

Digital soil survey and mapping underpinning inherent and dynamic soil attribute condition assessments

Brendan Malone^{*}, Uta Stockmann, Mark Glover, Gordan McLachlan, Sophia Engelhardt, Seiji Tuomi

CSIRO Agriculture and Food, Black Mountain, ACT, Australia

ARTICLE INFO

Keywords:

Farm scale soil mapping
Customization
Proximal soil sensing
Soil infrared spectroscopy
Gamma attenuation
Digital soil mapping
Soil condition
Soil pH
Digital agriculture

ABSTRACT

We set out in this work to create a suite of 3D-like digital soil attribute maps for a research farming enterprise situated in the South Eastern Agricultural Zone of NSW, Australia. These digital maps come about because of a reconnaissance soil survey to map out the farm's soil baseline conditions and bring to bear the tools and processes of the digital convergence. This entails on-the-go proximal soil sensing for rapid and granular farm landscape characterisation, which helped inform field-based soil survey and sampling, and later the soil mapping too. Soil infrared spectroscopy associated inference (vis-NIR) was then used to facilitate very detailed soil characterisations of key soil attributes that included soil pH, carbon, soil texture, bulk density, and cation exchange capacity amongst others. This work then enabled the ability to build bespoke spatial models to generate 3D-like inherent and dynamic digital soil attribute maps. We highlight some of the digital soil infrastructure features which include flexibility, customisation, and work pipelines that will ease the updating and improvement of maps. We also highlight some immediate goals which include a formal inclusion of measurement uncertainties brought about by soil spectral inference into the spatial modelling process. To date, some of the digital soil mapping has been used for applications such as lime requirement estimations to address subsoil acidity issues and therefore changing/improving the soil's capability at BARS. Overall, the intention of this work is to provide a detailed account of what is entailed in creating a comprehensive digital soils infrastructure for a farm including the likely costs to implement and the associated opportunities it provides.

Introduction

Disruptive digital technologies have the potential to re-frame and re-define how agri-businesses operate. The agricultural enterprise is one such business, that has embraced the recent technological revolution brought about by the ubiquity of digital data streams and sensing technologies coupled with unparalleled computing and data analysis systems (Sonka, 2021). This 'data-driven agriculture' promises an increase of on-farm productivity and sustainability gains through such things as optimal use of inputs, i.e. variable rate technologies, to granular insights on the biophysical status and function of the land, to real-time and forecasting insights of crop productivity and soil moisture fluxes brought about through sensor networks, detailed process-based modelling and forecasting simulation work. Data-driven agriculture then also promises to facilitate the rapid and cost-effective spatially explicit assessment of inherent and dynamic soil natural capital stocks and their condition and capability across a farm that can support natural

capital accounting (Rossiter et al., 2018).

One component of the agricultural system that has evolved immensely in the last 20 years is our ability to rapidly measure and comprehensively characterise soils. Developments in earth observation platforms, proximal soil sensing technologies, geostatistics, spatial data science and machine learning have collectively contributed to a powerful ability to get detailed and highly granular insights about the status and functions of soils (McBratney et al., 2018). For example, characterisation of soils has moved substantially onward from just coarse categorisations of soil type and suitability for a particular land-use. Now it is possible to get very detailed spatially explicit insights about individual soil properties such as pH, nitrogen, carbon, texture, CEC and soil water holding characteristics. These insights are underscored by digital soil mapping methodologies and include lateral, vertical, and even temporal characterisations (Searle et al., 2021).

Precision agriculture has been at the forefront of exploiting these new technologies by advancing the concept of site-specific crop

^{*} Corresponding author.

<https://doi.org/10.1016/j.soisec.2022.100048>

Received 5 October 2021; Received in revised form 20 January 2022; Accepted 30 January 2022

Available online 1 February 2022

2667-0062/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

management (Whelan, 2018), whereby in response to soil spatial variability, management practices such as fertiliser and lime applications can be customised accordingly (i.e. through variable rate application). Similarly, such detailed characterisations of soil enable efforts to identify and delineate soil constraints so that management and amelioration plans can be defined (Filippi et al., 2020). Given the substantially high costs of amelioration and the oftentimes, radical interventions required such as soil flipping (or inversion) and deep ripping (McBratney et al., 2016), a granular insight into soil spatial variability is keenly needed to reduce the high costs associated with these interventions. Other uses of detailed soil insights include on-farm soil carbon auditing via direct measurement methods which enables landholders to participate in the carbon economy and earn credits and potentially an alternate income stream simply through soil carbon sequestration (Malone et al., 2018a). This feeds into the emerging practice of natural capital accounting whereby the services and functions performed by soils are effectively

added to the balance sheet of a farming business enterprise (Guerry et al., 2015). This recognises the importance and value of natural resources to supporting agricultural productivity. Simply left unchecked or unsustainable, those key resources can potentially degrade in function and ultimately cannot support the kinds of agriculture those lands were previously capable of supporting. It is likely that general farm operations and data collection will further advance in sophistication as existing and new digital technologies are exploited to measure and monitor the natural assets of a farming enterprise.

The digital farms of the current time and into the future will likely be underpinned with rich datasets and monitoring technologies that will assist and augment on-farm decision making. This paper describes a set of processes which can easily be generalised and adapted to any agricultural landholding to establish a baseline spatial soil data infrastructure that over time can be updated, improved, refined, and expanded. More explicitly, here we describe how currently available digital

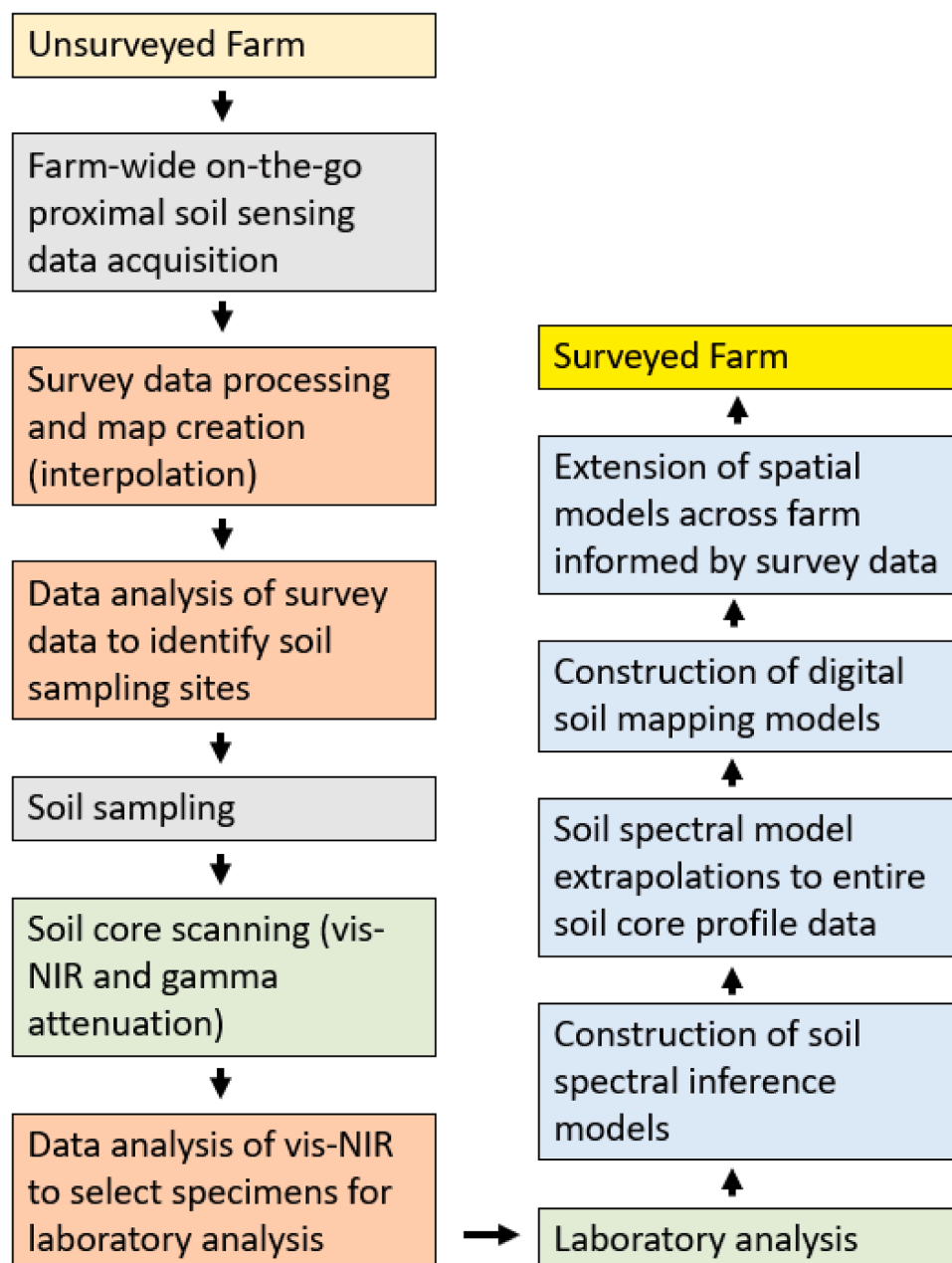


Fig. 1. General soil survey and digital mapping workflow. Grey boxes are fieldwork activities, orange boxes are data analysis steps, green boxes are laboratory-based activities, blue boxes are data modelling activities.

technologies and work methods can be used to create a suite of 3D-like soil condition baseline maps relevant for the farm enterprise. This work was carried out at the Commonwealth Scientific and Industrial Research Organisation Boorowa Agricultural Research farm (BARS), a mixed cropping enterprise in south eastern Australia.

With our eye on the possibilities of farm-scale digital soil mapping as demonstrated in Ramirez-Lopez et al. (2019) and Jones et al. (2022), the steps of surveying and mapping the soils of BARS uses tools and processes of digital agriculture carried out in a sequential manner from field-based optimal data acquisition and soil sampling, to soil measurement, modelling and finally mapping of the spatial distribution of key inherent and dynamic soil attributes across the extent of the farm in a comprehensive manner. This work sequence is executed as follows and summarised in a pictorial workflow in Fig. 1:

- Initial farm-wide proximal soil sensing survey using a suite of on-the-go measurement instruments.
- Spatially-explicit data-informed (by step 1) optimal site selection for soil core sampling with the intent to capture as much soil spatial variability as possible.
- Proximal soil sensing of all intact soil cores (<1.6 m) using infrared spectroscopy and gamma attenuation followed by soil infrared data-informed sub-sampling to select soil specimen for laboratory wet chemistry measurements. These laboratory-based soil measurements together with infrared spectral responses are then used to derive soil spectral calibration models, which can be extended to all soil specimen to deliver whole-soil characterisations of the given suite of measured soil variables.
- Whole soil profile characterisation data are then used together with processed data collected in step 1 to build bespoke spatial models to generate 3D-like digital soil attribute maps.

Materials and methods

CSIRO BARS site description

Boorowa Agricultural Research Station (BARS) is a 220 ha mixed cropping farming enterprise situated in south-eastern New South Wales, Australia Fig. 2. 3 km south of the town of Boorowa [34.4386S, 148.7231S], BARS is situated in the Boorowa River catchment located within the Lachlan Fold Belt. The terrain is comprised of gently undulating (1–3% slope) to undulating rises (3–10% slope) with local relief between 9 and 30 m above the average 600 m a.s.l. elevation across the area. The area experiences a temperate climate with long summers and cool to cold winters. Rainfall on average is 619 mm p.a. and is slightly winter dominated. The underlying geology at BARS are Silurian ignimbrites and tuffs with associated interbedded sediments of the Douro Group, which is dominated by the Hawkins Volcanics (Cas, 1983). On crests and slopes, soils are yellow to light reddish duplex (Texture Contrast >20% increase in Clay between A and B horizon) soils which classify out commonly to either Yellow or Red Chromosols or Kurosols depending on whether there is subsoil acidity. Mottling of the subsoil is common. Other soils include Red and Yellow Dermosols and Kandosols and Yellow Sodosols are often found near drainage lines (Hird, 1991).

Soil survey – proximal soil sensing

Data acquisition

An all-terrain vehicle (ATV) with precise RTK GPS navigation and recording capabilities, fitted with electromagnetic induction (EMI) and passive gamma-radiometric sensors were used for on-the-go data acquisition across BARS. The ATV was driven across BARS at 50 m line spacing intervals (5 km/h), with all digitally acquired data (GPS and proximal sensor outputs) streamed together into a custom-designed data acquisition software. The simultaneous recording of all sensor outputs was achieved at intervals of at least once per second while in operation.

A SMART6-L GNSS (NovAtel, Canada) digital GPS data receiver with centimetre level accuracy was used for the locational data acquisition. Data acquired by the GPS receiver are position and ground elevation. EMI data were acquired using a Geonics EM38 (Geonics Ltd, Mississauga, Ontario, Canada) in both the horizontal (EM38h) and vertical (EM38v) dipole modes of operation. In the horizontal mode, conductivity (mS/m) has the greatest sensitivity of measurement at the soil surface (~0.75 m) and declines with depth. In the vertical mode, conductivity is measured to an approximate depth of 1.5 m (McNeill, 1990). Gamma-radiometric data were collected using a RSX-1 gamma radiometric detector consisting of a 4 L Sodium-Iodine crystal (Radiation Solutions Inc., Mississauga, Ontario, Canada). This passive gamma sensor records the naturally occurring concentration of radioactive isotopes in the soil (top 30–50 cm) based on the principle that each gamma ray photon relates to a discrete energy window which is characteristic of the source isotope (Minty et al., 1998). Radioisotopes of potassium (40 K), uranium (238U-series) and thorium (232Th-series) produce high-energy gamma-rays with sufficient intensities to be picked up by the detector.

Processing of proximally sensed data

The raw data (GPS, radiometric, and EMI data) were pre-processed to clean up obvious acquisition errors and data outliers. The cleaned data was then processed to create maps on a regular 2.5 m grid across BARS. Here, we used local block kriging where block size was equivalent to the grid spacing, i.e. 2.5 m x 2.5 m. Digital elevation maps as derived from the GPS data were further processed using supplied terrain analysis algorithms from freely available GIS software (SAGA-GIS) to derive associated primary and secondary terrain attributes. Together, the suite of ‘environmental’ gridded data derived from the on-the-go proximal soil sensing survey included:

- Digital Elevation model

Elevation, slope gradient, slope curvatures (profile and plan), terrain wetness index, incoming solar radiation and flow direction and accumulations.

- Gamma radiometric data

Total count, gamma K, Th and U

- Bulk soil electrical conductivity

0–0.75 and 0–1.5 m

The raw survey transects together with interpolated surfaces of key attributes are shown in Fig. 3.

Soil survey – soil core sampling and measurement

Soil survey sampling design

Available resources permitted the collection and characterisation of soils at 300 sites across BARS. To select these sites, conditioned Latin Hypercube sampling (cLHS; Minasny and McBratney, 2006) was used, where the input data consisted of the on-the-go proximally-sensed data described in the previous Section. cLHS is a form of random-stratified sampling often used for soil survey delineation to support digital soil mapping as it is designed to optimally sample the spatial variation of its input variables. The algorithm has its origins in Latin hypercube sampling (LHS) first proposed by McKay et al. (1979), which is a demonstrably efficient way to reproduce an empirical distribution function, where the idea is to divide the empirical distribution function of a variable, X, into n equi-probable, non-overlapping strata, and then draw one random value from each stratum. In a multi-dimensional setting, for k variables, X1,X2,...,Xk, the n random values drawn for variable X1 are combined randomly (or in some order to maintain its correlation) with

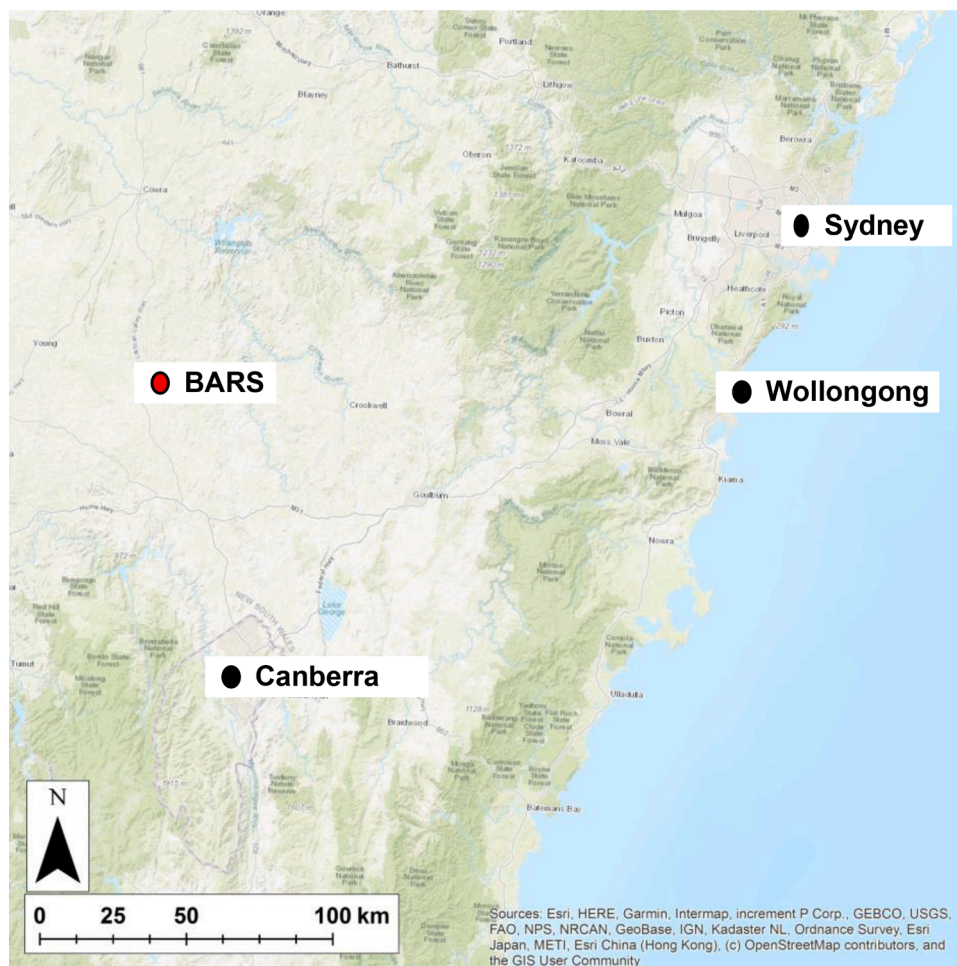


Fig. 2. Locality map of BARS in south-eastern NSW with reference to major metropolitan hubs of Sydney, Canberra and Wollongong.

the n random values drawn for variable X_2 , and so on until k -tuples are formed, that is, the Latin hypercube sample (Clifford et al., 2014). Minasny and McBratney (2006) recognized that some generalization of LHS sampling was required so that selected samples exist in the real world. Subsequently, they proposed a conditioning of the LHS, which is achieved by drawing an initial Latin hypercube sample from the ancillary information, then using simulated annealing to permute the sample in such a way that an objective function is minimized based on the matching of empirical distribution functions and correlations between samples and that of the gridded data (population). The outcome is a sample configuration of size n where there is a high confidence the samples capture the distribution of the variables used in the algorithm, and thus a greater confidence the samples capture the known spatial variability of soils, which ultimately improves the veracity of spatial models developed to map the soils across the site.

With 300 sites selected, soil cores were extracted at each of them using a vehicle mounted mechanical soil coring instrument (Geoprobe 7822DT). The cores were taken to a depth of 1.6 m or to the depth where consolidated material was reached whichever was first. Core diameter was 50 mm and each was encased in a plastic sleeve to minimise disturbances and contamination. At the time of sampling in March 2017, soils were in a dry condition. Fig. 4 shows the cLHS-derived soil coring locations across the extent of BARS.

Processing of soil cores

A field-deployable soil core sensing system (SCSS; Viscarra Rossel et al., 2017) equipped with a RGB camera, vis-NIR spectrometer and active gamma radiation source was used for acquiring comprehensive

digital data from which soil characteristics were derived or inferred. Down the length of each soil core at regular intervals (every 2.5 cm to 20 cm and every 5 cm from 20 cm to the end of the core), the roboticised and integrated scanning system collects an RGB image and a vis-NIR spectrum (300–2500 nm) from the soil surface face, together with measuring gamma attenuation as a gamma radiation source is pointed directly onto the soil. In the current work we do not explore an application for using the RGB imagery and only focus on the soil infrared and gamma attenuation properties collected.

The vis-NIR spectra were collected using a LabSpec spectrometer manufactured by Malvern Panalytical/ Analytical Spectral Devices (Boulder, CO, USA), which has spectral range of 350–2500 nm and a spectral resolution of 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. Measurements were made with a high-intensity contact probe illuminated by a halogen bulb (2901 ± 10 K). The contact probe measures a spot of 10 mm in diameter and is designed to minimize errors associated with stray light. At the beginning of each core measurement, the sensor is calibrated using a Spectralon® (Labsphere, North Sutton, NH, USA) white reference panel. Spectra were recorded with a sampling resolution of 1 nm so that each spectrum comprised reflectance at 2151 wavelengths.

Gamma attenuation was collected using an LB444 densitometer (Berthold Technologies GmbH, Bad Wildbad, Germany). The densitometer contains a radioactive source (^{137}Cs) with an activity of 185 MBq and a photon energy of 0.662 MeV. The detector is an LB5441–01 with an uncollimated NaI scintillation crystal of 25-mm diameter and 25-mm length (Berthold Technologies GmbH, Bad Wildbad, Germany).

The purpose of collecting soil infrared data down the length of all soil

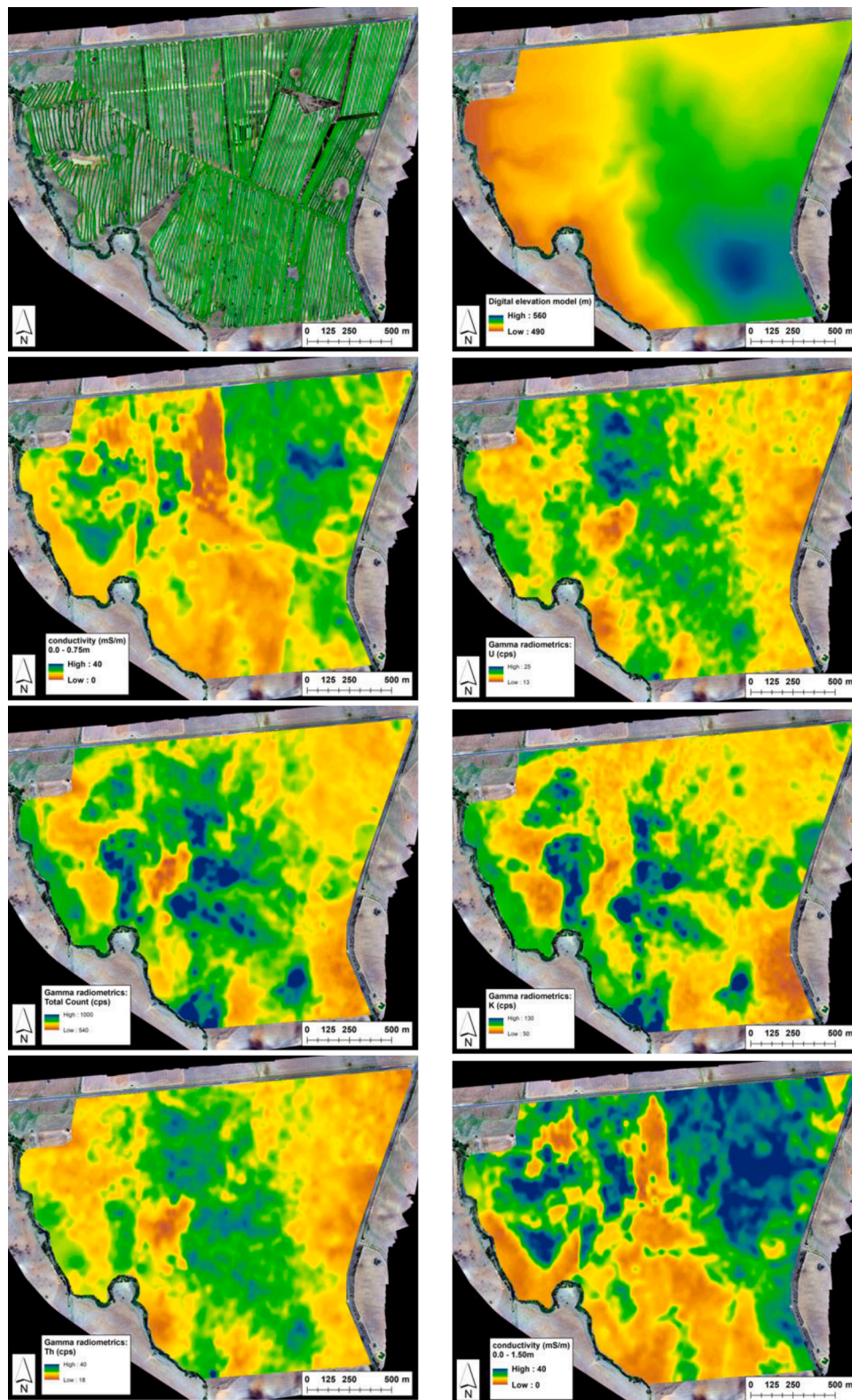


Fig. 3. Collected proximal soil survey data collected for BARS. This includes the drive transects as taken by the ATV (top left). The onboard sensor collected information related to spatial position, elevation, soil conductivity and gamma radiometrics. Geostatistical techniques were used to create maps of these variables to use for subsequent processes including planning soil survey and assisting in the creation of digital soil maps.

cores was to exploit functional relationships between the light interaction with soil and rapidly infer measurable soil attributes like soil carbon, texture, CEC and pH as examples. In general, optical sensors utilize various wavelengths from the electromagnetic spectrum to characterise

soil properties (visible (vis): 400–700 nm; near-infrared (NIR): 700–2500 nm; mid-infrared (MIR): 2500–25,000 nm). The absorption in the MIR region can be readily identified as it is related to the fundamental vibrations of molecules. The absorbance in the vis-NIR, however,

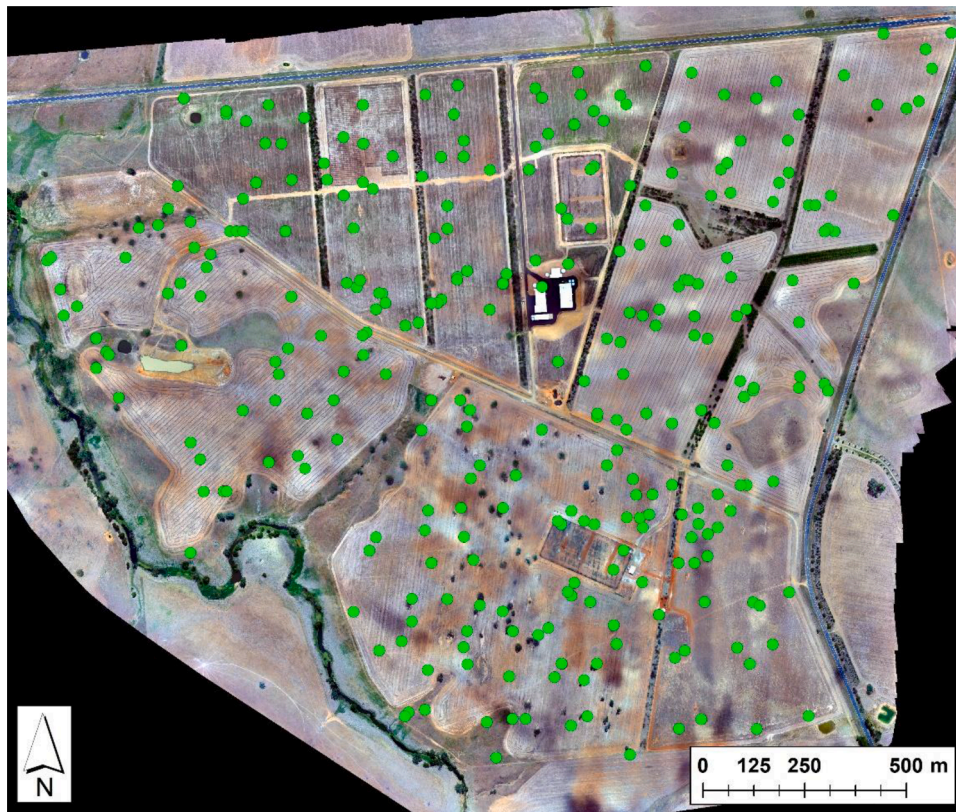


Fig. 4. Map of BARS showing the locations of the soil coring sites.

is harder to interpret as there are fewer and broader absorptions related to overtones and combinations of the fundamental vibrations in the MIR region (Stenberg et al., 2010). Absorptions in the visible region are associated with minerals that contain iron (Sherman and Waite, 1985; Mortimore et al., 2004) and organic matter, while absorptions in the NIR region (780–2500 nm) result from the overtones of OH, SO₄, and CO₃ groups, and combinations of fundamental features of H₂O and CO₂ (Clark, 1999). More of the underpinning theory and review of wide-ranging literature about soil infrared spectroscopy is found in Soriano-Disla et al. (2014).

The target measurement for gamma attenuation data is whole soil density. Lobsey and Viscarra Rossel (2019) describe the technique which is based on the fundamental physical relationships established through the Beer-Lambert Law where in the case of soil density, gamma attenuation is a function of its mass and the mass attenuation coefficients of soil and water in the attenuation path:

$$\frac{I}{I_0} = \exp[-x(\mu_s \rho_s + \mu_w \rho_w \theta)] \quad (1)$$

Where I is the incident radiation at the detector, I_0 is the unattenuated radiation emitted from the source (essentially the reading from the detector when nothing is in the path of the radiation beam), and x is the sample thickness in centimeters. Parameters μ_s and μ_w represent the mass attenuation coefficients of the soil and water, respectively. Parameter ρ_w is the density of water, which is 1 g cm^{-3} , and ρ_s is the density of the soil (which is to be estimated in our case). θ is the volumetric water content of the soil in $\text{cm}^{-3} \text{ cm}^{-3}$. The sensor used in Lobsey and Viscarra Rossel (2019) is the same used in this work. Therefore, we used the same attenuation coefficients established in that research, which were $0.0770 \text{ cm}^2 \text{ g}^{-1}$ and $0.0832 \text{ cm}^2 \text{ g}^{-1}$ for μ_s and μ_w , respectively. At the time of soil core scanning, an estimate of θ is made via a pre-calibrated vis-NIR soil spectral model based on data described in Baumann et al. (2022). In summary, that study used soils from 54 locations across Australian agricultural regions, from three depths: 0–15

cm, 15–30 cm and 30–60 cm. The volumetric water content of the samples and their vis-NIR spectra were measured at seven matric potentials from -1 kPa to -1500 kPa . From that data, the pre-calibrated soil spectral inference model composed of a bootstrap resampling partial least square's regression model of the collected spectra and measured θ data is used to infer θ at each measurement point down the soil core. With this estimate, the dry soil bulk density at the point of measurement is estimated as:

$$\rho_b = \frac{1}{x\mu_s} \ln\left(\frac{I}{I_0}\right) - \frac{\mu_w}{\mu_s} \rho_w \theta \quad (2)$$

Soil spectral inference

Laboratory analysis. While the dry soil bulk density can be estimated directly using Eq. (2) as the core scanning moves down the length of each soil core, the inference of other soil properties requires soil spectral model calibration which in this work were based on measurements of soil specimens taken from the soil cores that were put through the SCANS system. Collectively, across each of the cores and the associated depth intervals there were 7775 individual vis-NIR spectra taken. From these 7775 'specimens', 380 were selected to be subjected to laboratory analysis. The analytes that were measured included:

- Soil texture analysis to estimate the proportions of clay, silt and sand following the method of Gee and Bauder (1986).
- Total soil carbon (%), organic soil carbon (%), and total nitrogen (%) were measured via the dry combustion method using a LECO CNS analyser (LECO Corporation, MI, USA). Inorganic soil carbon was assessed as the difference between total and organic soil carbon. Organic soil carbon was measured after total carbon was measured by processing the specimens through a repeated hydrochloric acid washing routine, removing any inorganic carbon.

- Soil electrical conductivity (dS/m) was measured from a 1:5 soil/water extract indicated a method 3A1 from [Rayment and Lyons \(2010\)](#).
- Soil pH was measured both in 1:5 soil/water suspension (4A1) and soil/0.01 M calcium chloride extract (4B1) from [Rayment and Lyons \(2010\)](#).
- Bicarbonate-extractable or Colwell phosphorus (mg/kg) was measured following the standard 9B2 method ([Rayment and Lyons, 2010](#)).
- Exchangeable bases and cation exchange capacity (CEC) was measured using 1 M ammonium acetate at pH 7.0 with pre-treatment for soluble salts and using automatic extractor (15D2; [Rayment and Lyons, 2010](#)). Units are cmol^+/kg .
- Exchangeable acidity (hydrogen and aluminium) was measured by 1 M potassium chloride (15G1; [Rayment and Lyons, 2010](#)). Units are cmol^+/kg .

Soil specimen selection. To select the 380 specimens, all 7775 vis-NIR spectra were compiled, then subjected to dimension reduction using a principal component analysis (PCA). 6 principal components were able to summarise over 97% of the spectral variability which simplified the next step of specimen selection using the cLHS sampling algorithm as discussed previously in the context of site identification for soil survey. This general method for specimen selection is described in greater detail in [Wadoux et al. \(2021\)](#). Once selected, the specimen were extracted from their locations in their respective cores, dried, then grounded and sieved to <2 mm. The gravel and rock fraction were set aside for weighing.

Soil spectral modelling. Soil spectral modelling proceeded by generating one model per soil analyte, except in the case of soil texture where a compositional data analysis method based on log-ratio transformation of the original data was used as the target variable. Compositional data analysis ensures the integrity of the simplex, i.e. the fractions of clay, sand, and silt sum to 1 following any sort of modelling procedure. In this work we used isometric log-ratio transformations which decomposes the original 3 vector tuple to 2 log-ratio vectors ([Egozcue et al., 2003](#)). Each of these log ratio vectors is set as a target variable for soil spectral modelling. After this, modelling is performed for each of the 2 vectors, the inverse of the log-ratio is then calculated to return the data into their original soil texture units.

A soil spectral modelling approach using partial least squares regression modelling incorporated into a bootstrap resampling (with replacement) framework was used here. The number of bootstraps was set to 50, and cases not included in a particular model iteration (out-of-bag cases) were used for model validation. The useful property of this bootstrap model is that predictions can be made for each data case using each fitted model to generate an empirical distribution function (EDF), which can be used to derive key statistical moments such as the median and mean and variance. However, the EDF variance contributes only one part of the overall variance. The additional source of prediction variance in the modelling framework is what [Rossel et al. \(2015\)](#) describe as the discrepancies brought about through systematic (bias) and random (imprecision) errors in the modelling. This is expressed as the cross-validation mean squared error (MSE_{cv}) and is calculated as the average of the mean squared error values from each model iteration on out-of-bag data. The variance of the EDF (right-hand side of [Eq. \(3\)](#)) accounts for errors in the deterministic component of the models. The other terms in [Eq. \(3\)](#) include n which is the number of iterations which is equal to 50, x_i is the prediction for a case with a given model i . μ is the mean prediction for the case over all 50 iterations.

$$VAR_{tot} = MSE_{cv} + \frac{1}{1-n} \sum_{i=1}^n (x_i - \mu)^2 \quad (3)$$

With VAR_{tot} defined, it is then a relatively straightforward approach to define prediction intervals for any defined confidence level. This is

done by taking the square root of VAR_{tot} then multiplying it by $1 - \alpha$, where α is the quantile function of a normal distribution (upper tail) for a defined confidence level. For example, for a 90% prediction interval, the quantile function is approximately 1.64 ($\alpha = 0.95$). The multiplication of the square root of VAR_{tot} and the specified quantile function equates to a standard error which is added to and subtracted from the mean estimate of the EPD to derive the upper and lower prediction intervals, respectively. For all soil spectral models, we derive prediction intervals with 90% level of confidence about the mean estimates

Each of the models were validated in terms of the out-of-bag data withheld from each model iteration. Here, we report the root mean square error normalised by the interquartile range of the data (NRMSE) and Lin's concordance coefficient (CCC) to quantify the prediction accuracy and fidelity about the 1:1 line when comparing predictions with associated observations, respectively. The predictions here are the averaged values derived for each case.

All spectra were pre-processed prior to model fitting to increase signal-to-noise ratio through a sequence of steps which included: spectral trimming, Savitzky-Golay smoothing, and spectra baseline correction using the standard normal variate transformation all of which are described in more detail in [Wadoux et al. \(2021\)](#).

Soil spectral model extension. A soil information database containing all scanned depth intervals for all soil cores was populated via model extension to the pre-processed vis-NIR spectra collected, i.e. inferred at every 2.5 cm to 20 cm, and then every 5 cm to the full length of the soil core. In this work, we derived the mean predictions, but in later studies the inclusion of spectral model prediction uncertainties into downstream processes such as digital soil mapping may be considered. The mean soil attribute predictions formed the input data needed for creating digital soil maps for BARS.

Digital soil mapping

Digital soil maps were derived for those soil analytes described in the laboratory analysis section, plus dried soil bulk density, which was directly inferred via gamma attenuation as previously described. The same digital soil mapping workflows were used across all soil analytes, except for soil texture where similarly to the soil spectral inference work, a compositional data analysis framework was used. Final digital soil attribute maps were derived for several soil depths including maps of corresponding estimates of prediction uncertainty.

In general, 3D-like representations of soil attribute maps are produced via creation of maps for specified depth intervals of soil, in effect generating different layers for sequential depths down a profile to some defined maximum depth. Some guidance comes from the global soil mapping specifications for example where the recommendation is to output layers for the following depths: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm ([Arrouays et al., 2014](#)). Soil profile data is usually harmonised using one of a variety of soil depth functions including splines, decay functions and other mathematically defined functions which are integrated to the target harmonised depth intervals ([Malone et al., 2018b](#)). In this work we used a mass-preserving soil depth function ([Bishop et al., 1999](#)) due to its flexible behaviour and wide use among DSM practitioners. More detail about the underpinning theory and application of the mass-preserving spline can be found in either [Bishop et al. \(1999\)](#) or [Malone et al. \(2009\)](#).

The general use case for splines is to apply them to a collection of soil profile data where no common depth interval representation is present, such as a collection of profiles that have been described and characterised according to pedogenic horizons. And the spline is useful in this case because once a common depth support is determined, all soil profiles can be harmonised to these intervals without loss of the original information due to the mass preserving properties of the spline. The digital soil mapping workflow then proceeds where the target variable is modelled for each depth interval as a function of an exhaustive suite of covariates or environmental variables. Sometimes, depth itself may be

included in the model prediction variables, meaning that all available data can be included into a single predictive function instead of a separate one pertaining to each depth interval. The former approach (modelling each depth interval separately) is probably the most common approach. When including depth interval as a predictive variable, one needs to be careful about the selection of the modelling algorithm because artefacts generally will occur at the layer boundaries (Ma et al., 2021).

For the current work, all the characterisations for each soil core correspond to a common depth support as outlined previously, which means that the step of fitting splines could be negated here, and one could move immediately to the spatial modelling work. However, because spline fitting to soil profile data is an operationally simple task to carry out, and to demonstrate that digital soil maps can be customised to any specification set by a map user, we opted to set (arbitrarily) the depth support to the following intervals: 0–10 cm, 10–20 cm, 20–40 cm, 40–60 cm, 60–80 cm, 80–100 cm, 100–120 cm, 120–140 cm, 140–160 cm, and 160–180 cm. For example, at BARS a customized DSM depth support was required for two use cases; the estimation of lime requirement in certain paddocks which required finely resolved increment prediction intervals in the top 30 cm to detect banding of subsurface soil acidity, and soil moisture mapping. More on the lime requirement use case is discussed later on, whereas the soil moisture mapping work is discussed in a follow up research article.

Note that the spline function does not extrapolate beyond the maximum thickness of a soil core or profile. Also note that a relative important parameter of the spline (λ) controls the flexibility of the spline and consequently the fidelity of fitted spline to the observed data. There is a tendency when using very low values of λ that overfitting occurs which can manifest by providing unreliable point-based estimates down a soil profile. Therefore, from a series of visual experiments performed on the data we accepted a λ value of 0.1 to ensure a good balance between fidelity and spline error. More information about optimal tuning of the λ parameter is provided in Malone et al. (2009).

With the target variable data in the specified depth support, this information is then spatially intersected (via the nearest neighbour intersection method) with the suite of variables derived from the on-the-go proximal soil sensing work described earlier.

Spatial modelling was undertaken and was underpinned by the random forest (RF) data modelling algorithm (Breiman, 2001), specifically the ‘Ranger’ R package implementation (Wright and Ziegler, 2017). RF is used to model the structural component of the data which is the relationship between the target variable and a suite of prediction variables. This is followed up by residual modelling via variograms and kriging to account for additional spatial variability through the auto-correlated RF model errors which may or may not exhibit definitive spatial patterns. In this typical ‘regression kriging’ modelling approach the structural and residual components of the model predictions are summed together. Regression kriging models were derived for each target variable with the regression component conducted for all depths simultaneously whereby soil depth layers were used as a model predictor variable. Variogram modelling of the RF model residuals was performed for each depth interval. For variogram modelling, rather than visual observation to determine the best type, we opted to use auto-fitted variograms (Hiemstra et al., 2009) where up to 5 different types of model (Spherical, Exponential, Gaussian, Matern and Matern with Stein’s parametrisations) are tested simultaneously to determine the best one for the specific data configuration on hand. For the RF modelling we optimised the RF hyperparameter ‘mtry’ (number of variables to possibly split in each node of the random forest model) using a purpose built cross-validation scheme that is facilitated in the ‘caret’ R package (Kuhn et al., 2019). The number of trees to grow was not optimised and set to 500 as this was more computationally efficient.

After the model hyperparameter optimisation, 50 iterations of the regression kriging model system were run using a bootstrap data

resampling system in a similar fashion and purpose as described earlier in the soil spectral modelling section. The intent here is to quantify prediction uncertainties and to provide model validation diagnostics based on data not included in the modelling. We subsequently report the NRMSE and CCC indices to assess model goodness of fit.

The bootstrapped regression kriging modelling system is then used to output maps to the same extent and resolution of the proximally sensed data. Digital soil attribute maps were produced for the bootstrap regression kriging prediction average and the upper and lower 95th and 5th prediction limits (constituting a 90% confidence interval).

Method implementation

To carry out the workflows described in this study, numerous bespoke R code scripts were developed. Some of the main R packages used in this work include ‘ranger’ (Wright and Ziegler, 2017) for model fitting and both ‘raster’ (Hijmans, 2019) and ‘sp’ (Bivand et al., 2013) for specific spatial data analysis tasks. Variogram modelling was facilitated by the ‘automap’ package (Hiemstra et al., 2009).

Results and discussion

Results of the modelling, both soil spectral inference (Table 1) and digital soil mapping (Table 2) are presented in a similar way. Along with each soil attribute inferred or mapped is a description of whether and what data transformation was used in model fitting. For example, with each of the soil texture fraction, the isometric log ratio was used, for soil organic carbon the natural logarithm was used to normalise the data while for pH no data transformation was used at all. Transformation used for soil spectral modelling were also used for the digital soil mapping, but a check was made, in any case, prior to the spatial modelling. We then also show the NRMSE and concordance metrics for data being ‘in-bag’ and ‘out-of-bag’ which just means the model goodness of fit was performed with data included in model calibration or not. The ‘out-of-bag’ assessments will give a better indication of whether models are useful or not. The use of the NRMSE metric in this study was to assist in providing fair comparisons of the efficacy of models for each soil attribute. Values close to 0 indicate very low error while higher values indicate increasing levels or error. As our assessments of models are normalised by the interquartile range of the input data, it is possible to have NRMSE values above 1 which would indicate the model does not have any useful predictive skill. The model goodness of fit measures were performed on the data in their natural units as this information is much more useful for users of the model outputs.

Soil spectral inference to characterise whole soil cores

It is a general expectation that model in-bag goodness of fit measures are less good compared with out-of-bag, and this is apparent with the model outcomes shown in Table 1. Soil texture fractions except for silt appear predictable with vis-NIR spectroscopy in this case. Relatively less predictable were nitrogen, soil carbon (total and organic), pH and ECEC. Available phosphorous and EC were not predictable at all with our data. This scale of model predictive ability aligns quite well with previous research efforts using vis-NIR for soil spectral inference. For example, infrared analysis is well suited for soil carbon and total nitrogen analysis because of its sensitivity to the C–H, C–O, and C–N functional groups that dominate in organic matter. Similarly, clay mineralogy can be readily estimated, too as there are established spectral signatures for suites of primary and secondary clay minerals in both the vis-NIR and MIR regions (Soriano-Disla et al., 2014), and because of this, attributes like cation exchange capacity are relatively well predicted, too. Soil texture (at least clay and sand proportions) is also well predicted as infrared spectra is sensitive to the surface and solid composition of the soil. Attributes related to the soil solution rather than to the chemistry of the soil matrix are generally not going to be well predicted using

Table 1

Model goodness of fit outcomes from soil infrared spectroscopy for each of the target soil attributes.

Soil Attribute	Data transformation	Model in-bag goodness of fit		Model out-of-bag goodness of fit	
		NRMSE	Lin's Concordance	NRMSE	Lin's Concordance
Clay (%)	isometric log transformation	0.35	0.81	0.40	0.76
Sand (%)	isometric log transformation	0.40	0.76	0.46	0.70
Silt (%)	isometric log transformation	0.76	0.39	0.78	0.35
Total Nitrogen (%)	Square root	0.62	0.76	0.75	0.67
Total carbon (%)	Natural logarithm	0.65	0.82	0.82	0.71
Soil organic carbon (%)	Natural logarithm	0.67	0.77	0.71	0.68
EC (dS/m)	Natural logarithm	1.62	0.56	1.78	0.43
pH (1:5 H ₂ O)	none	0.63	0.49	0.68	0.41
pH (1:5 CaCl ₂)	none	0.62	0.57	0.71	0.42
Colwell P (mg/kg)	Natural logarithm	1.16	0.67	1.21	0.59
ECEC (cmol ⁺ /kg)	Square root	0.77	0.71	0.82	0.67

Table 2

Model goodness of fit outcomes from digital soil mapping for each of the target soil attributes.

Soil Attribute	Data transformation	Model in-bag goodness of fit		Model out-of-bag goodness of fit	
		NRMSE	Lin's Concordance	NRMSE	Lin's Concordance
Bulk Density (kg/m ³)	none	0.21	0.95	0.40	0.83
Clay (%)	isometric log ratio	0.21	0.94	0.42	0.80
Sand (%)	isometric log ratio	0.22	0.95	0.44	0.80
Silt (%)	isometric log ratio	0.22	0.94	0.42	0.79
Soil organic carbon (%)	Natural logarithm	0.17	0.97	0.34	0.87
pH (1:5 H ₂ O)	none	0.16	0.97	0.32	0.86
pH (1:5 CaCl ₂)	none	0.19	0.95	0.38	0.81
Colwell P (mg/kg)	Natural logarithm	0.15	0.97	0.30	0.88
ECEC (cmol ⁺ /kg)	Square root	0.22	0.96	0.43	0.84

infrared soil spectroscopy (Janik et al. 1998) and this was the case for Colwell P and EC. Soil pH on the other had was relatively well predicted which is indicative more of the correlation of these attributes with others such as soil texture, which in this work was quite well predicted.

Digital soil mapping

The relative change in model goodness of fit between in-bag and out-of-bag is greater for digital soil mapping compared with those from the soil spectral inference. This relationship is due to model selection. Random Forest modelling was used for digital soil mapping and does tend to deliver very strong model fitting capability based on the calibration or in-bag data. Therefore, it is necessary to assess the model using the out-of-bag data because we need to evaluate whether the model generalises well. From the results in Table 2, it is observed that the out-of-bag NRMSE values for all attributes is under 0.5 indicating very satisfactory and skilful spatial models. We did not carry out spatial modelling of the attributes: Total carbon, total nitrogen, and EC. In our case SOC is highly correlated with total nitrogen and carbon, so in this current work mapping on SOC seems appropriate. For EC, as this was not that well predicted with soil infrared spectroscopy, we decided not to proceed with the digital mapping step. Although Colwell P was also not very well inferred from vis-NIR spectroscopy, we carried out the spatial model to demonstrate a point of discussion later about certain and uncertain data. Despite relatively poor vis-NIR spectral model prediction, the spatial model of Colwell P is skilful as indicated by lower out-of-bag NRMSE and higher concordance.

Digital soil maps for all the soil attributes summarised in Table 2 can be shared upon request to the authors. For each attribute, maps are given for each of the output depth intervals: 0–10 cm, 10–20 cm, 20–40 cm, 40–60 cm, 60–80 cm, 80–100 cm, 100–120 cm, 120–140 cm, 140–160 cm, and 160–180 cm. These maps are presented with associated 90% confidence upper and lower prediction intervals.

We reserve a thorough interpretation of the soil-landscape relationships for a separate piece of research and discussion. Other than the reliable outcomes of the spatial modelling, these digital soil maps have been assessed in an informal manner by expert knowledge associated

with BARS to verify and confirm general soil patterns across the farm.

Using the digital mapping of soil pH (1:5 CaCl₂) as a discussion example, at the whole farm extent, areas from relatively acidic to neutral can be observed in the topsoil. With increasing soil depth, some areas of acidity persist but there is a general trend of increasing pH with increasing soil depth overall. This general pattern of soil pH is already well established for the Boorowa region where the farm is situated (Hird, 1991), but in this case has been explicitly defined in the spatial context using the digital mapping approaches that were implemented.

A phenomenon that has been observed and documented through the cropping zone of south eastern Australia is the occurrence of pH stratification within the top 20 cm of soil (Condon et al., 2021). Cropping practices lead to pH stratification where OH⁻ is removed with crop biomass and replacement lime is only applied at the surface and not fully incorporated via tillage. In addition to this, locally in the region around BARS some topsoil A1 and many A2 horizons are naturally acid due to the leaching action associated with illuviation and subsurface lateral flow conditions moving solutes (OH⁻), into or across B horizons in texture contrast soils. The ubiquitous down profile pH trends occur where surface soils are more acid and is fundamentally a function of cropping practices and soil age where OH⁻ is stripped from upper layers and deposited in lower B horizons over time. It is likely many of the A2 horizons in texture contrast soils were acidic prior to Pre-European settlement. Both processes result in pH stratification, a common feature of BARS soils particularly in lower slope positions and in areas where water accumulates and A2 horizons are thickest.

Consequently, pH stratification has had negative impacts on cropping, grazing and mixed farming enterprises across the south eastern agricultural region. One of the powerful aspects of the digital soil mapping workflow that has been used in this study and could potentially be used elsewhere is the ability to detect and delineate phenomena such as this pH stratification. This can be done for example through careful selection of the depth intervals to be output from the soil depth function depth which processes the raw soil profile information. As has been established in works such as Filippi et al. (2020) and Roudier et al. (2020) methodologies that have been described in this research and similar can achieve digital mapping outputs even to 1 cm depth

resolution to a nominated maximum depth. This level of granularity is particularly useful for the identification of soil constraints occurrence and estimation of soil natural capital stocks, for example, but could be computationally prohibitive to apply everywhere even with today's excellent computer capabilities. To demonstrate the capability of DSM for granularity of soil characterisation in the vertical dimension, we therefore also generated the same set of digital soil attribute maps following the workflows described, but at a depth support of 5 cm depth intervals between 0–40 cm.

At the farm scale, it is only possible to see subtle differences between maps given the slight change in depth support for the topsoil, i.e., from the original 0–10 cm layer to the 0–5 cm layer as shown in Fig. 5. However, scaling down to the paddock boundary level of a newly established long-term cropping trial at BARS (~11 ha), the usefulness to increase the granularity of depth support becomes evident. In addition, Fig. 6 shows to 60 cm the digital mapping using the original depth support compared to the depth support for every 5 cm to 40 cm in Fig. 7. In both Figures, the western side of the bounded field has a lower soil pH compared with the eastern side of the field. This surface pH is generally higher than the next depth layer below. What we see in Fig. 6 though is that there is a clear decrease in soil pH for the 10–20 cm layer and then an increase for 20–40 cm, a trend that continues for the 40–60 cm layer. From, this original mapping interval representation we might infer that there is this phenomenon of pH stratification in the 10–20 cm layer as described above for south eastern Australia. However, the interval mapping in Fig. 7 shows that this stratification is really occurring between 5 and 30 cm. This level of depth granularity will help aid decision making about how to best ameliorate the soil acidity issue and manage the soil's capability, e.g. whether surface or subsoil incorporation of lime should be used, and if subsoil incorporation, then to which specific depth. Had the maps in Fig. 6 been used to assist with an amelioration plan at BARS, then perhaps the correct treatment and equipment would not have been selected, which could potentially have led to inadequate amelioration and use of resources. The noteworthy point to make here is that the underlying soil data from the soil cores were the same for both mapping contexts, and only the data processing procedure was changed. The here demonstrated customisation or bespoke digital map production ability is a clear point of difference between conventional approaches to creating soil maps and marks a shift in the interaction between soil information and end-users. In our example the bespoke digital soil attribute vertical mapping approach empowered the end users to receive the soil information needed to underpin decision making at their own specifications.

General discussion

Costs of digital soil survey

We have costed up items performed in this work as if each were carried out by commercial operators. For reference though, the proximal soil survey and data modelling were carried out internally by CSIRO, leaving the soil sampling and laboratory analysis to commercial operators. The costs for each item are as follows:

- On-the-go proximal soil survey: AUD\$5000 given survey density and vehicle speed
- Soil Sampling: AUD\$25,348. Inclusive of all travel and associated incidentals
- Soil analysis: AUD\$76,000. For the selected analytes this costs out to \$200 per sample
- Soil Core Scanning: AUD\$17,500. This is based on labour costs of \$250 per hour
- Data Analysis and modelling: AUD\$17,500. Labour costs at same rate as above

On a cost per hectare basis, we estimate the cost of this survey to be about \$640. We would consider these costs to be a capital investment as they have fulfilled the task of providing a detailed assessment of soil variability across the farm. The usefulness of the established soil information infrastructure is that it is very much dynamic in the case that new data can be added seamlessly, and new models can be created accordingly. The nature of ongoing costs will be relatively small compared with the original capital investment of baselining the soil conditions at BARS. However, the value of the information in terms of granularity of detail, precision and accuracy will increase significantly. Much of this granularity in terms of both lateral and vertical soil characterisation has been driven by the incorporation of soil infrared spectroscopy into the soil survey process. Without doubt, soil spectroscopy in terms of measurement error is potentially prone to greater levels of uncertainty compared with a laboratory-base wet chemistry analysis (Soriano-Disla et al., 2014), but this is not always the case as reported in the literature and our own spectral predictions of clay and sand for example are demonstrable of the power of this digital technology. We discuss issues about measurement errors further on but want to make the point here that had we had submitted all soil subsamples to the laboratory for wet chemistry analysis to achieve the same digital mapping outcomes we have presented here, then total costs of the soil analysis would be in the region of

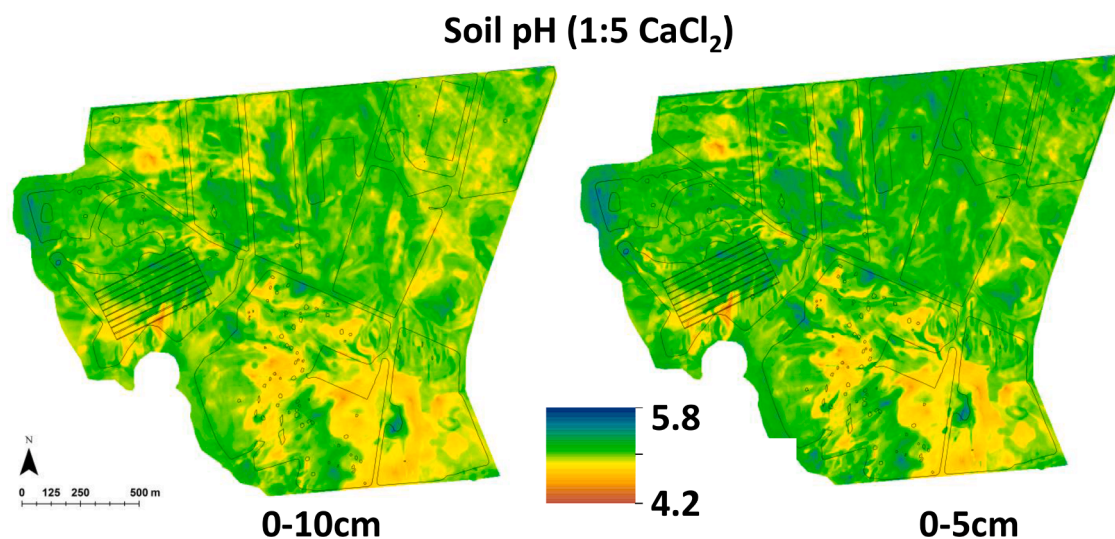


Fig. 5. BARS maps of soil pH (1:5 CaCl₂) output to 0–10 cm and 0–5 cm, respectively. This customisation was enabled through specification of the outputs from the soil depth function used to process in available soil profile data.

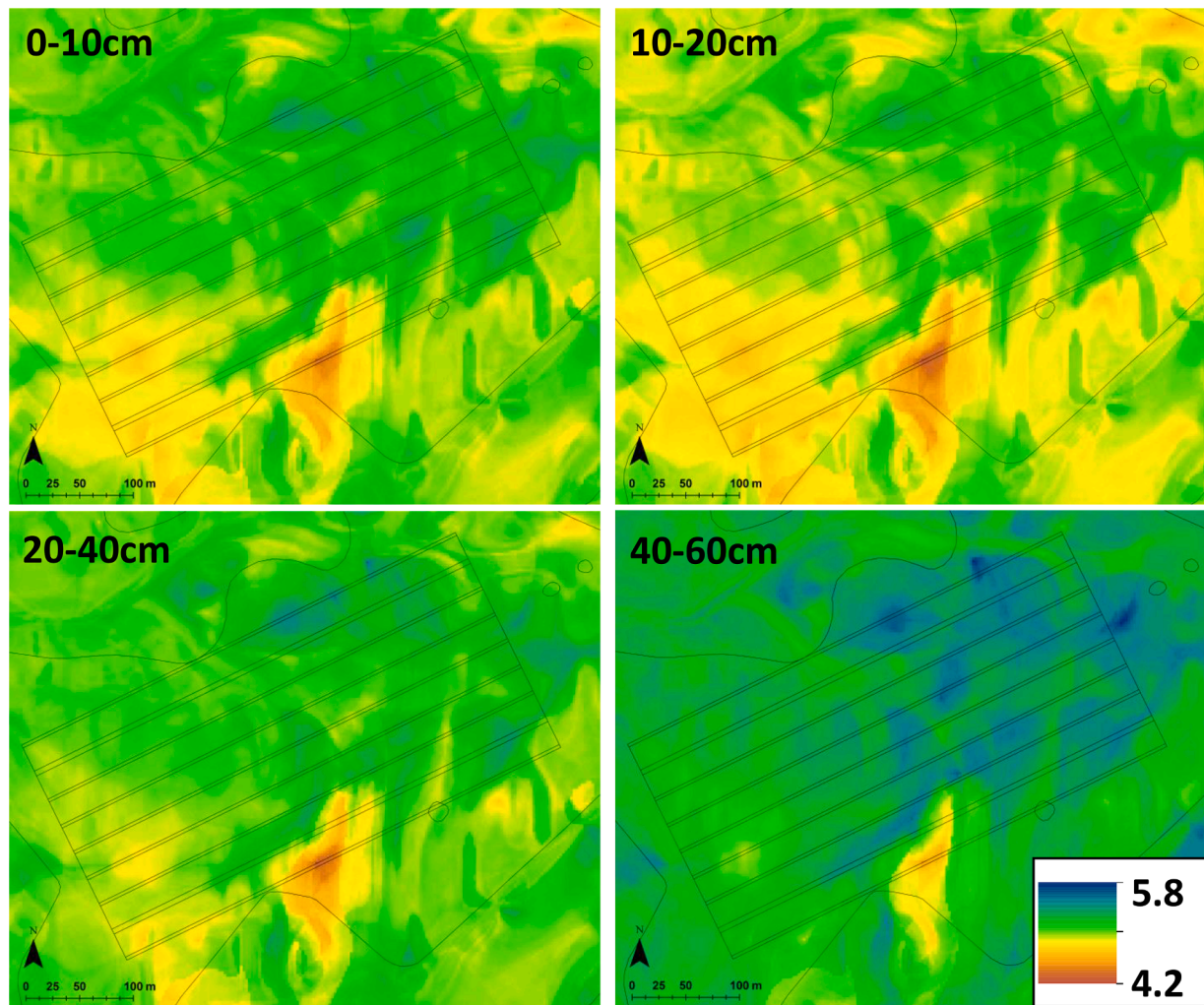


Fig. 6. BARS maps of soil pH (1:5 CaCl₂) focussed onto the location of the long-term cropping trials (~11Ha). The maps show the pH for the 0–10 cm, 10–20 cm, 20–40 cm and 40–60 cm depth intervals.

AUD\$600,000. Keeping to only traditional soil analysis and not advancing with using soil infrared spectroscopy, we may not have captured the spatial nor vertical variation of the BARS soils, resulting in inferior and even unsuitable outputs.

Possibilities of the soils spatial data infrastructure

Before discussing some caveats of the important outcomes from the digital soil survey of BARS presented in this work, we want to cover off on some of the important features and possibilities of the digital soils infrastructure that has been developed.

Ability to get highly resolved spatial soil data

Digital soil mapping enables site-specific characterisation of soil variability. Subsequently, any farming task related to variable rate, precision agriculture, soil natural capital stock accounting or any task related to differential management of the farming system can be informed and underpinned by outputs from digital soil mapping. Moreover, the specificity of the spatial soil information can be leveraged in better ways and into richly-designed agricultural simulation models which are known to be demanding of detailed soil information such as APSIM (Agricultural Production Systems sIMulator; [Holzworth et al., 2018](#)).

Flexibility of modelling approach to generate outputs to the desired resolution and depth support

The granularity of digital soil mapping outputs is often controlled by the grid cell resolution. The selection of this is often the choosing of the map producer or user which is often guided by management priorities. Too coarse a resolution and spatial variability can get overlooked, while too fine a resolution can lead to redundancies or a characterisation of spatial variability that is not particularly important to agricultural productivity outcomes. Anecdotally, management of a farm field on 10 m grids is feasible for broadacre cropping systems. Irrespective of what is best for which context, the digital soil mapping process can be adapted and customised accordingly. As stated earlier, this same customisation applies also in our case regarding depth support and the granularity of vertical soil variability. Adapting a fine characterisation of soil pH down the soil profile, at least in the top 40 or 50 cm is suitable for BARS and likely across the region the farm is situated, but this does not mean that this approach should be adopted exactly for other soil properties or in other regions. This flexibility and bespoke nature to deriving digital soil attribute maps is invaluable and provided an enhanced ability to manage soils across a variety of contexts, albeit with the same underpinning source soil data. Taking this to an operational situation, follow up work with regarding the soil moisture mapping across BARS is another example of soil map customisation where in the first instance the vertical support of the mapping is determined by the sensor depths on the soil probe network across the farm.

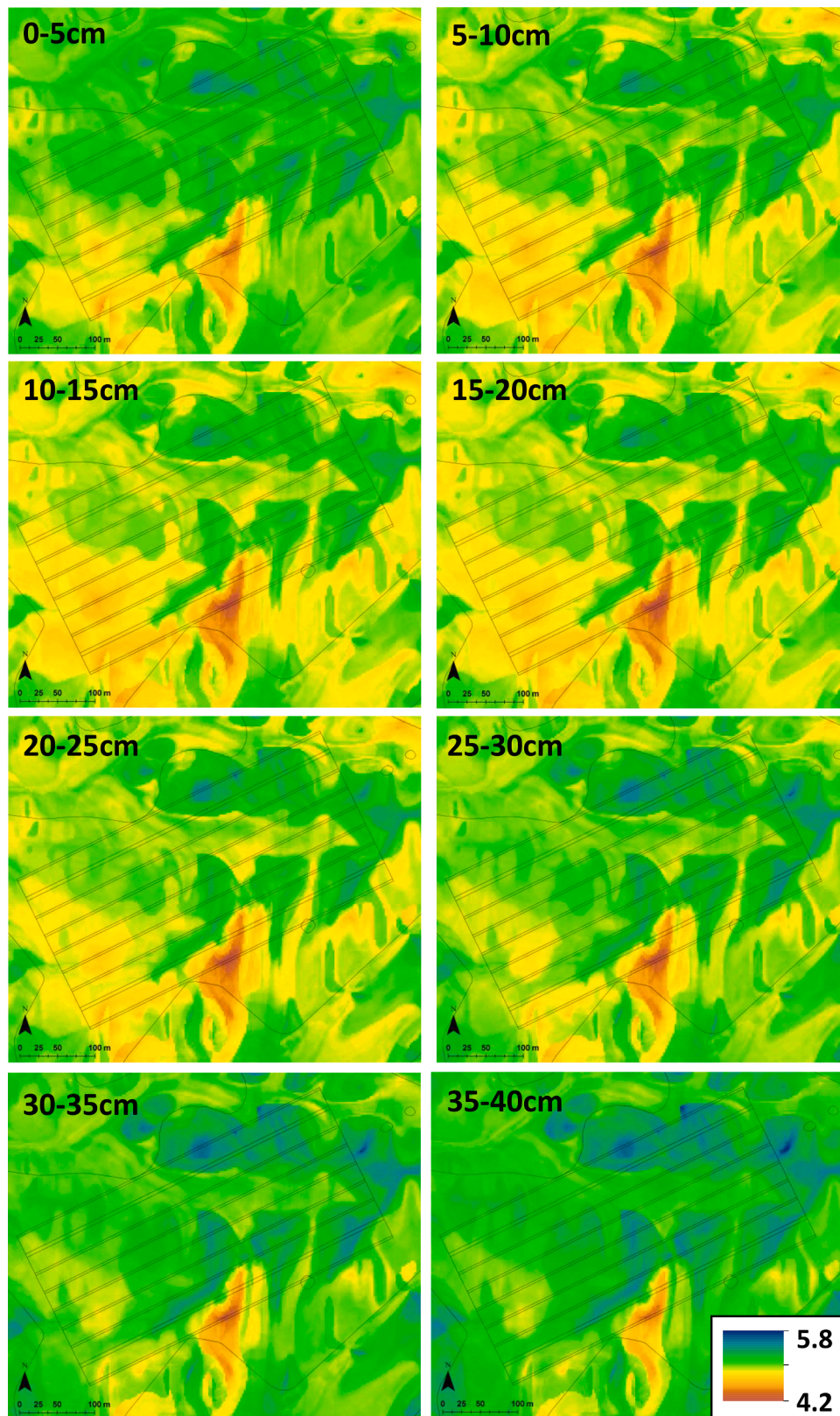


Fig. 7. BARS maps of soil pH (1:5 CaCl₂) focussed onto the location of the long-term cropping trials (~11Ha). The maps show the pH at 5 cm depth intervals to 40 cm.

Efficiency to update and improve models through time

It was not demonstrated in this study, but a possibility and increasingly common scenario is to update digital soil maps with the intention to improve upon previous iterations. Updating a digital map can take various forms that include adapting a new model type to existing data, incorporating new data into the modelling that were not available at the time of the previous version. This could include new site or soil profile data or improved spatial covariate information. Often a new model type and additional data could constitute an updated version. Moreover, the points described above about contextualising the spatial and vertical support of maps would be considered an update or improvement, too. This dynamic nature of the soil mapping process can be invaluable, but it does also necessitate skilled personnel to be on hand and evolve the soil information infrastructure, too. In effect there will need to be an ongoing resourcing need. While for BARS there is no current field plans to explicitly improve upon what has been created to date in terms of whole farm soil mapping, soil information is and will continue to be collected for various contexts for other projects undertaken there. Such data can be used either for a validation in the first instance of the existing products where possible, but eventually would be incorporated into ongoing mapping work.

Expansion of predictive suite for via soil spectral model calibrations and pedotransfer functions

The work carried out in what we would consider to be a reconnaissance soil survey delivered outputs of the primary attributes investigated. The modality to infer a greater suite of soil attributes in addition to the primary ones can be done via processes such as soil spectral inference and pedotransfer functions (Van Looy et al. 2017; McBratney et al., 2006). Both processes are related in that the existing primary information on hand can be used to infer other soil attributes that are oftentimes very expensive or time consuming (or both) to analyse. For example, the measurement of soil carbon fractions which are often used to understand soil carbon turnover, or soil moisture and hydraulic characteristics which require specialised equipment and long times in the field in order just to characterise a limited number of single point sites.

In the case of pedotransfer functions, the general concept is that with relatively easier to measure soil attributes, more difficult to measure ones can be inferred using empirical calibration with defined model structure or type. The relevant soil attributes here for BARS are those related to soil moisture which include attributes such as drained upper and crop lower limits and infiltration rate. The inputs to predict these variables often included soil texture, bulk density, soil carbon and cation exchange capacity. The soil hydraulic attributes are particularly important for example in the calibration of soil moisture sensors and then mapping soil moisture across the farm on a given time step, or for predicting crop yields via APSIM. Using published pedotransfer functions such as in Gasch et al. (2017) or Vervoort et al. (2006), these attributes can be defined relatively easily. Obviously, there can potentially be limits to using pedotransfer functions as they are known to be relevant to a given geographical context or calibrated on certain subset of soils or soil types (Van Looy et al. 2017) making them potentially ineffectual in the context where these models are extended too. So naturally, some in field validation would be required. Ultimately this cycle of estimation and verification would continue (and drawing upon earlier discussion about model updating and improvement) until the mapping products satisfy their requirements.

Soil spectral inference may be coupled with pedotransfer functions in that spectral models could predict primary soil attributes as has been done in this work, and then these predictions are processed through an existing pedotransfer function. This is indeed a possibility but one that we have not explored yet for BARS. The use case that needs elaboration however is the idea of soil spectral model extrapolation, whereby an existing model calibrated with some other unrelated data can be extended to data pertinent to the BARS data. Data here is the soil vis-NIR

spectral library and as has already been established, the spectral response of a soil specimen can provide a significant amount of information about the properties of that soil. If we take the example of vis-NIR modelling of soil carbon fraction as demonstrated in Viscarra Rosel and Hicks (2015), and accounting for instrument and soil differences through harmonisations and corrections, it is feasible to extend those models to the BARS vis-NIR spectral library and others. This work to date for soil carbon fraction mapping across BARS is still in process with the main task being the verification of the spectral model predictions. While some digital mapping of carbon fractions across BARS has been done to date, we do not share them in this instance until some further checking and verification is performed.

Ultimately with the use cases around soil moisture and carbon fractions, it is intended to illustrate the invaluable nature of the current soil information infrastructure for BARS in terms of inferring additional soil attributes to what were initially targeted.

Caveats

Including depth in spatial modelling

Work by Ma et al. (2020) demonstrated there are potential issues with including some information about depth together with the usual suite of predictor variables that are considered in digital soil mapping. They found that with certain algorithms such as random forest models and other machine learners, that resulting predicted depth function can appear stepped rather than continuous. This stepped feature is obviously not a pedological feature, but an artefact of the modelling. As this current work uses the random forest algorithm within the spatial modelling process, without a detailed checking, we would assume there are potentially some step artefacts in the soil maps we have produced. Our motivation to include some indicator of depth in our spatial models was to follow multiple lines of enquiry through trial-and-error which also included comparison against models without using depth increments as a predictor variable. We found substantial improvements if depth was used and are relatively confident, given the nature of our out-bag validation procedure, that any artefacts will be minimal in our case. The advantage of including depth in the spatial model using the modelling approach selected is that co-related soil information (hereby the same soil attribute at other depths in the same profile) provides useful information about what the attribute value may be somewhere else in the profile. The other potential advantage of including depth as a predictor variable is that the modelling complexity can be reduced significantly. Effectively, rather than a model for each targeted depth interval, only one would be needed. While we did implement residual variogram modelling on a depth-wise basis, the time saving device of reducing the number of models to fit allows outputs to be run in a timelier manner.

Incorporation of soil spectral model uncertainty in spatial predictions

In this research we have applied spatial models without consideration of measurement errors of the input data. We know from out-of-bag validation that the soil spectral models were not without error. Moreover, we were able to quantify prediction uncertainties by expressing the true value of each target variable to be within a defined interval which we established with 90% confidence. Surprisingly, laboratory measured data like spectral model predictions are also prone to measurement errors, though are not reported as frequently. Irrespective of this, we have not explored the types of models needed to include this uncertainty into the spatial modelling. Some of these model frameworks include filtered kriging (Christensen 2011), which is entirely posed within a geo-statistical modelling theory. Machine learning adaptations to measurement errors have also been investigated in Malone and Searle (2021) and Czarnecki et al. (2013). These types of models will be explored in further iterations of mapping. For this initial work our first motivation was to establish some general workflows and codebase to accomplish a

functioning soil data infrastructure for BARS.

What we can expect by including measurement errors into the spatial modelling are smoother looking maps and wider prediction intervals (Somarathna et al., 2018). This may affect some of the mapping for example Colwell P where spectral predictions were not ideal. Although the spatial modelling of Colwell P appeared relatively more accurate, this needs to be discounted by the less accurate spectral modelling outcomes. We would not expect a significant change in the overall spatial pattern, but currently the prediction intervals would be misleading i.e., too narrow. Consequently, decisions around P management would need to be adjusted accordingly. For the other soil attributes investigated we would expect more subtle differences given the relatively less spectral model prediction uncertainties.

Conclusion

Digital technologies or what Wadoux and McBratney (2021) describe as the digital convergence has brought undeniable new capacity to analyse and study the soil. This has been demonstrated in this work where we have used a suite of field sensing and lab sensing technologies coupled with powerful data processing and modelling capability to create a suite of comprehensive vertically granular digital soil attribute maps for BARS. We detailed a clear set of steps and their associated technologies so that this type of work can be extended to other sites and contexts. We have likened the costs of the soil survey and digital soil map creation as a capital investment into improved decision support capabilities which enables site-specific management to improve productivity outcomes which can be realised through such things as precise identification and delineation of soil constraints and opportunities, leading to more effective outcomes for securing soils.

The major shift in the creation of the 'farm soil map' amidst the digital convergence is not just the improved management-specific resolution of soil information and soil attribute specificity, but an ability to create customised or bespoke mapping to suit a particular purpose or assessment as demonstrated in the real example of soil pH mapping. Moreover, the digital soils infrastructure and workflows can also be readily updated, with the view to improving them. We also think the ability to analyse the collected data such as from the on-the-go proximal soil sensing to plan soil survey incorporates a collaboration between expert soil survey expertise and data driven decision making, bringing together both analogue and digital concepts.

Acknowledgements

The authors wish to acknowledge the CSIRO Agriculture and Food business unit internal digital platforms strategic investment for the operating and personnel support that made this research possible. We also want to acknowledge CSIRO colleagues Ross Searle and James Moloney for their thoughtful reviews during the drafting of this document.

References

- Proceedings of the 1st GlobalSoilMap Conference. Eds Arrouays, D., McBratney, A., Minasny, B., Hempel, J., Heuvelink, G., Macmillan, R.A., Hartemink, A., Lagacherie, P., McKenzie, N., Arrouays, D., McKenzie, N., Hempel, J., de-Forges, A. R., McBratney, A., 2014. GlobalSoilMap: Basis of the Global Spatial Soil Information System. CRS Press/Balkema, pp. 9–12. Proceedings of the 1st GlobalSoilMap Conference. Edspp.
- Baumann, P., Lee, J., Behrens, T., Biswas, A., Six, J., McLachlan, G. and Viscarra Rossel, R.A. (2022), Modelling soil water retention and water-holding capacity with visiblenear infrared spectra and machine learning. *Eur J Soil Sci.* Accepted Author Manuscript e13220. <https://doi.org/10.1111/ejss.13220>.
- 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), 27–45.
- Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. *Applied Spatial Data Analysis with R*, 2nd ed. Springer.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Cas, R., 1983. A review of the palaeogeographic and tectonic development of the Palaeozoic Lachlan Fold Belt of southeastern Australia, 10. Geological Society of Australia, Special Publication.
- Clark, R.N., Rencz, A.N., 1999. *Manual of Remote Sensing*. John Wiley and Sons, New York, pp. 3–58 pp.
- Clifford, D., Payne, J.E., Pringle, M.J., Searle, R., Butler, N., 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. *Comput. Geosci.* 67, 62–68.
- Condon, J., Burns, H., Li, G., 2021. The extent, significance and amelioration of subsurface acidity in southern New South Wales, Australia. *Soil Res.* 59 (1), 1–11.
- Czarnecki, W.M., Podolak, Saeed, I.T.K., Chaki, R., Cortesi, A., Wierzchoń, S., 2013. 'Machine Learning with Known Input Data Uncertainty Measure, Computer Information Systems and Industrial Management'. eds. (Springer Berlin Heidelberg), Berlin, Heidelberg.
- Egozcue, J.J., Pawłowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric Logratio transformations for compositional data analysis. *Math. Geol.* 35 (3), 279–300.
- Filippi, P., Jones, E.J., Bishop, T.F.A., 2020. Catchment-scale 3D mapping of depth to soil sodicity constraints through combining public and on-farm soil databases - a potential tool for on-farm management. *Geoderma* 374, 114396.
- Gasch, C.K., Brown, D.J., Brooks, E.S., Yourek, M., Poggio, M., Cobos, D.R., Campbell, C. S., 2017. A pragmatic, automated approach for retroactive calibration of soil moisture sensors using a two-step, soil-specific correction. *Comput. Electron. Agric.* 137, 29–40.
- Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. 'Methods of Soil Analysis'. (John Wiley & Sons, Ltd), pp. 383–411.
- Guerry, A.D., Polasky, S., Lubchenco, J., Chaplin-Kramer, R., Daily, G.C., Griffin, R., Ruckelshaus, M., Bateman, I.J., Duraiappah, A., Elmqvist, T., Feldman, M.W., Folke, C., Hoekstra, J., Kareiva, P.M., Keeler, B.L., Li, S., McKenzie, E., Ouyang, Z., Reyers, B., Ricketts, T.H., Rockström, J., Tallis, H., Vira, B., 2015. Natural capital and ecosystem services informing decisions: from promise to practice. *Proc. Natl. Acad. Sci. U.S.A.* 112 (24), 7348.
- Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J.W., Heuvelink, G.B.M., 2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput. Geosci.* 35 (8), 1711–1721.
- Hijmans, R.J. (2019) 'Raster: geographic data analysis and modeling.' (R package version 2.9–5: (<https://CRAN.R-project.org/package=raster>).
- Hird, C. (1991) 'Soil landscapes of the Goulburn 1:250,000 sheet map and report.' (Soil Conservation Service of NSW.
- Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., Snow, V., 2018. APSIM next generation: overcoming challenges in modernising a farming systems model. *Environ. Model. Softw.* 103, 43–51.
- Jones, E., Bishop, T.F.A., Malone, B.P., Hulme, P., Whelan, B.M., Filippi, P., 2022. Identifying causes of crop yield variability with interpretive machine learning. *Comput. Electron. Agric.* 192, 106632.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T. (2019) 'Caret: classification and regression training.' (R package version 6.0-84 (<https://CRAN.R-project.org/package=caret>).
- Lobsey, C.R., Viscarra Rossel, R.A., 2016. Sensing of soil bulk density for more accurate carbon accounting. *Eur. J. Soil Sci.* 67 (4), 504–513.
- Ma, Y., Minasny, B., McBratney, A., Poggio, L., Fajardo, M., 2021. Predicting soil properties in 3D: should depth be a covariate? *Geoderma* 383, 114794.
- Malone, B., Hedley, C., Roudier, P., Minasny, B., Jones, E., McBratney, A., 2018a. Auditing on-farm soil carbon stocks using downscaled national mapping products: examples from Australia and New Zealand. *Geod. Reg.* 13, 1–14.
- Malone, B., Searle, R., 2021. Updating the Australian digital soil texture mapping (Part 2): spatial modelling of merged field and lab measurements. *Soil Res.* 59 (5), 435–451.
- 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), 138–152.
- Malone, B.P., Odgers, N.P., Stockmann, U., Minasny, B., McBratney, A.B., McBratney, A. B., Minasny, B., Stockmann, U., 2018b. Digital mapping of soil classes and continuous soil properties. *Pedometrics*. Springer International Publishing, Cham, pp. 373–413.
- McBratney, A., Koppi, T., Field, D.J., 2016. Radical soil management for Australia: a rejuvenation process. *Geod. Reg.* 7 (2), 132–136.
- McBratney, A.B., Minasny, B., Stockmann, U., 2018. 'Pedometrics. Springer International Publishing, Cham.
- McBratney, A.B., Minasny, B., Viscarra Rossel, R., 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. In: *Geoderma*, 136, pp. 272–278.
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21 (2), 239–245.
- McNeill, J.D., 1990. Geonics EM38 Ground Conductivity Meter: Operating manual for the EM38 instrument. Geonics Limited, Ontario, Canada. Ontario, Canada.
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* 32 (9), 1378–1388.
- Minty, B., McFadden, P., Kennett, B., 1998. Multichannel processing for airborne gamma-ray spectrometry. *Geophysics* 63.
- Mortimore, J.L., Marshall, L.J.R., Almond, M.J., Hollins, P., Matthews, W., 2004. Analysis of red and yellow ochre samples from Clearwell Caves and Çatalhöyük by vibrational spectroscopy and other techniques. *Spectrochim. Acta Part A* 60 (5), 1179–1188.

- Ramirez-Lopez, L., Wadoux, A.M.J.C., Franceschini, M.H.D., Terra, F.S., Marques, K.P.P., Sayão, V.M., Demattê, J.A.M., 2019. Robust soil mapping at the farm scale with vis-NIR spectroscopy. *Eur. J. Soil Sci.* 70 (2), 378–393.
- Rayment, G., Lyons, D.J., 2010. *Soil Chemical Methods - Australasia*. CSIRO Publishing.
- Rossiter, D.G., Hewitt, A.E., Dominati, E.J., McBratney, A.B., Minasny, B., Stockmann, U., 2018. Pedometric valuation of the soil resource (Eds.). *Pedometrics. Progress in Soil Science*. Springer International Publishing, Cham, pp. 521–546.
- Roudier, P., Burge, O.R., Richardson, S.J., McCarthy, J.K., Grealish, G.J., Ausseil, A.G., 2020. National scale 3D mapping of soil pH using a data augmentation approach. *Remote Sens.* 12.
- Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafyllis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., Orton, T., Odgers, N., O'Brien, L., Minasny, B., Bennett, J.M., Liddicoat, C., Jones, E., Holmes, K., Harms, B., Gray, J., Bui, E., Andrews, K., 2021. Digital soil mapping and assessment for Australia and beyond: a propitious future. *Geod. Reg.* 24, e00359.
- Sherman, D.M., Waite, T.D., 1985. Electronic spectra of Fe³⁺ oxides and oxide hydroxides in the near IR to near UV. *Am. Mineral.* 70 (11–12), 1262–1269.
- Somarathna, P.D.S.N., Minasny, B., Malone, B.P., Stockmann, U., McBratney, A.B., 2018. Accounting for the measurement error of spectroscopically inferred soil carbon data for improved precision of spatial predictions. *Sci. Total Environ.* 377–389, 631–632.
- Sonka, S.T., Campos, H., 2021. Digital technologies, big data, and agricultural innovation. *The Innovation Revolution in Agriculture: A Roadmap to Value Creation*. (Springer International Publishing: Cham, pp. 207–226.
- Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., Macdonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.* 49 (2), 139–186.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., Sparks, D.L., 2010. Chapter Five - Visible and Near Infrared Spectroscopy in Soil Science, 107. Academic Press, pp. 163–215.
- Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer functions in earth system science: challenges and perspectives. *Rev. Geophys.* 55 (4), 1199–1256.
- Vervoort, R.W., Minasny, B., Cattle, S.R., 2006. The hydrology of Vertosols used for cotton production: II. Pedotransfer functions to predict hydraulic properties. *Aust. J. Soil Res.* 44 (5), 479–486.
- Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: australia's contribution to the GlobalSoilMap project. *Soil Res.* 53 (8), 845–864.
- Viscarra Rossel, R.A., Hicks, W.S., 2015. Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. *Eur. J. Soil Sci.* 66 (3), 438–450.
- Viscarra Rossel, R.A., Lobsey, C.R., Sharman, C., Flick, P., McLachlan, G., 2017. Novel proximal sensing for monitoring soil organic C stocks and condition. *Environ. Sci. Technol.* 51 (10), 5630–5641.
- Wadoux, A., Malone, B., Minasny, B., Fajardo, M., McBratney, A., 2021. *Soil Spectral Inference with R*. Springer, Cham.
- Wadoux, A.M.J.C., McBratney, A.B., 2021. Digital soil science and beyond. *Soil Sci. Soc. Am. J.* 85 (5), 1313–1331.
- Whelan, B., McBratney, A.B., Minasny, B., Stockmann, U., 2018. Site-specific crop management. *Pedometrics*. (Springer International Publishing: Cham), pp. 597–622 pp.
- Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* 1 (Issue 1) (2017).