Mapping Continuous Soil Depth Functions in the Edgeroi District, NSW, Australia, Using Terrain Attributes and Other Environmental Factors

B. P. Malone, B. Minasny, A. B. McBratney

Faculty of Agriculture, Food & Natural Resources
The University of Sydney
Telephone: 0011 +61 9568 4113
Email: b.malone@usyd.edu.au

1. Introduction

To benefit from the ecological and economical functions of soil, 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. To facilitate this need, we must first comprehend the functions and pertinent factors contributing to not only soil variability across a landscape but also soil variability with depth down a profile.

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.

A 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). 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 various types of soil properties.

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 for mapping soil variability. The response to this demand has been answered partly in the way of digital soil mapping, where soil properties are mapped based on their relationship with environmental variables (Minasny et al. 2008). The *scorpan* factors 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 variation of soil properties across a defined area. In this paper we propose a novel method for predicting the vertical and lateral variation of soil properties in areas where limited soil data exists. Using soil carbon as our exemplar soil property we want to firstly determine whether terrain attributes alone are feasible for predicting its lateral and vertical variation or whether it is better described with the inclusion of other environmental factors relating to parent materials and landuse into the predictive models. With the most parsimonious model we want to map carbon storage to a depth of 1m in our defined study area and then demonstrate the functionality of the underlying soil geo-database for data enquiry by mapping the depth at which soil carbon falls below 1%.

2. Methods

2.1 Study Area and Soil Data

The study site (1500km^2) is situated near Narrabri (30.32S 149.78E), 500km NNW of Sydney, NSW, Australia. Agricultural enterprises such as cropping and pastoral farming are predominant in the area. There are also some significant tracts of land covered with remnant vegetation (Fig. 1).

The soil dataset consists of 341 soil profiles (Fig. 1). The dataset describes and quantifies various soil morphological, physical and chemical attributes at depth intervals of 0–0.1, 0.1–0.2, 0.3–0.4, 0.7–0.8, 1.2–1.3 and 2.5–2.6m (McGarry et al. 1989). Soil carbon storage is of greatest interest in this study. For our subsequent analyses, carbon was assessed on a volume basis (kg m^{-3}) of which was derived from the measured organic carbon percentages using pedo-transfer functions.

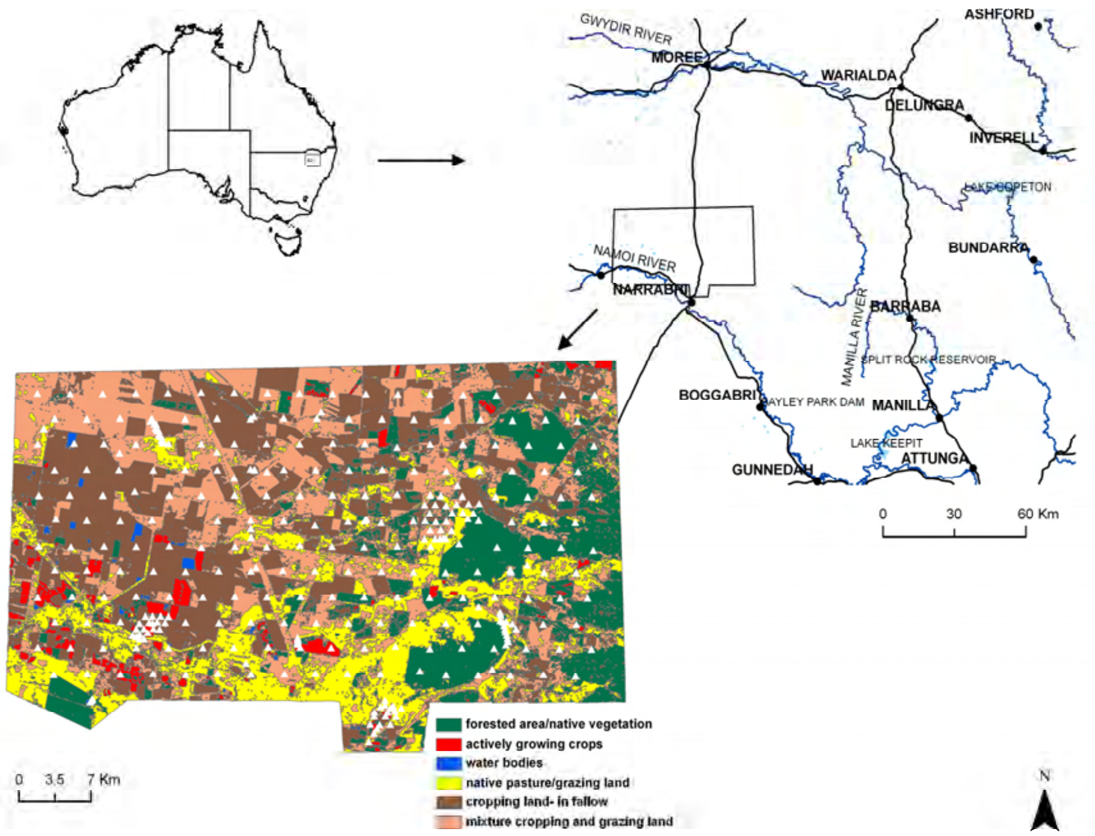

Figure 1. The Edgeroi study area

2.2 Environmental Data

A number of environmental covariates were sourced and interpolated onto a common grid of 90m resolution. These included:

- 3 arc-second (90m) digital elevation model (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.

- Landsat 7 ETM+ images from 2003. The Landsat bands were used for the estimation of land cover and land use. Vegetation cover and type was estimated using the Normalised Difference Vegetation Index (NDVI). Furthermore, the band ratios or more commonly, soil enhancement ratios of b3/b2, b3/b7 and b5/b7 were also derived.
- Gamma-radiometric survey data (Geosciences Australia 2008), which provides a measure of the spatial distribution of three radioactive elements (potassium-K, thorium-Th and uranium-U) in the top 30-45 cm of the earth's crust. This data was used to determine the distribution of various parent materials over the landscape.

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.

Splines were fitted to each of the 341 point using code written in Matlab (Mathworks 2005). For modelling purposes, the mean values at depth intervals; 0–10, 10–20, 20–30, 30–40, 40–50, 50–70, 70–80 and 80–100cm were derived from the splines.

2.4 Data Analysis and Modelling

JMP software (SAS) was used to construct neural networks for the depth-wise modelling of carbon storage. Prior to this, 50 data points were randomly selected and omitted from the dataset for the purpose of model validation.

Firstly, a neural network was constructed using only the DEM data and derived terrain attributes as predictor variables. Secondly, using the best combination of environmental factors as determined from stepwise regression another neural network was constructed. Profile formulae for each depth interval were saved for later use to predict in areas where data observations were not available. Residuals were calculated and then kriged (local) onto the common 90m grid.

For model validation, the profile formulae were applied to the 50 withheld data points. Residuals, calculated by inverse kriging from the 90m grid of residuals were added to the estimated depth values to give an amended prediction. Splines were then reconstructed and then compared to the raw data values as derived from the inputs provided by McGarry et al. (1989).

Using the best neural network, the profile formulae were applied to the common 90m grid geo-database in order to map total carbon storage for the study area. The kriged residuals were added to the predictions. For demonstration of functionality, the resulting geo-database of soil information generated in this study was queried to determine the depth at which carbon storage decreases to below 1% using computer programming code written in Matlab (Mathworks 2005).

3. Results

Table 1 shows the best combination of terrain attributes used for each layer. In the top depth interval (0–10cm) all terrain attributes except flow length were pertinent where

35% of variation could be explained. Elevation is critical at all depth intervals except at 80–100cm. Other important factors include altitude above channel network, stream power index and terrain wetness index.

Depth Interval (cm)	Terrain Attributes	R^2
0–10	Elevation, aspect, slope, terrain wetness index (TWI), altitude above channel network (AOCN), LS-factor, stream power index (SPI)	35%
10–20	Elevation, AOCN, LS-factor, SPI	22%
20–30	Elevation, AOCN, SPI	12%
30–40	Elevation, AOCN, SPI	11%
40–50	Elevation, AOCN, TWI, SPI	12%
50–70	Elevation, AOCN, TWI, flow length, SPI	15%
70–80	Elevation, AOCN, TWI, flow length, SPI	16%
80–100	TWI, AOCN, flow length, SPI	12%

Table 1. Depth wise variation in terrain attributes to predict soil carbon storage

With the inclusion of other environmental factors, the most parsimonious combination was found to comprise of elevation, slope, and altitude above channel network, stream power index, potassium (from radiometric survey), and Landsat bands #3, #4 and #5 and the band ratios 3/7and 5/7. On average 35% of the variation could be explained with these factors.

Using all the available terrain attributes as factors, the neural network for carbon prediction was found to explain 57% of the variation. Upon the automated cross-validation procedure, 0% of dataset could be accurately modelled. Conversely, using the most parsimonious set of environmental factors, the neural network model explained 60% of the variation. 4% of the variation was explained when the model was cross-validated.

Five data points were selected at random from the 50 validation points to graphically represent model predictions with the raw values derived from McGarry et al. (1989) (Fig. 2) As expected, the fitted splines, fit closely to the raw data. When using the terrain attributes only or the best combination of factors, the predictions display a fair agreement with the raw data values. The fits from both neural networks at *ed002*, *ed147* and *ed218* (Figures 2a, c, d) are reasonably similar. However, at *ed044* and *ed340* it appears that using the most parsimonious set of factors generates more accurate results (Figures 2b, e).

Fitted splines to raw data	Fitted splines predicted with only terrain attributes	Fitted splines predicted with best set of environmental covariates

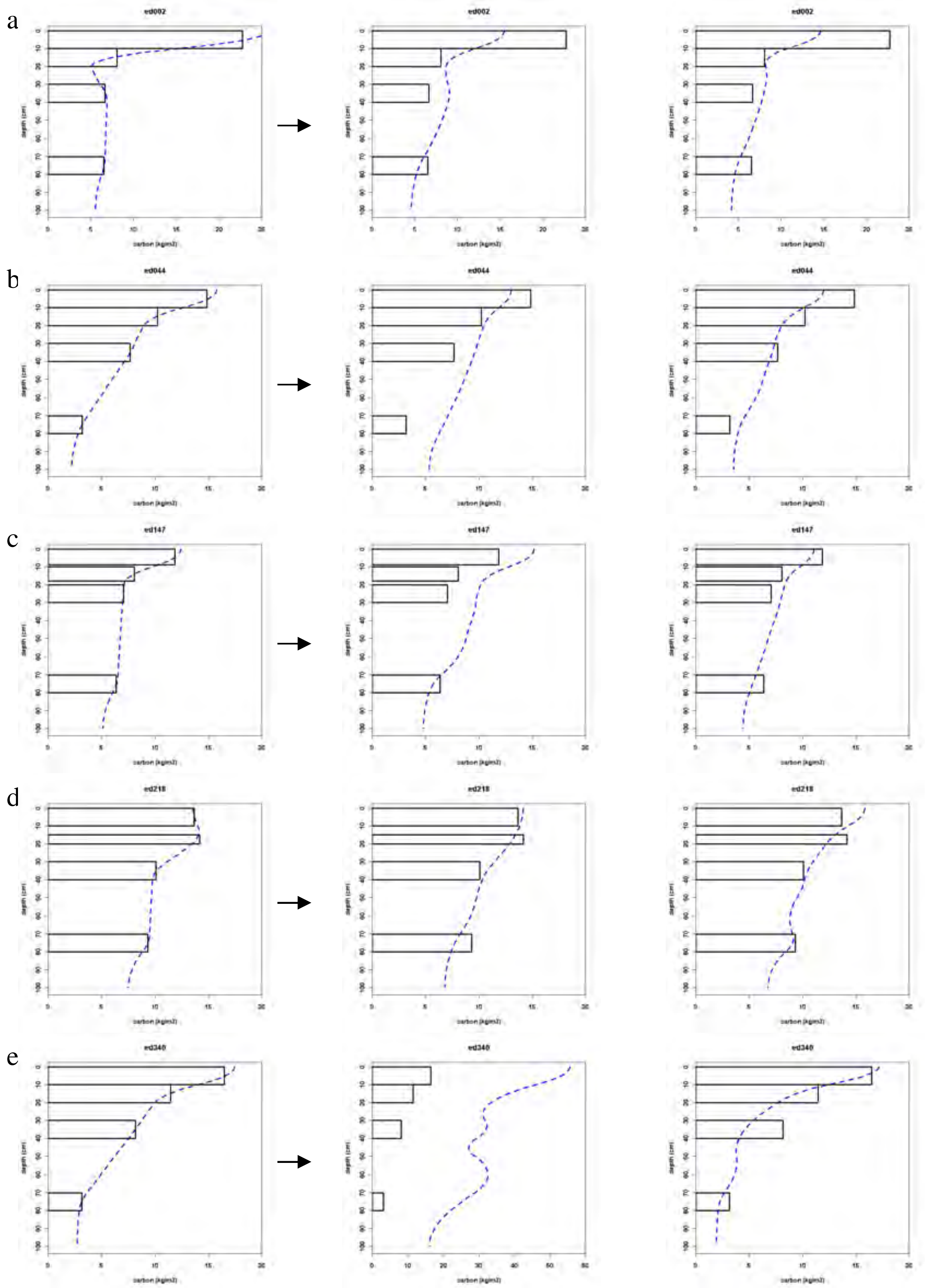

Figure 2. Fitted splines (dashed lines) of raw data and carbon predictions using either terrain attributes only or the most parsimonious set of environmental factors. (Polygons are the raw data values from McGarry et al. 1989).

The model with the best combination of environmental factors was used to map total carbon storage in the soil across the study area. Soil carbon ranged between 1–80kg m^{-2} to a depth of 1m (Figure 3). The total average carbon storage was 9.5kg m^{-2}, with the highest levels found to the eastern and southern sections of the area (8–80kg m^{-2}). These areas coincide with landuse 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 (1–7kg m^{-2}).

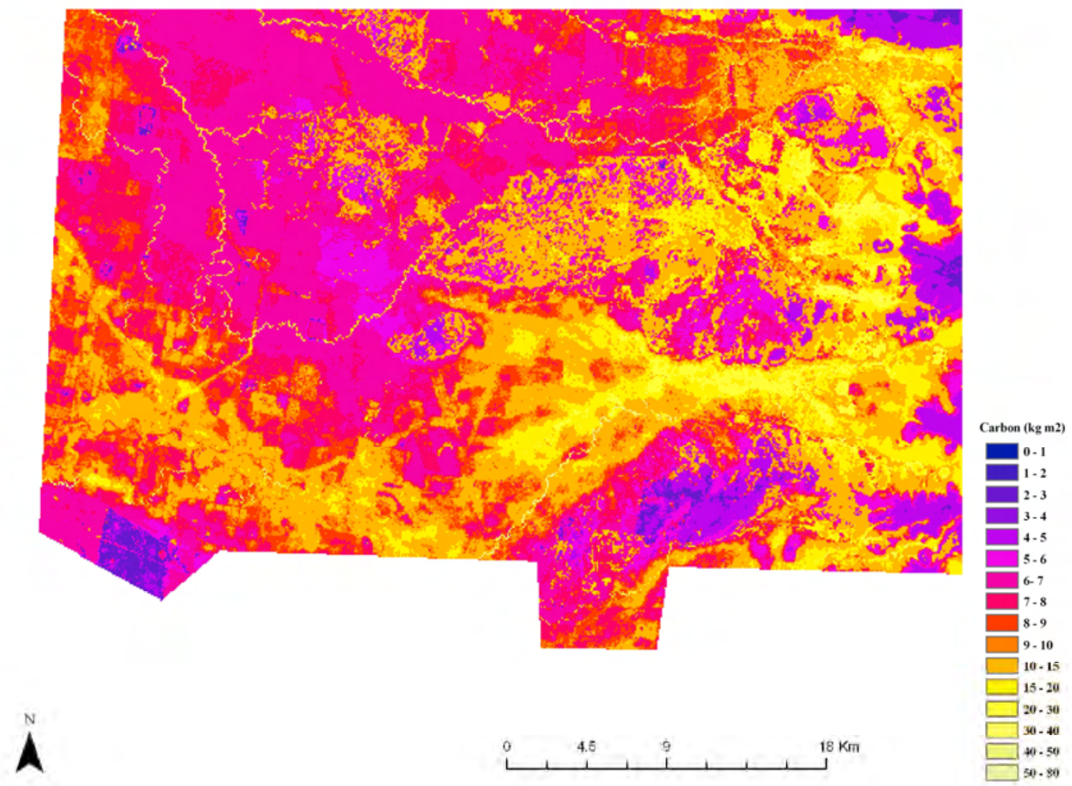

Figure 3. Total carbon storage across the Edgeroi study area.

As can be observed in Figure 4 the depth at which soil carbon drops below 1% is quite variable across the study area where it ranges from 1cm to over 1m, with the average depth at 21cm. The cropping areas situated mostly to the western areas of the study area have the highest concentration of soils where in the top 5cm, soil carbon falls 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.

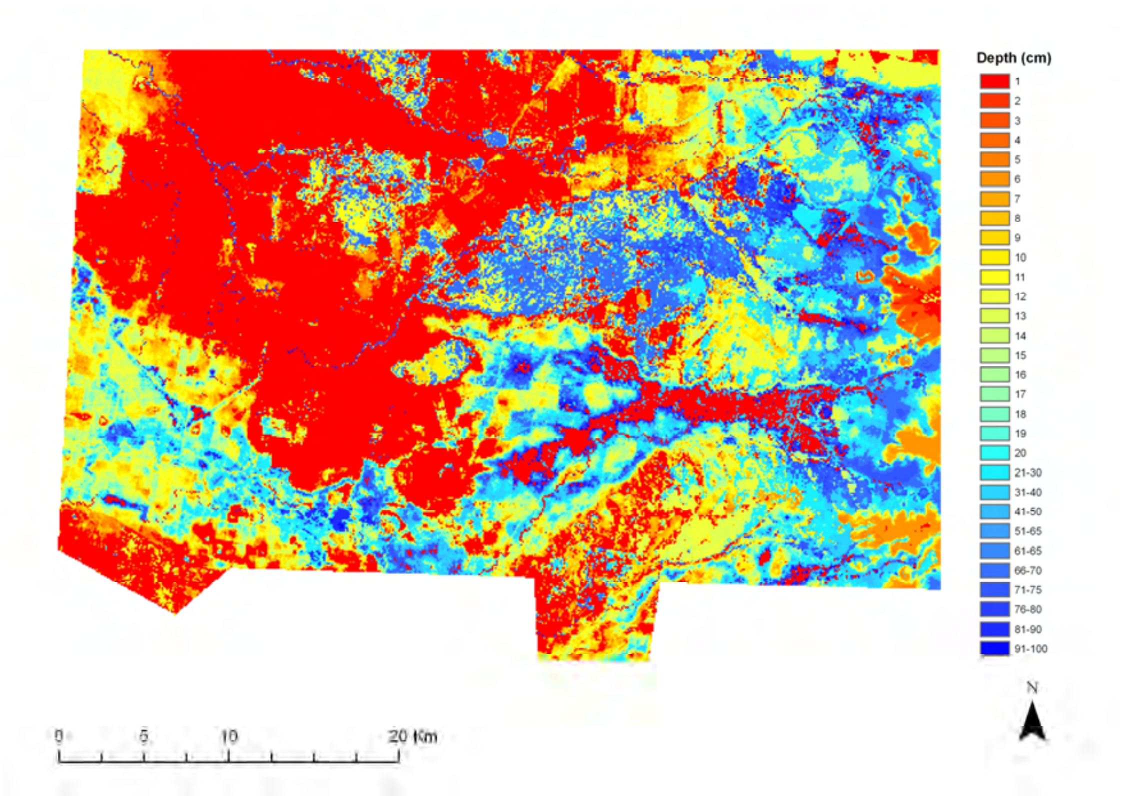

Figure 4. Depth at which soil carbon decreases to below 1% across the Edgeroi study area.

4. Discussion and Conclusions

Previous studies have identified the impact that land use has on soil carbon storage where areas under remnant vegetation have considerably more soil carbon than areas that have been cleared and used for agricultural pursuits (Knowles and Singh 2003). A similar trend was observed for this particular project. Additionally, the generated results of total carbon storage in the Edgeroi area are reflective of those of Minasny et al. (2006) who used an exponential decay to model soil carbon storage.

Alone, the terrain attributes provide a valid framework in which soil carbon can be mapped both vertically and laterally as observed by the re-constructed splines at certain points. This would indicate that the geomorphology of the study area is partly deterministic of the variation of soil carbon where elevation, slope, altitude above channel network and stream power index are important attributes. However, without the inclusion of other environmental factors the predictive accuracy of the model framework is reduced by comparison.

Nevertheless, this study provides an example where a rich soil attribute geo-database can be generated from a limited soil dataset. The functionality of this database for enquiry will suit the purposes of people concerned with the management of land. However, future work will attempt to address some of the structural and metrical uncertainty identified in this study.

References

Bishop TFA, McBratney AB, Laslett GM, 1999, Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma*, 91(1-2): 27-45.

Erh KT, 1972, Application of spline functions to soil science. *Soil Science*, 114(5): 333–338.

Knowles TA, Singh B, 2003, Carbon storage in cotton soils of northern New South Wales. *Australian Journal of Soil Research*, 41(5): 889-903.

Lagacherie P, 2008, Digital soil mapping: a state of the art. In: AE Hartemink, AB McBratney, ML Mendonca-Santos (Editors), *Digital Soil Mapping with Limited Data*. Springer Science, Australia, pp. 3–14.

McBratney AB, Mendonca-Santos ML, Minasny B, 2003, On digital soil mapping. *Geoderma*, 117(1-2): 3-52.

McGarry D, Ward WT, McBratney AB, 1986, Soil studies in the Lower Namoi Valley: methods and data. The Edgeroi Dataset. CSIRO Division of Soils, Adelaide.

Minasny B, McBratney AB, Mendonca-Santos ML, Odeh IOA, Guyon B, 2006, Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Australian Journal of Soil Research*, 44(3): 233-244.

Minasny B, McBratney AB, Lark MR, 2008, Digital soil mapping technologies for countries with sparse data infrastructures. In: AE Hartemink, AB McBratney, ML Mendonca-Santos (Editors), *Digital Soil Mapping with Limited Data*. Springer Science, Australia, pp. 15–30.

Ponce-Hernandez R, Marriott FHC, Beckett PHT, 1986, An improved method for reconstructing a soil-profile from analysis of a small number of samples. *Journal of Soil Science*, 37(3): 455-467.