CSIRO PUBLISHING
Soil Research, 2015, **53**, 932–955
http://dx.doi.org/10.1071/SR14268

Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia

Darren Kidd[A,B,C], *Mathew Webb*[A,B], *Brendan Malone*[B], *Budiman Minasny*[B], *and Alex McBratney*[B]

[A]Sustainable Landscapes Branch, Department of Primary Industries, Parks, Water and Environment, 171 Westbury Road, Prospect, Tas. 7250, Australia.
[B]Faculty of Agriculture and Environment, University of Sydney, 1 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015, Australia.
[C]Corresponding author. Email: darren.kidd@dpipwe.tas.gov.au

Abstract. Until recently, Tasmanian environmental modelling and assessments requiring important soil inputs relied on conventionally derived soil polygons that were mapped up to 75 years ago. In the 'Wealth from Water' project, digital soil mapping (DSM) was used in a pilot project to map the suitability of 20 different agricultural enterprises over 70 000 ha. Following on from this, the Tasmanian Department of Primary Industries Parks Water and Environment has applied DSM to existing soil datasets to develop enterprise suitability predictions across the whole state in response to further expansion of irrigation schemes. The soil surfaces generated have conformed and contributed to the Terrestrial Ecosystem Research Network Soil and Landscape Grid of Australia, a superset of GlobalSoilMap.net specifications. The surfaces were generated at 80-m resolution for six standard depths and 13 soil properties (e.g. pH, EC, organic carbon, sand and silt percentages and coarse fragments), in addition to several Tasmanian enterprise-suitability soil-attribute parameters.

The modelling used soil site data with available explanatory state-wide spatial variables, including the Shuttle Radar Topography Mission digital elevation model and derivatives, gamma-radiometrics, surface geology, and multi-spectral satellite imagery. The DSM has delivered realistic mapping for most attributes, with acceptable validation diagnostics and relatively low uncertainty ranges in data-rich areas, but performed marginally in terms of uncertainty ranges in areas such as the World Heritage-listed Southwest of the state, with a low existing soil site density. Version 1.0 soil-attribute maps form the foundations of a dynamic and evolving new infrastructure that will be improved and re-run with the future collection of new soil data. The Tasmanian mapping has provided a localised integration with the National Soil and Landscape Grid of Australia, and it will guide future investment in soil information capture by quantitatively targeting areas with both high uncertainties and important ecological or agricultural value.

Additional keywords: digital soil mapping, legacy data, radiometrics, regression trees, SRTM-DEM, TERN, terrain, uncertainty.

Received 25 September 2014, accepted 13 February 2015, published online 13 October 2015

Introduction

Until recently, Tasmanian environmental modelling and assessments requiring important soil inputs relied on subjectively derived soil polygons that were mapped up to 75 years ago. Commencing in 2009, numerous irrigation schemes commissioned by the state government have been initiated across much of Tasmania's agricultural land, primarily to intensify and diversify agricultural and horticultural production, and capitalise on the state's favourable climate and soils to ensure food security and economic prosperity (Kidd *et al.* 2012*b*, 2014*a*, 2014*b*). This current and impending land-use change is driving the need for improved spatial soils data as functional modelling parameters to assess suitability, and identify potential environmental degradation hazards. Most modellers require two-dimensional, continuously varying representations of soil attributes known as surfaces. These have historically been derived from the 'legacy' soil mapping polygons, with values extracted from modal profiles or classes where qualitative soil description with soil chemical and physical properties has been subjectively associated to similar landscapes. However, improved computing power and spatial modelling techniques have allowed substantial enhancements and generation of three-dimensional (3D) soil-attribute grids, which have now been developed for the whole state.

Digital soil mapping

In 2010, the Tasmanian Department of Primary Industries Parks Water and Environment (DPIPWE), in conjunction with the Tasmanian Institute of Agriculture (TIA) and the University of Sydney, undertook a quantitative enterprise suitability

assessment (ESA) for 20 different enterprises in two pilot areas totalling 70 000 ha as part of the 'Wealth from Water (WfW)' project (Kidd *et al*. 2012*b*, 2014*b*; Webb *et al*. 2014) (http://dpipwe.tas.gov.au/agriculture/investing-in-irrigation). The suitability rule-sets required detailed soil-attribute and climate inputs identifying the most limiting factor (Klingebiel and Montgomery 1961) to derive four suitability classes. Owing to the inappropriate scale, quality and format of the available legacy-soil information, it was necessary to collect new spatial soil information at the appropriate resolution and in a format that better provides soil-attribute values, rather than type or class.

A digital soil mapping (DSM) methodology was chosen as the optimum approach to generate this new soil resource, enabling a quantitative assessment and reduced subjectivity and associated uncertainties of prediction (McBratney *et al*. 2003). There is now sufficient published literature outlining the benefits and appropriate methodologies of DSM to make this a valid scientific approach for development of operational government products. The success and interest generated by the WfW ESA has led to the generation of new soil-attribute mapping for the whole of Tasmania using the DSM '*scorpan*' approach (McBratney *et al*. 2003), based on existing legacy-soil site data and available spatial *scorpan* soil-forming factors. The *scorpan* environmental correlation premise is defined as:

$$S_P = f(S, C, O, R, P, A, N) \quad (1)$$

where the soil attribute of interest at various depths (the soil property at a given site, S_P), is a function (f) of the available spatial soil-forming factors (covariates), where S is available soil data, C is climate (rainfall and temperature), O is influences of organisms (land use and management, vegetation), R is relief (terrain shape and elevation), P is parent material (geology), A is landscape history or age (geological age), and N is the spatial location of the calibration points.

New soil attribute surfaces were generated as Version 1 raster-based maps of a planned, evolving suite of products to be updated as new soil information is collected. The maps were produced at 80-m resolution (equivalent to the 3-s Shuttle Radar Topography Mission (SRTM) digital elevation model; Gallant *et al*. 2011) for standard depths and soil attributes with upper and lower predictions (Table 1), and comply with the Terrestrial Ecosystem Research Network (TERN) Soil and Landscape Grid of Australia (www.tern.org.au/), and Globalsoilmap.net (GSM) programs (Arrouays *et al*. 2014; Grundy *et al*. 2012). They have been uploaded as a regional, stand-alone contribution to the National Soil and Landscape Grid of Australia, and integrated with the national grids by prioritising the areas for inclusion where predictions have the lower uncertainty (www.csiro.au/soil-and-landscape-grid). The suite of products will inform state-wide ESA as well as a range of current and future environmental modelling scenarios. By using the size and distribution of the uncertainties, the spatial reliability of the surfaces can be assessed to encourage and guide future investment in the collection of land resource and soil data by targeting important environmental or agricultural productivity areas with high uncertainties.

The aims of this study are therefore to: (*i*) generate a suite of multi-depth soil attribute surfaces and mapped estimates of uncertainty across the whole of Tasmania at 80-m resolution; and (*ii*) present the methodology and associated modelling diagnostics as accompanying documentation to the Version 1.0 products.

Methods and materials

Study area

Tasmania, as Australia's southern-most and only island state, has a cool-temperate climate, with mean annual rainfall averaging >1800 mm year^{-1} in the west, to <450 mm year^{-1} in the central Midlands (Australian Bureau of Meteorology 2014), driven by the central-plateau rain-shadow effect

Table 1. Tasmanian spatial covariates

	Scale, resolution	Reference, source
	Remote-sensing	
LandSat persistent green areas	80 m (processed)	SAGA GIS (System for Automated Geoscientific Analyses, www.saga-gis.org), LandSat Imagery 2009
Integrated gamma radiometrics—geology (radioactive nuclides: K, U, Th, total dose)	80 m (processed)	Base products (Mineral Resources Tasmania and Geoscience Australia)
	Climate	
Mean annual rainfall	80 m (processed)	DPIPWE, Bureau of Meteorology, Australia
Mean annual temperature	80 m (processed)	DPIPWE, Bureau of Meteorology, Australia
	Terrain	
SRTM DEM-S	80 m	3 Arc Second Digital Elevation Model, adaptively smoothed (Geosciences Australia)
Slope, eastness index, northness index, curvatures (plan and profile), topographic wetness index (TWI), multi-resolution valley bottom flatness (MR), multi-resolution ridge top flatness (MRRTF), vertical desistance to chanel network (VDCN), altitude above channel network (AACN), TCI_Low (lowland exaggeration), topographic position index (TPI), mid-slope position (MSP), terrain ruggedness index (TRI), SAGA wetness index (SWI)	80 m	SAGA GIS (System for Automated Geoscientific Analyses, www.saga-gis.org)

(Davies 1967). Population is ~500 000, with agriculture being one of the most economically important activities. Area is 68 401 km^2, with a diverse range of soils and landscapes and associated native flora and fauna.

Dominant soils and land uses

Some of the most productive soils in Australia are derived from Tertiary basalt on the north-west coast, and the north-east around Scottsdale, used for intensive vegetable and alkaloid poppy cropping and some dairying. These Red Ferrosols (Isbell 2002; Nitisols or Acrisols, IUSS Working Group WRB 2007) are fertile, well structured and freely draining (Spanswick and Kidd 2000), and relatively high in organic carbon (Sparrow *et al*. 1999; Cotching *et al*. 2009; Cotching and Kidd 2010; Cotching 2012). The Midlands (from Launceston to Hobart) is another important agricultural area for Tasmania, supporting cereal cropping, alkaloid poppies, and grazing beef and sheep. The area is predominantly associated with duplex soils (sharp change in texture between the A and B horizons), many of which are sodic (exchangeable sodium percentage >6). These classify as Sodosols (Isbell 2002; Solonetz or Lixisols, IUSS Working Group WRB 2007). Primary salinity is evident in small, localised break-in-slope and depression areas in the lowest rainfall areas of the Midlands (Kidd 2003).

Soils formed from Jurassic Dolerite cover much of the state (Kirkpatrick 1981), consisting of undulating low hills and mountainous areas of stony Brown Dermosols (Isbell 2002; Lixisols, IUSS Working Group WRB 2007) supporting grazing on foot-slopes, native and plantation forestry, and conservation (Cotching *et al*. 2009). Sandy coastal plains provide grazing, dairy and cropping in the far north-west and north-east, forming Aeric, Acquic and Semi-acquic Podosols (Isbell 2002; Podzols, IUSS Working Group WRB 2007) (Cotching *et al*. 2009). Perennial horticulture (mainly apples) is common in the Huon Valley (south of Hobart), and is proliferating as emerging stone-fruit and viticulture industries in many other parts of the state.

The state's west and south-west have large areas of ecologically important conservation land, much of this with World Heritage Area (WHA) listing. These are mainly wilderness areas of rainforest, peatlands and moorlands, from button-grass plains to rocky skeletal mountain ranges. The areas contain vast areas of peat soils, extremely high in organic carbon and matter (Organosols, Isbell 2002; Histosols, IUSS Working Group WRB 2007).

Legacy soil information

Much of Tasmania's historical soil information takes the form of reconnaissance-level soil surveys undertaken by CSIRO Division of Soils, Adelaide, between 1940 and 1967, consisting of soil mapping at a scale of 1 : 63 360, reports, site descriptions and analytical samples. These maps and reports were updated and correlated by the DPIPWE between 1997 and 2001, and re-published at a scale of 1 : 100 000 (Spanswick and Kidd 2001). Additional soil mapping was undertaken by DPIPWE in 1993 for a 1 : 100 000 map sheet in the South Esk region (Doyle 1993), and as 1 : 100 000 scaled land-capability mapping of the important agricultural areas through most of the 1990s (Grose 1999). Additional ad hoc 1 : 100 000 surveys have been undertaken by Forestry Tasmania in some of the state-forest areas (Forth, Pipers and Forester map sheets), as well as several minor, more detailed surveys in various agricultural parts of the state. Most of the state's legacy-soil mapping has involved assigning soil type (as the dominant soil profile class, i.e. a grouping of similar soil properties, described values, parent material and topographic position into a modal or typical conceptual soil based on soil attribute ranges) or soil associations, where a dominant soil is assigned to a polygon, described as in association with other unmapped minor soils, based on a regularly repeating landscape pattern (Spanswick and Kidd 2001; McKenzie *et al*. 2008). Figure 1 shows the extent of the correlated 1 : 100 000 soil maps, and existing soil database sites.

Most of this mapping was on agricultural land; however, vast but very important ecologically sensitive areas of the Southwest WHA remain relatively unmapped or sampled. These areas are vulnerable to land-use and climate change in terms of threatened species and carbon storage (Tasmanian Climate Change Office 2012). In addition, the agriculturally important north-west Ferrosols are under-represented in the legacy mapping.

The DPIPWE soil database holds ~5500 soil sites, descriptions, analytical data and field observations of varying quality. These sites formed the basis for the soil survey descriptions and associated mapping, as well as other *ad hoc* monitoring or environmental assessments.

The only other available soil-related mapping is Land Systems of Tasmania, available for the entire state at a nominal scale of 1 : 250 000, a series of mapping and reports developed in the 1980s based on existing soil mapping, geology, terrain, rainfall and vegetation (Richley 1978; Pinkard and Richley 1982; Davies 1988; Pemberton 1989). This is essentially in accordance with the SOTER (World Soils and Terrain Digital Soils Database) approach (Land and Water Development Division 1993; Oldeman and Van Engelen 1993), where each land-system polygon is conceptually delineated on the basis of these repeating environmental characteristics, with minor components split on topographic position, vegetation and/or brief soil descriptions. Through an expert process, DPIPWE have assigned modal soil profiles to these minor unmapped components, which have been attributed and uploaded to the Australian Soil Resources Information System (ASRIS) (www.asris.csiro.au) as most likely soil properties of standard depths for percentage area estimates of minor components.

For any Tasmanian environmental modelling or assessments requiring important soil attribute information as inputs, the 1 : 100 000 polygonal soil mapping was the only major source of soil information available in many agricultural areas. Elsewhere, it was necessary to rely on the coarse and conceptual land systems. Where soil types or associations were mapped, it was first necessary to determine the range or averaged soil property or descriptive value from the conceptual soil type or profile class, and then determine an area-weighted-mean by each polygon, for each major and minor unmapped soil (subjectively estimated) component. This was difficult where no estimate was available of minor soil component area.

The age of the Tasmanian legacy soil mapping and its continued usage by decision makers confirms that investment

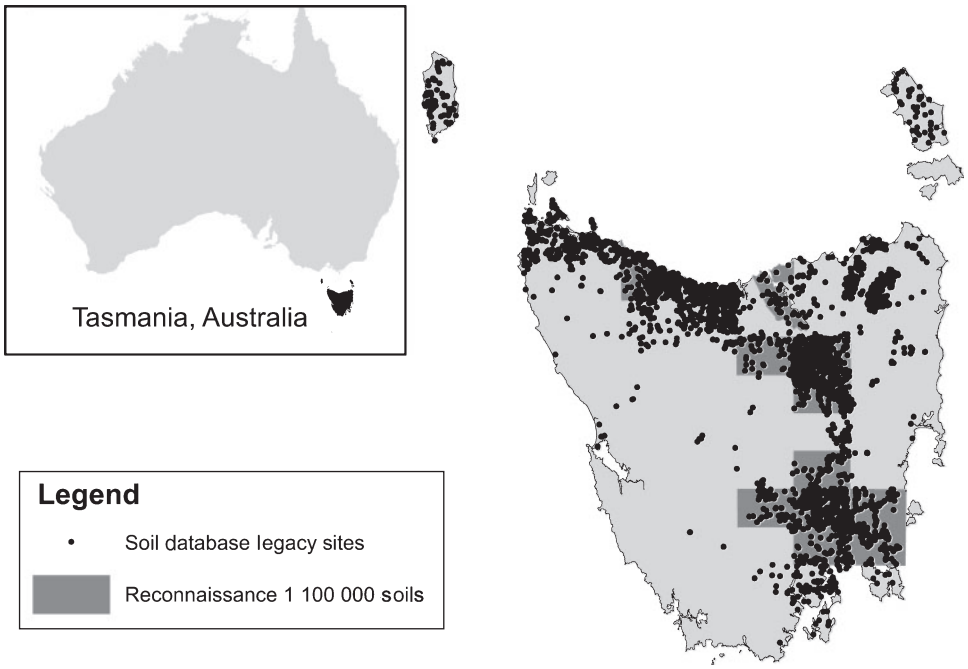

Fig. 1. Tasmania—location and legacy soils data.

in soil information infrastructure is worthwhile, and of positive cost–benefit.

Calibration sites

Site data, including spatial reference, soil attribute of interest, and upper and lower depths, were extracted from the DPIPWE Natural Values Atlas (http://dpipwe.tas.gov.au/conservation/development-planning-conservation-assessment/tools/natural-values-atlas) soils database, and cleaned to remove obvious errors, (e.g. invalid attribute values, depths, or coordinates). Database sites were sourced from a variety of different projects, areas and uses and over a wide temporal range. For example, sites from CSIRO soil reconnaissance mapping from the 1930s to 1950s, land-capability sites from the 1990s and 2000s, and the more recent ESA (Kidd *et al*. 2012*a*, 2012*b*, 2014*b*). Consequently, the remaining sites have a wide range of spatial precision, chemical analyses methodology, and surveyor descriptions. It was important therefore to ensure that analytical methodology was consistent, removing unreferenced sources and applying transfer-functions where known methodology relationships have been developed. Temporal variability was not considered for the Version 1.0 outputs; hence, they essentially show the average soil-property condition over time in Tasmania, as per GSM specifications (Arrouays *et al*. 2014). It is acknowledged that there would be high temporal variability for surface soil attributes such as pH, electrical conductivity (EC) and organic carbon percentage, which are highly affected by land use and management. Subsoil values are less prone to change (McKenzie *et al*. 2002), therefore producing more stable modelling. However, site numbers were insufficient to use more recent data (e.g. over the last decade); this will be re-assessed for future version updates as additional legacy data are incorporated, or from new field-sampling campaigns.

Spatial clustering may also be evident with the majority of database sites, most of which were located using a purposive 'free-survey' approach (National Committee on Soil and Terrain 2009) and could therefore not adequately represent the entire covariate feature space (Carré *et al*. 2007*b*). In cases where the underlying range of covariates is not adequately sampled, de-clustering approaches are generally not effective; a de-biasing approach is more beneficial (Pyrcz and Deutsch 2003). For the Version 1.0 undertaking, no attempt was made to remove sites because of clustering or bias. It was assumed that more intensively sampled areas would provide the opportunity to develop better target covariate relationships, potentially lowering uncertainties in these areas. Modelling bias towards more intensively sampled areas is inevitable in these situations but is intuitively less problematic where a data mining approach is used, because there is no geostatistical component within the modelling process.

An average nearest neighbour analysis (ANNA) of an example dataset (coarse fragments) (using ESRI ArcGIS 10.2) resulted in a nearest neighbour ratio (NNR, observed mean distance divided by expected (random) mean distance); Clark and Evans 1954; Ebdon 1985; Mitchell 2005; Pinder and Witherick 1972) of <1.0 (0.25), implying that the site data are not random, but spatially clustered (as expected). It is expected that other soil-attribute training datasets will also be spatially clustered because many of these are from the same field observations. De-clustering or de-biasing will therefore need to be considered in future Version 1.0+ updates as more data become available to train the DSM models, along with whether this is strictly necessary for a data mining, rather than a classical geostatistical approach.

Mass-preserving depth-splines (Malone *et al*. 2009, 2011) were fitted to the site data for each horizon sample to produce calibration data for the appropriate standard depths (0–5, 5–15,

15–30, 30–60, 60–100 and 100–200 cm), as per the Soil and Landscape Grid of Australia specifications, a superset of the GSM specifications (Arrouays *et al*. 2014); and 0–15 cm for the ESA requirements (Kidd *et al*. 2012*b*, 2014*b*).

Covariates

Table 1 shows the spatial covariates (*scorpan* soil-forming factors, McBratney *et al*. 2003) chosen to model each soil attribute. These were selected using those covariates most correlated (i.e. important in explaining the soil property value at a given location) in the original ESA DSM pilot project (Kidd *et al*. 2014*b*). However, this mapping had now encompassed the entire state, and covariates that were more globally relevant needed to be considered. Hence, mean annual rainfall and temperature were added. Rainfall was considered especially important for Tasmanian soil formation owing to the previously mentioned west–east rainfall trend across the state, and the associated diversity of soil formation (Cotching *et al*. 2009).

Terrain

For elevation and the associated terrain derivatives (R, relief, as in *scorpan*; McBratney *et al*. 2003), the 3-arc-second SRTM DEM was used (Gallant *et al*. 2011) and projected. This was re-sampled to 80-m resolution due to the southern latitudes of Tasmania, determined as the optimum resolution to re-project the surfaces accurately back into the required geographic coordinate system. It was necessary to produce the surfaces using the Australian Map Grid (GDA94, Zone 55) because some covariate algorithms did not work in the geographic system (e.g. SAGA Wetness Index, SAGA GIS 2013), and this was the standard coordinate system required for the Tasmanian publically accessible spatial internet portal (www.theLIST.tas.gov.au). Several additional terrain derivatives were incorporated into the state-wide modelling, including TCI-Low (SAGA GIS 2013), which exaggerates low-lying relief by relatively highlighting terrain detail in low-inclined regions (Bock *et al*. 2007). This was considered important for differentiating the subtle terrace formations existing in areas of the Launceston Tertiary Basin (Doyle 1993; Kidd 2003). Eastness and northness indices were also generated and incorporated into the modelling to avoid the potential 'confusion' where values such as 359° and 1° are spatially very close but at opposite end of the covariate value range in terms of modelling inputs.

Remote sensing

Gamma radiometrics and geology

Gamma radiometrics were shown to be an important predictor of many soil properties within the ESA pilot work (Kidd *et al*. 2014*b*), as well as DSM activities elsewhere (Cook *et al*. 1996; McKenzie and Ryan 1999; Dobos *et al*. 2000; Viscarra Rossel *et al*. 2014). The Tasmanian products show, in addition to total count (TC), the proportions of radiometric uranium (U), potassium (K) and thorium (Th), which in combination can help to identify areas of deposition (e.g. alluvial) areas, as well as areas of denudation (e.g. mountain ranges) (Pain *et al*. 1999; Taylor *et al*. 2002; Erbe *et al*. 2010; Herrmann *et al*. 2010). This effectively relates to the parent material (P, from *scorpan*; McBratney *et al*. 2003), and the landscape history (A from *scorpan*; McBratney *et al*. 2003).

However, only partial radiometric coverage existed for Tasmania, covering ~50% of the state (Fig. 2). In addition, the other important parent material covariate, geology, was only available at a scale of 1 : 250 000 as a state-wide coverage (Fig. 2), producing mapping 'artefacts' (unrealistic mapping anomalies, see *Discussion*). A large representation of the state's geology was covered by the existing radiometrics; therefore, it was decided to 'model' and extrapolate the existing products into unmapped areas to allow its use as a potential spatial covariate. Initially, this was undertaken by regression tree modelling (Cubist, RuleQuest Research, Empire Bay, NSW; Quinlan 2005), using terrain derivatives as covariates, and TC, U, K, and Th as separate calibration datasets from the existing radiometric coverage, using each raster-cell as a training point; 30% of pixels were 'held-back' to use as validation data.

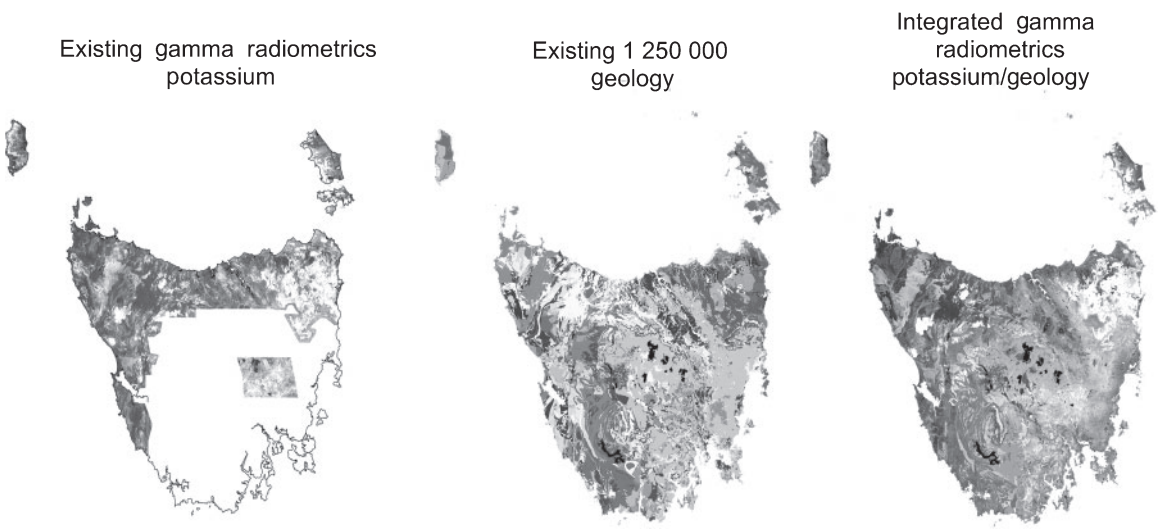

Fig. 2. Existing and extrapolated gamma-radiometrics, Tasmania (potassium).

However, initial surfaces did not adequately reflect some known geological formations in the extrapolation zones, for example, granitic landscapes in mid-west Tasmania. The 1 : 250 000 geology (Mineral Resources Tasmania 2008) was incorporated as an additional covariate into the regression tree modelling, which produced more realistic geological extrapolation. The geology class was used as conditions or partitioning rules for all surfaces (TC, U, K, Th) (Fig. 2, extrapolated K). The final surfaces were tested as both a 'stand-alone' product, introducing an integrated 'geology-radiometrics' covariate, and also by 'stitching' the original radiometrics back into each surface, and tested in initial DSM modelling as a covariate. Improved DSM outputs were achieved by using the integrated geology-radiometrics surfaces in their entirety as covariates and replacement for the 1 : 250 000 geology, producing realistic DSM modelling outputs in terms of known soil–landscape relationships, also with improvements to modelling diagnostics. The benefits of this approach meant that we were able to use the existing radiometric-terrain–geology relationships, extrapolate these to non-mapped parts of the state, and reduce the mapping artefacts produced by using the broad-scale geological mapping (see *Discussion*). It could be argued that this might introduce potential circularity and modelling weakness in the DSM because terrain derivatives were used as spatial covariates in the DSM modelling as well as in the radiometric extrapolation. However, the radiometric extrapolation was able to provide a measure of the terrain and associated parent material relationship that would otherwise be missed by using terrain alone as a modelling covariate, and generally improved validation diagnostics.

Vegetation: persistent greenness

Persistent greenness, that is, areas that highlight where vegetation is 'green' for longer periods of the year were generated as an index using LandSat imagery (Yang *et al.* 2001) and re-sampled to 80-m resolution. This not only explains the vegetation components of the soil-forming factors (O, organism in *scorpan*), but is also useful in identifying 'land use', which has also been shown to explain the variability of soil-property mapping using DSM (McBratney *et al.* 2003). This covariate could explain soils and properties that have a higher nutrient status or water-holding capacity.

Climate

Mean annual temperature and rainfall were generated by using existing Bureau of Meteorology and ESA climate loggers (Webb *et al.* 2014) and incorporated as the climate soil-forming factor covariates (C in *scorpan*). This was undertaken using terrain covariates intersected with 20-year average rainfall and temperature values to form the training dataset, and regression-kriging to estimate the values spatially. Again, these covariates were generated using terrain (raising the potential conundrum of data circularity); however, they were also found to be important explanatory datasets and provided model inputs in terms of topographic variations of temperature and rainfall with improved modelling diagnostics. Where modelling artefacts (see *Discussion*) were introduced as a result of rainfall 'banding', variations in prevailing weather patterns, in terms of rainfall and terrain, were investigated, with rainfall divided by windward–leeward wind effects (SAGA GIS 2013) found to be a good explanatory soil-forming variable for organic carbon. This approach reduced mapping artefacts while maintaining strong modelling diagnostics.

Modelling

A raster stack of all covariates was generated and the target variable (each soil property and depth) individually intersected with the covariate values to provide the calibration and validation data. All modelling was undertaken in R (R Development Core Team 2014), using regression tree (specifically the Cubist R package (Quinlan 2005; Kuhn *et al.* 2012, 2013). The regression tree method is a popular modelling approach for many disciplines (Breiman *et al.* 1984), and has been widely used with DSM (McKenzie and Ryan 1999; Grunwald 2009; Kidd *et al.* 2014*a*). The Cubist package develops the regression trees by first applying a data-mining approach to partition the calibration and explanatory covariate values into a set of structured 'classifier' data. The tree structure is developed by repeatedly partitioning the data into linear models until no significant measure of difference in the calibration data is determined (McBratney *et al.* 2003). A series of covariate-based rules (conditions) is developed, and the linear model corresponding to the covariate conditions is applied to produce the final modelled surface. For this modelling exercise, the model controls were set to allow the Cubist algorithm to determine the optimum number of rules to generate.

A perceived benefit of the regression tree (Cubist) approach is that there is no need to select the most important covariates before modelling (e.g. by stepwise linear regression). This is because only those covariates that have some covariance with the target variable are chosen by the Cubist data mining, with non-correlated covariates excluded from the regression tree conditions and linear models within the partitions. This is a useful time-saving measure when predicting multiple soil attributes from the same covariates. Similarly, principal component analysis (PCA), often used to de-correlate covariates in some modelling approaches (Hengl *et al.* 2007), was not deemed necessary, due to the Cubist data-mining capabilities. Use of PCA of covariates would also diminish the regression-tree model interpretability; that is, end-users are able to observe how each covariate is used in the models. Testing has also indicated little need to 'normalise' or transform target data to normal distribution with the Cubist methodology, making little difference to outputs and diagnostics, again mainly due to the powerful data mining capabilities.

Uncertainty

Leave-one-out cross-validation (LOOCV) was applied to the Cubist model to generate rule-based uncertainties, using only those covariates forming the conditional partitioning of each rule, following Malone *et al.* (2014). LOOCV can be beneficial for smaller datasets (Kohavi 1995), and therefore useful within this DSM exercise, because some regression-tree rule-based conditions might not contain sufficient data points for use with alternative cross-validation approaches (such as random holdback). The LOOCV, applied to an individual Cubist model

for each rule, effectively produced a mean value for each regression-tree partition, with the upper and lower 5% and 95% quantiles of the prediction variation providing the lower and upper prediction uncertainty values, respectively, at the 90% prediction interval (PI). An example regression-tree rule is shown below (Rule 1, for clay percentage, 30–60 cm), with 'n' data points meeting the Rule 1 condition.

If Th ≤ 3.69, and DEM ≤ 198, and MrRTF ≤ 4.85, then:

$$\begin{aligned}\text{Clay}_{n-1} = {} & (-0.19 \times \text{TC}) + (-2.1 \times \text{Kpc}) + (-0.7 \times \text{MrRTF}) \\ & + (-0.13 \times \text{MrVBF}) + (-0.385 \times \text{PG}) + 0.26 \\ & \times \text{Slope} + (-28 \times \text{TCI\_Low}) + (-0.26 \times \text{TRI}) \\ & + 0.23 \times \text{TWI} + 1.44 \times \text{Th} + (-0.7 \times \text{Uppm}) + 57.43\end{aligned}$$

where Clay is clay (%), TC is total radiometric count, Kpc is radiometric K (%), MrRTF is multi-resolution ridge-top flatness and MrVBF is multi-resolution valley-bottom flatness (Gallant and Dowling 2003), PG is persistent-greenness, Slope is slope (%), TCI_Low is topographic classification index (lowlands), TRI is terrain ruggedness index, TWI is topographic wetness index, Th is radiometric Th (ppm) and Uppm is radiometric U (ppm), and each data-point held back is sequentially applied for validation of each loop.

Initially, a random hold-back of 30% of the training data was used for validation; however, re-running the models with different random hold-backs produced variations in predictions, uncertainties and modelling diagnostics, implying model sensitivity to the data variance. To reduce this potential modelling bias, a k-fold cross-validation approach was implemented (Rodriguez *et al.* 2010), where one-tenth of the data was randomly held back, and the modelling looped 10 times using a different tithe of the data held back for validation of each iteration. The k-fold cross-validation approach has been widely used in DSM when available training data are limited or no independent validation data are resourced (Grimm *et al.* 2008; Hengl *et al.* 2014; Martin *et al.* 2011). Each data point is held back only once, meaning that every item of the training data is tested. The final prediction and upper and lower values for each surface cell are then produced. This is done by taking the mean from each of the ten k-fold model outputs, as well as the mean validation diagnostics, determining R^2, root-mean-square error (RMSE), bias and concordance (Lin 1989), and the percentage of validation values within 5% and 95% PI (i.e. the 'prediction interval coverage probability', expected to be at 90% where modelling uncertainty is optimal; Malone *et al.* 2014). This approach effectively reduces bias and tests modelling variance, with studies showing that 10-fold cross-validation is the optimum number of k-folds to test adequately all parts of the training data and model sensitivity to the full training-data range (Kohavi 1995). It is anticipated that generating the rule-based estimates of uncertainty within each regression-tree partition, then averaging by k-fold cross-validation to reduce modelling bias, will produce a better understanding of which landscapes have better predictions of soil property variability than relying on an average k-fold cross-validation uncertainty estimate across all regression tree partitions and covariates.

Three 80-m resolution raster surfaces of mean prediction with mean upper and lower predictions were generated for each soil property at the 90% PI, for each depth. Diagnostics for each model k-fold were recorded and averaged, as well as the individual regression-tree models, documenting variable usage, rule-sets, and linear model coefficients.

Continuous and categorical data

The regression-tree modelling was used for continuous datasets and soil properties, such as clay and sand percentages, pH, organic carbon percentage, and EC (1 : 5 soil–water suspension; Rayment and Lyons 2011). The method was also used for qualitative description data, such as coarse fragment (CF) (>2 mm) class estimates and soil drainage class, as per Kidd *et al.* (2014*a*), where the ordinal categorical classes were treated as a continuous data. Where the CF classes (National Committee on Soil and Terrain 2009) correspond to stone percentage ranges (Table 2), the final raster surfaces were stretched between each class range to correspond to the percentage range. For example, Class 2, corresponding to a continuous modelled range 1.5–2.5, was stretched between these values to a range of 2–10%, using the R Raster Package (Hijmans and van Etten 2012) (Table 2). For CF, this approach produced better modelling diagnostics and mapping outputs than modelling median CF percentage values as the target variable, or using decision trees DT class modelling.

Regression kriging

To reduce the unexplained spatial variability of the DSM modelling, regression kriging (RK) was tested to model residual spatial autocorrelation. RK is effectively a hybridised modelling approach that incorporates regression modelling with the interpolated model residuals, which has been shown to improve model performance in DSM (Odeh *et al.* 1995; McKenzie and Ryan 1999; Hengl *et al.* 2004, 2007). For this study, residual model estimates from the regression-tree procedures underwent simple kriging and the output was incorporated into the final surfaces. However, testing the spatial semi-variance of the regression-tree output residuals for many soil properties did not show strong spatial autocorrelation. Various modelling types and sill and nugget ranges applied to the semi-variogram settings did not produce good semi-variogram fits. The RK approach also drastically increased model processing time, needing to krige the entire state individually for >10 000 000 cells for each soil property and depth, in addition to the time taken to fit each variogram model manually. Because of the increase in modelling time, offset against the marginal improvements in testing surface validations, it was decided to desist with RK for the Version 1.0 surfaces.

Table 2. Coarse-fragment (CF) class index with percentage stretch

CF class	CF per cent range	Continuous index raster range	New 'stretched value'
0	0	0	0
1	<2	0–1.5	0–2
2	2–10	1.5–2.5	2–10
3	10–20	2.5–3.5	10–20
4	20–50	3.5–4.5	20–50
5	50–90	4.5–5.5	50–90
6	>90	5.5–6	90–100

Pedotransfer functions

Pedotransfer functions (PTFs) are correlation relationships developed to predict a soil property from other existing soil property datasets (McBratney *et al*. 2002), and were used where there was insufficient training data for certain soil attributes. The PTFs were applied to predicted surface values (and upper and lower predictions), rather than applying the PTFs to the individual points as modelling target variables. This approach was favoured, mainly to reduce DSM modelling errors due to the incorporation of the PTFs unexplained soil attribute variability into the RT process; and because many sites did not necessarily have all required soil property PTF inputs, which would ultimately reduce the number of training points available for the RT DSM modelling.

Electrical conductivity of saturated paste

Very few available sites have data for the required soil property EC_{se} (EC of a saturated paste, 1 : 1 soil–water); hence, this was generated by applying the PTF from Peverill *et al*. (1999) (Eqn 1):

$$EC_{se} = EC_{1:5} \times (500 + 6 \times 0.59 + 0.016 \times (clay\%^{1.5}))/ (30.34 + 6.57 \times 0.59 + 0.016 \times (clay\%^{1.5})) \quad (1)$$

where $EC_{1:5}$ is EC in a 1 : 5 soil–water suspension (Rayment and Lyons 2011), and clay% corresponds to the predicted clay values for each cell.

Bulk density

There was also very few available data points with any bulk density (BD) values. A PTF calibrated using Australian data from Tranter *et al*. (2007) was used, which incorporates the predicted sand and organic carbon percentages for each cell value (Eqns 2 and 3). First, a mineral density was predicted as a function of sand and depth:

$$BD_{min} = 0.842 + 0.097 \times \log(depth) + 0.0057 \times sand + (sand - 44.72)^2 \times (-0.0000845) \quad (2)$$

where BD_{min} is BD of the mineral soil fraction (g cm^{-3}), depth is mid-depth of layer (cm), and sand is sand percentage. The final BD estimate is determined by incorporating the effect of soil organic matter through Eqn 3 (Adams 1973):

$$BD = 100/(OM/0.223 + (100 - OM/BD_{min}) \quad (3)$$

where BD is final BD estimate, and OM is organic matter content, estimated from:

$$OM = 1.72 \times OC \quad (4)$$

where OC is predicted organic carbon percentage. This does not take into account any land-management influences on BD (such as compaction), but is considered a reasonable approximation of the most likely state, as influenced by the mineral, overburden, and organic matter (Tranter *et al*. 2007).

Silt content

Silt percentage was initially modelled for all standard depths using the DSM regression tree approach, and compared against calculating the predicted silt percentage value for each raster cell by subtracting clay and sand percentages from 100 (Eqn 5):

$$Silt\% = 100 - (sand\% + clay\%) \quad (5)$$

It was decided to use the calculated silt percentage surface from Eqn 5 as the final Version 1.0 products, because the sand and clay modelling diagnostics were generally superior to the silt modelling, and would also remove the potential problem whereby the combined predicted particle-size products were >100%.

pH

Available pH measurements were used as a 1 : 5 soil–water suspension (Rayment and Lyons 2011), with insufficient data using the $CaCl_2$ suspension to form state-wide models based on these measurements. The pH in $CaCl_2$ can also be predicted from the pH in water surfaces by using PTFs, such as from Henderson and Bui (2002) and Minasny *et al*. (2011), which incorporate information on soil EC.

Effective soil depth and depth to rock

Effective soil depth (or plant-exploitable depth) (Arrouays *et al*. 2014) was considered as the depth of soil-database descriptive sites to the upper value of any layer that corresponded to a C horizon (weathered substrate), rock, or hard pan (National Committee on Soil and Terrain 2009). The values were used as continuous target variables (in cm) within the standard regression-tree approach. Depth to rock was modelled as above, using depth to any horizon with an 'R' (rock) designation.

Expert validation and data release

All surfaces were assessed within DPIPWE by departmental soil scientists to determine whether there was general agreement with historical mapping and state-wide soil–landscape knowledge. Figure 3 shows an example map (Burnie Map Sheet, Spanswick and Kidd 2000) with polygons generally aligning with surface sand percentage. The surfaces are publically available on the TERN web portal (www.clw.csiro.au/aclep/soilandlandscapegrid/index.html), where they can be further appraised by relevant soil–landscape experts around the country. Table 3 summarises the produced DSM surfaces and methodology for predictions.

Results

The DSM outputs and modelling diagnostics are presented here as individual soil attributes, with brief surface and subsoil comments.

Clay content

Clay percentage surfaces were generated using site data with particle size analyses (PSA) values for each horizon. In total, 1288 sites were available with clay percentage PSA, with values generated by the depth-spline interpolations for most horizons. The averaged k-fold modelling diagnostics are shown in Table 4.

For surface layers (0–5 cm), modelling diagnostics were fair, with concordance values of 0.51 and 0.36 and RMSE 10.6% and 12.1% for calibration and validation, respectively.

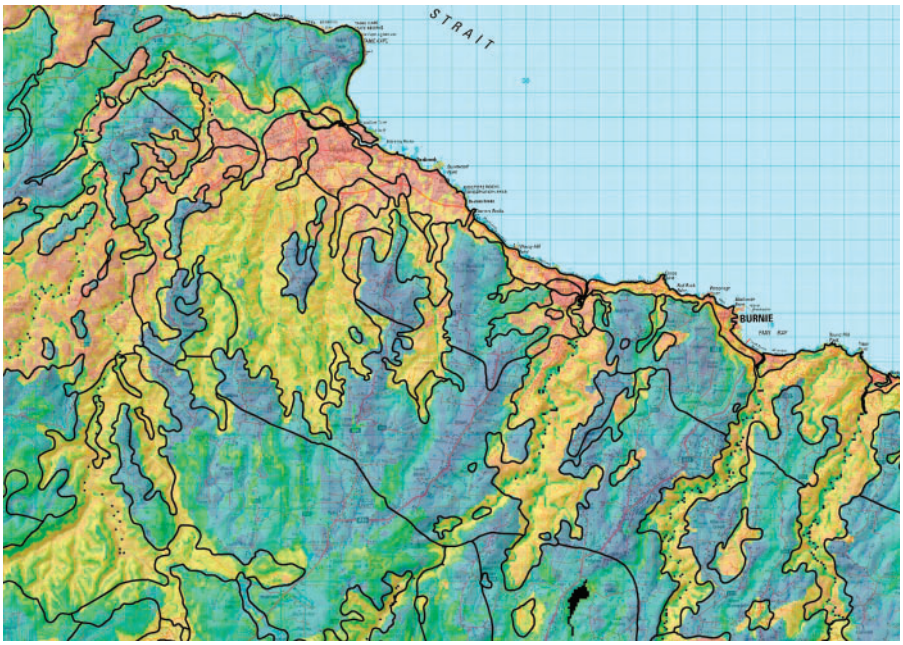

Fig. 3. Variations in surface sand percentage, and correlation with existing mapping (Burnie).

Table 3. Summary of digital soil-mapping surfaces

RT, Regression tree; PTF, pedotransfer function. Standard depths (cm): 0–5, 5–15, 15–30, 30–60, 60–100, 100–200. Uncertainties are to the 90% prediction interval (5th and 95th per cent quantile)

Soil property	No. of depths (cm)	Value	Method	No. of surfaces
pH	7 (standard + 0–15)	pH units, predicted, lower, upper	RT	21
EC	7 (standard + 0–15)	dS m^{-1}, predicted, lower, upper	RT	21
EC_{se}	7 (standard + 0–15)	dS m^{-1}, predicted, lower, upper	PTF	21
Sand	6 (standard)	%, predicted, lower, upper	RT	18
Clay	7 (standard + 0–15)	%, predicted, lower, upper	RT	21
Silt	7 (standard + 0–15)	%, predicted, lower, upper	PTF	18
Organic carbon (OC)	7 (standard + 0–15)	%, predicted, lower, upper	RT	21
Coarse fraction (CF)	7 (standard + 0–15)	%, >2 mm, 2–200 mm, >60 mm, >200 mm, predicted, lower, upper	RT	30
Effective depth	1 (depth to)	cm, predicted, lower, upper	RT	3
Available water content (AWC)	7 (standard + total profile)	m^{-3} m^{-3}, predicted, lower, upper	PTF	11
Bulk density (BD)	6 (standard)	Mg m^{-3}, predicted, lower, upper	PTF	18
ExCa	1 (0–15)	cmol kg^{-1}, predicted, lower, upper	RT	3
ExMg	1 (0–15)	cmol kg^{-1}, predicted, lower, upper	RT	3
Drainage	1 (total profile)	Class, predicted, lower, upper	RT	3
Depth to sodic layer	1 (depth to)	cm, predicted, lower, upper	RT	3
Depth to duplex clay	1 (depth to)	cm, predicted, lower, upper	RT	3
Total				218

However, validation diagnostics were better for subsoil predictions (60–100 cm), with 0.28 and 17.0% for concordance and RMSE, respectively. Validation values were generally at or near expected prediction interval ranges (at the 90% confidence limit (CL)), with 89% validating within these limits for both example depths (or within 90% when accounting for standard deviations). The validation RMSE standard deviations were 1.4% and 1.5%, respectively, for these surface and subsoil depths (~12% of the mean value), implying that a broad range of training and validation values has marginal effect on the k-fold model variations and diagnostic outputs.

Figure 4 shows surface (0–5 cm) clay percentages for the state, which generally agrees with known regional soil–landscape relationships, for example, low clay in sandy coastal areas, and higher surface clay percentages in the clay-loam topsoils of the north-west Ferrosols (Isbell 2002). From the k-fold diagnostics, many of the terrain derivatives, including elevation (DEM), altitude above channel network (AACN), valley depth, multi-resolution valley bottom flatness (MrVBF),

Table 4. Clay percentage modelling diagnostics (averaged k-folds)
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

Depth (cm)		Calibration RMSE	Calibration R^2	Calibration Bias	Calibration CC	Validation RMSE	Validation R^2	Validation Bias	Validation CC	% Within 90% CL
0–5	Mean	10.6	0.36	–0.95	0.51	12.1	0.19	–1.08	0.36	88.7
	s.d.	0.3	0.04	0.14	0.04	1.4	0.09	1.07	0.07	3.0
0–15	Mean	11.2	0.32	–1.18	0.48	12.4	0.18	–1.29	0.35	88.6
	s.d.	0.5	0.06	0.25	0.07	1.2	0.08	1.59	0.08	3.3
5–15	Mean	11.7	0.31	–1.34	0.46	13.0	0.16	–1.42	0.33	89.4
	s.d.	0.3	0.04	0.13	0.05	0.9	0.07	1.28	0.07	2.5
15–30	Mean	14.9	0.31	–1.56	0.46	16.4	0.18	–1.71	0.34	88.7
	s.d.	0.3	0.03	0.18	0.03	1.0	0.06	1.79	0.06	2.8
30–60	Mean	16.0	0.25	–0.23	0.40	17.6	0.13	–0.48	0.28	89.1
	s.d.	0.4	0.04	0.28	0.05	1.4	0.07	1.45	0.08	3.9
60–100	Mean	15.7	0.26	–0.19	0.40	17.0	0.14	–0.09	0.28	89.4
	s.d.	0.4	0.04	0.22	0.06	1.5	0.09	1.89	0.11	2.4
100–200	Mean	14.4	0.29	–0.57	0.45	16.1	0.14	–0.59	0.30	88.6
	s.d.	0.4	0.03	0.29	0.05	1.5	0.08	2.31	0.08	3.6

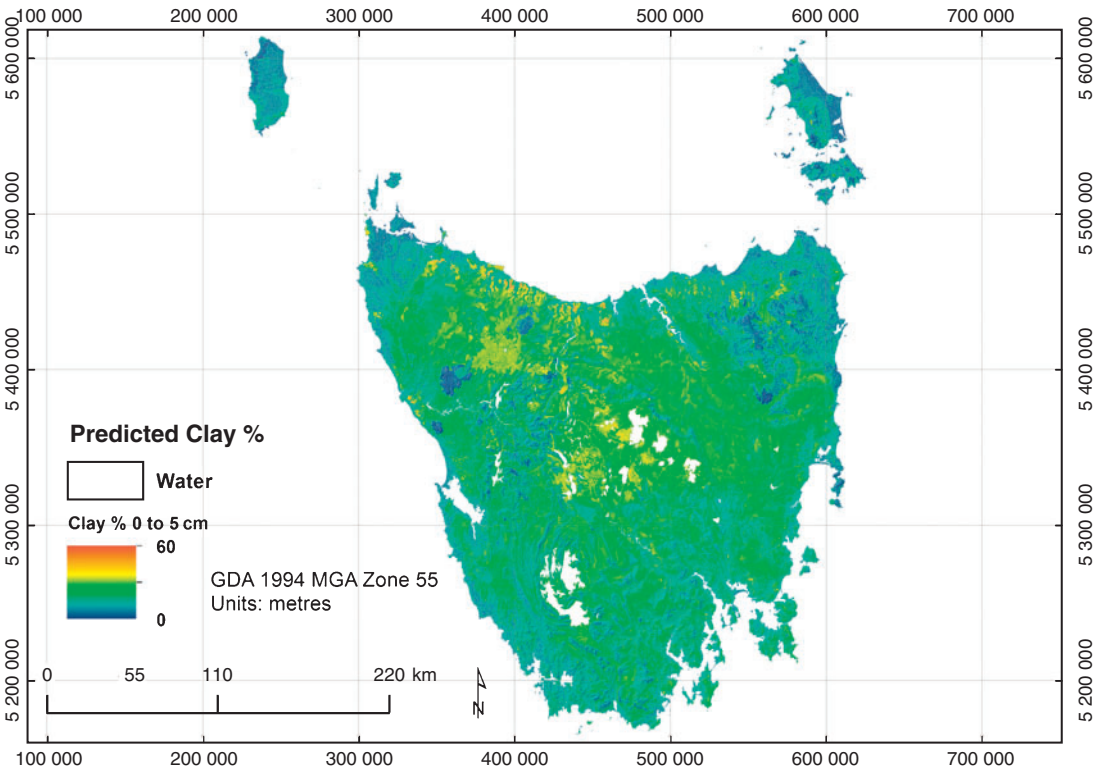

Fig. 4. Surface (0–5 cm) clay percentage.

and northness are important predictors of surface-soil clay percentage. The integrated radiometrics-geological layers are also important explanatory variables, especially K and Th. This is demonstrated, for example, by the seventh k-fold model variable usage, with similar usage statistics in other iterations and depths (Fig. 5). Rainfall was initially found to be an important predictor, but was removed from the clay modelling because of the introduction of unrealistic mapping artefacts within the prediction surfaces for most depths (see *Discussion*).

Sand content

There were 461 sites available with PSA for sand percentage. Modelling of sand percentage produced slightly better calibration–validation diagnostics than clay percentage in terms of concordance. For example, surface sand percentage (0–5 cm) had values of 0.71 and 0.54 for calibration and validation, respectively; however, RMSE was slightly higher, with 17.3% and 21.1% for calibration and validation. This implies that the modelled data fitted better around the observed *v.* predicted 1 : 1

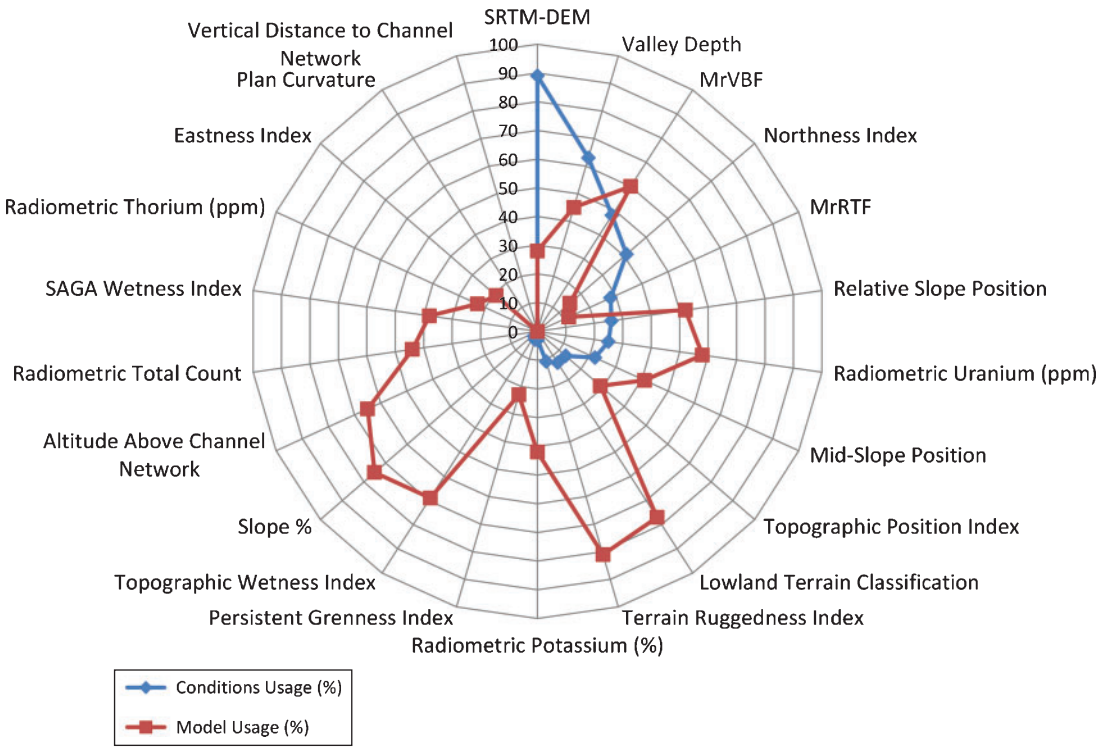

Fig. 5. Example covariate usage, clay percentage.

Table 5. Sand percentage modelling diagnostics (averaged k-fold)
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

Depth (cm)		Calibration				Validation				% Within
		RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	90% CL
0–5	Mean	17.3	0.55	1.21	0.71	21.1	0.34	1.89	0.54	85.0
	s.d.	1.3	0.07	0.81	0.05	2.3	0.13	2.58	0.11	7.4
5–15	Mean	17.8	0.53	0.96	0.69	22.3	0.29	0.46	0.50	85.0
	s.d.	1.4	0.07	0.79	0.06	1.8	0.13	3.11	0.12	4.5
15–30	Mean	20.2	0.47	1.34	0.64	23.8	0.28	1.08	0.48	86.0
	s.d.	1.4	0.07	0.93	0.06	3.2	0.12	2.54	0.11	7.2
30–60	Mean	20.4	0.47	–1.03	0.64	24.4	0.25	–2.10	0.45	88.5
	s.d.	1.3	0.07	0.95	0.07	2.6	0.14	5.07	0.13	3.5
60–100	Mean	21.6	0.40	–1.41	0.57	24.6	0.23	–1.56	0.42	89.6
	s.d.	1.0	0.05	1.08	0.06	2.2	0.10	4.28	0.09	5.3
100–200	Mean	22.4	0.37	0.47	0.54	27.9	0.08	–0.15	0.26	85.9
	s.d.	1.9	0.10	1.26	0.10	4.3	0.09	5.62	0.11	8.1

line of fit (Lin 1989) but were more dispersed around this line, resulting in higher RMSE values (Table 5). Sand percentage diagnostics were generally similar with all depths.

As expected, the sand percentage is inverse in appearance to the clay percentage mapping, being relatively high in coastal zones, and low in areas of expected high-clay soils, as per the clay percentage mapping examples (Fig. 6). Some under-prediction of sand percentage might be evident in beach areas where close to 100% is expected, mainly due to the lack of available coastal sites with PSA.

In terms of covariate usage, the DEM and several derivatives were important explanatory variables, as well as radiometric K. Model performance in terms of validation values within the upper and lower PI were slightly worse than clay percentage, ranging from 85.0% to 89.6% (90% CL), but were all within the 90% range if taking standard deviation into account. A standard deviation of 7.4% for validation within the 90% CL implies that moderate modelling sensitivity to the calibration data, due in part to the smaller sample size, and potential data outliers.

Silt content

Silt percentages for all depths was calculated from the clay and sand percentage surfaces, and is therefore reliant on the modelling diagnostics of those surfaces.

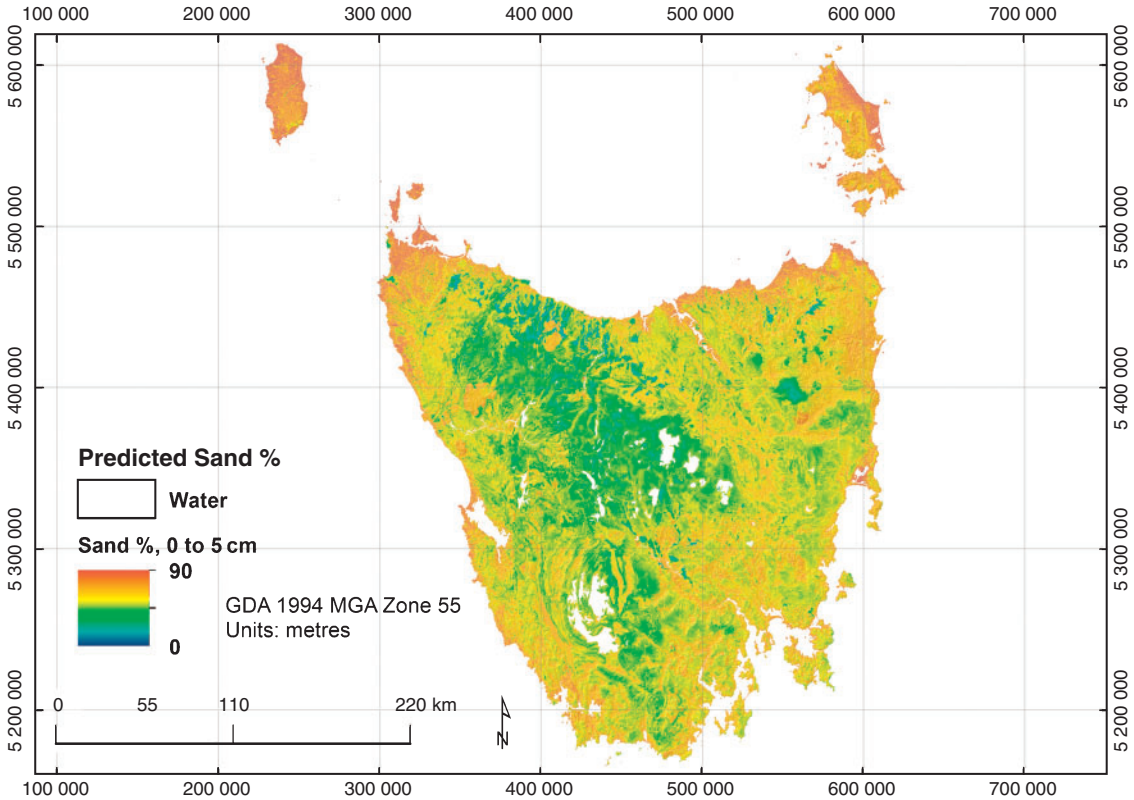

Fig. 6. Surface (0–5 cm) sand percentage.

Table 6. pH modelling diagnostics (averaged k-fold)
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

Depth (cm)		Calibration				Validation				% Within 90% CL
		RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	
0–5	Mean	0.6	0.19	–0.04	0.30	0.7	0.05	–0.03	0.16	88.1
	s.d.	0.0	0.04	0.01	0.06	0.1	0.04	0.05	0.06	2.6
0–15	Mean	0.6	0.22	–0.05	0.33	0.6	0.09	–0.05	0.22	88.9
	s.d.	0.0	0.08	0.01	0.09	0.1	0.04	0.04	0.07	3.8
5–15	Mean	0.6	0.18	–0.05	0.30	0.7	0.08	–0.06	0.21	89.8
	s.d.	0.0	0.02	0.01	0.02	0.0	0.03	0.05	0.03	2.5
15–30	Mean	0.6	0.42	–0.02	0.59	0.7	0.23	–0.02	0.43	88.9
	s.d.	0.0	0.03	0.01	0.03	0.1	0.10	0.05	0.09	2.6
30–60	Mean	0.7	0.55	–0.01	0.71	0.8	0.42	0.00	0.61	90.0
	s.d.	0.0	0.04	0.01	0.03	0.1	0.09	0.09	0.07	2.0
60–100	Mean	0.8	0.60	0.00	0.75	1.0	0.45	0.01	0.65	88.8
	s.d.	0.0	0.04	0.02	0.03	0.1	0.06	0.09	0.05	3.7
100–200	Mean	0.9	0.60	–0.03	0.75	1.1	0.41	–0.05	0.61	87.2
	s.d.	0.0	0.04	0.03	0.03	0.1	0.07	0.14	0.05	3.4

pH

There were 1440 sites with laboratory pH available (Rayment and Lyons 2011) for all or some horizons. Surface-modelling diagnostics were generally poor; for example, the 0–5 cm surface had a concordance of 0.30 and 0.16, and RMSE of 0.6 and 0.7, for calibration and validation, respectively. However, modelling diagnostics generally improved with depth in terms of concordance, with calibration–validation values of 0.75 and 0.65 at a depth of 60–100 cm, (Table 6). The models generally validated within the 90% CL, most ~89%.

Visually, there is a prominent west–east trend in pH, with lower values (more acidic) in the high-rainfall western areas, and higher values (more neutral to alkaline) in lower rainfall areas (in the central Midlands rain-shadow). This is reflected in the covariate model usage for all k-folds, with rainfall being one of the most important variables in terms of conditions and model

usage. High pH values were also evident around some coastal areas, due to seashell-fragment deposition. Figure 7 shows subsoil pH (60–100 cm).

Electrical conductivity

There were 3522 sites available with EC of a 1 : 5 soil–water suspension (Rayment and Lyons 2011). Surface-modelling diagnostics (0–5 cm) were very poor, with calibration and validation concordance both 0.02, and RMSE of 0.30 dS m^{-1}. Subsoil modelling (60–100 cm) was an improvement, with a concordance of 0.64 and 0.47 for calibration and validation, and RMSE of 0.30 and 0.29 dS m^{-1} respectively. The subsoil EC values were higher than surface values; hence, the RMSE were not as large in relative terms. Most surfaces validated at or near the required 90% CL (Table 7).

Visually, there was relatively little variation in surface EC across the state, with small, localised areas of higher EC showing

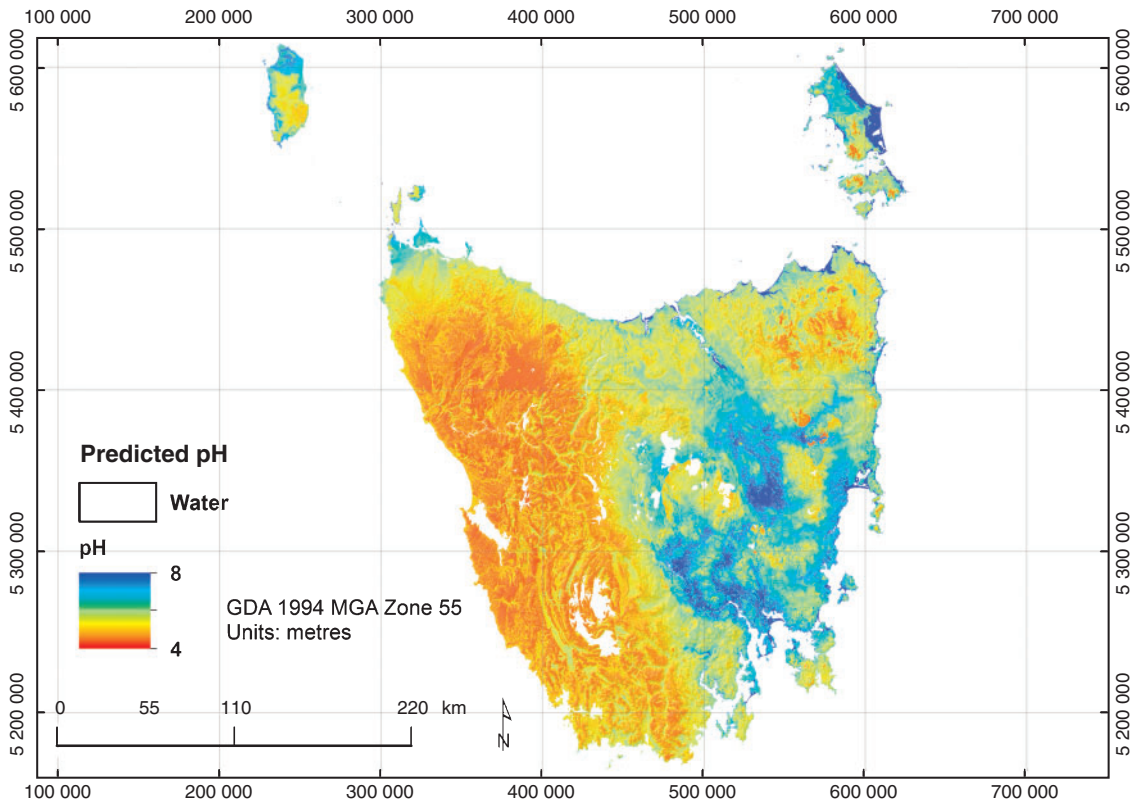

Fig. 7. Subsoil (60–100 cm) pH.

Table 7. Electrical conductivity (dS m^{-1}) modelling diagnostics (averaged k-folds)
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

Depth (cm)		Calibration				Validation				% Within 90% CL
		RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	
0–5	Mean	0.3	0.01	–0.05	0.02	0.3	0.01	–0.05	0.02	89.9
	s.d.	0.0	0.00	0.00	0.01	0.1	0.01	0.02	0.01	2.0
0–15	Mean	0.3	0.13	–0.04	0.11	0.3	0.06	–0.05	0.02	90.7
	s.d.	0.0	0.15	0.00	0.11	0.1	0.15	0.02	0.02	2.3
5–15	Mean	0.3	0.12	–0.04	0.15	0.3	0.04	–0.04	0.06	89.7
	s.d.	0.0	0.11	0.00	0.14	0.1	0.04	0.01	0.06	1.8
15–30	Mean	0.2	0.25	–0.04	0.33	0.3	0.08	–0.03	0.18	89.6
	s.d.	0.0	0.09	0.00	0.11	0.1	0.06	0.02	0.09	2.2
30–60	Mean	0.3	0.43	–0.04	0.53	0.3	0.17	–0.04	0.31	89.0
	s.d.	0.0	0.09	0.00	0.09	0.1	0.09	0.02	0.11	1.9
60–100	Mean	0.3	0.50	–0.04	0.64	0.3	0.29	–0.04	0.47	89.1
	s.d.	0.0	0.03	0.00	0.03	0.1	0.12	0.02	0.13	2.1
100–200	Mean	0.3	0.67	–0.04	0.79	0.4	0.31	–0.04	0.52	88.5
	s.d.	0.0	0.04	0.01	0.03	0.1	0.15	0.04	0.12	3.0

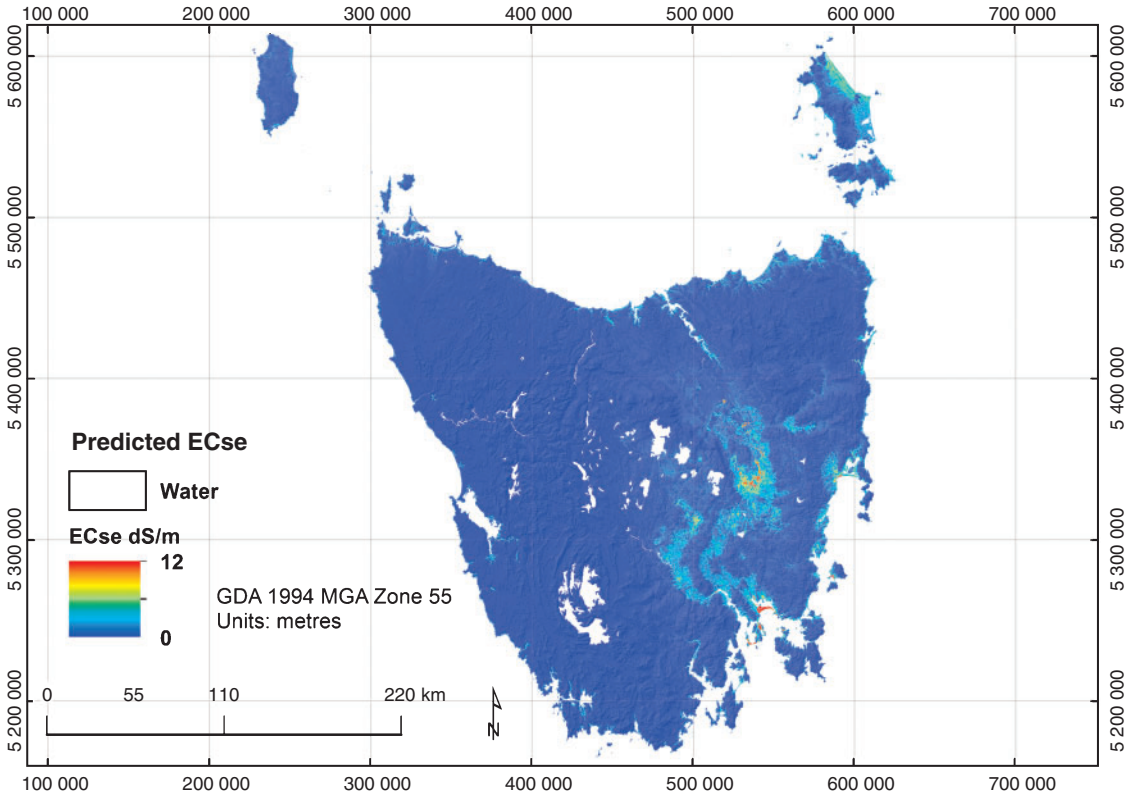

Fig. 8. Subsurface electrical conductivity of a saturated extract (EC_{se}, 60–100 cm).

surface-expression in evaporation basins and break-of-slope areas, concentrated in the low-rainfall areas of the central Midlands, as expected (Kidd 2003). Some coastal areas were also realistically highlighted as higher EC, and therefore saline zones. Subsoil EC was generally higher, also highlighting the well-known, central Midlands primary salinity-prone areas and naturally occurring saltpans. In terms of covariate usage, most k-fold iterations showed that elevation, moisture-simulation terrain derivatives such as topographic wetness index (TWI), and gamma-radiometric K, were important predictors, along with mean annual rainfall.

Electrical conductivity (saturated extract)

As per the PTF methodology, EC_{se} for all depths was calculated by using the clay and EC outputs, and it is therefore reliant on the modelling diagnostics of those surfaces. Mapping showed environmentally realistic patterns similar to the EC layers. Figure 8 highlights the high-level subsoil salinity evident in the low-rainfall central Midlands.

Soil organic carbon content

There were 1623 available sites with soil organic carbon percentage (OC) data. These surfaces modelled very well in terms of calibration and validation diagnostics, with surface (0–5 cm) concordance values of 0.88 and 0.72, respectively. RMSE values were 3.5% and 5.0%. Subsoil (60–100 cm) values for calibration and validation were poor, with concordances of 0.15 and 0.05, and RMSE values of 1.4% and 1.2% (Table 8).

In terms of mapping, OC values were dominated by the Southwest WHA, which, according to Cotching *et al*. (2009), is known to contain very high carbon levels in well-formed peat soils (Organosols, Isbell 2002). Maximum modelled values were up to 70% OC in these peats (Fig. 9); however, very few sites were available within these remote areas. This is a very high value for the organic carbon component, which implies that modelling could be slightly over-predicting in these areas. The most important covariates in most k-folds were rainfall and terrain-related products. Most depths validated within the 90% CL with respect to the standard deviation around the averaged k-fold validation percentages. Future work needs to identify and map out the peat areas separately.

Coarse fragments content

There were 3469 sites available with CF class estimates (>2 mm), which were modelled as continuous data. Modelling diagnostics were moderate, producing surface (0–5 cm) calibration and validation diagnostics for concordance of 0.49 and 0.26, respectively, and RMSE of 1.2% and 1.4%. Subsoil (60–100 cm) diagnostics were slightly poorer, with RMSE of calibration and validation of 1.5% and 1.6% (Table 9).

Visually, surface maps (once class estimates were stretched to corresponding percentage values) showed much higher stone content in the central highlands and mountainous areas, most consisting of weathering-resistant Jurassic Dolerite (Fig. 10). The more important explanatory variables were again radiometrics, elevation and terrain.

Table 8. Organic carbon percentage modelling diagnostics (averaged k-folds)
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

Depth (cm)		Calibration				Validation				% Within 90% CL
		RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	
0–5	Mean	3.5	0.88	–0.33	0.93	5.0	0.72	–0.36	0.83	89.6
	s.d.	0.3	0.02	0.04	0.01	1.8	0.17	0.38	0.10	1.9
0–15	Mean	3.1	0.90	–0.29	0.95	4.4	0.78	–0.37	0.87	89.2
	s.d.	0.3	0.02	0.03	0.01	1.7	0.13	0.30	0.08	2.6
5–15	Mean	3.3	0.89	–0.29	0.94	5.3	0.66	–0.24	0.78	89.1
	s.d.	0.5	0.03	0.06	0.02	2.4	0.25	0.53	0.18	3.5
15–30	Mean	3.0	0.91	–0.23	0.95	4.4	0.75	–0.19	0.84	88.6
	s.d.	0.3	0.02	0.03	0.01	2.3	0.23	0.37	0.15	2.2
30–60	Mean	1.3	0.41	–0.17	0.51	1.4	0.25	–0.18	0.34	89.0
	s.d.	0.2	0.18	0.03	0.18	0.8	0.24	0.13	0.22	3.5
60–100	Mean	1.4	0.10	–0.16	0.15	1.2	0.02	–0.15	0.05	89.9
	s.d.	0.2	0.06	0.02	0.10	0.9	0.03	0.08	0.07	2.4
100–200	Mean	0.9	0.14	–0.10	0.15	0.8	0.09	–0.07	0.16	90.2
	s.d.	0.2	0.21	0.02	0.22	0.9	0.06	0.14	0.12	2.6

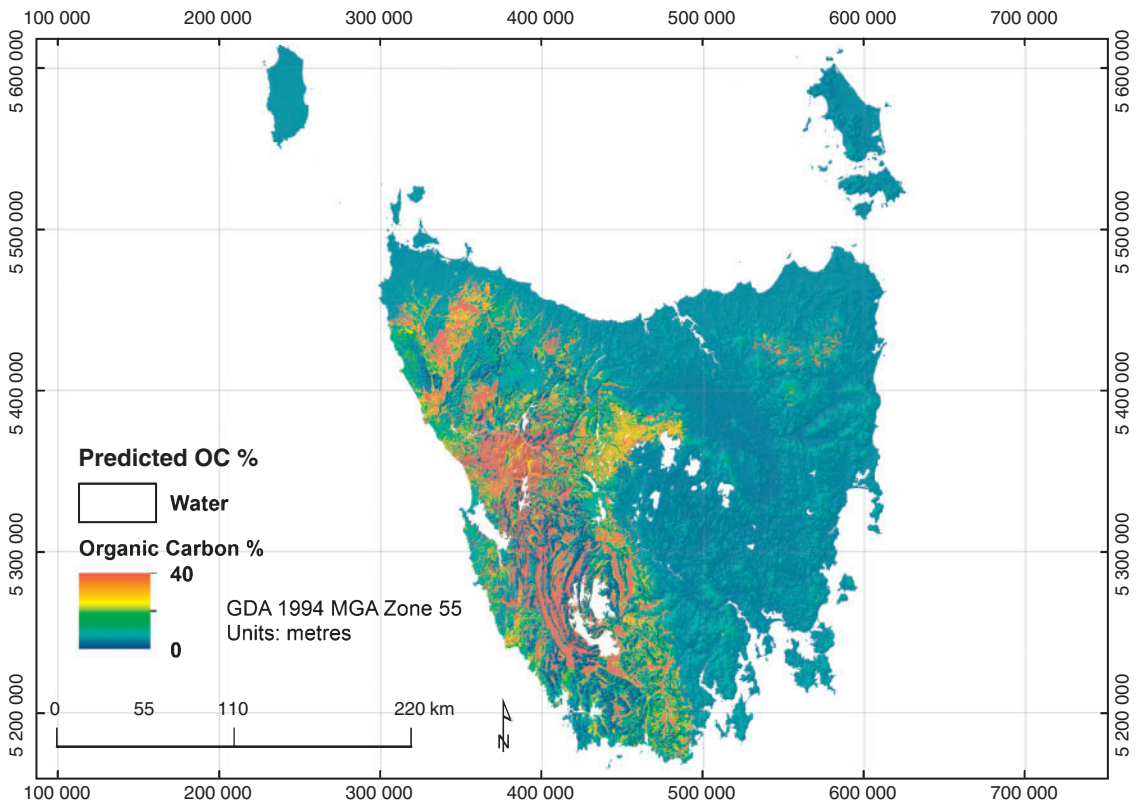

Fig. 9. Surface organic carbon percentage (0–5 cm).

Effective soil depth

There were 1149 database sites available with an effective soil depth estimation. Moderate modelling diagnostics were achieved, with concordances for calibration and validation of 0.45 and 0.30, and RMSE of 43 and 47 cm, respectively (Table 10). Most k-folds were within the 90% CL for validation.

Visually, mapping showed realistic terrain-related depth, with shallower soils on ridge-tops and mountain ranges, with the deepest soils showing as the northern Midlands part of the Launceston Tertiary Basin, consisting of deep Tertiary sediments (Fig. 11). Variable usage by the Cubist regression-tree approach was dominated by most terrain derivatives for all k-folds, most notably valley depth and TCI-Low.

Additional enterprise suitability surfaces

Additional surfaces were generated for the state-wide ESA: exchangeable calcium 0–15 cm (exCa), exchangeable

Table 9. Coarse fragment percentage diagnostics (averaged k-folds)
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation; the 100–200 cm layer not applicable to this parameter

Depth (cm)		Calibration				Validation				% Within
		RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	90% CL
0–5	Mean	1.2	0.31	–0.20	0.49	1.4	0.09	–0.21	0.26	88.1
	s.d.	0.0	0.05	0.03	0.06	0.1	0.05	0.07	0.08	1.9
0–15	Mean	1.2	0.31	–0.18	0.49	1.4	0.11	–0.17	0.30	88.3
	s.d.	0.0	0.04	0.03	0.04	0.1	0.03	0.12	0.05	1.9
5–15	Mean	1.2	0.32	–0.17	0.50	1.4	0.10	–0.15	0.28	87.5
	s.d.	0.1	0.06	0.04	0.07	0.0	0.03	0.11	0.04	2.3
15–30	Mean	1.3	0.28	–0.19	0.45	1.5	0.09	–0.19	0.26	88.7
	s.d.	0.0	0.02	0.02	0.03	0.1	0.03	0.11	0.05	1.7
30–60	Mean	1.4	0.22	–0.25	0.37	1.5	0.06	–0.24	0.19	89.3
	s.d.	0.0	0.04	0.05	0.06	0.1	0.03	0.10	0.05	2.8
60–100	Mean	1.5	0.15	–0.36	0.26	1.6	0.04	–0.36	0.12	89.1
	s.d.	0.0	0.05	0.04	0.07	0.1	0.03	0.16	0.07	3.4

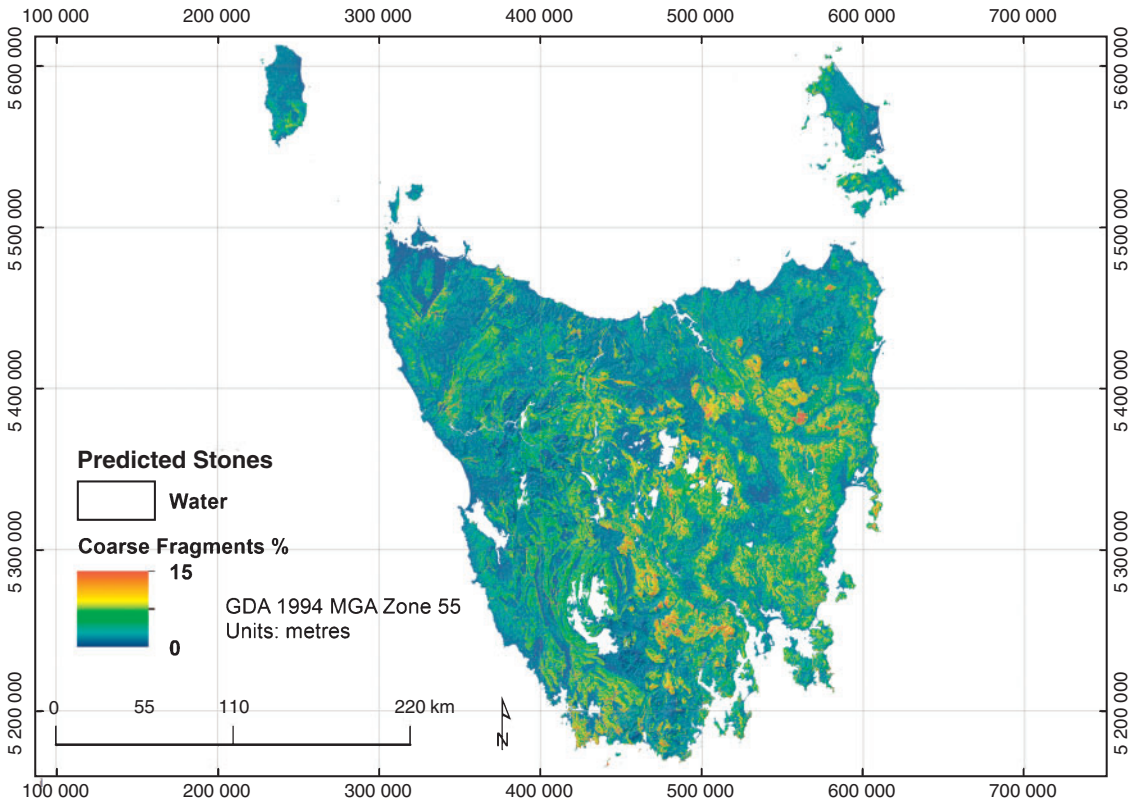

Fig. 10. Surface coarse fragments (0–5 cm).

magnesium 0–15 cm (exMg), and depth to sodic layer (exchangeable sodium percentage (ESP) >6%; Kidd *et al.* 2014*b*). Concordances for calibration and validation were 0.49 and 0.33 for exCa, 0.61 and 0.35 for depth to sodic layer, and slightly poorer at 0.28 and 0.17 for exMg (Table 11). An additional soil drainage index surface was modelled, as per Kidd *et al.* (2014*a*), based on the qualitative soil drainage expert-estimate at each site. Concordance was 0.48 and 0.38 for training and validation, and showed good agreement with expert knowledge of relative soil–landscape drainage patterns around the state.

Poorly predicted soil attributes

Depth to rock and ECEC (effective cation exchange capacity) modelled very poorly, with no correlation between the target variables and available covariates; hence, these surfaces were not released, and they will require future research to develop.

Table 10. Effective soil depth modelling diagnostics
RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

k-fold	Calibration				Validation				% Within
no.	RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	90% CL
K1	42.4	0.35	–7.00	0.48	40.5	0.17	–1.84	0.38	0.91
K2	46.8	0.20	–6.86	0.31	41.0	0.06	–3.29	0.18	0.94
K3	38.8	0.43	–5.59	0.58	58.2	0.02	–8.38	0.16	0.83
K4	44.9	0.23	–7.27	0.34	47.4	0.26	–9.61	0.39	0.86
K5	43.3	0.28	–6.35	0.42	45.4	0.33	–6.13	0.42	0.87
K6	42.6	0.35	–6.94	0.49	37.0	0.17	–0.49	0.37	0.90
K7	41.5	0.33	–5.83	0.49	54.9	0.10	–12.83	0.21	0.90
K8	39.1	0.39	–6.22	0.54	61.7	0.03	–8.37	0.12	0.91
K9	42.6	0.30	–5.52	0.45	50.7	0.13	–7.90	0.28	0.84
K10	45	0.26	–6.78	0.37	37.2	0.32	–5.11	0.48	0.93
Mean	42.7	0.31	–6.44	0.45	47.4	0.16	–6.39	0.30	0.89
s.d.	2.51	0.07	0.63	0.09	8.8	0.11	3.78	0.12	0.04

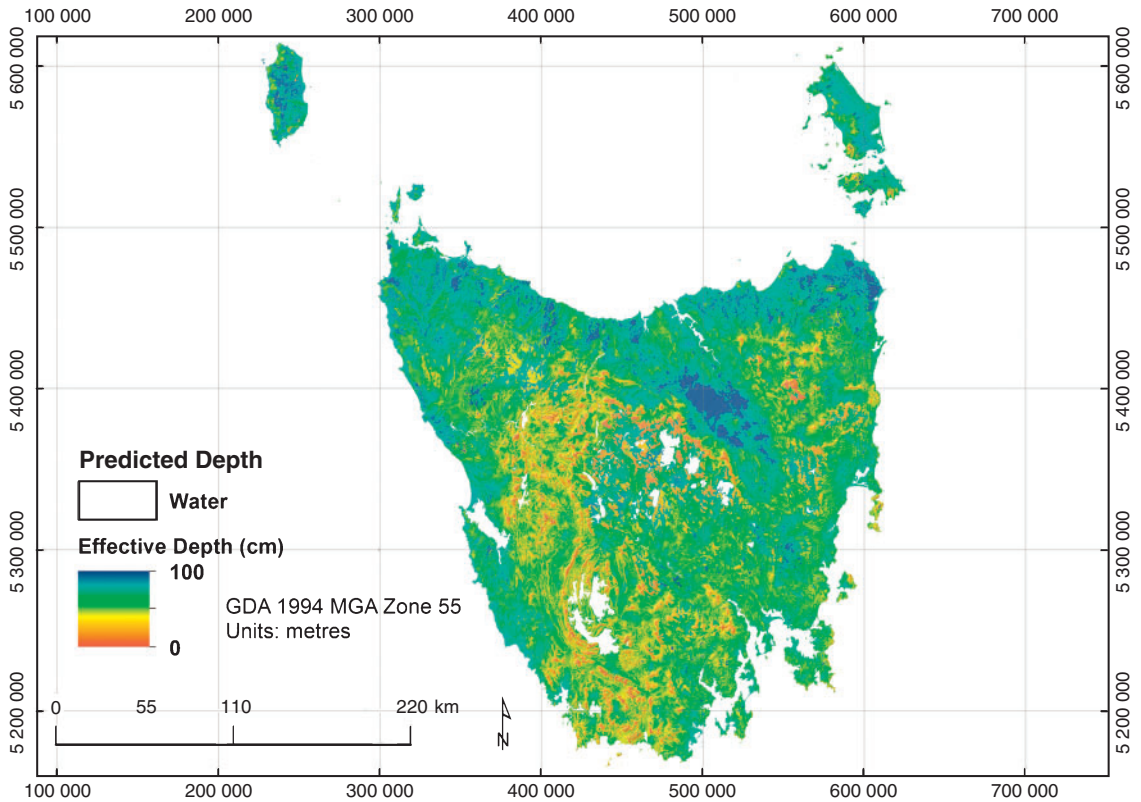

Fig. 11. Effective soil depth (cm).

These soil properties are not required for the current ESA rule-sets for Tasmania.

Discussion

The Version 1.0 Tasmanian soil-attribute maps were developed using a regression-tree modelling process that has produced reasonable diagnostics, and realistic mapping in terms of topographic variation and extent. The regression-tree rule-based LOOCV approach (Malone *et al.* 2014) has effectively taken into account the sensitivity of the linear modelling approach to the covariate-based conditions, using the variation in modelling due to the data variance to develop the upper and lower prediction limits, with 90% confidence. The k-fold cross-validation has also reduced any modelling bias by using different parts of the available target data both to calibrate and to validate the modelling, averaging the outputs to 'smooth-out' any extreme model output variations due to data 'outliers'.

The Version 1.0 products have been constructed with no initial attempt to test the environmental conditions (covariate feature space) that are represented by the existing soil attribute datasets, or to consider the uncertainties produced by the

Table 11. Modelling diagnostics for exchangeable calcium and exchangeable magnesium (cmol kg^{-1}), depth to sodic layer and drainage (averaged k-folds)

RMSE, Root-mean-square error; CC, concordance; CL, confidence limit; s.d., standard deviation

		Calibration				Validation				% Within
		RMSE	R^2	Bias	CC	RMSE	R^2	Bias	CC	90% CL
ExCa (0–15 cm)	Mean	6.3	0.32	–0.81	0.49	7.2	0.15	–0.65	0.33	86.7
	s.d.	0.4	0.06	0.15	0.06	1.9	0.09	0.54	0.12	4.4
ExMg (0–15 cm)	Mean	4.4	0.19	–1.16	0.28	4.7	0.08	–1.17	0.17	90.3
	s.d.	0.2	0.08	0.12	0.11	0.6	0.03	0.46	0.05	2.1
Depth to sodic layer (cm)	Mean	0.2	0.45	–0.03	0.61	0.3	0.15	–0.03	0.35	95.5
	s.d.	0.0	0.02	0.00	0.02	0.0	0.04	0.01	0.05	1.7
Drainage index (whole profile)	Mean	1.0	0.29	0.00	0.48	1.0	0.18	–0.01	0.38	89.3
	s.d.	0.0	0.03	0.01	0.03	0.0	0.02	0.05	0.02	1.5

temporal range of the training data. The effects of land use and management on some soil properties were also not considered because of lack of available data at the time of modelling, other than the use of the 'persistent greenness' satellite covariate, which effectively showed land-use patterns in some areas.

Temporal variability

The modelling uncertainty due to the temporal range of the training data was most apparent as poor modelling diagnostics and high uncertainty ranges for pH and EC in the top 30 cm of the output surfaces (0–5, 5–15 and 15–30 cm). The top 30 cm is generally more variable for many soil properties (McKenzie *et al*. 2002) and is more prone to the effects of climate and land management inputs than deeper subsoil (as most of these impacts are initially at or near the surface). Hence, the older site data will not be representative of the conditions identified by newer, nearby sites, introducing additional unexplained variability into the modelling. The subsoil diagnostics and uncertainty ranges were better for pH and EC because these soil horizons are generally less spatially and temporally variable, and more 'static' than the surface horizons. The temporal range of the subsoil training data will therefore be less prone to introducing temporal uncertainty into the models.

Future versions of the products would benefit by introducing a temporal component into the modelling, for example, only using soil samples from the past decade, or modelling by decade, and comparing model diagnostics to determine whether temporal instability is contributing to the unexplained variability. However, there were insufficient data for some soil attributes to provide meaningful training data across such a large area, which could be addressed by the targeting and collection of new soils data, and the incorporation of recently accessed additional legacy data.

Mapping artefacts

For some soil property surfaces, especially those strongly explained by rainfall, good modelling diagnostics were achieved, but 'unrealistic' mapping artefacts were produced; that is, a sharp change in the continuous attribute was evident at the boundary of a rainfall isohyet. This was caused by: (*i*) the strongly evident west–east trend in mean annual rainfall; (*ii*) the relatively sharp change in rainfall with respect to distance, due to the rain-shadow effects of the central plateau; (*iii*) the strong influence of rainfall on Tasmanian soil formation; and (*iv*) the data-partitioning effects of the regression-tree approach.

It was decided to test the modelling by removing the rainfall covariate where these artefacts were being produced, for example, soil OC percentage. However, in this case, modelling diagnostics were considerably worse when rainfall was removed. In an attempt to allow the effects of rainfall to be incorporated into the regression-tree DSM, covariates were tested that would better explain the target OC percentage variability due to rainfall, but without the isohyet effects, and with better variation with terrain. The index produced by dividing rainfall by dominant prevailing wind (windward-leeward, SAGA GIS 2013) effects (to accentuate the rain-shadow areas of the state) was found to be an important explanatory dataset, and was effectively able to reduce mapping anomalies, producing more realistic mapping products showing carbon changing by terrain, rather than the rainfall 'smooth-curves'.

For clay percentage, rainfall (as an important covariate for partitioning the regression trees) also introduced some 'naturally unrealistic' mapping artefacts (Fig. 12), which were still evident when using the above rainfall–wind effect index. By removing rainfall altogether as a covariate, these artefacts were eliminated without overly affecting the modelling diagnostics (i.e. the model calibration-validation quality was not significantly reduced). For example, the clay percentage predictions for 0–15 cm had an RMSE difference of 0.07% and R^2 difference of 0.01 for calibration, and an RMSE difference of 0.12% and R^2 difference of 0.01 for validation. These comparisons could be as a result of the incidental rainfall formation influences already inherent within the other covariates used (e.g. terrain, persistent greenness and radiometrics).

In similar cases, it is necessary to weigh up the modelling diagnostics and co-variable usage against the final mapping appearance. Unnatural appearing DSM products could potentially lose 'credibility' with end-users (especially considering the early resistance to adoption of this science by the traditional soil science community); therefore, new covariates will need to be developed that will still capture strong co-variance without producing artefacts. If reasonably strong and comparable modelling diagnostics can still be achieved after removing the covariate in question while producing more 'naturally appearing' mapping, it could be argued that this approach is warranted, and that the other soil-forming factors are still able to explain enough variability. Another potential solution is to use an alternative

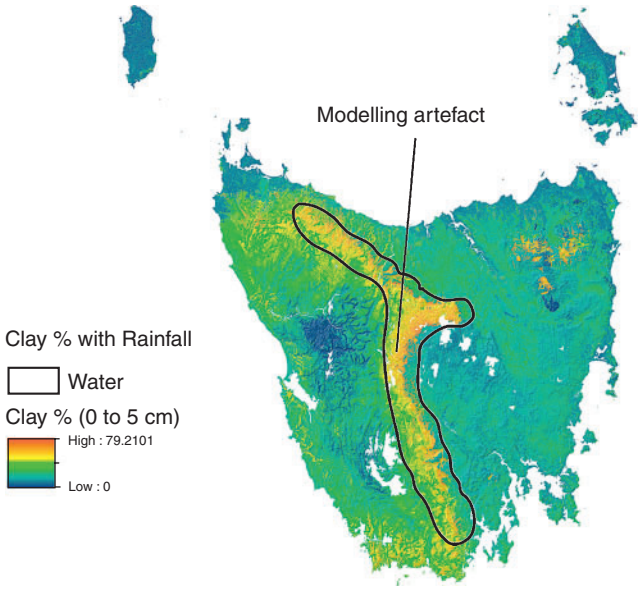

Fig. 12. Clay rainfall artefacts.

modelling approach to regression trees, where the models are continuous and artefacts due to data partitioning are minimised. Such artefacts are also discussed in the work of Padarian *et al.* (2014), who suggested a balance between numerical performance and a visual representation without artefacts.

Uncertainties

The model diagnostics reported are averaged across all regression-tree 'partitions'; therefore, some areas of the state will have better predictions and lower uncertainties than others. The relative magnitude of the uncertainties produced for the different soil attributes at their various depths were reasonable considering the data density and spatial spread available. A benefit of the regression-tree rule-based LOOCV approach is that uncertainties can be viewed spatially, so that end-users can determine which parts of the landscape have better soil-attribute predictions. For example, Fig. 13 shows the uncertainty (upper–lower prediction range) for clay percentage in the top 5 cm. The mapping shows that greater uncertainties (darker shading, up to 54%, i.e. ±27% from the predicted value) are evident in some coastal areas (where clay percentage is general lower, and sand percentage is generally higher), whereas lighter shaded areas have uncertainties as low as 12% (±6% from the predicted value). The lower uncertainties generally correspond to parts of the state where more soil-site data exist, as expected. However, some parts of the state that have low uncertainties (such as the Central Plateau) also have very few site data, implying similar environmental (covariate) conditions to the more data-dense parts of the state, informing these modelled areas. Based on these similar conditions, the soil-attribute modelled relationships are extrapolated into data-poor areas, similar to the 'homosoil' concept of extrapolating soil properties on a global scale (Mallavan *et al.* 2010).

There would also be inherent uncertainties in each of the PTFs, which were not considered as part of the Version 1.0 mapping. For future (Version 2.0) surfaces, these will be

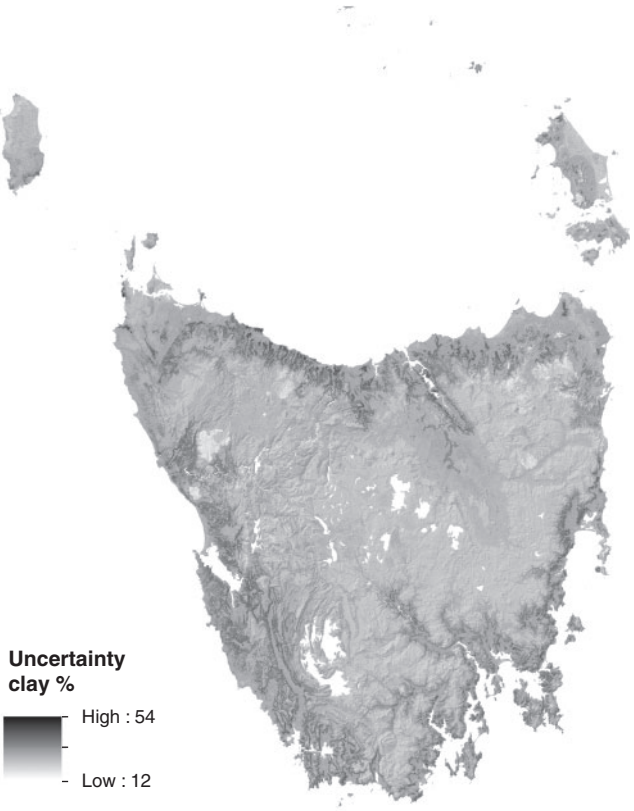

Fig. 13. Surface clay percentage (0–5 cm) uncertainties.

incorporated into the spatial modelling uncertainties for each of the contributing attributes.

The uncertainty mapping can provide a tool for targeting future soil-sampling exercises, whereby areas of high uncertainty could be prioritised for sampling if also environmentally or agriculturally important. However, the spatial distribution of existing site density should also be considered, ensuring that the entire Tasmanian covariate-feature space is well represented (as per Brungard and Boettinger 2010), and that data-poor areas with low uncertainties are tested for validation and future refinement of models if necessary.

Some of the Version 1.0 products can have relatively high uncertainties in some data-poor areas. However, a high uncertainty (in terms of a raster cell having a relatively large difference between the upper and lower PI) can still be useful for environmental modelling or digital soil assessments (Carré *et al.* 2007*a*), depending on where the threshold of interest occurs within the confidence limits. If a threshold value is outside the PI range, the end-user can have good confidence (90% in this case) that the value is higher or lower than the PI range. However, situations where a threshold value occurs around the predicted value (between the upper and lower PI) will introduce a higher level of uncertainty into the end-user product.

There has been much discussion regarding the development of standard approaches for generating estimates of uncertainty within the DSM and GSM community (Heuvelink 2014). As

such, continued testing and research are still required within this important element of DSM. The regression-tree rule-based uncertainty approach used for the development of Version 1.0 Tasmanian products is a preliminary attempt at developing meaningful uncertainty estimates for Tasmanian soil-attribute spatial variability, which will also be tested and refined during future version modelling.

Soil analyses and predictions

All database analytical data were assessed to ensure that the methodology and units were comparable. The cumulative distribution of the datasets was also assessed to identify and remove obvious data errors. For soil OC, all available data used were analysed by the Walkley–Black extraction method (Walkley and Black 1934), or MIR prediction was calibrated by this measurement. However, this method under-predicts the OC soil fraction, especially in higher concentrations in Tasmanian soils (McDonald *et al*. 2009). This indicates that potential OC could be underestimated for many of the Tasmanian forest sites at these locations, resulting in underestimation of spatial predictions; however, modelling could be over-predicting OC in peat areas, as observed with the high values (>60%) obtained in the Southwest WHA landscapes. It would therefore be advantageous to delineate the peat areas and model them separately from minerals soils because the environmental factors affecting OC in peat and mineral soils are different. Future versions of the DSM products would also benefit from the incorporation of newly collected OC analyses using the dry combustion method, and/or developing PTFs to convert the Walkley–Black OC data to dry combustion methods such as LECO (Wang and Anderson 1998).

Qualitative estimates

Although most of the surfaces generated were based on quantitative measurements of soil properties, several soil properties such as depth-related estimates, CF and drainage relied on qualitative descriptive data. This was necessary because inadequate data existed with direct measurements such as hydraulic conductivity and stone counts. Despite this, the qualitative integration of expert-based field estimates, even though from a variety of sources, produced reasonable modelling diagnostics and meaningful and realistic spatial variation in terms of soil–landscape relationships. Although not necessarily linear in relationship, the CF and drainage ordinal classes can be effectively captured as a continuous surface index using the regression-tree approach, as demonstrated by Kidd *et al*. (2014*a*), with reasonable validation demonstrating that the modelling can effectively account for any non-linearity. Applying the non-linear stretch of the CF percentage ranges to the 'indexed-class' values also produced meaningful patterns of CF abundance (as discussed in the *Results*); however, further validation could benefit from actual stone-count percentage values and testing within the 90% CL.

National v. regional DSM

The regional Tasmanian Version 1.0 surfaces have been modelled over a range and distribution of soil properties and covariate soil-forming factors different from the national TERN products, and should therefore show different spatial detail and PI values. All covariates were generated as regional Tasmanian products, and would potentially have values different from the national covariates because many terrain derivatives are produced in relative or index terms, stretched over the differences and distributions of elevation found within Tasmania. The differences in local *v.* national range of each target variable could also influence model formulation; local DSM products could have the advantage of forming models within the local range of conditions, and consequently show more local variability. However, national models could have the advantage of extrapolation of additional soil-training data in similar environmental conditions; for example, the lack of OC data in Tasmania's south-west peat areas could be better informed by the additional carbon site data elsewhere in similar parts of the country. Further research would inform whether the national and local products would each benefit from splitting the country into stratified environmental zones, for example, Tasmania and Victoria, and re-running the point-driven DSM process within the more homogeneous environments.

Future work

Legacy data

The Version 1.0 Tasmanian surfaces are considered the genesis of an evolving product, with modelling scripts written to automate the addition of site and covariate data. DPIPWE has undertaken a substantial effort in identifying, digitising and cleaning a wide range of legacy soil data from a variety of historical sources, targeting good-quality analytical data, and areas with a paucity of good site data. To date, ~3500 sites of varying quality have been identified and will be integrated into new DSM model re-runs (Version 2.0) as these data are processed. It is hoped that comparison of newly created Version 2.0 surfaces against Version 1.0 surfaces, in terms of mapping differences, uncertainties and model diagnostics, will clearly demonstrate the value of additional data and potentially stimulate further investment in collecting new soils data.

Covariates

The integration of the radiometrics and geology was shown to be an important predictor in many soil properties and demonstrates the importance of good remotely sensed data, especially related to parent material. Future work will also explore the development and integration of improved covariate layers, including potential LIDAR elevation models and multi-spectral satellite imagery and derivatives. Incorporation of fractional groundcover (Muir 2011) and fractional dynamic land cover (Armston *et al*. 2009) covariates would also be beneficial for quantifying potential spatial variations in soil properties, and as an additional explanatory variable for impacts of land use on soil attributes. Testing will be done to determine whether currently used modelling hardware infrastructure can cope with producing the products at 1-arc-second (30-m) resolution. Alternative testing will involve building the regression-tree models with 30-m covariates to increase the chances of applying an accurate covariate value allocation at each point, but applying the model to the 80-m covariates to reduce processing time.

Modelling

As mentioned as a possible solution to reducing mapping artefacts, alternative modelling approaches will also be tested, however, regression tree (Cubist) is strongly favoured because of the interpretive benefits and transparent outputs. End-users can clearly see how each covariate contributed to the modelled soil attributes and better understand the soil-forming soil–landscape processes occurring in different parts of the environment. This is lacking in approaches such as artificial neural networks in soil-property prediction (Zhao *et al*. 2009) and random forests (Liaw and Wiener 2002), where model outputs are less easily interpreted.

Another potential approach is to test the disaggregation of land-systems mapping, the only state-wide polygon product available in some areas, which could be split into minor spatial components of modal soil properties by using an approach consistent with the DSMART methodology developed by Odgers *et al*. (2014). A model-ensemble approach could be integrated to average the disaggregation outputs with the point-source DSM modelling, to potentially better inform areas with no or few soil-site data; this has been beneficial elsewhere (Malone *et al*. 2014).

The predictive approach used for the Version 1.0 surfaces fitted models to each standard depth separately (following Arrouays *et al*. 2014), and these are considered 3D in that there are spatial soil-attribute predictions across the state through all standard depths to 2 m. However, no integration of vertical data trend was considered or incorporated into a true 3D modelling process, as described by Hengl *et al*. (2014); future modelling could benefit from testing such an approach.

Sampling

As an example of how the uncertainties could be used to help guide future sampling, Fig. 14 shows the combined uncertainty values for several important soil attributes for an ESA in the Great Forester–Brid Irrigation Scheme, in the north-east of Tasmania. Surface soil (0–5 cm) and subsoil (60–100 cm) uncertainty ranges for pH, clay percentage, EC_{se} and CF were calculated by subtracting the lower PI from the upper PI values, then standardised to a range of 0–100 to give an indication of relative error across both topsoil and subsoil predictions. Values were then averaged to provide an indication of where in the landscape uncertainties were highest for more soil attributes. Figure 14 shows that generally in lower elevations corresponding to coastal plains and dissected valley systems (Quaternary alluvium), uncertainties are larger than on the upper slopes around Scottsdale. This would be due in part to these areas often containing extreme prediction values, that is, low clay, low CF, high pH, and high EC, as well as low site-data density. Future site sampling would be prioritised to areas of high DSM uncertainties, but ensuring the sampling distribution is still representative of the covariate distribution. This could be achieved using a purposive sampling approach such as Conditioned Latin Hypercube Sampling (Minasny and McBratney 2006*a*, 2006*b*), which could be effectively constrained following the methodologies described by Clifford *et al*. (2014) and Roudier *et al*. (2012), where the sampling constraint would be the areas of high DSM uncertainty, rather than access (distance to roads). Clustering of covariates for a stratified-random approach, taking into account the covariate distribution of the existing site data in conjunction with higher uncertainties, would be another approach, as per Kidd *et al*. (2015).

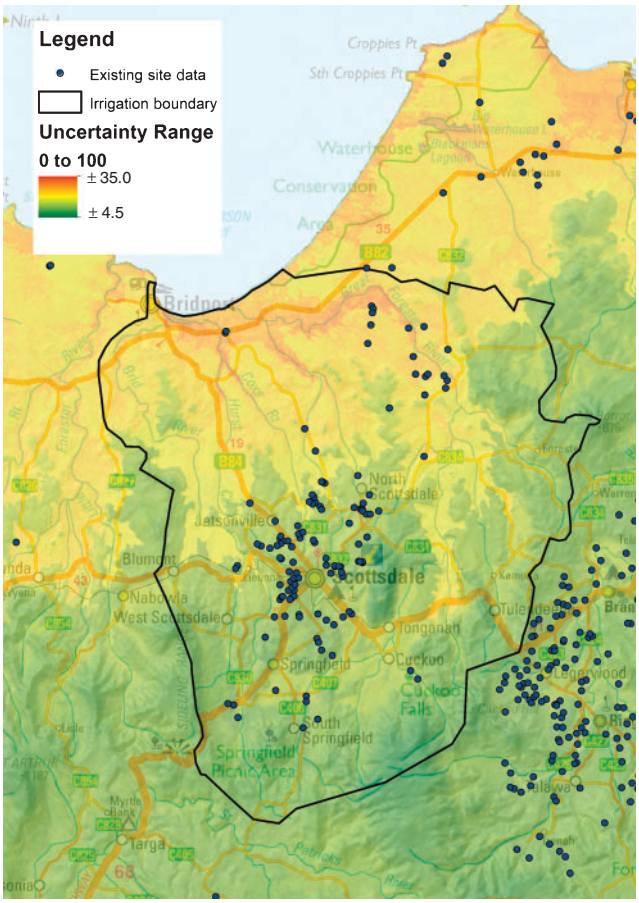

Fig. 14. Sampling scenario based on uncertainties.

Standardised uncertainties could be averaged across all depths and all soil attributes to guide a sampling campaign aimed at improving the Version 1.0 products across all areas and attributes.

Initial uses

After acknowledging the limitations of some areas and attributes of the Version 1.0 DSM surfaces, some products have already been requested and incorporated into various environmental or agricultural modelling scenarios. For example, the clay percentage and drainage surfaces were used to identify areas of high 'pugging' risk (soil structural damage from cattle in wet conditions), and ryegrass suitability was modelled using Tasmanian ESA rule-sets (Kidd *et al*. 2014*b*) to identify areas suitable for 'winter-finishing' of beef cattle in Tasmania (Davey 2014).

Importantly, the Version 1.0 surfaces provide consistent inputs to environmental modelling and assessment in areas outside the legacy-soil mapped areas that were previously not available (without relying on conceptual land systems), with the additional benefit of providing uncertainty estimates. They are a

first attempt at developing a quantitative spatial soil-attribute product for all of Tasmania. The authors acknowledge that the Version 1.0 products should be improved with the addition of appropriate soil and covariate data; however, the products are considered an important, foundational soil-infrastructure dataset for the state, quantifying where soil information uncertainty is highest, which can guide future investment in data capture.

Conclusions

The Version 1.0 digital soil maps of soil attributes and uncertainties produced for Tasmania are an important first step in developing a comprehensive soil infrastructure to deliver quantitative soil-attribute predictions and modelled uncertainties at a useful resolution for farm enterprise and environmental planning. Most soil surfaces were produced with acceptable modelling diagnostics and uncertainty ranges, delivering realistic soil–landscape spatial patterns extrapolated into unsampled areas. The maps have been produced to allow continuous improvements, with models that have been automated to accept newly collected soil data and covariates to generate new versions as required, which should improve diagnostics and uncertainties in some areas. It is the first attempt at quantifying the soil properties of Tasmania based on existing data, which will help to guide future investment in soil data collection and provide consistent soil-attribute data with uncertainties to environmental modelling and assessment activities.

Acknowledgements

The authors acknowledge the ARC Linkage project LP110200731 for supporting the Wealth from Water project. Modelling was undertaken on the NCI National Facility in Canberra, Australia, which is supported by the Australian Commonwealth Government. The authors also acknowledge the TERN team (Terrestrial Ecosystem Research Network Soil and Landscape Grid of Australia) including Ross Searle, Raphael Viscarra-Rossel and Mike Grundy (CSIRO) for advice and collaboration; Rob Moreton and Chris Grose (DPIPWE) for review and expert comment on surfaces; Rob Moreton's collection of additional legacy data for future model re-runs; and Rhys Stickler and Peter Voller (DPIPWE) for project management and support.

References

Adams W (1973) The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. *Journal of Soil Science* **24**, 10–17. doi:10.1111/j.1365-2389.1973.tb00737.x

Armston JD, Danaher TJ, Scarth PF, Moffiet TN, Denham RJ (2009) Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. *Journal of Applied Remote Sensing* **3**, 033540–033540–28. doi:10.1117/1.3216031

Arrouays D, McBratney A, Minasny B, Hempel J, Heuvelink G, MacMillan R, Hartemink A, Lagacherie P, McKenzie N (2014) The GlobalSoilMap project specifications. In 'GlobalSoilMap. Basis of the global spatial soil information system'. (Eds D Arrouays, NJ McKenzie, JW Hempel, AR de Forges, AB McBratney) pp. 9–12. (CRC Press/Balkema: Leiden, The Netherlands)

Australian Bureau of Meteorology (2014) Climate statistics for Australian sites—Tasmania. Bureau of Meteorology Australia, Climate Data Online. Available at: www.bom.gov.au/climate/data/ (accessed June 2014).

Bock M, Böhner J, Conrad O, Köthe R, Ringeler A (2007) XV. Methods for creating functional soil databases and applying digital soil mapping with SAGA GIS. In 'Status and prospect of soil information in south-eastern Europe: soil databases, projects and applications'. (Eds T Hengl, P Panagos, A Jones, G Toth) (Office for Official Publications of the European Communities: Luxemburg)

Breiman L, Friedman J, Olshen R, Stone C (1984) 'Classification and regression trees (CART).' (Wadsworth: Monterey, CA, USA)

Brungard CW, Boettinger JL (2010) Conditioned Latin hypercube sampling: Optimal sample size for digital soil mapping of arid rangelands in Utah, USA. In 'Digital soil mapping. Vol. 2'. (Eds JL Boettinger, DW Howell, AC Moore, AE Hartemink, S Kienast-Brown) pp. 67–75. (Springer: Dordrecht, The Netherlands)

Carré F, McBratney AB, Mayr T, Montanarella L (2007*a*) Digital soil assessments: Beyond DSM. *Geoderma* **142**, 69–79. doi:10.1016/j.geoderma.2007.08.015

Carré F, McBratney AB, Minasny B (2007*b*) Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma* **141**, 1–14. doi:10.1016/j.geoderma.2007.01.018

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **35**, 445–453. doi:10.2307/1931034

Clifford D, Payne JE, Pringle MJ, Searle R, Butler N (2014) Pragmatic soil survey design using flexible Latin hypercube sampling. *Computers & Geosciences* **67**, 62–68. doi:10.1016/j.cageo.2014.03.005

Cook SE, Corner RJ, Groves PR, Grealish GJ (1996) Use of airborne gamma radiometric data for soil mapping. *Soil Research* **34**, 183–194. doi:10.1071/SR9960183

Cotching W (2012) Carbon stocks in Tasmanian soils. *Soil Research* **50**, 83–90.

Cotching WE, Kidd DB (2010) Soil quality evaluation and the interaction with land use and soil order in Tasmania, Australia. *Agriculture, Ecosystems & Environment* **137**, 358–366. doi:10.1016/j.agee.2010.03.006

Cotching WE, Lynch S, Kidd DB (2009) Dominant soil orders in Tasmania: distribution and selected properties. *Soil Research* **47**, 537–548. doi:10.1071/SR08239

Davey L (2014) Tasmanian Red Meat Strategy. Macquarie Franklin, Meat and Livestock Australia Limited, Sydney.

Davies JL (1967) Tasmanian landforms and Quaternary climates. In 'Landform studies from Australia and New Guinea'. pp. 1–25. (Australian National University Press: Canberra, ACT)

Davies JB (1988) 'Land Systems of Tasmania. Region 6: South, East and Midlands—a resource classification survey.' (Department of Agriculture Tasmania: Hobart, Tas.)

Dobos E, Micheli E, Baumgardner MF, Biehl L, Helt T (2000) Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma* **97**, 367–391. doi:10.1016/S0016-7061(00)00046-X

Doyle R (1993) 'Soils of the South Esk Sheet, Tasmania (southern half).' (Department of Primary Industry and Fisheries: Hobart, Tas.)

Ebdon D (1985) 'Statistics in geography.' (Wiley: Hoboken, NJ, USA)

Erbe P, Schuler U, Wicharuck S, Rangubpit W, Stahr K, Herrmann L, Gilkes R (2010) Creating soil degradation maps using gamma-ray signatures. In 'Soil solutions for a changing world. Proceedings 19th World Congress of Soil Science'. 1–6 August, Brisbane, Qld. (International Union of Soil Sciences) Available at: www.iuss.org/19th%20WCSS/Symposium/pdf/0407.pdf

Gallant JC, Dowling TI (2003) A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* **39**, 1347–1360. doi:10.1029/2002WR001426

Gallant J, Dowling TI, Read A, Wilson N, Tickle P (2011) '1 second SRTM Derived Digital Elevation Models User Guide.' (Geoscience Australia: Canberra, ACT)

Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma* **146**, 102–113. doi:10.1016/j.geoderma.2008.05.008

Grose C (1999) 'Land capability handbook. Guidelines for the Classification of Agricultural Land in Tasmania.' 2nd edn (Department of Primary Industries, Water and Environment: Hobart, Tas.) Available at: http://dpipwe.tas.gov.au/Documents/Land_Cap_Revised-handbook.pdf

Grundy M, Searle R, Robinson J, Minasny B (2012) An Australian soil grid: infrastructure and function. In 'Digital soil assessments and beyond.' (Eds B Minasny, BP Malone, AB McBratney) pp. 459–464. (CRC Press/Balkema: Leiden, The Netherlands)

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

Henderson B, Bui EN (2002) An improved calibration curve between soil pH measured in waterand $CaCl_2$. *Soil Research* **40**, 1399–1405. doi:10.1071/SR01098

Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **120**, 75–93. doi:10.1016/j.geoderma.2003.08.018

Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: From equations to case studies. *Computers & Geosciences* **33**, 1301–1315. doi:10.1016/j.cageo.2007.05.001

Hengl T, de Jesus J, MacMillan R, Batjes N, Heuvelink G, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars J, Walsh M, Gonzalez M (2014) SoilGrids1km—Global Soil Information based on automated mapping. *PLoS One* **9**, e105992. doi:10.1371/journal.pone.0105992

Herrmann L, Schuler U, Rangubpit W, Erbe P, Surinkum A, Zarei M, Stahr K, Gilkes R (2010) The potential of gamma-ray spectrometry for soil mapping. In 'Soil solutions for a changing world. Proceedings 19th World Congress of Soil Science'. 1–6 August, Brisbane, Qld. (International Union of Soil Sciences) Available at: www.iuss.org/19th%20WCSS/Symposium/pdf/WG1.5.pdf

Heuvelink GB (2014) Uncertainty quantification of GlobalSoilMap products. In 'GlobalSoilMap, Basis of the global spatial soil information system'. (Eds D Arrouays, NJ McKenzie, JW Hempel, AR de Forges, AB McBratney) pp. 327–332. (CRC Press/Balkema: Leiden, The Netherlands)

Hijmans RJ, van Etten J (2012) raster: Geographic analysis and modeling with raster data. R package version 1. 9–92. The R Foundation, Vienna. Available at: https://cran.r-project.org/web/packages/raster/

Isbell R (2002) 'The Australian Soil Classification.' Revised edn. AustralianSoil and Land Survey Handbook Series 4. (CSIRO Publishing: Melbourne)

IUSS Working Group WRB (2007) 'World Reference Base for soil resources 2006.' First update 2007. World Soil Resources Report No. 103. (FAO: Rome)

Kidd DB (2003) 'Land degradation and salinity risk investigation in the Tunbridge District, Tasmanian Midlands.' (Department of Primary Industries, Parks, Water and the Environment: Hobart, Tas.)

Kidd DB, Malone BP, McBratney AB, Minasny B, Webb M, Grose CJ, Moreton RM, Viscarra Rossel RA, Cotching WE, Sparrow LA, Smith R (2012*a*) Using digital soil mapping for enterprise suitability assessment in support of Tasmanian irrigation development. In 'Proceedings 5th Joint Australian and New Zealand Soil Science Conference: Soil solutions for diverse landscapes'. Hobart, Tas. (Eds LL Burkitt, LA Sparrow) pp. 636–640. (Australian Society of Soil Science Inc.)

Kidd DB, Webb MA, Grose CJ, Moreton RM, Malone BP, McBratney AB, Misnasny B, Viscarra-Rossel RA, Cotching WE, Sparrow LA, Smith R (2012*b*) Digital soil assessment: Guiding irrigation expansion in Tasmania, Australia. In 'Digital soil assessments and beyond: Proceedings 5th Global Workshop on Digital Soil Mapping 2012'. Sydney, Australia. (Eds B Minasny, BP Malone, AB McBratney) pp. 3–8. (CRC Press/Balkema: Leiden, The Netherlands)

Kidd DB, Malone BP, McBratney AB, Minasny B, Webb MA (2014*a*) Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. *Soil Research* **52**, 107–119. doi:10.1071/SR13100

Kidd DB, Webb MA, McBratney AB, Minasny B, Malone BP, Grose CJ, Moreton RM (2014*b*) Operational digital soil assessment for enterprise suitability in Tasmania, Australia. In 'GlobalSoilMap: Basis of the global soil spatial information system'. (Eds D Arrouays, NJ McKenzie, JW Hempel, AR de Forges, AB McBratney) pp. 113–121. (CRC Press/Balkema: Leiden, The Netherlands)

Kidd D, Malone B, McBratney A, Minasny B, Webb M (2015) Operational sampling challenges to digital soil mapping in Tasmania, Australia. *Geoderma Regional* **4**, 1–10. doi:10.1016/j.geodrs.2014.11.002

Kirkpatrick J (1981) A transect study of forests and woodlands on dolerite in the Eastern Tiers, Tasmania. *Vegetatio* **44**, 155–163. doi:10.1007/BF00156625

Klingebiel AA, Montgomery PH (1961) 'Land capability classification.' Agriculture Handbook No. 210. (USDA, Soil Conservation Service: Washington, DC)

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In 'Proceedings 14th International Joint Conference on Artificial Intelligence'. Vol. 2. pp. 1137–1143. (Morgan Kaufmann Publishers Inc.: San Francisco, CA)

Kuhn M, Weston S, Keefer C, Coulter N (2012) Cubist Models for regression in R. The R Foundation, Vienna. Available at: https://cran.r-project.org/web/packages/Cubist/vignettes/cubist.pdf

Kuhn M, Weston S, Keefer C, Coulter N, Quinlan R (2013) Cubist: Rule- and Instance-Based Regression Modeling. R package version 0.0.15. Available at: http://CRAN.R-project.org/package=Cubist.

Land and Water Development Division (1993) 'Global and national soils and terrain digital databases (SOTER): Procedures manual.' (FAO: Rome)

Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* **2**, 18–22.

Lin LIK (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268. doi:10.2307/2532051

Mallavan B, Minasny B, McBratney A (2010) Homosoil, a methodology for quantitative extrapolation of soil information across the globe. In 'Digital soil mapping'. (Eds JL Boettinger, DW Howell, AC Moore, AE Hartemink, S Kienast-Brown) pp. 137–150. (Springer: Dordrecht, The Netherlands)

Malone BP, McBratney AB, Minasny B, Laslett GM (2009) Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **154**, 138–152. doi:10.1016/j.geoderma.2009.10.007

Malone BP, McBratney AB, Minasny B (2011) Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* **160**, 614–626. doi:10.1016/j.geoderma.2010.11.013

Malone BP, Minasny B, Odgers NP, McBratney AB (2014) Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* **232–234**, 34–44. doi:10.1016/j.geoderma.2014.04.033

Martin M, Wattenbach M, Smith P, Meersmans J, Jolivet C, Boulonne L, Arrouays D (2011) Spatial distribution of soil organic carbon stocks in France. *Biogeosciences* **8**, 1053–1065. doi:10.5194/bg-8-1053-2011

McBratney AB, Minasny B, Cattle SR, Vervoort RW (2002) From pedotransfer functions to soil inference systems. *Geoderma* **109**, 41–73. doi:10.1016/S0016-7061(02)00139-8

McBratney AB, Mendonça Santos ML, Minasny B (2003) On digital soil mapping. *Geoderma* **117**, 3–52. doi:10.1016/S0016-7061(03)00223-4

McDonald D, Baldock J, Kidd D (2009) 'Cradle Coast organic carbon monitoring trial. National Land and Water Resources Audit.' (Department of Primary Industry and Fisheries: Hobart, Tas.)

McKenzie NJ, Ryan PJ (1999) Spatial prediction of soil properties using environmental correlation. *Geoderma* **89**, 67–94. doi:10.1016/S0016-7061(98)00137-2

McKenzie NJ, Henderson B, McDonald W (2002) Monitoring soil change: principles and practices for Australian conditions. CSIRO Land & Water, Technical Report 18/02, May 2002, Canberra, ACT.

McKenzie N, Grundy M, Webster R, Ringrose-Voase A (2008) 'Guidelines for surveying soil and land resources.' (CSIRO Publishing: Melbourne)

Minasny B, McBratney AB (2006*a*) Latin hypercube sampling as a tool for digital soil mapping. In 'Developments in soil science'. Vol. 31, Ch. 12. (Eds ABMP Lagacherie, M Voltz) pp. 153–606. (Elsevier: Amsterdam)

Minasny B, McBratney AB (2006*b*) A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* **32**, 1378–1388.

Minasny B, McBratney A, Brough D, Jacquier D (2011) Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. *European Journal of Soil Science* **62**, 728–732. doi:10.1111/j.1365-2389.2011.01386.x

Mineral Resources Tasmania (2008) '1 : 25,000 Digital Geology.' (Mineral Resources Tasmania: Hobart, Tas.)

Mitchell A (2005) 'The ESRI guide to GIS analysis: Spatial measurements and statistics.' Vol. 2. (ESRI Press: Redlands, CA, USA)

Muir J (2011) 'Field measurement of fractional ground cover: a technical handbook supporting ground cover monitoring for Australia.' (ABARES: Canberra, ACT)

National Committee on Soil and Terrain (2009) 'Australian soil and land survey field handbook.' 3rd edn. Australian Soil and Land Survey Handbooks Series 1. (CSIRO Publishing: Melbourne)

Odeh IOA, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* **67**, 215–226. doi:10.1016/0016-7061(95)00007-B

Odgers NP, Sun W, McBratney AB, Minasny B, Clifford D (2014) Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma* **214–215**, 91–100. doi:10.1016/j.geoderma.2013.09.024

Oldeman L, Van Engelen V (1993) A world soils and terrain digital database (SOTER)—An improved assessment of land resources. *Geoderma* **60**, 309–325. doi:10.1016/0016-7061(93)90033-H

Padarian J, McBratney AB, Minasny B, Dalgliesh N (2014) Predicting and mapping the soil available water capacity of Australian soils wheatbelt. *Geoderma Regional* **2–3**, 110–118. doi:10.1016/j.geodrs.2014.09.005

Pain C, Wilford J, Dohrenwend J (1999) Regolith-landform evolution on Cape York Peninsula. Implications for Mineral Exploration. In 'New approaches to an old continent. Proceedings 3rd Australian Regolith Conference'. (Eds AF Britt, L Bettenay) pp. 55–66. (CRC LEME)

Pemberton M (1989) 'Land Systems of Tasmania. Region 7, South West.' (Department of Agriculture: Hobart, Tas.)

Peverill KI, Sparrow L, Reuter DJ (1999) 'Soil analysis: an interpretation manual.' (CSIRO Publishing: Melbourne)

Pinder D, Witherick M (1972) The principles, practice and pitfalls of nearest-neighbour analysis. *Geography* **57**, 277–288.

Pinkard GJ, Richley LR (1982) 'Land systems of Tasmania, Region 2: A Report.' (Tasmanian Department of Agriculture: Hobart, Tas.)

Pyrcz MJ, Deutsch CV (2003) Declustering and debiasing. Newsletter 19, October 2003. (Ed. S Searston) Geostatistical Association of Australia, Melbourne.

Quinlan J (2005) An overview of Cubist. RuleQuest Research Data Mining Tools. Available at: www.rulequest.com/cubist-info.html

R Development Core Team (2014) 'R: A language and environment for statistical computing.' (R Foundation for Statistical Computing: Vienna)

Rayment GE, Lyons DJ (2011) 'Soil chemical methods: Australasia.' (CSIRO Publishing: Melbourne)

Richley LR (1978) 'Land systems of Tasmania, Region 3: A report.' (Tasmanian Department of Agriculture: Hobart, Tas.)

Rodriguez JD, Perez A, Lozano JA (2010) Sensitivity analysis of kappa-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 569–575.

Roudier P, Beaudette D, Hewitt A (2012) A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In 'Digital soil assessments and beyond. Proceedings 5th Global Workshop on Digital Soil Mapping 2012'. Sydney, Australia. (Eds B Minasny, BP Malone, AB McBratney) pp. 227–232. (CRC Press/Balkema: Leiden, The Netherlands)

SAGA GIS (2013) System for Automated Geoscientific Analyses. Available at: www.saga-gis.org/en/index.html

Spanswick S, Kidd D (2000) 'Burnie soil report—a revised edition.' Reconnaissance Soil Map Series of Tasmania. (Department of Primary Industries, Parks, Water and the Environment: Hobart, Tas.)

Spanswick S, Kidd D (2001) 'Oatlands soil report—a revised edition.' Reconnaissance Soil Map Series of Tasmania. (Department of Primary Industries, Parks, Water and the Environment: Hobart, Tas.)

Sparrow LA, Cotching WE, Cooper J, Rowley W (1999) Attributes of Tasmanian ferrosols under different agricultural management. *Australian Journal of Soil Research* **37**, 603–622. doi:10.1071/SR98108

Tasmanian Climate Change Office (2012) Adapting to climate change in Tasmania. Issues Paper. Tasmanian Climate Change Office, Hobart, Tas.

Taylor MJ, Smettem K, Pracilio G, Verboom W (2002) Relationships between soil properties and high-resolution radiometrics, central eastern wheatbelt, Western Australia. *Exploration Geophysics* **33**, 95–102. doi:10.1071/EG02095

Tranter G, Minasny B, McBratney A, Murphy B, McKenzie N, Grundy M, Brough D (2007) Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use and Management* **23**, 437–443. doi:10.1111/j.1475-2743.2007.00092.x

Viscarra Rossel RA, Webster R, Kidd D (2014) Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random forest kriging. *Earth Surface Processes and Landforms* **39**, 735–748. doi:10.1002/esp.3476

Walkley A, Black IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* **37**, 29–38. doi:10.1097/00010694-193401000-00003

Wang D, Anderson DW (1998) Direct measurement of organic carbon content in soils by the Leco CR-12 carbon analyzer. *Communications in Soil Science and Plant Analysis* **29**, 15–21. doi:10.1080/00103629809369925

Webb M, Kidd D, Grose C, Moreton R, Malone B, McBratney A, Minasny B (2014) Integrating climate into the Digital Soil Assessment framework to assess land suitability. In 'GlobalSoilMap: Basis of the global spatial soil information system'. (Eds D Arrouays, NJ McKenzie, JW Hempel, AR de Forges, AB McBratney) pp. 393–399. (CRC Press/Balkema: Leiden, The Netherlands)

Yang L, Homer C, Hegge K, Huang C, Wylie B, Reed B (2001) A Landsat 7 scene selection strategy for a National Land Cover Database. In 'Proceedings International Geoscience and Remote Sensing Symposium'. pp. 1123–1125. (Institute of Electrical and Electronics Engineers)

Zhao Z, Chow TL, Rees HW, Yang Q, Xing Z, Meng F-R (2009) Predict soil texture distributions using an artificial neural network model. *Computers and Electronics in Agriculture* **65**, 36–48. doi:10.1016/j.compag.2008.07.008