Geoderma 154 (2009) 138–152

Contents lists available at ScienceDirect

Geoderma

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

Mapping continuous depth functions of soil carbon storage and available water capacity

B.P. Malone [a,*], A.B. McBratney [a,1], B. Minasny [a,2], G.M. Laslett [b,3]

[a] Faculty of Agriculture, Food & Natural Resources, The University of Sydney, JRA McMillan Building A05, NSW 2006, Australia
[b] CSIRO Mathematical and Information Sciences, Private Bag 10, Clayton South MDC, Vic 3169, Australia

ARTICLE INFO

Article history:
Received 25 April 2009
Received in revised form 2 October 2009
Accepted 13 October 2009

Keywords:
Digital soil mapping
Equal-area spline functions
Carbon storage
Available water capacity
Neural networks

ABSTRACT

There is a need for accurate, quantitative soil information for natural resource planning and management. This information shapes the way decisions are made as to how soil resources are assessed and managed. This paper proposes a novel method for whole-soil profile predictions (to 1 m) across user-defined study areas where limited soil information exists. Using the Edgeroi district in north-western NSW as the test site, we combined equal-area spline depth functions with digital soil mapping techniques to predict the vertical and lateral variations of carbon storage and available water capacity (AWC) across the 1500 km^2 area. Neural network models were constructed for both soil attributes to model their relationship with a suite of environmental factors derived from a digital elevation model, radiometric data and Landsat imagery. Subsequent fits of the models resulted in an R^2 of 44% for both carbon and AWC. For validation at selected model depths, R^2 values ranged between 20 and 27% for carbon prediction (RMSE: 0.30–0.52 log (kg/m^3)) and between 8 and 29% for AWC prediction (RMSE: 0.01 m/m). Visually, reconstruction of splines at selected validation data points indicated an average fit with raw data values. In order to improve upon our model and validation results there is a need to address some of the structural and metrical uncertainties identified in this study. Nevertheless, the resulting geo-database of quantitative soil information describing its spatial and vertical variations is an example of what can be generated with this proposed methodology. We also demonstrate the functionality of this geo-database in terms of data enquiry for user-defined queries.

1. Introduction

In order to benefit from the ecological and economical functions of soil in a sustainable way; land holders, corporate stakeholders and, governmental departments need access to quantitative soil information. Such information confers weight to decisions regarding the management of the land and soil resources. Even in most environmental and agricultural research, accurate, continuous soil attribute data is becoming increasingly important in computer simulation models and for the assessment and monitoring of soil resources (Hempel et al., 2008). To facilitate this need, this paper introduces a novel method for the prediction of user-defined, continuous soil properties to a given soil depth, across landscapes (at a fixed resolution) where only limited soil data exists.

The variation of soil properties down a profile is usually continuous (Ponce-Hernandez et al., 1986). Soil depth functions are often created to represent the depth-wise variation of soil properties. However, with traditional sampling of soil profile horizons, it is often assumed that the horizon value of a particular attribute represents the average value for that attribute for the depth interval of that horizon. With this paradigm, in effect what should be a continuous function, the data often appears discontinuous or stepped. Similarly, current digital soil mapping techniques are limited to map soil properties at specified depths or a combination of depth intervals (see Grimm et al., 2008; Stoorvogel et al., 2009 as recent examples).

Bishop et al. (1999) pointed out, the discontinuity of depth functions derived from bulk horizon data lends to inaccuracies when attempting to predict the value of an attribute at specific depths within a soil profile. Subsequently, many attempts have been made to derive continuous depth functions of soil attributes. The earliest known attempt was by Jenny (1941) who drew freehand curves between attribute data points that corresponded to the mid-point value of a horizon. Over time, more sophisticated methods have evolved for constructing continuous soil depth functions such as using exponential decay functions (Russell and Moore, 1968). Minasny et al. (2006) demonstrated that fitting exponential decay functions to carbon profile data resulted in an adequate quality of fit when attempting to

* Corresponding author. Tel.: +61 2 95684113.
E-mail addresses: b.malone@usyd.edu.au (B.P. Malone), A.McBratney@usyd.edu.au (A.B. McBratney), b.minasny@usyd.edu.au (B. Minasny), Geoff.Laslett@cmis.csiro.au (G.M. Laslett).
[1] Tel.: +61 2 93513214.
[2] Tel.: +61 2 90369043.
[3] Tel.: +61 3 95458018.

0016-7061/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2009.10.007

map carbon storage in the Lower Namoi Valley, NSW. Linear regression and polynomials to the *n*-th degree by least-squares fitting have also been additional methods for deriving continuous soil depth functions (Colwell, 1970; Moore et al., 1972). However, the disadvantage of these novel procedures is that the value of a property at any depth affects the form of the fitted function at all depths (Ponce-Hernandez et al., 1986). As a consequence, the inflexibility of these functions results in a varied quality of fit (Webster, 1978).

Irrespective of the identity of soil attribute, a more flexible and accurate method for fitting continuous functions of soil data is the use of smoothing splines (Erh, 1972) and equal-area spline functions as proposed by Ponce-Hernandez et al. (1986). Both Ponce-Hernandez et al. (1986) and Bishop et al. (1999) provide a good mathematical explanation of the operation of spline functions. Essentially, a spline function is a set of local quadratic functions tied together with 'knots' that describe a smooth curve through a set of points. Bishop et al. (1999) demonstrated their superiority over other continuous soil depth functions when they predicted a number of soil properties including soil pH, electrical conductivity (EC), clay content, organic carbon content, and gravimetric water content.

Clearly, continuous soil depth functions such as equal-area splines are advantageous for prediction of soil properties at specific depths. However, in a spatial context, a collection of spline functions for individual site observations will ultimately lead only to point observation data sets. To the parties concerned, such data will be of little use when they require continuous estimates of soil property variation across defined study areas or landscapes. The response to this demand has been answered partly in the way of digital soil mapping, where soil properties are mapped digitally based on their relationship with environmental variables (Minasny et al., 2008). The *scorpan* factors or environmental covariates as proposed by McBratney et al. (2003) provide a valuable predictive framework for determining soil variability in areas with limited soil data.

Given the predictive capabilities of soil depth functions and an explosion in the capabilities of digital soil mapping in areas with limited data (Lagacherie, 2008), it seems only logical for there to be an amalgam of both methods to quantitatively predict the vertical and lateral variations of soil properties across a defined area. Using agronomically important soil properties-soil carbon storage and available water capacity (AWC) as the subjects for prediction, this paper proposes a novel method of predicting both their spatial and depth-wise variation. This is achieved in a number of stages:

(i) Fitting of equal-area spline functions to soil carbon and AWC profile data
(ii) Assembly of a geo-database of environmental or *scorpan* factors for a defined study area where the point observations exist.
(iii) Derive a neural network model using the best available set of *scorpan* factors to predict the depth-wise variation of the two soil properties.
(iv) Extrapolate the spline parameters onto the wider study area where soil observations do not exist.
(v) Map the carbon storage and AWC of the entire study area to a depth of 1 m.
(vi) Demonstrate the functionality of the resulting soil geo-database for data enquiry.

2. Material and methods

2.1. Study area

The study site is situated in the lower valley of the Namoi River, near Narrabri (30.32S 149.78E), approximately 500 km NNW of Sydney, NSW, Australia. Within this area which covers approximately 1500 km^2, agriculture is the major landuse with irrigated cotton, wheat and pastoral farming being the predominant enterprises. Significant areas of native vegetation are also present, where it is mostly concentrated on the lower foothills of the Nandewar Range on the eastern flanks of the study site.

2.2. Environmental data

For the purpose of digital soil mapping, a number of environmental indices were sourced and interpolated onto a common grid of 90 m resolution, encapsulating the study area. These included:

- Landsat 7 ETM+ images from 2003: b1 (0.45–0.52 μm), b2 (0.52–0.60 μm), b3 (0.63–0.69 μm), b4 (0.78–0.90 μm), b5 (1.55–1.75 μm), and b7 (2.09–2.35 μm). The Landsat bands were used to approximate the biosphere as a soil forming factor in terms of generalised land cover and land use. Vegetation cover and type was approximated using the Normalised Difference Vegetation Index (NDVI) determined by using bands b3 and b4, where:

$$NDVI = (b4 - b3) / (b4 + b3).$$

Furthermore, the band ratios or more commonly, soil enhancement ratios of b3/b2, b3/b7 and b5/b7 were derived. It has been proposed that these soil enhancement ratios can accentuate carbonate radicals, ferrous iron, and hydroxyl radicals respectively in exposed soil (Saunders and Boettinger, 2007).

- Gamma-radiometric survey data (Geosciences Australia, 2008). The method of gamma-radiometric survey estimates the abundances of potassium (^{40}K), thorium (^{232}Th) and uranium (^{238}U) gamma-ray radiation emitted from the earth's surface. Cook et al. (1996) demonstrated that variations in the gamma-ray radiation of earth surfaces correspond with the distribution of various parent materials over the landscape.
- Digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) terrain data. From the DEM, first and second derivatives, namely: slope, aspect, terrain wetness index (TWI), flow length, slope length factor (LS-factor), area above channel network (AOCN) and stream power index (SPI) were determined. Moore et al. (1993) and McKenzie and Ryan (1999) provide exemplar studies where some or all of the derivatives have been used to derive relationships with the spatial distribution various soil properties.

2.3. The equal-area smoothing spline

The spline model we used is a generalisation of the quadratic spline model of Bishop et al. (1999). The model by Bishop et al. (1999) is when data are averages over adjacent horizons or layers in a soil profile. The model used in this paper is more general where the data are again averages of soil layers, but the supports of the data are not adjacent. Given measurements for soil properties at *n* layers in a soil profile, the boundaries of the layers are given in increments (u_1, v_1), (u_2, v_2),… (u_n, v_n), given that

$$u_1 < v_1 \leq_{u2} < v_2 \leq \ldots \leq_{un} < v_n.$$

The measurement of the bulk sample from layer i is assumed to reflect the mean attribute level, apart from measurement error. Mathematically, the measurements are modelled as

$$y_i = \overline{f}_i + e_i.$$

It is assumed that the true soil attribute values vary smoothly with depth. This is translated into mathematical terms. We denote depth by x, and the depth function describing the true attribute values by $f(x)$; which mean that $f(x)$ and its first derivative $f'(x)$ are both continuous, and that $f'(x)$ is square integrable. The depths of the boundaries of the n layers are given by $x_n < x_n, \ldots < x_n$, and $\overline{\text{fi}}$ is the

mean value of $f(x)$ over the interval (x_{i-1}, x_i) and e_i are measurement errors with mean 0 and variance σ^2. $f(x)$ represents a spline function, which can be found by minimising:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{f}_i)^2 + \lambda\int_{x_o}^{x_n}[f^m(x)]^2dx.$$

The first term represents the fit to the data, the second term measures the roughness of function $f(x)$, expressed by its first derivative $f'(x)$. Parameter λ controls the trade-off between the fit and the roughness penalty. The solution is a linear-quadratic smoothing spline, linear between layers, and quadratic within layers. See Appendix A for derivation.

2.4. Mapping the smoothing spline soil depth function

1. Collate legacy profile descriptions. They can have any kind of variation in any depth increments as long as they describe some of the variation in properties with depth. The soil dataset we used consists of 341 soil profiles. 210 were sampled on a systematic, equilateral triangular grid with a spacing of 2.8 km between sites (McGarry et al., 1989). The further 131 soil profiles are distributed more irregularly or on transects (Fig. 1). The dataset describes and quantifies various soil morphological, physical and chemical attributes at depth increments of 0–0.1, 0.1–0.2, 0.3–0.4, 0.7–0.8, 1.2–1.3 and 2.5–2.6 m.

The focus of this study examines the vertical and lateral variabilities of carbon storage and AWC across the Edgeroi area at the time of measurement (1985–1987). The units of measurement used for carbon and AWC are kg m^{-3} and m/m respectively for each depth interval. Soil carbon was measured and calculated by McGarry et al. (1989). Meanwhile, AWC was derived from a pedo-transfer function using sand, clay and organic carbon as predictors (Minasny et al., 2006).

2. Fit the spline to the values for each property. This generates a continuous profile description for each legacy soil profile. The maximum depth of our fitted splines to the legacy profile data was 1 m. The spline function depends on a smoothing parameter lambda (λ). For this parameter lambda values of 10, 1, 0.1, 0.01, 0.001, 0.0001 and 0.00001 were tried. The 'best' lambda value coincided with the spline that had the highest prediction quality i.e. the lowest root mean square error (RMSE). The 'best' lambda values were recorded and then assessed to determine the most frequently occurring value for each variable. This value was then used as a blanket value to re-fit splines for all data points.
3. From the fitted spline, derive the mean value of the soil property within defined depth increments. For this study, the mean values at depth increments; 0–10, 10–20, 20–30, 30–40, 40–50, 50–70,

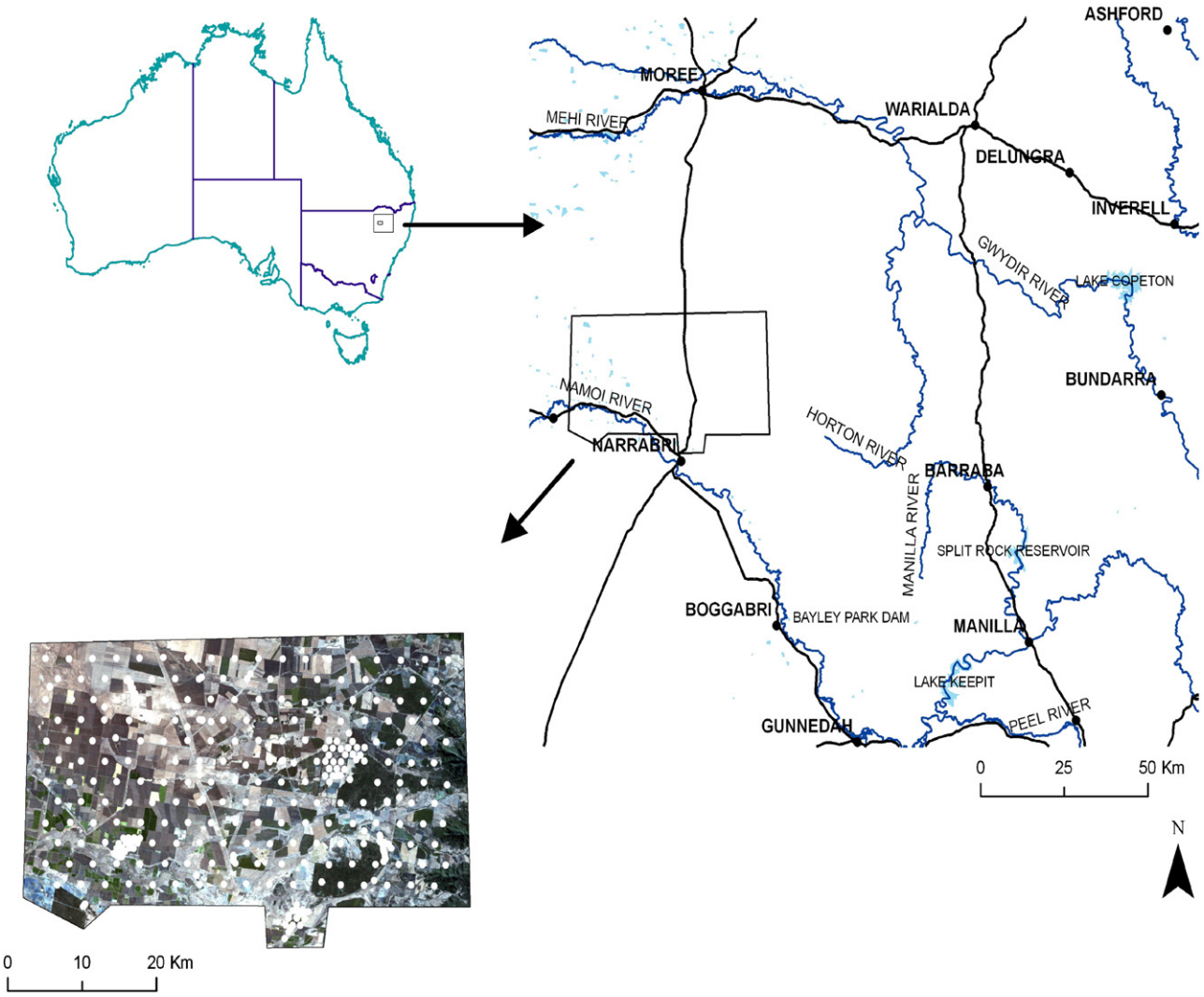


Fig. 1. The Edgeroi study area.

Table 1
Accuracy of constructed neural networks for AWC with respect to iterative changes to the number of nodes.

Nodes	R^2	Cross-validated R^2
1	9%	38%
2	34%	19%
3	44%	15%
4	52%	4%
5	59%	0%

70–80 and 80–100 cm were derived from the splines. The values for each property for each of the depth increments become the input for training data sets that are modelled against environmental covariates. The expectation is that there will be a number of surfaces (in our case — 8) for any region of interest. One surface for each of the proposed depth increments.

4. Implement a model framework to derive relationship between the training data set and environmental covariates. Firstly, we joined the model inputs (based on their spatial location) to the environmental data using the nearest neighbour method. Stepwise regression was used to determine the best combination of environmental variables to predict both carbon content and AWC.

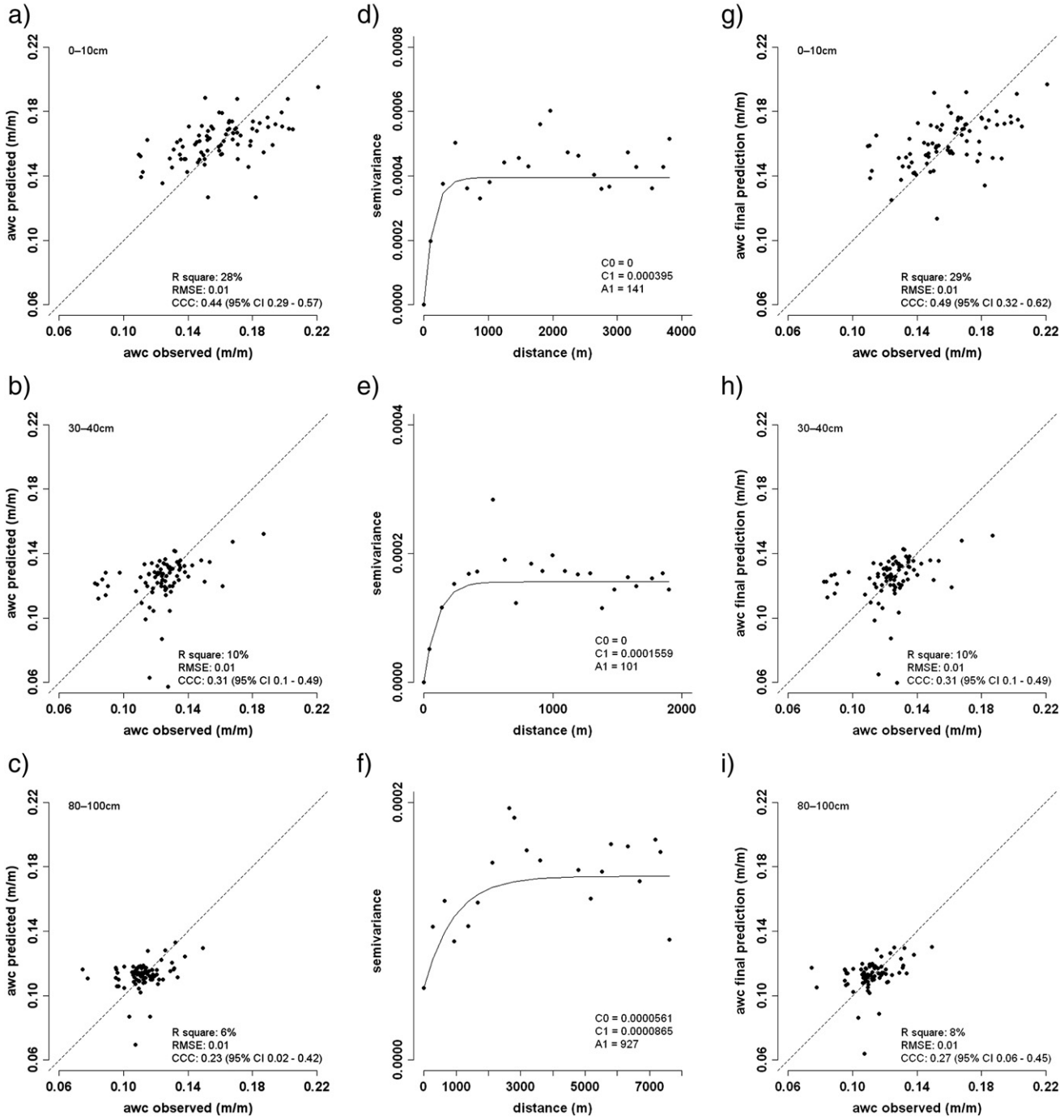


Fig. 2. Carbon neural network prediction vs. observed plots from 80 randomly selected validation points at a) 0–10 cm b) 30–40 cm and c) 80–100 cm. Semi-variogram models of the residuals at each prediction depth (d–f). Final prediction (model prediction + residual) vs. observed plots at each prediction depth (g–i). RMSE: root mean square error. CCC: Lin's concordance correlation co-efficient.

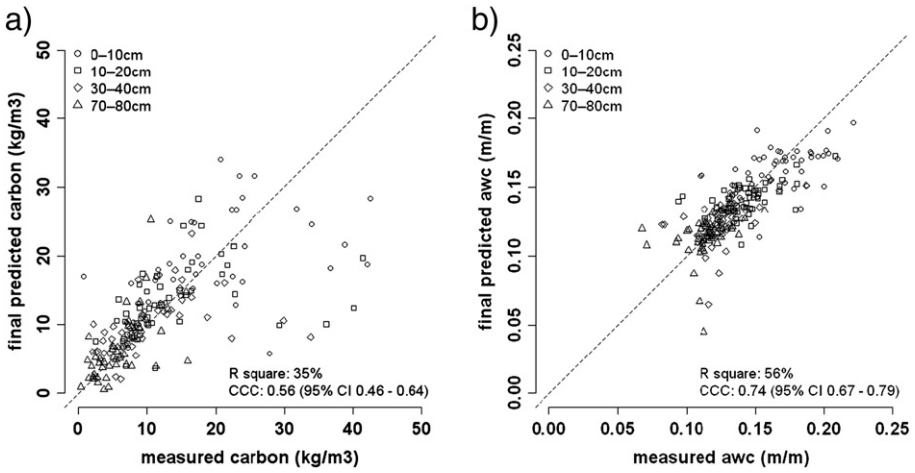


Fig. 3. Validation of measured soil attribute (McGarry et al., 1989) vs. final predictions at 0–10 cm, 10–20 cm, 30–40 cm, and 70–80 cm for a) carbon and b) AWC. CCC: Lin's concordance correlation co-efficient.

A systematic approach was implemented for modelling the training data against the environmental covariates. Successive neural networks were constructed for a training dataset each soil attribute, each time adjusting model complexity by way of increasing the number of hidden nodes. We used neural networks on the basis that predictions can be made simultaneously at each depth interval for each iteration. For each iteration, 33% of the training data was used for cross-validation purposes in order to assess model over-fitting. Therefore, without compromising model predictive capability, we opted for models that also had a reasonable predictive capability for the cross-validated training data. Table 1 below shows this systematic model construction and selection process (using AWC as the example). Here we chose the third option (3 nodes).

Once selected, formulae for each model depth interval were saved for later use to predict in areas where data observations were not available. Residuals (raw depth interval value derived from spline minus modelled value) were calculated and then kriged (local) onto the common 90 m grid as used for the environmental factors to determine the spatial pattern random error.

5. Interpolate model rules or formulae onto the study area where information only relating to the environmental covariates exists. The kriged residuals (from Step 4 of general procedure) were added to the predictions. Ultimately this resulted in a final prediction for each modelled depth interval (0–10, 10–20, 20–30, 30–40, 40–50, 50–70, 70–80 and 80–100 cm) at each pixel within the study area. For each soil attribute splines were reconstructed using the lambda parameter (from Step 2) and predicted value of each soil attribute at each depth interval as inputs. From this, maps displaying the total cumulative value of each soil attribute were produced.

Additionally, for demonstration of functionality, the resulting geodatabase of soil information generated in this study was queried for the following three scenarios:

1) At what depth does soil carbon first decrease to below 1%?
2) At what depth in the soil can we find the cumulative sum of carbon equal 5 kg m^{-2}?
3) What is the lowest depth at which total AWC equal 100 mm?

Maps were produced in order to visualise the results of each scenario.

2.5. Model validation

For model validation, the profile formulae were applied to 80 validation data points selected randomly from the original dataset. Residuals, coinciding with the location of each validation point were extracted from the 90 m grid of residuals then added to the estimated depth values, equating to a final prediction. To visually assess the fit of predictions against observed legacy soil information, splines for selected validation data points were reconstructed using the 8 predicted depth increments and defined lambda parameter (from Step 2) as inputs.

3. Results

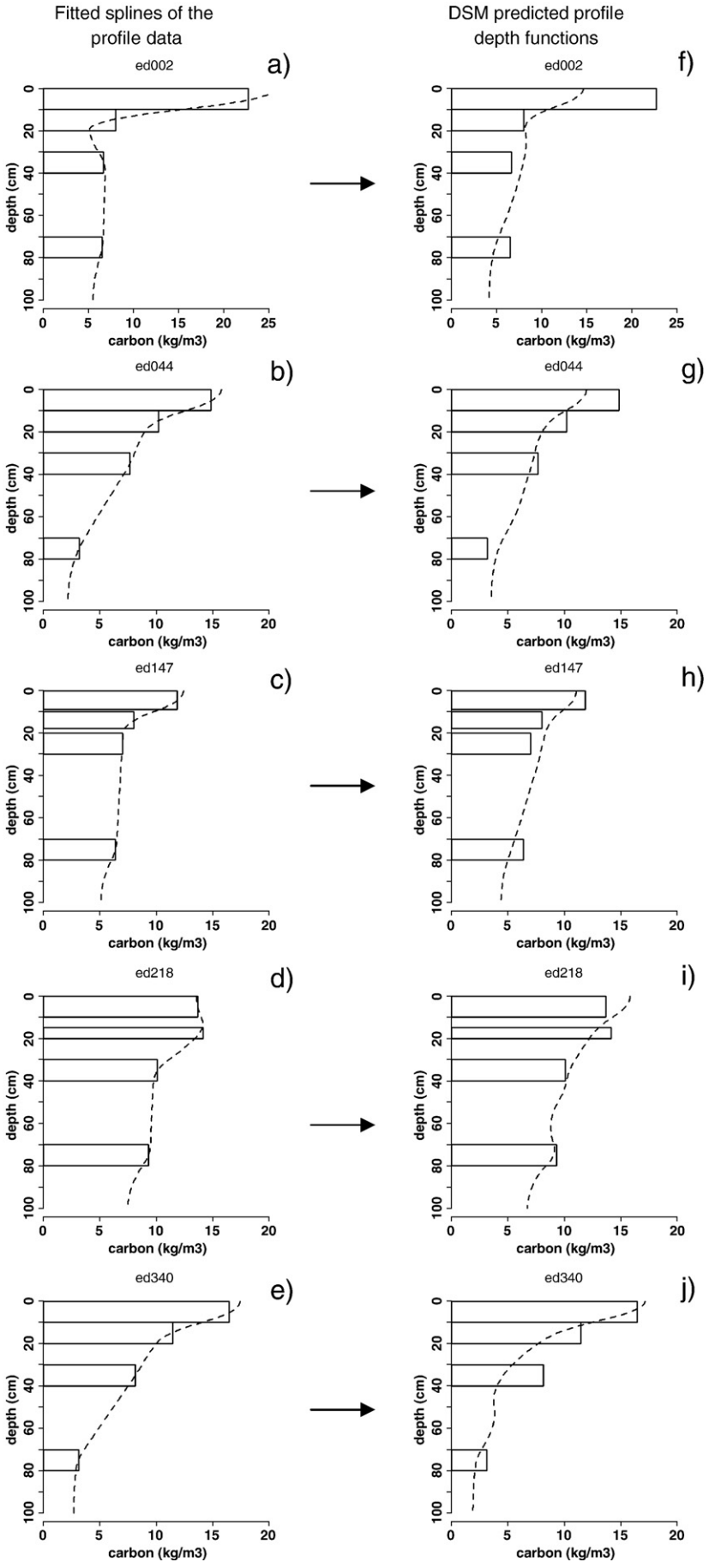
3.1. Fitting equal-area splines to the dataset

The raw carbon data displayed a log-normal distribution and subsequently was log-transformed prior to fitting the splines. The data for AWC did not require any transformation. For carbon the standard deviation of the log-transformed data was found to be 0.89 with data values ranging from 0.0–4.7 log(kg m^{-3}). Overall, the best fitting splines were found to have a lambda (λ) value of 0.01. AWC values ranged from 0.05 to 0.23 m m^{-1} with a standard deviation of 0.03. The best fitting splines for AWC were also found to have a λ value of 0.01.

3.2. Stepwise regression of environmental factors

Elevation, slope, radiometric *K*, band 5 and band ratios 3/7 and 5/7 were found to be strong covariates for both carbon and AWC predictions. Other environmental factors of importance for carbon prediction included altitude above channel network, stream power index and bands 3 and 4. While for AWC, LS-factor, terrain wetness index and NDVI were strong prediction covariates.

Fig. 4. Fitted splines (dashed lines) of observed carbon profile data (polygons) at five randomly selected sites (a–e). Digital soil map prediction depth functions of carbon (dashed lines) and observed carbon profile data (polygons) at same selected sites (f–j).


Fitted splines of the profile data
DSM predicted profile depth functions
ed002
a)
f)
ed044
b)
g)
ed147
c)
h)
ed218
d)
i)
ed340
e)
j)
depth (cm)
carbon (kg/m3)

3.3. Neural networks for prediction of carbon and AWC from environmental factors

From the systematic approach of neural network training, we selected models for carbon and AWC that featured 2 and 3 hidden nodes respectively as being the most appropriate (compromise between model predictive capability and over-fitting). Both neural networks resulted in fits where the R^2 value was 44%. Cross-validation of the training datasets resulted in R^2 values of 10% and 14% for carbon and AWC respectively. Depth wise, the prediction of carbon was best between 20 and 70 cm where the RMSE ranged between 0.27 and 0.30. The top of the soil profile was adequately predicted where 50% of the data variation for that layer could be explained. The predictions of the depth function for AWC were similar in that predictions were best between 20 and 70 cm. The least predictive estimates were found at the top of the profile (0–10 cm) followed by predictions at the bottom of the profile (80–100 cm).

Model residuals at each depth interval were kriged using local neighbourhood prediction models. For both carbon and AWC, whilst there was some spatial patterning of residuals looking at each depth increment independently, there was no similarity in the spatial distribution of residuals when comparisons were made between each depth increment. Overall, a general observation was that there was only a slight degree of spatial auto-correlation of residuals for both soil attributes.

3.4. Model validation with the 80 withheld data points

Validation of AWC indicated that model fits were significantly better than in the top 20 cm of soil compared to the rest. The R^2 values for the top two depth increments prior to the addition of residuals were 28% (0–10 cm) and 25% (10–20 cm). For the remaining soil profile (20–100 cm), R^2 values ranged between 6 and 12%. Fig. 2a–c illustrate the observed vs. fitted plots at the selected depth increments of 0–10 cm, 30–40 cm and 80–100 cm respectively prior to the addition of residuals. At these depths, Lin's concordance coefficients (CCC) range between 0.23 and 0.44 indicating a moderate agreement, with the strongest agreement for the 0–10 cm depth increment. While there was some spatial pattern in the distribution of residuals (Fig. 2d–f), their addition to predicted estimates of AWC made only a little improvement on the final predictions where R^2 values ranged between 8 and 29% (RMSE: 0.01). CCC also indicate a modest improvement in predictions resulting from the addition of residuals (CCC: 0.27–0.49). Fig. 2g–i illustrate the observed vs. final prediction plots at the aforementioned selected depth increments.

Prior to the addition of residuals, validation of the neural networks on the 80 withheld data points indicates that the accuracy of carbon prediction decreased for each depth interval. For the top 40 cm of the soil profiles, model fits resulted in R^2 values between 17 and 24%. For the bottom 60 cm, R^2 values ranged between 13 and 15%. Despite the fact that there was not a very well defined spatial distribution of residuals at any significant separation distances >160 m (data not shown), the addition of residuals to predictions had an overall improvement on model fits at all depth increments where R^2 values ranged between 20% and 27% (RMSE: 0.30–0.52).

For an additional validation; the majority of the raw data of each soil variable were measured at specified depth ranges, for example, at 0–10 cm, 10–20 cm, 30–40 cm and 70–80 cm among others. Validation results at these specific depth increments indicate good fits for both carbon and AWC (Fig. 3a–b). For carbon (R^2: 35%), the strongest agreements between the measured and final predicted carbon were in the 30–40 cm and 70–80 cm depth ranges. Conversely for the 0–10 cm and 10–20 cm it can be seen that there are a greater proportion of systematic deviations from the 45° line. The resulting CCC of 0.56 is indicative of this. For AWC (R^2: 56%), there was overall a good agreement between the measured and final predicted values (CCC: 0.74), with no obvious deviations at specific depth increments.

Five data points were selected at random from the 80 validation points to graphically represent model predictions with observed data. These representations are shown in Figs. 4f–j and 5f–j for carbon and AWC respectively. The polygons represent the measured value at the specified depth increment. These figures are compared to the spline fits of the observed data calculated in the first stage of this study (Figs. 4a–e and 5a–e). Carbon values were back-transformed after the construction of the modelled spline estimates. Comparing the modelled spline functions fitted to the raw data indicate that there is an average agreement between the predictions and the observed values. While not fitting exactly to the raw data, the splines are sensitive to actual changes in carbon down the profile and follow the general trend of carbon distribution as from the raw data. For these graphical examples, the greatest errors of prediction were found at the surface, particularly at *ed002* and *ed044* (Fig. 4f and g). At these sites the actual vs. predicted values of volumetric carbon differed by up the 7 kg m^{-3}.

For AWC prediction there was also only a fair agreement between the predicted spline depth functions and raw values (Fig. 5f–j). In general the largest disagreement between predicted and actual values was found at the soil surface, particularly at *ed033*, *ed039*, and *ed157* (Fig. 5f, g, and i) where AWC differed by up to 0.05 m/m.

3.5. Mapping carbon storage and available water capacity

Total carbon storage in the soil of the study area ranged between 1 and 50 kg m^{-2} to a depth of 1 m (Fig. 6a). The total average carbon storage was 9.5 kg m^{-2}, with the highest levels found to the eastern and southern sections of the area (8–50 kg m^{-2}). From a generalised perspective these areas coincide with particular land uses not dedicated to cropping for example in forested areas, along watercourses and grazing areas. The cropping areas, situated in the northern and western sections of the area have the lowest carbon storage (1–7 kg m^{-2}). This trend is similar for the description of the spatial variability of AWC. Here, AWC ranged between 76 and 169 mm to a depth of 1 m, with an average of 125 mm (Fig. 6b). For the cropping areas to the north-west, AWC ranged mostly between 91 and 120 mm.

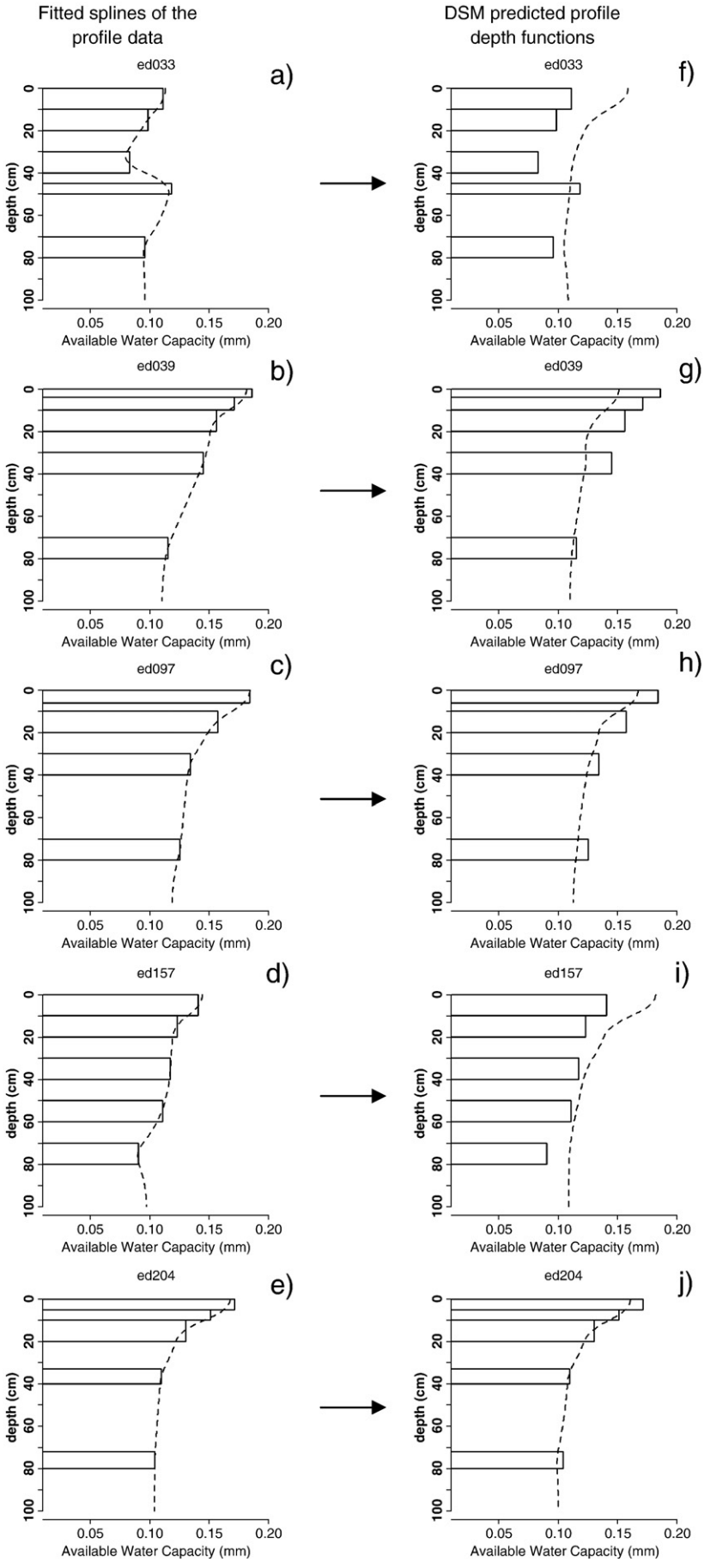
The whole-soil profile maps (displayed as depth increments) of carbon storage and available water capacity are shown in Figs. 7 and 8. It can be observed that carbon is more abundant in the surface layers than in the sub-surface layers. The amount of carbon in the surface layer (0–10 cm) of the cropping soils ranges between 6 and 9 kg m^{-3}. While for areas with natural vegetation, volumetric carbon ranges between 15 and 51 kg m^{-3} at the soil surface.

AWC in the 0–10 cm layer across the study area is relatively homogeneous where it ranges between 0.12 and 0.205 m m^{-1}, the lowest values occurring where cropping is practiced in the east of the study area. In all cases AWC decreases with the increase in soil depth. However, the rate of decrease in soil water is higher in areas where the land cover is either under native vegetation or pasture.

3.6. Scenario-based queries of the generated soil geo-database

By calculation, the geo-databases generated in this study each contain over 200 million bits of information (210370 lateral or grid

Fig. 5. Fitted splines (dashed lines) of observed AWC profile data (polygons) at five randomly selected sites (a–e). Digital soil map prediction depth functions of AWC (dashed lines) and observed AWC profile data (polygons) at same selected sites (f–j).


Fitted splines of the profile data
DSM predicted profile depth functions
ed033
a)
f)
ed039
b)
g)
ed097
c)
h)
ed157
d)
i)
ed204
e)
j)
depth (cm)
0
20
40
60
80
100
0.05
0.10
0.15
0.20
Available Water Capacity (mm)

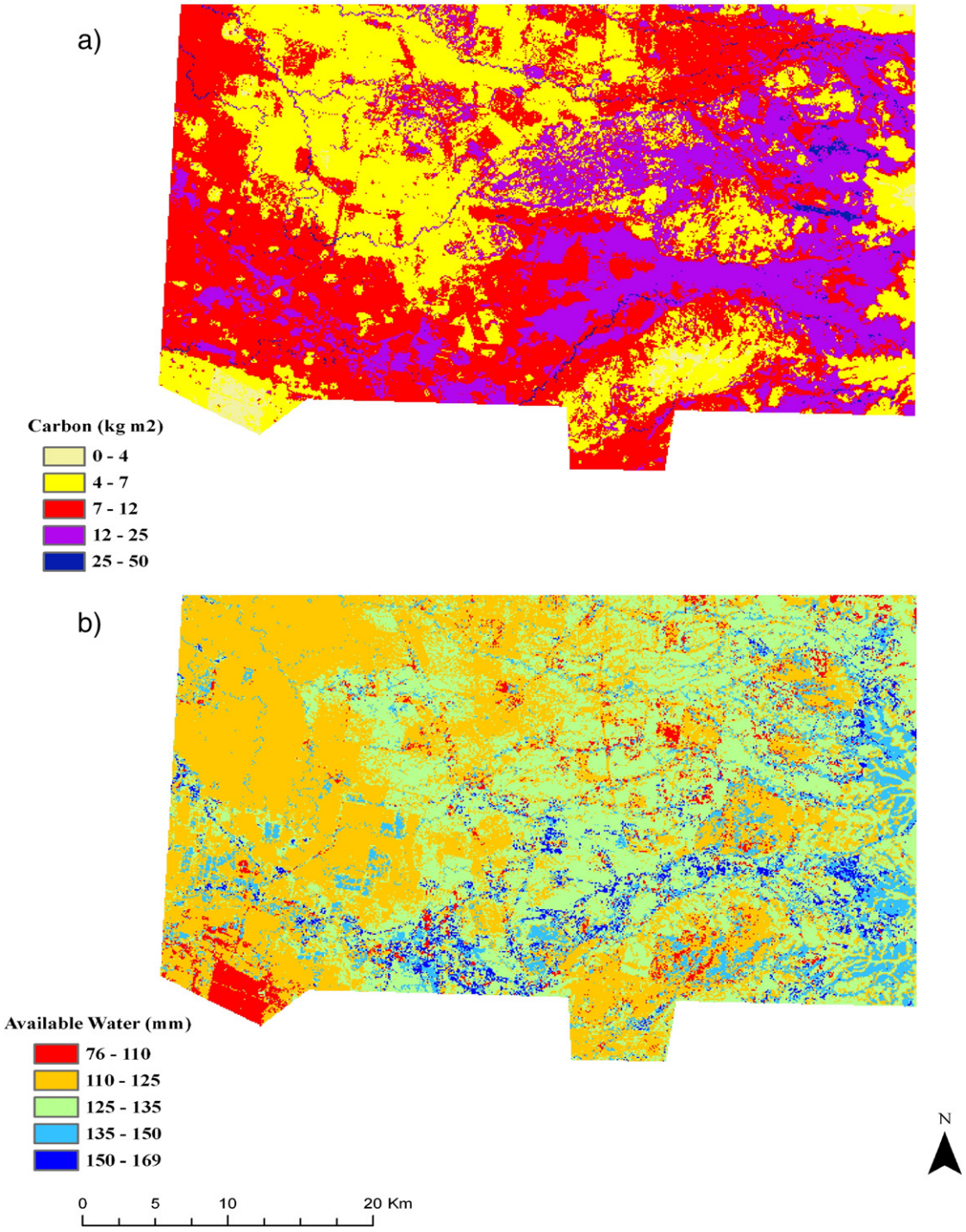


Fig. 6. Predicted total carbon (kg m^{-2}) and available water capacity (mm) to a depth of 1 m across the Edgeroi study area.

square observations × 100 depth observations) that describe the vertical and lateral variabilities of carbon storage and AWC in the Edgeroi area.

As can be observed in Fig. 9a the depth at which soil carbon drops below 1% is quite variable across the study area where it ranges from 1 cm to over 1 m, with the average depth at 21 cm. The cropping areas situated mostly to the western areas of the study area tend to have the highest concentration of soils where in the top 5 cm of soil, soil carbon falls below 1%. In most cases in these areas, even at 2 cm below the surface, soil carbon has already decreased to below 1%. Conversely, the areas that do not appear to be cropped maintain soil carbon levels above 1% to greater depths. The range of depths at which soil carbon decreases to below 1% is much larger than that observed in the areas where cropping is practiced and would be predominantly due to land use (grazing as apposed to dense vegetation etc) and other factors such as parent materials and proximity to waterways.

There is some correlation between the depth at which soil carbon decreases to below 1% and depth at which the cumulative sum of carbon equals 5 kg m^{-2} (Fig. 9b). This simple relationship highlights the negative exponential distribution of carbon in a soil profile, where carbon is most concentrated at the surface but decreases exponentially

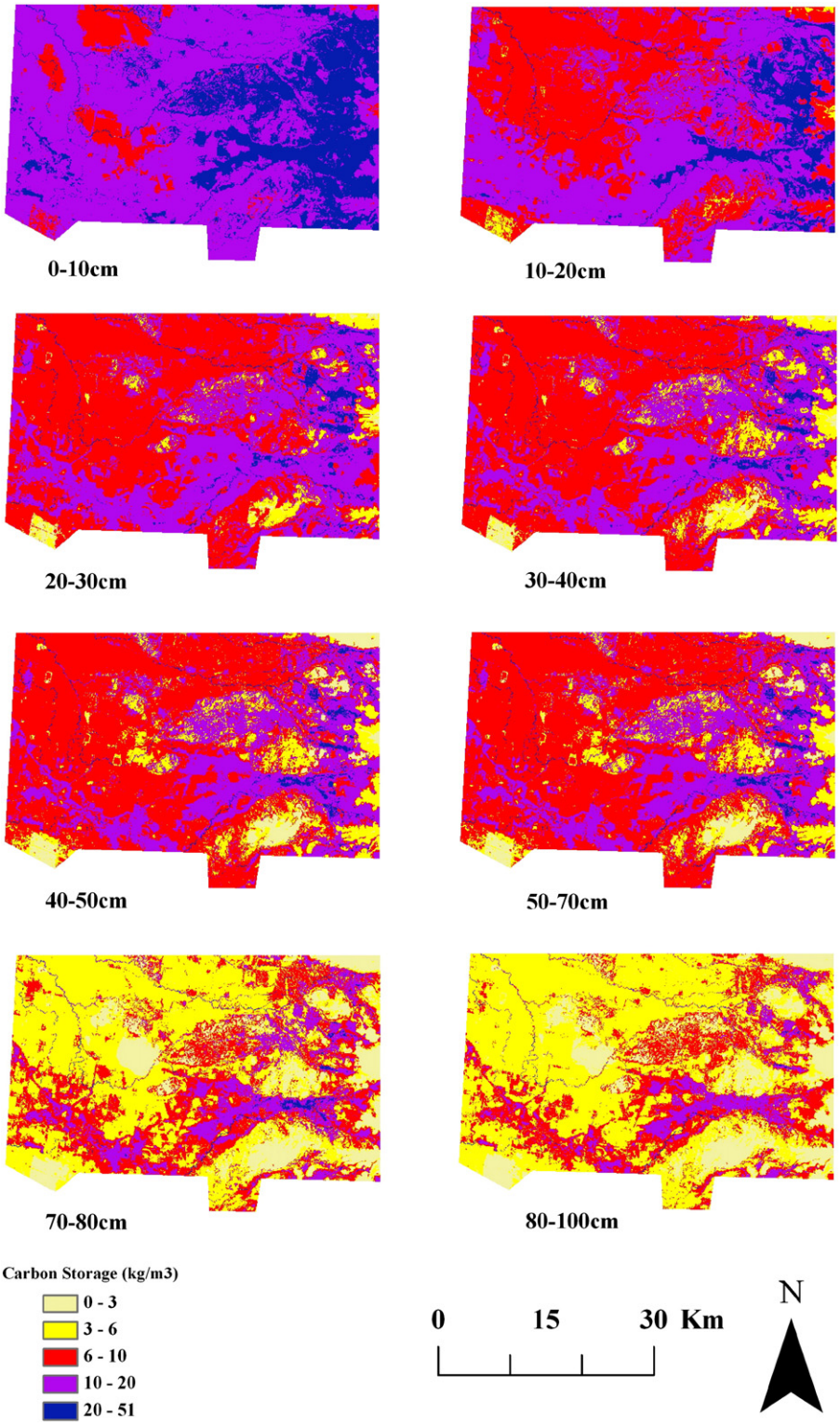


Fig. 7. Predicted soil profile carbon (kg/m^3) to 1 m displayed in 8 profile layers.

with increased soil depth. Therefore, the soils with low carbon storage i.e. greater depths required to attain 5 kg m^{-2} are the ones that have initially low concentrations of carbon at the soil surface. Despite a number factors determining the variability of carbon across the study area the average depth required to attain 5 kg m^{-2} was found to be 50 cm.

In terms of the soil depth required to attain 100 mm of AWC (Fig. 9c), the results indicated that the shallowest depth needed was 55 cm. However the frequency of this phenomenon occurring was relatively sparse and tended to be concentrated close to waterways or sources of water. Nevertheless the average depth required to attain 100 mm of AWC was 79 cm. There were some sparse areas where greater than 1 m was required. While variability in depth required does not appear significant, which could be due to a similarity in climate, a pattern of land use effect appears evident. Here the areas that have cropping have marginally less AWC than those that do not, reflecting both an increased pressure on AWC to sustain crops and an increase in evaporation due to cultivation effects.

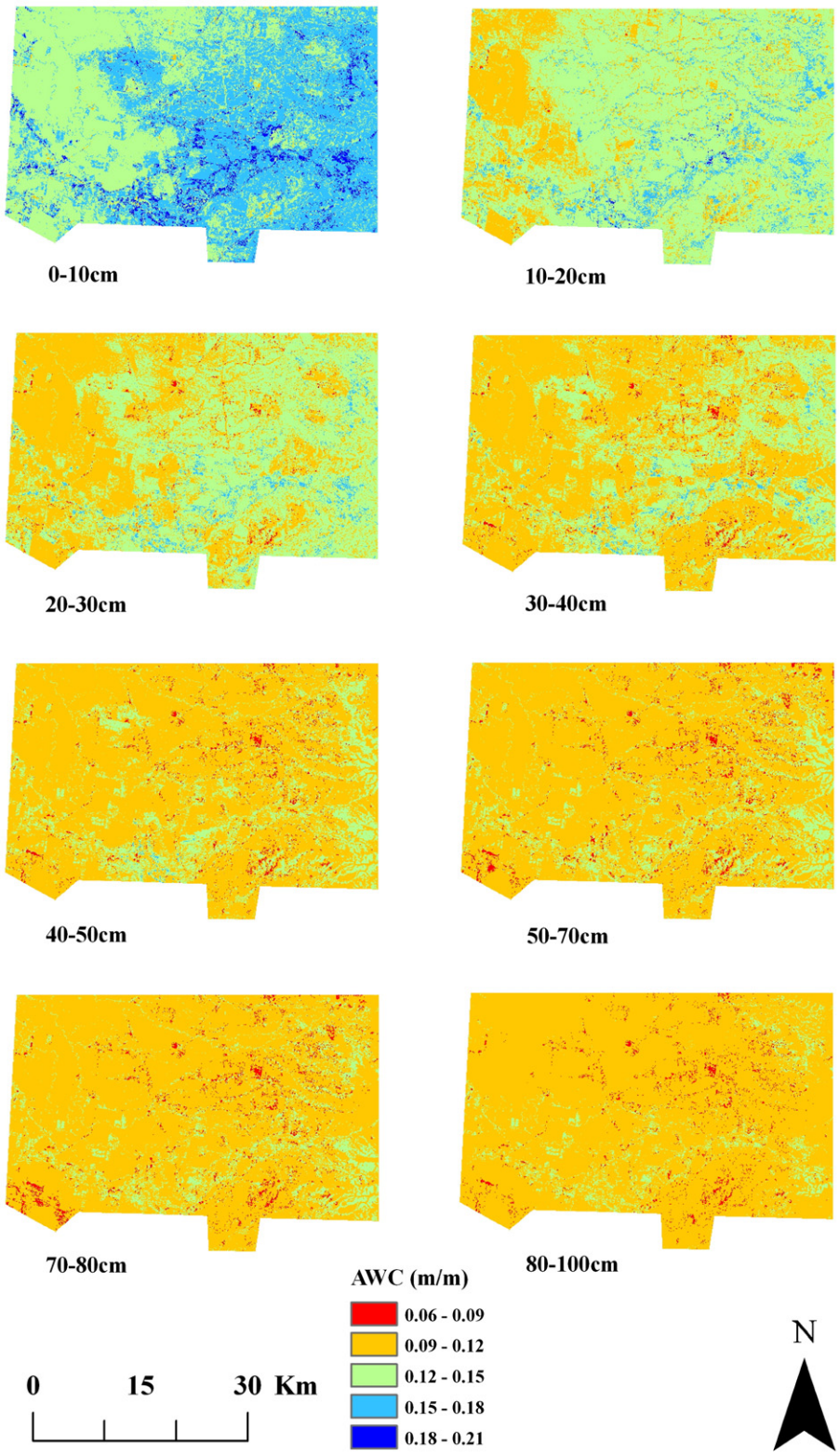
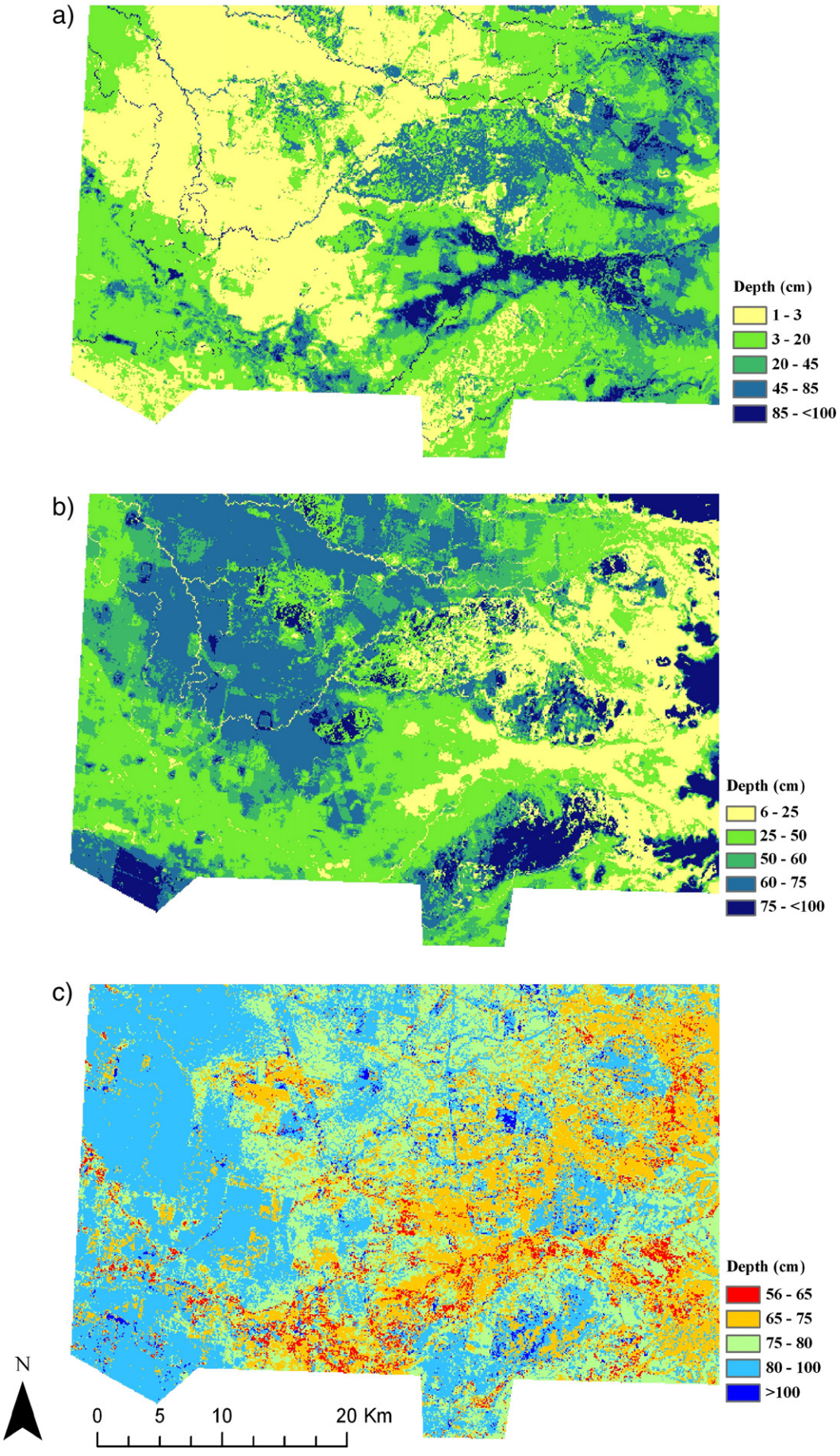


Fig. 8. Predicted soil profile available water capacity (m/m) to 1 m displayed in 8 profile layers.

4. Discussion

The Edgeroi area represents an ideal location for testing the methodologies of this study. First, the placement of survey sites in a mostly equilateral grid means that site location is random with respect to the topography, landuse and soil type (McGarry et al., 1989). Secondly, the density of observations at each site ensures that the vertical distribution of soil properties is sufficiently represented.

Fig. 9. Maps of scenario-based queries. a) Depth at which soil carbon decreases to below 1%. b) Depth at which cumulative total of soil carbon equals 5 kg m^{-2}. c) Depth at which cumulative sum of AWC equals 100 mm.


a)
Depth (cm)
1 - 3
3 - 20
20 - 45
45 - 85
85 - <100
b)
Depth (cm)
6 - 25
25 - 50
50 - 60
60 - 75
75 - <100
c)
Depth (cm)
56 - 65
65 - 75
75 - 80
80 - 100
>100
N
0 5 10 20 Km

Our prediction models of carbon and AWC were able to account for 44% of the variation of these properties across the study area. Similar accuracy assessments of models have also been reported for other digital soil mapping studies, for example, Ryan et al. (2000) and Florinsky et al. (2002). Broadly, these results are acceptable given that for quantitative soil spatial models, R^2 over 70% are unusual and values of 50% or less are common (Beckett and Webster, 1971). Our studies show that while the shape of the predicted spline depth functions is determined by the modelled depth increments, their flexible and sensitive nature makes them quite conducive for digital soil mapping purposes as we can fit them to any type of soil property and then derive information from them to use as model inputs. Post-modelling, the reconstruction of splines results in densely populated datasets with fine vertical resolution (1 cm) which can be queried as per the intentions of the user.

Validation results (R^2) of between 20–27% and 8–29% for carbon and AWC indicated that the predictions were not as good as those generated by the prediction model. These results are confounding, given that our predictions at specific measured depths of the soil profile were of much greater agreement. Nevertheless, we did not expect to improve upon the results of the calibration models as we are trying to predict soil depth functions just from a suite of environmental variables. Many studies of mapping continuous soil properties are rarely validated. Studies that have, such as Minasny et al. (2006), ($R^2 = 50\%$) and Stoorvogel et al. (2009) ($R^2 = 8$–23%) as examples, also reported average validation results. In this study, average validation results could have stemmed from the fact that the models performed better for predicting surface soil conditions compared to sub-surface conditions (Minasny et al., 2006). Particularly for AWC, there was significant error propagation with depth. While the environmental variables that were used to predict carbon and AWC are important sources of variability, it is likely we have not fully realised other sources of variability. In other words, we may have captured most of the variation of these soil properties at the soil surface using the existing data sources but we need to explore and seek out current and new data sources that will also explain their variation in the soil sub-surface. Additionally, we also found that the spatial distribution of residuals displayed very little patterning. This affected the final model predictions in that only a slight improvement was achieved when they were incorporated into predictions. Other studies such as Odeh et al. (1995) and Stacey et al. (2006) have reported greater success using the regression kriging approach we implemented for this study.

Given that we need to address the predictive performance of our models, there should not be an expectation that modelled observations will later fit exactly with the raw data. For each step in this methodology, there has been propagation of some degree of uncertainty or error in prediction. This uncertainty could be translated as being a combination of both metrical and structural uncertainty (Rowe, 1994). Metrical uncertainty is unavoidable because our reliance on models to define real objects. For example, calculating the various terrain attributes from a digital elevation model is derived from a series of polynomial equations of various orders fitted using least-squares (Skidmore, 1989). However, structural uncertainty in this study would be the most prevalent due to the fact that we are modelling environmental phenomena. Natural systems are inherently complex and difficult to define (Young, 1998); a collection of environmental factors alone will never account for the entire variability of natural processes. In a lot of cases, models of natural systems do not account for interactions between factors, or more often there are a whole suite of other factors which are not considered or difficult to interpret (Young, 1998). This brings us back to the point discussed previously where new and alternative data sources are needed if we want to capture more of the soil variability in both lateral and vertical spaces.

While it is likely that improvements can be made if we follow these suggestions, what we really want is to be able to account for the known and unknown structural and metrical uncertainties within our predictions. Estimation of the uncertainty for model outputs can be approached via a number of alternatives which include forecasting model outputs probabistically; analysing the statistical properties model outputs and observed data; usage of simulation and resampling based techniques; or via methodologies based on fuzzy theory and machine learning techniques (Shrestha and Solomatine, 2006). The study completed by Shrestha and Solomatine (2006) propose a novel method whereby they express uncertainty in the form of two quantiles (prediction interval). Future studies will have to explore these ideas further, but expressing a prediction within a defined interval rather than having a single estimate seems an appropriate route to follow when we are modelling difficult environmental processes over large spatial scales.

Overall, the maps of carbon storage and AWC are the 'end-product' of a richly populated dataset of their variability with depth and across the Edgeroi area. At first glance, interpretations can be made to describe the pattern of variability. More importantly, behind these maps is an invaluable geo-database of quantitative soil information suited to the requirements of end-users for the assessment and monitoring of soil resources. The versatility of this data was demonstrated by the three scenarios that queried the underlying geo-database.

5. Conclusions

- Spline functions are sensitive and flexible to the variation of both carbon and AWC with soil depth and are thus quite amendable to use within the digital soil mapping framework.
- This study identifies two types of predictive uncertainty — structural and metrical. Our validation results indicate there is a need to address these forms of uncertainty. By incorporating a measure of uncertainty within predictions, improving the model calibration process and using new and existing alternative data sources as model variables, we envisage that more reliable estimates can be generated to describe the lateral and vertical variations of soil in prescribed study areas.
- This study provides an example where a rich soil attribute geo-database can be generated from a limited soil dataset. We highlighted the functionality of this geo-database in terms of data enquiry for user-defined purposes.

Appendix A

The following is the derivation for the quadratic smoothing spline:

Given measurements for soil properties at n layers in a soil profile, the boundaries of the layers are given in increments (u_1, v_1), (u_2, v_2),… (u_n, v_n), given that

$$u_1 < v_1 \leq u_2 < v_2 \leq \ldots \leq u_n < v_n.$$

The measurement of the bulk sample from layer i is assumed to reflect the mean attribute level, apart from measurement error. Mathematically, the measurements are modelled as

$$y_i = \overline{f_i} + e_i. \quad (1)$$

It is assumed that the true soil attribute values vary smoothly with depth. This is translated into mathematical terms. We denote depth by x, and the depth function describing the true attribute values by $f(x)$; which mean that $f(x)$ and its first derivative $f'(x)$ are both continuous, and that $f'(x)$ is square integrable.

The depths of the boundaries of the n layers are given by $x_n < x_n$,… $< x_n$. Where $\overline{f}_i$ is the mean value of $f(x)$ over the interval (x_{i-1}, x_i) and

e_i are measurement errors with mean 0 and variance σ_2. $f(x)$ represents a spline function, which can be found by minimising:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{f}_i)^2 + \lambda\int_{x_o}^{x_n}[f'(x)]^2dx. \quad (2)$$

The quadratic spline

We define quadratic spline $s(x)$, in each layer, it conforms to a quadratic polynomial $p(x)$. The polynomials $p_i(x)$ and $p_{i+1}(x)$ for two adjacent layers meet smoothly at the boundary. The curve is given by

$$s(x) = p_i(x) \text{ for } x_{i-1} \le x \le x_i, i = 1, 2, \ldots, n.$$

The smoothness conditions are

$$p_i(x_i) = p_{i+1}(x_i)$$

$$p_i'(x_i) = p_{i+1}'(x_i)$$

for $i=1, 2,\ldots, n-1$, and

$$p_1(x_0) = 0$$

$$p_n'(x_n) = 0.$$

The latter two conditions mean that $s(x)$ is a natural spline. The points $(x_i, s(x_i))$ at the layer boundaries are called knots, and each x_i is referred to as a knot location. We define

$$f_i = s(x_i) = p_i(x_i).$$

$$b_i = s'(x_i) = p_i'(x_i).$$

for $i=1, 2,\ldots, n$.

Quadratic polynomials

A quadratic polynomial can be written as:

$$p(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

with coefficients $\beta_0, \beta_1, \beta_2$. In the case where the polynomial is over depth interval (t, u) where $u>t$, then the coefficients can be determined from $p'(t), p'(u), \overline{p} = \int_t^u p(x)dx / (u-t)$.

This is expressed as:

$$p(x) = \overline{p} - \frac{p'(u) + 2p'(t)}{6}\Delta + p'(t)(x-t) + \frac{p'(u) + 2p'(t)}{2\Delta}(x-t)^2 \quad (3)$$

for $t \le x \le u$, where $\Delta = (u-t)$.

Smoothing quadratic splines

Since $f(x)$ is represented as a natural quadratic spline $s(x)$, $f'(x)$ is a linear function between knots. For any linear function $l(x)$, we have

$$\int_u^t [f'(x)]^2 dx = \frac{u-t}{3}(l(t)^2 + l(u)l(t) + l(u)^2).$$

Hence

$$\int_{x_o}^{x_n}[f'(x)]^2dx = \sum_{i=1}^{n}\frac{x_i-x_{i-1}}{3}(b_{i-1}^n + b_{i-1}b_i + b_i^2).$$

The condition that $f(x)$ is continuous at the internal knots yields:

$$p_i(x_i) = p_{i+1}(x_i)$$

for $i=1, 2,\ldots, n-1$. Using Eq. (3), this translates into a set of equations:

$$b_{i-1}(x_i-x_{i-1}) + 2b_i(x_{i+1}-x_{i-1}) + b_{i+1}(x_{i+1}-x_i) = 6(\overline{f}_{i+1}-\overline{f}_i)$$

for $i=1, 2,\ldots, n-1$.

This can be expressed in a matrix form. Let $\mathbf{R}$ be the $(n-1)\times(n-1)$ symmetric tridiagonal matrix with diagonal elements $R_{ii} = 2(x_{i+1}-x_{i-1})$ and off-diagonal elements $R_{i+1,i} = R_{i,i+1} = x_{i+1}-x_i$. Then

$$\int_{x_o}^{x_n}[f'(x)]^2dx = \frac{1}{6}\mathbf{b}'\,\mathbf{R}\mathbf{b}.$$

Eq. (2) becomes

$$\frac{1}{n}(\mathbf{y}-\overline{\mathbf{f}})'(\mathbf{y}-\overline{\mathbf{f}}) + \frac{\lambda}{6}\mathbf{b}'\,\mathbf{R}\mathbf{b}.$$

Minimising with respect to $\overline{\mathbf{f}}$, the solution is represented as

$$[I + 6n\lambda(\mathbf{R}^{-1}\mathbf{Q}')'\,\mathbf{R}(\mathbf{R}^{-1}\mathbf{Q}')']\,\hat{\overline{\mathbf{f}}} = \mathbf{y}$$

where I is the identity matrix, $\mathbf{Q}$ is a $(n-1)\times n$ matrix with $Q_{ii} = -1$, $Q_{i,i+1} = 1$ and $Q_{ij} = 0$ otherwise. Solving this equation yield the fitted layer values $\hat{\overline{\mathbf{f}}}$. The fitted values at the knots can be obtained from

$$\hat{\mathbf{b}} = 6\mathbf{R}^{-1}\mathbf{Q}'\,\hat{\overline{\mathbf{f}}}$$

and (3).

References

Beckett, P.H.T., Webster, R., 1971. Soil variability: a review. Soils and Fertilizers 34, 1–15.

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45.

Colwell, J.D., 1970. A statistical–chemical characterization of four great soil groups in southern New South Wales based on orthogonal polynomials. Australian Journal of Soil Research 8, 221–238.

Cook, S.E., Corner, R.J., Groves, P.R., Grealish, G.J., 1996. Use of airborne gamma radiometric data for soil mapping. Australian Journal of Soil Research 34, 183–194.

Erh, K.T., 1972. Application of spline functions to soil science. Soil Science 114, 333–338.

Florinsky, I.V., Eilers, R.G., Manning, G.R., Fuller, L.G., 2002. Prediction of soil properties by digital terrain modelling. Environmental Modelling & Software 17, 295–311.

Geosciences Australia, 2008. Radiometric data of the Narrabri, Moree, Inverell and Manilla 1:250 000 topographic map sheets. In (Geophysical Archive Data Delivery System (GADDS) Website, date accessed 15/05/08. http://www.geoscience.gov.au/bin/mapserv36?map=/public/http/www/geoportal/gadds/gadds).

Grimm, R., Behrens, T., Marker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — digital soil mapping using Random Forests analysis. Geoderma 146, 102–113.

Hempel, J.W., Hammer, R.D., Moore, A.C., Bell, J.C., Thompson, J.A., Golden, M.L., 2008. Challenges to Digital Soil Mapping. In: Hartemink, A.E., McBratney, A.B., Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer Science, Australia, pp. 81–90.

Jenny, H., 1941. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, New York. 281 pp.

Lagacherie, P., 2008. Digital Soil Mapping: A State of the Art. In: Hartemink, A.E., McBratney, A.B., Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer Science, Australia, pp. 3–14.

McBratney, A.B., Mendonca-Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

McGarry, D., Ward, W.T., McBratney, A.B., 1989. Soil Studies in the Lower Namoi Valley: Methods and Data. The Edgeroi Dataset. CSIRO Division of Soils, Adelaide. 2 vols.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.

Minasny, B., McBratney, A.B., Mendonca-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Australian Journal of Soil Research 44, 233–244.

Minasny, B., McBratney, A.B., Lark, M.R., 2008. Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures. In: Hartemink, A.E., McBratney, A.B.,

Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer Science, Australia, pp. 15–30.
Moore, A.W., Russell, J.S., Ward, W.T., 1972. Numerical analysis of soils: a comparison of three soil profile models with field classification. Journal of Soil Science 23, 193–209.
Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57, 443–452.
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, 215–226.
Ponce-Hernandez, R., Marriott, F.H.C., Beckett, P.H.T., 1986. An improved method for reconstructing a soil-profile from analysis of a small number of samples. Journal of Soil Science 37, 455–467.
Rowe, W.D., 1994. Understanding uncertainty. Risk Analysis 14, 743–750.
Russell, J.S., Moore, A.W., 1968. Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data: Transactions of the 9th International Congress of Soil Science, vol. 4, pp. 205–213.
Ryan, P.J., McKenzie, N.J., O'Connell, D., Loughhead, A.N., Leppert, P.M., Jacquier, D., Ashton, L., 2000. Integrating forest soils information across scales: spatial prediction of soil properties under Australian forests. Forest Ecology and Management 138, 139–157.
Saunders, A.M., Boettinger, J.L., 2007. Incorporating Classification Trees into a Pedogenic Understanding Raster Classification Methodology, Green River Basin, Wyoming, USA. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science. Elsevier, Amsterdam, pp. 389–399.
Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction interval for the model output. Neural Networks 19, 225–235.
Skidmore, A.K., 1989. A comparison of techniques for calculating gradient and aspect from a gridded gigital elevation model. International Journal of Geographical Information Systems 323–334.
Stacey, K.F., Lark, R.M., Whitmore, A.P., Milne, A.E., 2006. Using a process model and regression kriging to improve predictions of nitrous oxide emissions from soil. Geoderma 135, 107–117.
Stoorvogel, J.J., Kempen, B., Heuvelink, G.B.M., de Bruin, S., 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. Geoderma 149, 161–170.
Webster, R., 1978. Mathematical treatment of soil information: Transactions of the 11th International Congress of Soil Science, vol. 3, pp. 161–190.
Young, P., 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. Environmental Modelling & Software 13, 105–122.