi	Narrabri	1493	334	0.22
Hunter Valley	Cessnock, Singleton	261	789	3.02

at each grid node of a 100 m grid. Also the model residuals were geostatistically modelled to krige values of the residuals (Hengl et al., 2004) at the same nodes of 100 m grid. Then these interpolated residuals were added to model predictions to form the final map.

2.5.2. Moving window approach

When it comes to mapping large areas, usually the entire region is mapped using a single model calibrated for the entire study area. As discussed above, one could argue that at least for some local areas such models are less accurate compared to a locally calibrated model, though this has seldom been tested. A previous study by Sun et al. (2012) proposed a local regression kriging approach, where by local regression kriging models were calibrated at each pixel or prediction location based on a defined number of closest observations. When dealing with areas of large extent, the approach by Sun et al. (2012) is not feasible. In this study, we used a combination of whole-area model and local models for locations where the sampling points are abundant and sparse, respectively for the spatial prediction of SOC content across the study area.

In this approach, first the whole NSW was divided into tiles with size 100 km × 100 km. The delineation of tiles had two offsets (t1, t2). "t1" was offset 50 km north and east form the "t2". We used these two offsets of tiles in a way that they create overlapping areas (Fig. 4). Then each covariate layer was spilt into 100 × 100 km tiles so that each tile for each offset has a separate set of covariate layers. Next, the observation data set was randomly split into 70:30 for model training and testing of the products. Then, each tile was mapped with a local or whole-area MLR model based on the number of observation points found within each window. The number of observations for each tile for modelling is a collection of the observations within the tile and the observations from nine nearest neighbour tiles. If the number of sampling points is greater than 50, the model was calibrated using local observations. Otherwise, the whole-area model was used for the predictions of the SOC in that particular window.

When the local model was being used, the residual kriging with a local variogram was applied (Sun et al., 2012; Whelan et al., 2003), and prediction variance and the kriging variance were also calculated and cumulated with the predictions and kriged residuals in order to increase the accuracy of the final product. Residual kriging was not applied for tiles where the whole-area model was applied, based on the assumption that accuracy of the final product will not significantly

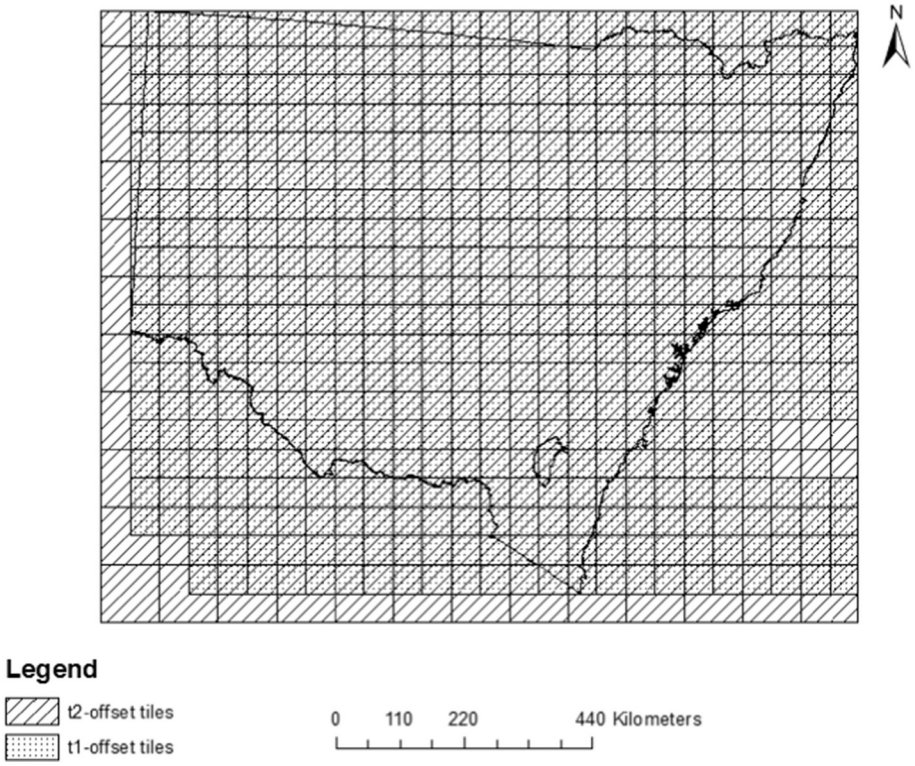

Fig. 4. Moving window tile delineation.

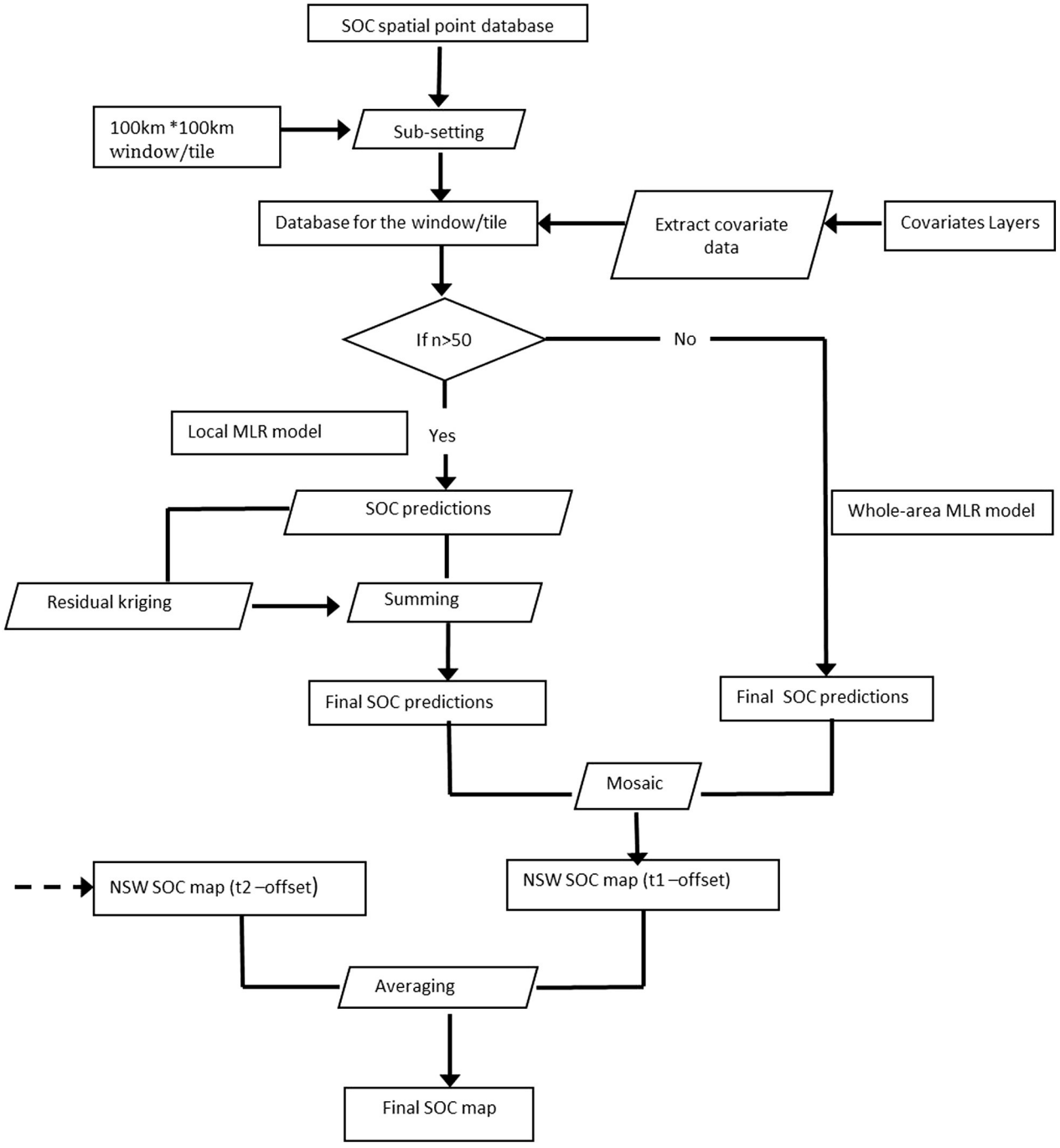

Fig. 5. Flowchart for moving window mapping.

improve since there are only fewer number of data points inside the tiles. The prediction of the t1 offset was based on local model predictions for 27 tiles and whole-area model predictions for 29 tiles and for t2 offset 30 tiles were locally predicted and 20 tiles were globally predicted. Therefore, around 50% of the study area was predicted using local models in the MW approach.

Table 2
Comparison of performance of MLR, Cubist and SVR using 50 fold cross validations (70:30 training:validation) with standard deviation for the whole-area models for all depth increments. Comparison is based on log (SOC) in log (g/100 g).

	MLR-whole-area			Cubist-whole-area			SVR-whole-area		
	RMSE	R^2	CCC	RMSE	R^2	CCC	RMSE	R^2	CCC
0–5 cm	0.82 ± 0.01	0.15 ± 0.06	0.26 ± 0.08	0.81 ± 0.01	0.19 ± 0.05	0.35 ± 0.07	0.79 ± 0.01	0.22 ± 0.05	0.38 ± 0.06
5–15 cm	0.79 ± 0.01	0.16 ± 0.05	0.29 ± 0.07	0.77 ± 0.01	0.20 ± 0.05	0.37 ± 0.07	0.75 ± 0.01	0.25 ± 0.04	0.41 ± 0.06
15–30 cm	0.92 ± 0.01	0.15 ± 0.07	0.26 ± 0.09	0.89 ± 0.01	0.20 ± 0.06	0.35 ± 0.08	0.88 ± 0.01	0.23 ± 0.05	0.39 ± 0.07
30–60 cm	0.96 ± 0.01	0.10 ± 0.07	0.19 ± 0.10	0.94 ± 0.01	0.15 ± 0.07	0.27 ± 0.09	0.93 ± 0.01	0.16 ± 0.06	0.30 ± 0.09
60–100 cm	0.96 ± 0.01	0.05 ± 0.08	0.10 ± 0.11	0.95 ± 0.01	0.08 ± 0.07	0.18 ± 0.10	0.93 ± 0.01	0.11 ± 0.07	0.22 ± 0.09
Profile average	**0.89 ± 0.01**	**0.12 ± 0.06**	**0.22 ± 0.09**	**0.87 ± 0.01**	**0.16 ± 0.06**	**0.30 ± 0.08**	**0.86 ± 0.01**	**0.20 ± 0.06**	**0.34 ± 0.08**

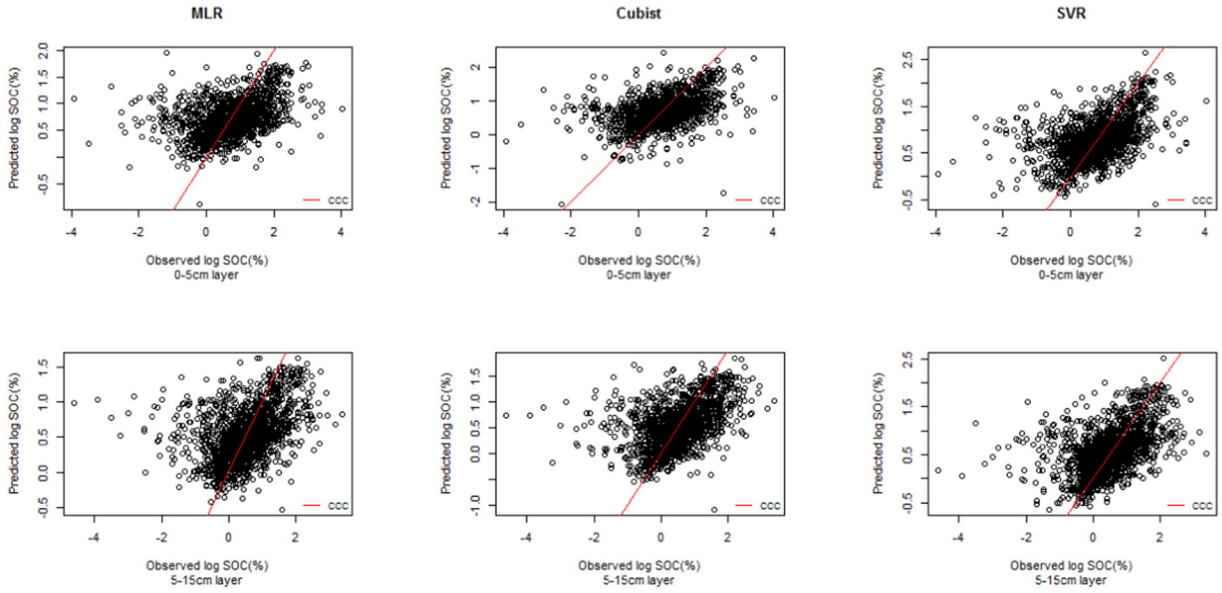

Fig. 6. Validation between measured and predicted SOC values for the whole-area models for the 0–5 cm, 5–15 cm depth intervals.

To create a map for each offset, all tiles from each offset were mosaiced together. Then the mosaiced SOC maps from each offset were averaged to produce the final output of moving window mapping. This process was repeated to produce SOC maps for all standardized depth layers. Then SOC maps for each depth were validated using the 30% validation data set. Fig. 5 provides an overview of the moving window approach that has been used in our study.

3. Results and discussion

3.1. Best smoothing parameter

For continuous soil depth functions, λ value of 0.01 gave the lowest corresponding MSE among the tested λ values. State-wide average SOC% values predicted for the harmonized depth intervals (0–5 cm,

Table 3
Average values of performance indicators along with its standard deviation (SD) of the local validation of the whole-area and local models: (a) Comparison of RMSE values, (b) comparison of R^2 values, (c) comparison of CCC values.

(a).

Average RMSE with SD	MLR-local	MLR-whole-area	Cubist-local	Cubist-whole-area	SVR-local	SVR-whole-area
Lismore	1.07 ± 0.07	1.21 ± 0.03	1.02 ± 0.07	1.12 ± 0.03	0.98 ± 0.04	1.08 ± 0.03
Kempsey	0.99 ± 0.01	1.17 ± 0.02	0.98 ± 0.01	1.10 ± 0.02	0.95 ± 0.01	1.07 ± 0.02
Liverpool Plains	0.66 ± 0.03	0.81 ± 0.01	0.67 ± 0.01	0.80 ± 0.01	0.66 ± 0.01	0.75 ± 0.01
Dungog	0.89 ± 0.05	0.94 ± 0.02	0.83 ± 0.02	0.92 ± 0.02	0.78 ± 0.02	0.91 ± 0.02
Bega Valley	0.97 ± 0.02	1.15 ± 0.02	0.96 ± 0.02	1.10 ± 0.02	0.94 ± 0.02	1.06 ± 0.02
Goulburn	0.78 ± .03	0.85 ± 0.02	0.72 ± 0.02	0.85 ± 0.02	0.71 ± 0.02	0.81 ± 0.02
Edgeroi	0.46 ± 0.02	0.58 ± 0.02	0.47 ± 0.02	0.56 ± 0.02	0.43 ± 0.02	0.51 ± 0.02
Hunter Valley	0.60 ± 0.01	0.63 ± 0.01	0.59 ± 0.01	0.59 ± 0.01	0.58 ± 0.01	0.57 ± 0.01

(b).

Average R^2 with SD	MLR-local	MLR-whole-area	Cubist-local	Cubist-whole-area	SVR-local	SVR-whole-area
Lismore	0.02 ± 0.03	0.02 ± 0.01	0.03 ± 0.02	0.07 ± 0.02	0.04 ± 0.02	0.07 ± 0.02
Kempsey	0.03 ± 0.01	0.02 ± 0.01	0.04 ± 0.01	0.12 ± 0.02	0.06 ± 0.01	0.18 ± 0.02
Liverpool Plains	0.07 ± 0.01	0.09 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.08 ± 0.01	0.16 ± 0.01
Dungog	0.08 ± 0.01	0.11 ± 0.01	0.06 ± 0.01	0.15 ± 0.02	0.12 ± 0.01	0.19 ± 0.01
Bega Valley	0.08 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.13 ± 0.01	0.10 ± 0.01	0.20 ± 0.01
Goulburn	0.03 ± 0.01	0.07 ± 0.01	0.04 ± 0.01	0.08 ± 0.01	0.07 ± 0.01	0.14 ± 0.01
Edgeroi	0.14 ± 0.02	0.06 ± 0.01	0.12 ± 0.02	0.12 ± 0.02	0.20 ± 0.02	0.25 ± 0.02
Hunter Valley	0.05 ± 0.01	0.04 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.08 ± 0.01	0.08 ± 0.01

(c).

Average CCC with SD	MLR-local	MLR-whole-area	Cubist-local	Cubist-whole-area	SVR-local	SVR-whole-area
Lismore	0.06 ± 0.01	0.05 ± 0.01	0.06 ± 0.01	0.20 ± 0.03	0.09 ± 0.01	0.23 ± 0.02
Kempsey	0.08 ± 0.01	0.03 ± 0.01	0.09 ± 0.01	0.19 ± 0.02	0.13 ± 0.01	0.24 ± 0.01
Liverpool Plains	0.16 ± 0.01	0.14 ± 0.01	0.12 ± 0.01	0.13 ± 0.01	0.17 ± 0.01	0.27 ± 0.01
Dungog	0.18 ± 0.02	0.16 ± 0.01	0.15 ± 0.02	0.22 ± 0.02	0.21 ± 0.02	0.26 ± 0.02
Bega Valley	0.19 ± 0.01	0.02 ± 0.01	0.16 ± 0.01	0.05 ± 0.01	0.21 ± 0.01	0.04 ± 0.01
Goulburn	0.09 ± 0.02	0.11 ± 0.01	0.07 ± 0.01	0.17 ± 0.02	0.14 ± 0.01	0.25 ± 0.01
Edgeroi	0.28 ± 0.02	0.11 ± 0.01	0.24 ± 0.02	0.19 ± 0.02	0.32 ± 0.02	0.34 ± 0.02
Hunter Valley	0.13 ± 0.01	0.05 ± 0.01	0.10 ± 0.01	0.09 ± 0.01	0.18 ± 0.01	0.13 ± 0.01

Table 4
Comparison of different mapping approaches using RMSE.

RMSE	SVR whole-area model (100 m)	SV regression kriging (100 m)	Moving window MLR (100 m)	Moving window MLRegression kriging (100 m)	TERN national grid (100 m)	ISRIC SoilGrid1k
0–5 cm	0.81	1.07	0.76	0.74	0.93	0.97
5–15 cm	0.73	0.91	0.73	0.71	0.77	0.81
15–30 cm	0.87	0.87	0.79	0.77	1.32	1.14
30–60 cm	0.91	1.10	0.89	0.88	0.96	1.13
60–100 cm	0.95	1.44	0.90	0.89	0.99	1.15
Profile average	0.85	1.08	0.81	0.80	0.99	1.04

5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm) using this λ value were 3.00, 2.37, 1.50, 0.85, and 0.57% respectively.

3.2. Model training and validation

3.2.1. The best predictive whole-area model

The calculated RMSE, R^2, and CCC values along with their standard deviation based on the 50 realizations show that SVR has the best predictive capabilities across all depth layers, while MLR displayed the least accuracy (Table 2).

Considering the layer-wise prediction, all models performed the best for the top three depth intervals. Between the three layers, the models have predicted SOC content in the 5–15 cm depth interval with the greatest accuracy. Scatter plots given in Table 2 illustrate this further, and highlight the superiority of SVR. Also, it is important to note that the prediction accuracy subsides for all three models when the prediction depth increases.

The profile average RMSE of all three whole-area models is more or less similar. This implies that the MLR, Cubist, and SVR whole-area models have similar prediction capabilities in terms of the prediction accuracy. However, with regard to both accuracy and the precision of the predictions, SVR whole-area model stands out since it has comparatively higher CCC value (34% ± 8%) than MLR (22% ± 6%), and Cubist (30% ± 8%). Fig. 6 further confirms that SVR model predictions also have stronger agreement with the observed values than the other two models. Therefore, SVR whole-area model was selected as the best performing whole-area model for the spatial prediction of SOC across the NSW.

3.2.2. Testing whole-area models against the local models

The RMSE values for all three local model types are constantly smaller than respective whole-area models (Table 3). Thus, the accuracy of local model predictions is higher than the whole-area model predictions. Further, all model predictions are comparable in terms of

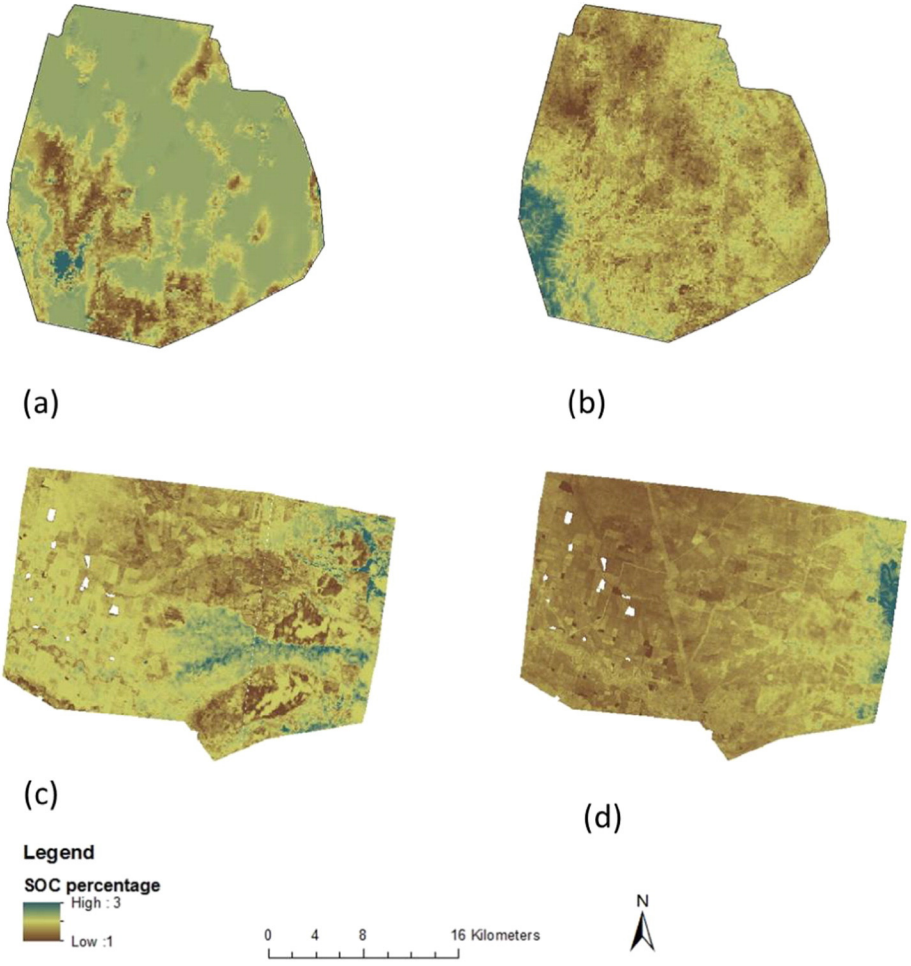

Fig. 7. Zoomed-in maps for local clusters, (a) whole-area SVR SOC map of 5–15 cm for Hunter Valley, (b) local MLR SOC map of 5–15 cm for Hunter Valley, (c) whole-area SVR SOC map of 5–15 cm for Edgeroi, (d) local MLR SOC map of 5–15 cm for Edgeroi.

prediction accuracy, though the SVR-local models perform slightly better than other models for all local areas.

However R^2 values of SVR-whole-area models are higher than the respective local SVR models for all local regions except for Hunter Valley, while Cubist and MLR models have mixed results. Kempsey, Bega Valley, Edgeroi and Hunter Valley areas display slightly higher agreement (R^2) between predictions and observed values for local MLR models than the respective MLR-whole-area models. Cubist-whole-area models have a better fit with the observed data for all local regions except Edgeroi and Hunter Valley which are having the similar R^2 values for both local and whole-area Cubist models.

The CCC values for MLR local models are greater than the respective MLR-whole-area models for all local regions whereas Cubist and SVR are having mixed results. CCC for Cubist local models for regions Bega Valley, Edgeroi, and Hunter Valley are higher than the Cubist-whole-area models. Most of the local areas except Goulburn and Hunter Valley, SVR-whole-area models have comparatively higher CCC. Overall, these results imply that locally calibrated models display a slightly better fit with the data though the whole-area model performances are comparable.

Throughout this validation process we also noted that the local regions are having area specific performances. In comparison between both local and whole-area model validations, Lismore, Kempsey, and Bega Valley areas exhibit a lower predictability, while Dungog, Edgeroi and Hunter Valley have a higher degree of accuracy of predictions despite the prediction model used. This confirms the dependency of model performance on local attributes in testing and training.

Therefore, these results imply that the importance of using local models whenever possible in DSM. Thus, we introduced the MW approach where we used a hybrid series of local and whole-area models to map the entire study area.

3.3. NSW SOC mapping and validation

Given that a) SVR has the highest overall performance, b) MLR has the highest sensitivity to local conditions, and c) it is always better to start with the most simple method when testing a new scenario, a hybrid series of locally and whole-area calibrated MLR and whole-area calibrated SVR were used separately to map SOC of entire NSW. The moving window approach was used as opposed to mapping the entire state at once. Generated maps from both scenarios show a similar spatial pattern of SOC distribution within the state (Fig. 8). It can be clearly seen that the pattern is highly correlated with the climatic, lithological, biological and anthropogenic factors of the state (Rossel et al., 2014). Climate is described as the most influential factor over the state as it determines productivity of biomass and the rate of decomposition of SOC (Wynn et al., 2006; Bui et al., 2009). This pattern is reflected through the generated maps where humid coastal areas have a higher percentage of SOC and it is gradually decreasing towards the arid central desert area. Considering the vertical distribution of SOC, the spatial distribution shows a similar pattern. There are artefacts of margins of windows in the moving window product and this effect is seen on tiles where the local models are being applied. This can be due to the lack of observations and covariate effects on the local models which are specific for each window.

The validation of SOC whole-area model mapping approach and the moving window approach shows that the MW approach produces more accurate predictions than the whole-area model. Also, it is clearly seen that the incorporation of residual kriging has improved the accuracy of the product in the MW approach (Table 4). The SVR whole-area model was also tested with the addition of residual kriging. However, the accuracy of SVR whole-area model predictions seems to diminish with the addition of residual kriging. This

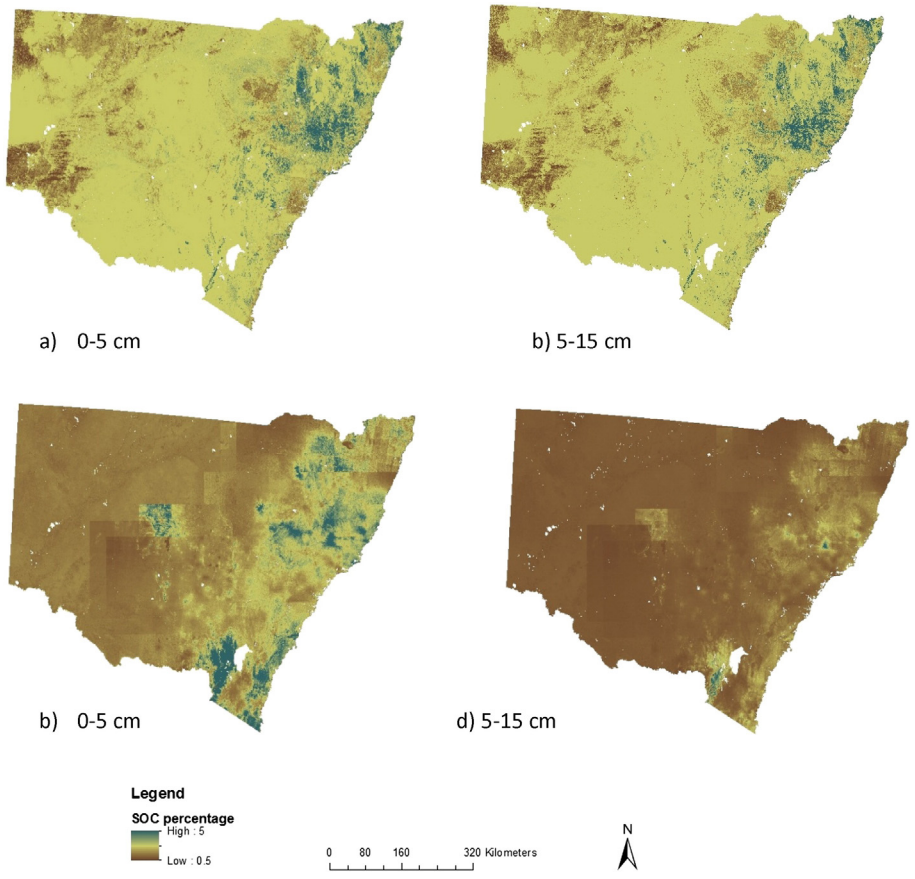

Fig. 8. SOC maps for the top two 0–5 cm, 5–15 cm standard depth layers; (a, b) SVR whole-area model products and (c, d) MLR kriging–moving window products.

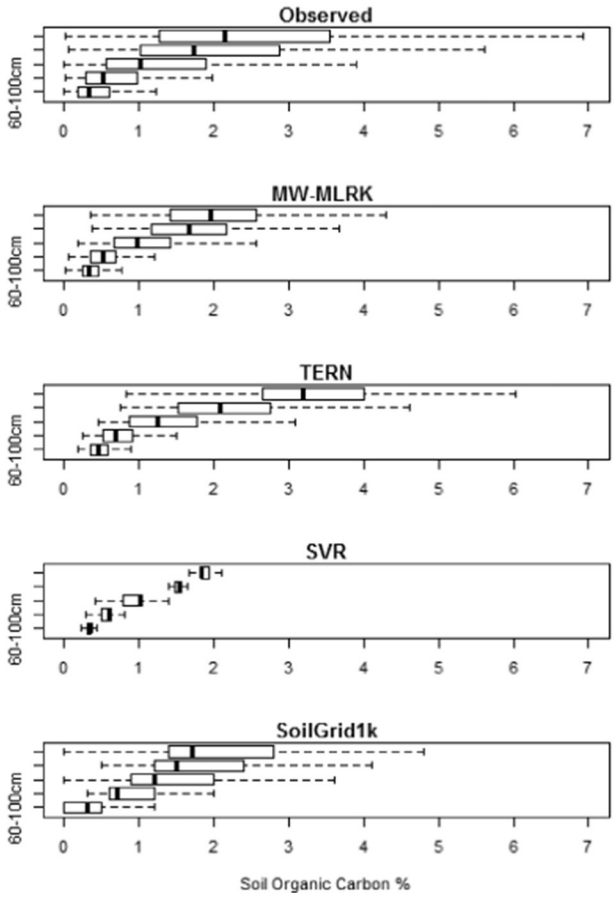

Fig. 9. Boxplots comparison of predicted SOC% of various map products with the observed data.

can be due to the non-stationary and patchy nature of the sample point data.

We also looked into zoomed-in maps of local areas. These maps show that those local models outperformed whole-area models. Fig. 7 depicts maps that are zoomed-in on the Hunter Valley and Edgeroi. The Hunter Valley local model product shows a more detailed appearance than the whole-area model product. It is proven by lower RMSE (0.53) for MLR local than the RMSE of SVR whole-area model (0.58) and Edgeroi is also having a lower RMSE for MLR local (0.41), than the SVR whole-area model (0.50). The higher accuracy for local areas can be due to the benefit gained by the local models with the residual kriging in moving window where the data are fairly stationary and evenly covering the covariate space. Therefore, for the areas with fairly high sampling density, the local model seems to be more accurate (Whelan et al., 2003) as expected. Since it is proven that the SVR models perform better than MLR models, the usage of SVR model in moving window approach may further increase the accuracy of the product.

Finally we compare our mapping approaches with maps that were produced at continental and global extents, i.e. the Soil and Landscape Grid National Soil Attributes (90 m) map which is known as TERN grid (Rossel et al., 2014) and SoilGrids1k (Hengl et al., 2014). The TERN grid presents the mass fraction of carbon by weight in the less than 2 mm soil material which is given as SOC% (g 100 g^{-1}). These maps have been produced using Cubist models with kriging of residuals; The SoilGrids1k is a product of General Linear Models (GLM's) with log-link function. The SoilGrids1k presents the SOC content in mid-depth of the standard GlobalSoilMap layers in g kg^{-1} of soil. The calculated RMSE for the validation data set shows (Tables 4 and 5) that the MW approach outperforms all other products and SVR whole-area model performance is also better than the TERN and SoilGrids1k grid. When the Mean Error (ME) which is the average difference of the sum between observed and predicted values, is considered the TERN and SoilGrids1k tend to have negative values indicating over prediction bias. The local model shows that ME values close to zero indicate the least biased predictions.

Further, we compared the distribution of 30% test data for the various map products with the observed data (Fig. 9). MW-MLRK prediction distribution closely matches the observed data. Also MW-MLRK has a smaller prediction variance compared to TERN and SoilGrid1k predictions. SVR model has the least prediction variance, as it is a single model applied for the whole area. The TERN prediction over predicts with the highest prediction variance as it is an ensemble of 100 bootstrap realizations. Meanwhile SoilGrid1k under predicts.

4. Conclusions

Overall, this study reveals that SVR models perform the best at predicting SOC at both whole-area and local extents across all depth intervals. The performance of Cubist is slightly below the SVR at both local and whole-area extents. MLR showed the least prediction powers in general.

Nevertheless, all three models were equally sensitive to local conditions, resulting in higher prediction accuracy for some local areas and lower accuracy for some other areas. Area specific unaccounted covariates in modelling could partly explain this behaviour. This study also confirms the previous research finding that the performance of models constricts when the depth of the prediction increases irrespective of the type of model used.

This study also reveals that the performance of models in mapping SOC can vary considerably depending on the type of model, the extent at which the model is trained, and the depth and extent of mapping. We recommend the use of locally calibrated models for areas with high sampling density, and whole-area calibrated models for areas with low sampling density. The moving window approach presented in this study can be used to map SOC of large areas using a hybrid of locally and whole-area calibrated models. Also, this moving window approach considerably reduces the computational time required to map SOC of large areas especially when the residual kriging is accompanied. It also improves the product accuracy and provides least biased predictions. There are artefacts produced by the moving window approach because of the un-even data distribution over space. Future work will attempt to merge the locally calibrated with the whole-area maps based on their uncertainty of prediction (Malone et al., 2011).

Table 5
Comparison of different mapping approaches using Mean Error (ME).

ME	SVR whole-area model (100 m)	SV regression kriging	Moving window MLR	Moving window MLRegression kriging (100 m)	TERN national grid (100 m)	ISRIC SoilGrid (1 km)
0–5 cm	−0.07	0.74	0.02	0.03	−0.44	0.04
5–15 cm	-0.03	0.52	0.01	0.02	-0.19	0.00
15–30 cm	0.05	-0.05	0.01	0.00	-0.75	-0.31
30–60 cm	0.02	-0.64	0.05	0.04	-0.26	-0.48
60–100 cm	-0.01	-1.08	0.01	-0.01	-0.28	-0.35

Acknowledgement

The authors are grateful to the Australian Department of Agriculture, Round 2 — Filling the Research Gap Program for supporting this research (grant number - 1194105-66) and we also extend our thanks to two anonymous reviewers whose perceptive comments improved the content presented in this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi: http://dx.doi.org/10.1016/j.geodrs.2015.12.002. These data include Google map of the most important areas described in this article.

References

Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H., 2014. Constructing a soil class map of Denmark based on the FAO legend using digital techniques. Geoderma 214, 101–113.
Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A.E., Lagacherie, P., McKenzie, N.J., 2014. The GlobalSoilMap project specifications. Globalsoilmap: Basis of the Global Spatial Soil Information System.
Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma 151 (3–4), 338–350.
Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69.
Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91 (1–2), 27–45.
Bui, E., Henderson, B., Viergever, K., 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.
Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2014. Soil organic carbon stocks in the Limpopo National Park, Mozambique: amount, spatial distribution and uncertainty. Geoderma 213, 46–56.
Cook, S.E., Corner, R.J., Groves, P.R., Grealish, G.J., 1996. Use of airborne gamma radiometric data for soil mapping. Aust. J. Soil Res. 34 (1), 183–194.
Dai, F.Q., Zhou, Q.G., Lv, Z.Q., Wang, X.M., Liu, G.C., 2014. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. Ecol. Indic. 45, 184–194.
de Brogniez, D., Ballabio, C., van Wesemael, B., Jones, R.J.A., Stevens, A., Montanarella, L., 2014. Topsoil organic carbon map of Europe. Soil Carbon.
Dorji, T., Odeh, I.O.A., Field, D.J., Baillie, I.C., 2014. Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, eastern Himalayas. For Ecol Manage 318, 91–102.
Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour Res 39 (12).
Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The elements of statistical learning: data mining, inference, and prediction. Springer, New York.
Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using, decision trees. Geoderma 124 (3–4), 383–398.
Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km—global soil information based on automated mapping. PLoS One 9 (8).
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10 (6).
Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120 (1–2), 75–93.
Holmes, G., Hall, M., Frank, E., 1999. Generating rule sets from model trees. In: Foo, N. (Ed.), Advanced Topics in Artificial Intelligence. Lecture Notes in Artificial Intelligence, pp. 1–12.
Jenny, H.A., 1941. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, New York.
Kanevski, M., Pozdnukhov, A., Canu, S., Maignan, M., 2002. Advanced spatial data analysis and modeling with Support Vector Machines. Int. J. Fuzzy Syst. 4 (1), 606–615.
Karunaratne, S.B., Bishop, T.F.A., Baldock, J.A., Odeh, I.O.A., 2014. Catchment scale mapping of measureable soil organic carbon fractions. Geoderma 219, 14–23.
Kidd, D.B., Malone, B.P., McBratney, A.B., Minasny, B., Webb, M.A., 2014. Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. Soil Res. 52 (2), 107–119.
Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311.
Lagacherie, P., 2008. Digital Soil Mapping: a state of the art. Digital Soil Mapping with Limited Data.
Lin, L.I., 1989. A concordance correlation-coefficient to evaluate reproducibility. Biometrics 45 (1), 255–268.
Malone, B.P., de Gruijter, J.J., McBratney, A.B., Minasny, B., Brus, D.J., 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.
Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154 (1–2), 138–152.
McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.
Miklos, M., Short, M.G., McBratney, A.B., Minasny, B., 2010. Mapping and comparing the distribution of soil carbon under cropping and grazing management practices in Narrabri, north-west New South Wales. Soil Res. 48 (3), 248–257.
Minasny, B., McBratney, A.B., 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J. 66 (2), 352–361.
Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. In: Sparks, D.L. (Ed.), Advances in AgronomyAdvances in Agronomy 118. Elsevier Academic Press Inc, San Diego, pp. 1–47.
Minasny, B., McBratney, A.B., Salvador-Blanes, S., 2008. Quantitative models for pedogenesis — a review. Geoderma 144 (1–2), 140–157.
Mishra, U., Lal, R., Liu, D.S., Van Meirvenne, M., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil Sci. Soc. Am. J. 74 (3), 906–914.
Moore, I.D., Lewis, A., Gallant, J.C., 1993. Terrain attributes: estimation methods and scale effects. Modelling Change in Environmental Systems.
Mora-Vallejo, A., Claessens, L., Stoorvogel, J., Heuvelink, G.B.M., 2008. Small scale digital soil mapping in southeastern Kenya. Catena 76 (1), 44–53.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes— heterotopic cokriging and regression-kriging. Geoderma 67 (3–4), 215–226.
O'Rourke, S.M., Angers, D.A., Holden, N.M., McBratney, A.B., 2015. Soil organic carbon across scales. Glob Chang Biol 21, 3561–3574.
Padarian, J., Minasny, B., McBratney, A.B., Dalgliesh, N., 2014. Predicting and mapping the soil available water capacity of Australian wheatbelt. Geoderma Reg. 2–3, 110–118.
Phachomphon, K., Dlamini, P., Chaplot, V., 2010. Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables. Geoderma 155 (3–4), 372–380.
Piccini, C., Marchetti, A., Francaviglia, R., 2014. Estimation of soil organic matter by geostatistical methods: use of auxiliary information in agricultural and environmental assessment. Ecol. Indic. 36, 301–314.
Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma 209, 1–14.
Rawlins, B.G., Scheib, A.J., Lark, R.M., Lister, T.R., 2009. Sampling and analytical plus subsampling variance components for five soil indicators observed at regional scale. Eur. J. Soil Sci. 60 (5), 740–747.
Quinlan, J.R., 1992. Learning with continuous classes. In Proc. of the Fifth Australian Joint Conference on Artificial IntelligenceWorld Scientific, Singapore, pp. 343–348.
Reeves, D.W., 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil Tillage Res. 43 (1–2), 131–167.
Rossel, R.A.V., Webster, R., Bui, E.N., Baldock, J.A., 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. Glob Chang Biol 20 (9), 2953–2970.
Smola, A.J., Scholkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.
Stevens, F., Bogaert, P., Van Oost, K., Doetterl, S., Van Wesemael, B., 2014. Regional-scale characterization of the geomorphic control of the spatial distribution of soil organic carbon in cropland. Eur. J. Soil Sci. 65 (4), 539–552.
Stormy weather, Bureau of Meteorology, Australia. available online at: http://www.bom.gov.au/nsw/sevwx/facts/stormy-weather.shtml. Retrieved August 12, 2015.
Subburayalu, S.K., Slater, B.K., 2013. Soil series mapping by knowledge discovery from an Ohio County soil map. Soil Sci. Soc. Am. J. 77 (4), 1254–1268.
Sun, W., Minasny, B., McBratney, A., 2012. Analysis and prediction of soil properties using local regression-kriging. Geoderma 171, 16–23.
Wang, Y., Witten, I.H., 1996. Induction of model trees for predicting continuous classes. (Working paper 96/23). Hamilton, New Zealand: University of Waikato, Department of Computer Science.
Wang, D., Wang, M., Qiao, X., 2009. Support vector machines regression and modeling of greenhouse environment. Comput. Electron. Agric. 66 (1), 46–52.
Whelan, B.M., McBratney, A.B., Minasny, B., 2003. VESPER 1.5 — spatial prediction software for precision agriculture. Proceedings of the 6th International Conference on Precision Agriculture and Other Precision Resources Management, Minneapolis, MN, USA, 14–17 July, p. 2002.
Wiesmeier, M., Barthold, F., Blank, B., Koegel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant and Soil 340 (1–2), 7–24.
Wynn, J.G., Bird, M.I., Vellen, L., Grand-Clement, E., Carter, J., Berry, S.L., 2006. Continental-scale measurement of the soil organic carbon pool with climatic, edaphic, and biotic controls. Global Biogeochem Cycles 20 (1).
Zhao, Z., Yang, Q., Benoy, G., Chow, T.L., Xing, Z., Rees, H.W., Meng, F.-R., 2010. Using artificial neural network models to produce soil organic carbon content distribution maps across landscapes. Can J Soil Sci 90 (1), 75–87.