

Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change

RAPHAEL A. VISCARRA ROSSEL¹, RICHARD WEBSTER², ELISABETH N. BUI¹ and JEFF A. BALDOCK³

¹Bruce E. Butler Laboratory, CSIRO Land and Water, PO Box 1666, Canberra, ACT 2601, Australia, ²Rothamsted Research, Harpenden, AL5 2JQ, UK, ³CSIRO Land and Water, APMB 2, Glen Osmond, SA 5064, Australia

Abstract

We can effectively monitor soil condition—and develop sound policies to offset the emissions of greenhouse gases—only with accurate data from which to define baselines. Currently, estimates of soil organic C for countries or continents are either unavailable or largely uncertain because they are derived from sparse data, with large gaps over many areas of the Earth. Here, we derive spatially explicit estimates, and their uncertainty, of the distribution and stock of organic C in the soil of Australia. We assembled and harmonized data from several sources to produce the most comprehensive set of data on the current stock of organic C in soil of the continent. Using them, we have produced a fine spatial resolution baseline map of organic C at the continental scale. We describe how we made it by combining the bootstrap, a decision tree with piecewise regression on environmental variables and geostatistical modelling of residuals. Values of stock were predicted at the nodes of a 3-arc-sec (approximately 90 m) grid and mapped together with their uncertainties. We then calculated baselines of soil organic C storage over the whole of Australia, its states and territories, and regions that define bioclimatic zones, vegetation classes and land use. The average amount of organic C in Australian topsoil is estimated to be 29.7 t ha⁻¹ with 95% confidence limits of 22.6 and 37.9 t ha⁻¹. The total stock of organic C in the 0–30 cm layer of soil for the continent is 24.97 Gt with 95% confidence limits of 19.04 and 31.83 Gt. This represents approximately 3.5% of the total stock in the upper 30 cm of soil worldwide. Australia occupies 5.2% of the global land area, so the total organic C stock of Australian soil makes an important contribution to the global carbon cycle, and it provides a significant potential for sequestration. As the most reliable approximation of the stock of organic C in Australian soil in 2010, our estimates have important applications. They could support Australia's National Carbon Accounting System, help guide the formulation of policy around carbon offset schemes, improve Australia's carbon balances, serve to direct future sampling for inventory, guide the design of monitoring networks and provide a benchmark against which to assess the impact of changes in land cover, land management and climate on the stock of C in Australia. In this way, these estimates would help us to develop strategies to adapt and mitigate the effects of climate change.

Keywords: sequestration, soil carbon baseline, soil carbon stock, soil organic carbon, spatial modelling

Received 4 December 2013 and accepted 1 February 2014

Introduction

Organic carbon in soil derives from living organisms in and above the soil that convert atmospheric carbon dioxide (CO₂) into a range of organic compounds and structures. Throughout their life cycles, the organisms synthesize organic matter, which later decomposes in the soil. As these organisms respire, most of the captured carbon is eventually returned to the atmosphere as CO₂.

The decomposition of organic matter in the soil releases significant quantities of nutrients, particularly nitrogen, that become available to plants, microorgan-

isms and fungi. The nutrients released can also be retained in soil, in store and on exchange sites, thereby augmenting the soil's buffering capacity. Organic matter also helps to aggregate soil particles and to develop soil structure, and it increases the storage of water and availability of that water for plants. In sum, the amount of organic carbon in the soil is an important attribute of the soil's condition.

The organic carbon content of the soil is seen as increasingly important for ecosystems, both natural and agricultural; the topic attracts interest both nationally and internationally (e.g., Bui *et al.*, 2009; Chaplot *et al.*, 2010; Lugato *et al.*, 2013). Capturing and retaining additional carbon in soil (sequestration) can mitigate the emissions of the greenhouse gases, carbon dioxide, methane, nitrous oxide and, at the same time, improve

Correspondence: Raphael A. Viscarra Rossel, tel. +61 2 6246 5945, fax +61 2 6246 5965, e-mail: raphael.viscarra-rossel@csiro.au

the quality and productivity of the soil to sustain food production. Around the world, governments are developing policies to increase or restore the organic C stored in the soil. One example is the Australian Government's Carbon Farming Initiative (Garnaut, 2011).

Soil holds the largest terrestrial store of organic C. Globally, estimated stores of organic C are approximately 684–724 Gt (1 Gt = 10^{15} g = 1 Pg) in the top 30 cm, 1200–1550 Gt in the uppermost metre of soil and around 2300–2450 Gt in the upper 2 or 3 m (Sombroek *et al.*, 1993; Eswaran *et al.*, 1995; Batjes, 1996; Jobbágy and Jackson, 2000; Lal, 2004). Comparative estimates of organic C contained in living biomass (550–560 Gt) and the atmosphere CO₂ (760–780 Gt) (Lal, 2004; Houghton, 2005) indicate that variations in the size of the soil's store of organic C could significantly alter the concentration of CO₂ in the atmosphere. In contrast, the amount of inorganic C in the soil is estimated to be around 720–930 Gt (Sombroek *et al.*, 1993; Batjes, 1996), with a more recent estimate of 947 Gt provided by Eswaran *et al.* (2000).

In Australia, modelling studies of the dynamics of organic C have produced estimates of about 20 and 26.9 Gt in the top 20 and 100 cm (Barrett, 2002, 2013), and 18.8 and 34.2 Gt in the top 30 and 100 cm of soil respectively (Grace *et al.*, 2006). However, the state and temporal trends of the stores of organic C in Australian soil and their spatial distributions are largely unknown or uncertain, mainly because there are too few data on the soil's organic C content and bulk density (BD) that can be used to provide estimates for the country. Conventional methods of soil survey and analyses are expensive.

It is thought that soil in many regions is losing C, but some scientists suggest that—with appropriate land management or changes in land use—the soil could store more organic C than it does now, particularly in the savannas of northern Australia (Cleugh *et al.*, 2011; Richards *et al.*, 2011) and the agricultural land of south-eastern Australia (Luo *et al.*, 2010; Zhao *et al.*, 2013). There is no consistent overall trend, however, and evidence is fragmentary. To improve the confidence and robustness of such assessments, we must investigate methods for obtaining spatially explicit estimates of organic C in the soil with estimates of uncertainty from data.

The aim of this study was to estimate at a fine spatial resolution the current stock of organic C in Australian soil in the 0–30 cm layer, which is the reference depth of the Intergovernmental Panel on Climate Change (IPCC), along with estimates of uncertainty, and to calculate baselines of organic C stock and their uncertainty over the whole of Australia, its states and territories, regions that define bioclimatic zones and vegetation and land-use types.

Soil inventory

The data on the soil's organic C content and BD recorded from 2000 to 2013 (median year 2009) came from three sources.

1. Australia's National Soil Carbon Research Programme (SCaRP). It was designed to quantify variations in the content, stock and composition of organic C in the 0–30 cm layer of soil due to agricultural management (Baldock *et al.*, 2013). The soil was sampled within 25 m × 25 m quadrats that were representative of predefined combinations of soil type and agricultural management. Soil was collected from each of 10 randomly selected locations from the intersection points of a 5 m × 5 m grid within each quadrat with a ≥40-mm diameter soil corer. Ten cores from each of the 0–10, 10–20 and 20–30 cm layers were composited into one bulk sample for each layer for laboratory analysis. At each site, the BD and gravel content of the soil was measured for each depth layer. Organic C content was measured in the laboratory on a dry-combustion Dumas elemental analyser (Rayment and Lyons, 2011). This gave us 4125 values of the organic C content and BD from sites in fields on commercial farms. We denote them as the variables C and D_B for our study.
2. Spectroscopic estimates of organic C and BD made with the Australian visible–near infrared database (Viscarrá Rossel and Webster, 2012) on soil samples collected for the National Geochemical Survey of Australia (NGSA) (de Caritat *et al.*, 2008). The samples were collected from across Australia after first dissection of the continent into drainage catchments and then selection of sampling sites at low points of the catchments but well above the water table in the lowest positions. At each site, samples were collected and bulked to produce two specimens from within two depth layers, 0–10 cm and 60–80 cm. The spectroscopic measurements were made on the fine earth fraction (<2 mm), and we used data from 1101 sites.
3. The Australian Soil Resource Information System (ASRIS), the Commonwealth Scientific and Industrial Research Organisation's (CSIRO's) central repository for soil data in Australia (Johnston *et al.*, 2003). We could use 491 soil profile data from ASRIS that had measurements of organic C content and BD from two or three depth layers. The samples originated primarily from agricultural soil in eastern and southern Australia, and the organic C contents had been determined by the Walkley–Black wet oxidation method and a LECO elemental

analyser (Rayment and Lyons, 2011). When recorded, the most common method of measuring BD of the soil was the one using intact cores as described in Cresswell and Hamilton (2002).

The combined data represent soil from all states and territories of Australia, all soil types present in the Australian Soil Classification (Isbell, 2002) and all land-use classes (DAFF, 2010) (Table 1 second column). Figure 1 shows the spatial distribution of the data.

Materials and methods

Harmonizing the data and calculating the 0–30 cm stock of carbon

Determining values of soil properties from bulk horizon data, which are often sparse, can produce inaccuracies when prediction is needed at a specific depth within a soil profile. In our case, we needed estimates of the organic C content for the 0–30 cm layer, but as above, our data originated from various sources on soil sampled at various depths. The soil from the SCaRP was sampled from within the three depth layers 0–10, 10–20, and 20–30 cm, and measurements of organic C, BD and gravel were made to represent the soil there. Soil from the geochemical survey from which the spectroscopic estimates of organic C content and BD were made came from within two depth layers, 0–10 and 60–80 cm, and data from ASRIS had at each sampling site, measurements from within the 0–30 cm but also from within two or three different depth layers. To derive estimates of the total amount of organic C integrated over the 0–30-cm layer for the modelling (below), we harmonized the data by interpolation using continuous depth functions as follows.

First, we calculated the carbon density, D_C , of each sample, i , from the three different sample sets:

$$D_{C_i} = (C_i \times D_{B,i}) \times (1 - g_i), \quad (1)$$

where C_i is the gravimetric proportion of organic C (%) in the <2-mm fraction, $D_{B,i}$ is the bulk density in g cm^{-3} , and g is the gravimetric proportion of gravel in the sample. This results in D_C in units of g cm^{-3} .

Second, we needed to estimate the total stock, S_C , for the 0–30 cm layer. For sites that had only two data points recorded from within specific depth layers, (i.e. largely the spectroscopic estimates, but also some of the ASRIS data), we fitted log–log models (Jobbágy and Jackson, 2000) to the data:

$$\log D_C = \beta \log d + \alpha, \quad (2)$$

where β and α are parameters of the model and d represents the depth. We then integrated the functions over the 0–30 cm to obtain estimates of the total amount of organic C to that depth.

For sites with three data points (i.e. largely data from the SCaRP and also some of the ASRIS data), we fitted natural cubic splines (Bartels *et al.*, 1987) to the carbon density data with depth by first generating a basis matrix with two degrees

Table 1 The number of data used in the spatial modelling, by State and Territory, Australian soil classification order and land-use class. Count_A is the total number of data, Count_{tr} is the number of data used to train the model, Count_{ts} is the number of independent test data used to assess the results

	Count_A	Count_{tr}	Count_{ts}
State or Territory			
New South Wales (NSW) & Australian Capital Territory (ACT)	1697	1224	473
Western Australia (WA)	1236	892	344
Victoria (Vic)	939	677	262
Queensland (Qld)	788	568	220
South Australia (SA)	415	299	116
Tasmania (Tas)	286	206	80
Northern Territory (NT)	227	163	64
Total	5588	4029	1559
Australian soil classification order			
Sodosol	1738	1255	483
Vertosol	829	598	231
Chromosol	576	416	160
Kandosol	538	388	150
Calcarosol	372	268	104
Tenosol	366	264	102
Kurosol	245	176	69
Dermosol	241	174	67
Hydosol	227	163	64
Ferrosol	199	143	56
Rudosol	143	103	40
Podosol	98	70	28
Organosol	13	9	4
Anthrosol	3	2	1
Total	5588	4029	1559
Land use			
Improved grazing	2608	1896	712
Cropping	1370	972	398
Grazing	736	525	211
Minimal use	632	455	177
Nature conservation	111	79	32
Irrigated cropping	84	71	13
Forestry	32	21	11
Horticulture	11	8	3
Irrigated horticulture	4	2	2
Total	5588	4029	1559

of freedom and boundary conditions that extend the range of depths in the data between 5 and 25 cm. This imposes the constraint that the function is to be linear (rather than cubic) beyond the boundary points. The coefficients of the function were then used to estimate the stock, S_C , every 1 cm from 0 to 30 cm so that we could obtain the total estimates of S_C to that depth.

Finally, our harmonized estimates of S_C were positively skewed, so for the modelling, we transformed the data to their common logarithms, i.e. $\log_{10}(S_C)$.

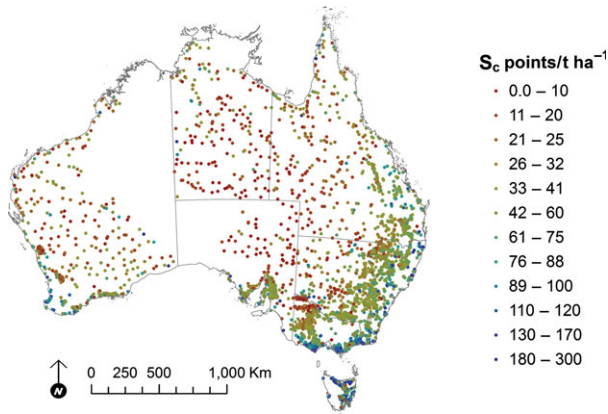


Fig. 1 Spatial distribution of the points from which data on the stock of soil organic C used in the spatial modelling.

Variogram of organic carbon

As a preliminary to the modelling, we wanted an idea of the spatial scale of variation in carbon stock across Australia. For this, we treated S_C as a stationary random process:

$$S_C(\mathbf{u}) = \mu(\mathbf{u}) + \varepsilon(\mathbf{u}) \quad (3)$$

Here $\mu(\mathbf{u})$ is a constant, the mean of the process, and $\varepsilon(\mathbf{u})$ is a spatially correlated random component with a mean of zero and variogram

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[\varepsilon(\mathbf{u}) - \varepsilon(\mathbf{u} + \mathbf{h})] = \frac{1}{2} E[\{\varepsilon(\mathbf{u}) - \varepsilon(\mathbf{u} + \mathbf{h})\}^2], \quad (4)$$

in which $\varepsilon(\mathbf{u})$ and $\varepsilon(\mathbf{u} + \mathbf{h})$ are values of the random variable at places \mathbf{u} and $\mathbf{u} + \mathbf{h}$ separated by the vector \mathbf{h} , and E denotes the expectation. We estimated and modelled a declustered variogram derived using the approach we described in Marchant *et al.* (2013) and which is an elaboration of Eqn (5) below.

We treated the variation as isotropic, so that \mathbf{h} became a scalar, h , in distance only. We fitted an isotropic double-

spherical-plus-nugget model to the resulting ordered set of $\hat{\gamma}(\mathbf{h})$ by weighted least-squares approximation using the FITNONLINEAR algorithm in GenStat (Payne, 2013).

The resulting variogram is shown in Fig. 2(a) in which the points are the individual estimates and the fitted double spherical function appears as the curve with variance parameters, nugget c_0 and correlated variances c_1 and c_2 , and ranges r_1 and r_2 . Their values are listed on the figure. Two distinct correlation ranges are evident; one of $r_1 = 73$ km, the other much longer at $r_2 = 1787$ km. We comment on them later.

Pedological inference and spatial distribution

For many years, pedologists have recognized that the soil as a whole and its individual properties are the outcomes of the climate, the biota and the landscape processes acting in concert on parent material. Some attributes of these general environmental factors are easier to observe and measure than the soil and can be used as surrogates from which to predict soil properties such as organic C. We took advantage of the experience by setting up a model in the form of a decision tree at the sites for which we had data and then using the model to predict S_C elsewhere. In this case, we assumed S_C to be a non-stationary process in which $\mu(\mathbf{u})$, Eqn (3), depends on \mathbf{u} and is a deterministic component, which could be described by our pedological model. The surrogates that we used to represent the environmental factors and their interactions in the model are listed in Table 2.

The model was the data mining algorithm CUBIST (Quinlan, 1992). CUBIST is a form of piecewise linear decision tree, which we have described in some detail elsewhere (Viscarrá Rossel and Webster, 2012). It partitions the response data into subsets within which their characteristics are similar with respect to the predictors. A series of rules derived using **if** and **else** define the partitions, and these rules are arranged in a hierarchy. A condition may be a simple one based on one predictor

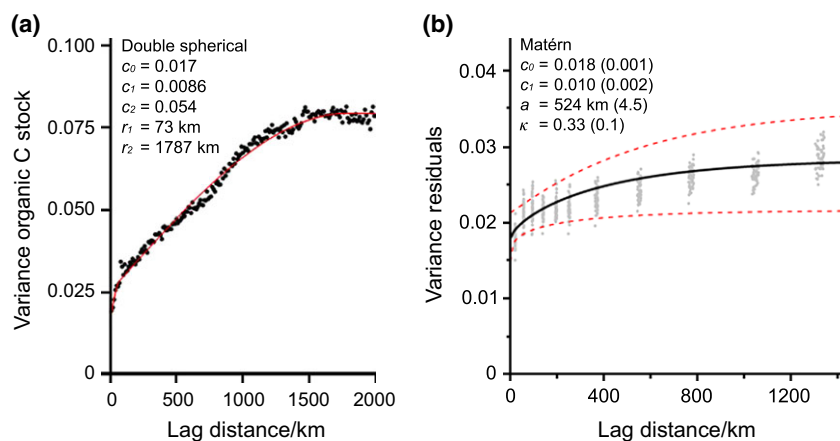


Fig. 2 Variograms of (a) the soil organic C stock data fitted with a double spherical function (red line) and (b) the set of 100 bootstrap residuals from the CUBIST model fitted with a Matérn function (black line) and twice their standard deviation (red lines). Values in parenthesis are standard deviations. The parameters of the double spherical function are the nugget c_0 and correlated variances c_1 and c_2 , and ranges r_1 and r_2 , those of the Matérn function are c_0 and c_1 , which are the nugget and correlated variances, a is the distance parameter of the model, and κ is a smoothness parameter.

Table 2 Proxies for the environmental factors used in the spatial modelling. They are likely to affect the content and distribution of carbon in Australian soil

Theme	Name of surrogate variable	Description	Environmental factor	Resolution* (m)	Source
Soil and parent material	Kaolinite, Illite, Smectite	Clay mineral abundances	soil, lithology, relief, climate	90	Viscarra Rossel (2011)
	PC1, PC2, PC3	First three principal components of vis-NIR spectra representing soil organic and mineral composition: PC1 represents soil with abundant haematite PC2 represents soil rich in organic matter PC3 represents soil with which abundant smectite, illite and goethite	soil, relief, climate	90	Viscarra Rossel and Chen (2011)
	Total dose, K, U and Th	Gamma radiometrics dose and concentrations in % and mg kg^{-1} , respectively	soil, lithology	100	Minty <i>et al.</i> (2009)
Climate	Magnetics	Magnetic anomalies in nanoTesla (nT) units	soil	90	Milligan <i>et al.</i> (2004)
	Gravity	Bouguer gravity anomaly, Gal	soil	90	GA (2009)
	Rainfall	Average annual rainfall, mm	climate, relief, biota, soil	90	Xu and Hutchinson (2011)
	Minimum temperature	Average annual minimum temperature / °C	climate, relief, biota, soil		
	Maximum temperature	Average annual maximum temperature / °C	climate, relief, biota, soil		
	Potential evapotranspiration	Mean annual evapotranspiration/mm	climate, relief, biota, soil		
	Solar radiation	Mean annual solar radiation / $\text{J m}^{-2} \text{yr}^{-1}$	climate, biota, relief, soil		
	Prescott	Prescott index	climate, relief, biota, soil	90	Prescott (1950)
	NDVI	Landsat Normalized Difference Vegetation Index (NDVI)	biota, management	30	Landsat (2010)
	Fpar-e and Fpar-r	Fraction of photosynthetically active radiation intercepted by the sunlit canopy of herbaceous and woody vegetation	biota	250	Donohue <i>et al.</i> (2009)
Terrain and landscape position	Landcover	MODIS derived landcover types and vegetation greenness	biota, management	250	Lymburner <i>et al.</i> (2011)
	DEM	SRTM 3 second digital elevation model with drainage enforcement, m	relief, climate, biota	90	Gallant <i>et al.</i> (2011)
	Slope and slope length	Rate of fall in elevation and slope lengths of 300 and 1000 m	relief, management		

Table 2 (continued)

Theme	Name of surrogate variable	Description	Environmental factor	Resolution* (m)	Source
Aspect	Aspect	Orientation of the line of steepest decent measured in degrees clockwise from north	relief, biota, climate		
Curvatures	Curvatures	Curvature of the surface in the down-slope direction (profile) and across the slope (plan)	relief		
Hill-shades	Hill-shades	The intensity of lighting on a surface given a light source at a particular location 0, 45, 90, 135	relief, climate		
Contributing area	Contributing area	Area in a drainage basin that contributes water to stream-flow	relief		
Relief	Relief	Terrain relief m	relief, biota management		
Topographic wetness index	Topographic wetness index	The propensity for a site to be saturated given its contributing area and local slope characteristics	relief, soil		
MfVBF	MfVBF	Multi-resolution Valley Bottom Flatness index	relief	90	Gallant and Dowling (2003)
Erosional surfaces	Erosional surfaces	Proportion of erosional surfaces %	relief, soil	90	Stein (2008)
Terrain roughness	Terrain roughness	Terrain roughness %	relief, soil	90	Stein (2008)

*Refers to the approximate pixel resolution of the covariates. Those not at 3 arc-sec (approximately 90 m) resolution were interpolated to that resolution in a geographic information system (GIS) using bilinear interpolation.

or, more often, it comprises several. If a condition is true, the next step, **then**, is the prediction of S_C by ordinary least-squares regression from the predictors within that partition. If the condition is false, then the rule defines the next node in the tree. The sequence **if, then, else** is repeated. The result is that the regression equations, although general in form, are local to the partitions and their errors smaller than they would otherwise be. CUBIST has been used effectively in several branches of research and for various purposes including soil mapping over large areas (e.g., Bui *et al.*, 2009; Viscarra Rossel and Chen, 2011; Viscarra Rossel, 2011).

There are inevitable discrepancies, or errors, between the predictions and the true values; these are represented by the residuals from CUBIST. They, like the data, are spatially correlated, and they too can be predicted, provided they can be modelled suitably. For this purpose, we treated the residuals, ε , as spatially correlated random variables with mean = 0 and variance defined as in Eqn (4). Values of semivariance, $\gamma(\mathbf{h})$ in that equation were estimated at a sequence of values of \mathbf{h} from the residuals to give an experimental variogram by the method of moments:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{j=1}^{m(\mathbf{h})} \{\varepsilon(\mathbf{u}_j) - \varepsilon(\mathbf{u}_j + \mathbf{h})\}^2, \quad (5)$$

in which $\varepsilon(\mathbf{u}_j)$ and $\varepsilon(\mathbf{u}_j + \mathbf{h})$ are the residuals at positions \mathbf{u}_j and $\mathbf{u}_j + \mathbf{h}$ and $m(\mathbf{h})$ is the number of comparisons contributing to the estimate at lag \mathbf{h} . This variogram was then modelled, and the fitted functions used for ordinary kriging to predict the residuals. To derive the final estimates of S_C , as in Eqn (3), the predictions from CUBIST and the kriging estimates are summed. We call these the CUBIST-kriging (CK) estimates of S_C .

Bootstrapping the spatial model

A shortcoming of the above approach is that the prediction variances are underestimated (Webster and Oliver, 2007); one cannot reliably use the kriging variances as measures of uncertainty. An alternative is to use the non-parametric bootstrap (Efron and Tibshirani, 1993) to provide independent sets of residuals on which to compute and model the random component, Eqn (3), and so prevent underestimation of the prediction variances.

We used the bootstrap also to assess the uncertainties in both the deterministic and random components of the spatial model. By repeated sampling with the bootstrap, and performing the spatial modelling on each bootstrap sample, one obtains probability distributions of the outcomes from the modelling. Robust estimates can be derived by averaging the bootstrap samples, and the uncertainty of the modelling can be quantified by computing confidence limits on the estimates, as we describe below. Viscarra Rossel *et al.* (2013) provide details on a similar approach.

We took 100 bootstrap samples of S_C and associated predictors (Table 2), and for each bootstrap, which we denote b , we implemented CUBIST. For each of the 100 bootstrap samples, the CUBIST model was then used to predict the

values at the sampling points not included in the bootstrap, i.e. the out-of-bag samples. The number of data in the out-of-bag samples was roughly one third of the original data, i.e. approximately 1500, which is ample for estimating a variogram (Webster and Oliver, 2007). The differences between the predictions and the observed values at these points, excluded at random from the bootstrap, provide a set of residuals that are assumed to be independent and on which we computed variograms by the usual method of moments, Eqn (5), and treating the variation as isotropic. We fitted Matérn models (Webster and Oliver, 2007) to each of the series and used the functions for ordinary punctual kriging (e.g., Goovaerts, 1997; Webster and Oliver, 2007) using the nearest 20 to 90 data points. The resulting 100 variograms of the residuals, and their uncertainty are summarized in Fig. 2b. We comment on the figure below. For each b we then added the CUBIST predictions to those from kriging to derive the CK estimates of S_C , as in Eqn (3).

Model training and validation

We selected at random two-thirds of the data from which to train the model and used the remaining third to test it. The States and Territories of Australia, the Australian soil classification orders and land-use classes were represented in both data sets (Table 1, third and fourth columns). The spatial modelling was done with the training set, and we assessed it using a 10-fold cross-validation and the bootstrap out-of-bag samples. Both the CUBIST and the CK estimates were validated independently from the modelling by comparison of their predictions of organic C from each b th bootstrap sample with the values of organic C in the test set of data. Thus, we could quantify the improvements in the modelling by CK compared with our use of CUBIST alone. The assessment statistics that we used were the concordance correlation coefficient, ρ_c (Lin, 1989) to assess covariation and correspondence between our predictions and the original data, the root mean squared error (RMSE) of the predictions to quantify their inaccuracy, the standard deviation of the error (SDE) to quantify their imprecision and the mean error (ME) their bias. We note that the inaccuracy embraces both the bias and imprecision, so that $\text{RMSE}^2 = \text{ME}^2 + \text{SDE}^2$ and that ρ_c combines measures of both precision and bias to determine how far the observed data deviate from the line of perfect concordance, which is the 1:1 line. ρ_c ranges from -1 to $+1$. A value of $+1$ denotes perfect agreement, values >0.9 suggest near perfect agreement, values between 0.8 and 0.9 substantial agreement, between 0.65 and 0.8 moderate agreement and values <0.65 poor agreement. The 100 bootstraps enabled us to also derive distributions for these validations and their statistics, which we report in a Table (Table 4).

Mapping the organic C stock and its uncertainty

For each bootstrap sample, b , the CUBIST model was used to predict values of S_C at the nodes, \mathbf{u}_0 , of the 3-arc-sec grid. We denote these predictions $\hat{\mu}_C^b(\mathbf{u}_0)$. Note that to map at this

resolution, the predictors described in Table 2 were all resampled to a 3-arc-sec resolution, which is the same as the Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM).

To improve on these predictions, any residual variance not accounted for by CUBIST was modelled geostatistically, as above, and the parameters of the variograms were used together with the residuals to krig values of the residuals at the same nodes of the 3-arc-sec grid. We denote these predictions $\hat{\varepsilon}^b(\mathbf{u}_0)$. At each node, the two values were added to give our final CK predictions of the organic C stock, $\hat{S}_C^b(\mathbf{u}_0)$, for the b th bootstrap sample:

$$\hat{S}_C^b(\mathbf{u}_0) = \hat{\mu}_C^b(\mathbf{u}_0) + \hat{\varepsilon}^b(\mathbf{u}_0). \quad (6)$$

As this sequence was repeated over all 100 bootstrap samples, the outcomes were probability distributions of the predictions at each and every grid node. We averaged the estimates of $\hat{S}_C^b(\mathbf{u}_0)$ from the $B = 100$ bootstrap samples to obtain our map of the most likely and robust estimates of the stock of organic C:

$$\hat{S}_C^m(\mathbf{u}_0) = \frac{1}{B} \sum_{b=1}^B \hat{S}_C^b(\mathbf{u}_0). \quad (7)$$

The overall uncertainty, from the 100 bootstraps, was calculated by summation of the variance in the CK estimates and the average kriging variances of the residuals, $\bar{\sigma}_{OK}^2$:

$$\hat{V}(\mathbf{u}_0) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B \left\{ \hat{S}_C^b(\mathbf{u}_0) - \hat{S}_C^m(\mathbf{u}_0) \right\}^2 + \bar{\sigma}_{OK}^2}. \quad (8)$$

We back-transformed $\hat{S}_C^m(\mathbf{u}_0)$, which was on the \log_{10} scale, to the original scale by

$$\exp \left\{ \ln(10) \times \hat{S}_C^m(\mathbf{u}_0) + \ln(10) \times 0.5 \text{var} \left[\hat{S}_C^m(\mathbf{u}_0) \right] \right\} \quad (9)$$

We used $\hat{V}(\mathbf{u}_0)$ to compute 95% confidence intervals, on the logarithmic scale and back-transformed them by Cox's method, which is described in Zhou and Gao (1997):

$$\exp \left\{ \ln(10) \times \left(\hat{S}_C^m(\mathbf{u}_0) + \ln(10) \times 0.5 \hat{V}^2(\mathbf{u}_0) \right) \pm \xi_{1-\alpha/2} \sqrt{\frac{\hat{V}^2(\mathbf{u}_0)}{B} + \frac{\hat{V}^4(\mathbf{u}_0)}{2(B-1)}} \right\}, \quad (10)$$

where ξ is the standard normal deviate for the chosen probability $\alpha = 0.05$. We expressed the uncertainty of our estimates in standardized form as the range of the 95% confidence intervals divided by their mean, $\hat{S}_C^m(\mathbf{u}_0)$.

CUBIST also provided estimates of the frequency of use of the predictors in the conditions and regressions of each rule set. We mapped the rule sets to assess their spatial pattern and gain insights into the soil and environmental factors that characterize the spatial distribution of the organic C stock in the soil of the continent.

We compared this map of the CUBIST rule sets with the Regional Carbon Cycle Assessment and Process (RECCAP) bioclimatic classification of Australia based on Hutchinson *et al.* (2005).

The mean and total organic carbon stock of Australia

The final step in the procedure was to calculate the mean and total stock of organic C in the soil of Australia as a whole and in the individual States and Territories, together with uncertainties on those estimates. These were computed from the back-transformed estimates $\hat{S}_C^m(\mathbf{u}_0)$ and their upper and lower 95% confidence intervals. Each mean was calculated as the average of the predictions at the grid nodes within its area, and the total stock was computed as the sum:

$$S_C^T = \sum_{i=1}^N \hat{S}_C^m(\mathbf{u}_0) \times A, \quad (11)$$

where A is the area of Australia or that of a state or territory and N is the total number of pixels in the area.

To interpret our results and to evaluate the estimates, we also intersected spatial data not used in the spatial modelling with the back-transformed estimates, $\hat{S}_C^m(\mathbf{u}_0)$ and confidence intervals. The data that we used include a map of land use in the years 2005–2006 (DAFF, 2010), a map of native vegetation classes (DEH, 2006) and RECCAP zones. As above, we computed the mean and total stock, Eqn (11), for the classes in each map, and we present the data in Tables and graphs.

Results

Table 3 summarizes the statistical distribution of the data. The distribution of the stock of C, S_C , was positively skewed. Its mean was 49 t ha⁻¹, its median 40 t ha⁻¹ and its range was from 0.3 to 300 t ha⁻¹ (Table 3). Ninety per cent of the S_C values were smaller than 94 t ha⁻¹. Table 3 also lists statistics for measurements of organic C and BD, which were used to derive S_C in the 0–30 cm layer.

We have already drawn attention to the variogram of S_C across the continent (Fig. 2a) with its two distinct structures. The first, with its range of 73 km represents variation on the regional scale in eastern and coastal Australia. The second, with an estimated range of 1787 km, characterizes the variation across large areas in the centre and west of Australia.

Figure 2b shows the experimental variograms of the bootstrap samples with twice the standard deviation of the fitted model in red. All could be fitted with single Matérn functions plus nugget variances. Comparison of the two variograms in Fig. 2 shows that CUBIST has taken into account on average more than 84% of the spatially correlated variance: the quantity c_1 of the residuals is 0.010 compared with $c_1 + c_2 = 0.0626$ for the raw data. The nugget variance is almost unchanged at 0.018. This component includes measurement error, which of course remains irrespective of the statistical analysis, plus variation over much shorter distances than those between observation points. Two spatial structures are no longer distinguishable; the change of

Table 3 Summary statistics for the data used in the modelling with statistics for all the data, those used to train the model and those used to test the predictions. The variables listed are the content of organic C in the soil (C) and bulk density (D_B) used to derive the carbon densities from which the stock of organic C, S_C , for the 0–30-cm layer was calculated

	Mean	SD	Minimum	10%	1st Quartile	Median	3rd Quartile	90%	Maximum	Skew
All data $N = 5588$										
$C / \%$	1.28	1.01	0.01	0.36	0.62	0.98	1.61	2.54	8.53	2.12
$D_B / \text{g cm}^{-3}$	1.43	0.17	0.59	1.2	1.33	1.44	1.54	1.62	1.99	-0.76
$S_C / \text{t ha}^{-1}$	49.25	33.55	0.33	16.37	26.26	40.09	63.44	94.47	299.58	1.62
Training $N = 4029$										
$C / \%$	1.29	1.02	0.01	0.37	0.63	0.98	1.62	2.57	8.53	2.15
$D_B / \text{g cm}^{-3}$	1.42	0.18	0.59	1.2	1.33	1.44	1.54	1.62	1.99	-0.75
$S_C / \text{t ha}^{-1}$	49.77	34.05	0.33	16.49	26.42	40.29	64.33	95.93	299.58	1.63
Test $N = 1559$										
$C / \%$	1.24	0.98	0.07	0.35	0.6	0.96	1.57	2.45	7.99	1.98
$D_B / \text{g cm}^{-3}$	1.43	0.17	0.6	1.21	1.34	1.45	1.54	1.62	1.95	-0.77
$S_C / \text{t ha}^{-1}$	48.4	33.01	0.5	16.2	25.86	39.81	62.82	2.76	288.16	1.66

Table 4. Cross-, out-of-bag (OOB) and independent test set validation statistics for the spatial model of S_C . Assessment with the concordance correlation coefficient (ρ_c) and the root mean square error (RMSE), the mean error (ME) and the standard deviation of the error (SDE). The latter three are in $\log_{10}(S_C)/\%$ units. Note that the RMSE embraces both the ME and SDE, such that $\text{RMSE}^2 = \text{ME}^2 + \text{SDE}^2$

	Mean	SD	Minimum	1st quartile	Median	3rd quartile	Maximum
Cross validation							
ρ_c	0.784	0.032	0.749	0.773	0.789	0.796	0.803
RMSE	0.175	0.009	0.165	0.169	0.172	0.181	0.195
OOB validation							
ρ_c	0.803	0.007	0.782	0.782	0.803	0.807	0.818
RMSE	0.165	0.003	0.158	0.163	0.165	0.167	0.173
Test validation							
ρ_c	0.812	0.004	0.802	0.810	0.813	0.815	0.821
RMSE	0.165	0.001	0.162	0.164	0.166	0.166	0.168
ME	-0.001	0.004	-0.009	-0.003	-0.001	0.002	0.007
SDE	0.165	0.002	0.162	0.164	0.165	0.166	0.168

slope at around 73 m evident in the model fitted to the variogram of the raw data (Fig. 2a) is replaced by the smoother curve (Fig. 2b), with a somewhat shorter effective range of 1316 km ($\approx 2.5 \times a$).

Spatial modelling and mapping of carbon stock and its uncertainty

The validation statistics of the CUBIST model calculated from 10-fold cross- and out-of-bag validations suggest good predictability, with values of ρ_c ranging from 0.75 to 0.82 (Table 4). The validation of the CK model on the independent set of test data ($\rho_c = 0.812$) was better than that of CUBIST alone ($\rho_c = 0.759$). The validation statistics of the CK model are shown in Table 4. They are similar to the internal validations above, with values of ρ_c ranging between 0.8 and 0.82. Evidently, the model and its predictions were robust. In the validations of

the test set, the primary contribution to the RMSE was from the SDE and not the ME, as our predictions were almost unbiased (Table 4).

Figure 3a is a map of carbon stock, S_C . Values on it range from around 5.9 t ha⁻¹ in the centre of Australia, and increase gradually towards the coast in the north, south-west and east, to around 230 t ha⁻¹ in the high-temperate regions in south-eastern Australia and in the cool, wet regions of western Tasmania (Fig. 3a). The fine-resolution maps show detailed expressions of the multi-scale spatial variation of S_C across Australia (Fig. 3a) and provide estimates that might also be used to derive baselines for regions and catchments.

The uncertainties of the estimates, expressed as the standardized range of the 95% confidence intervals (Fig. 3b) were generally small where observations were dense—in Australia's agricultural regions (Fig. 1). The uncertainties were small too where the soil and

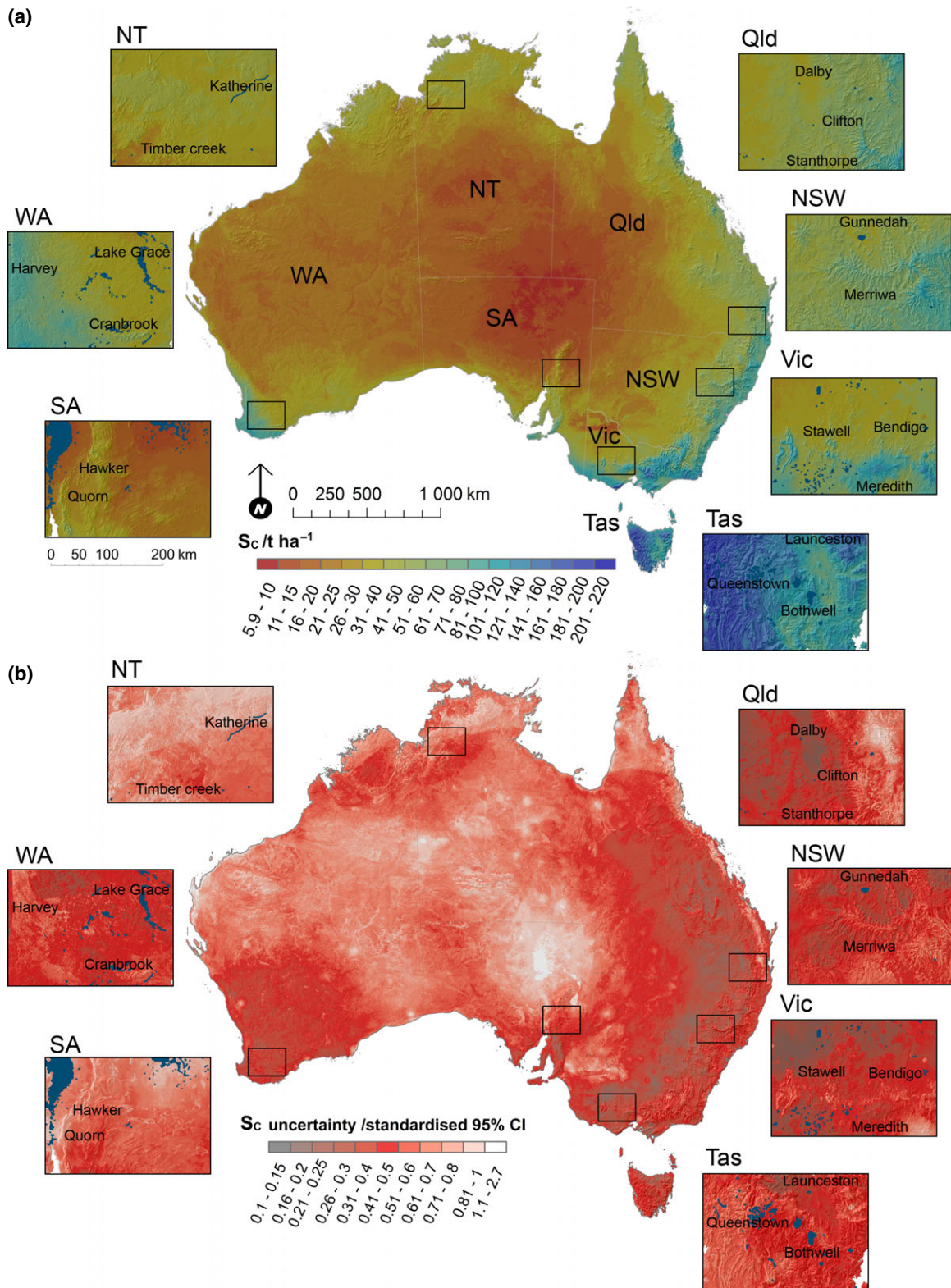
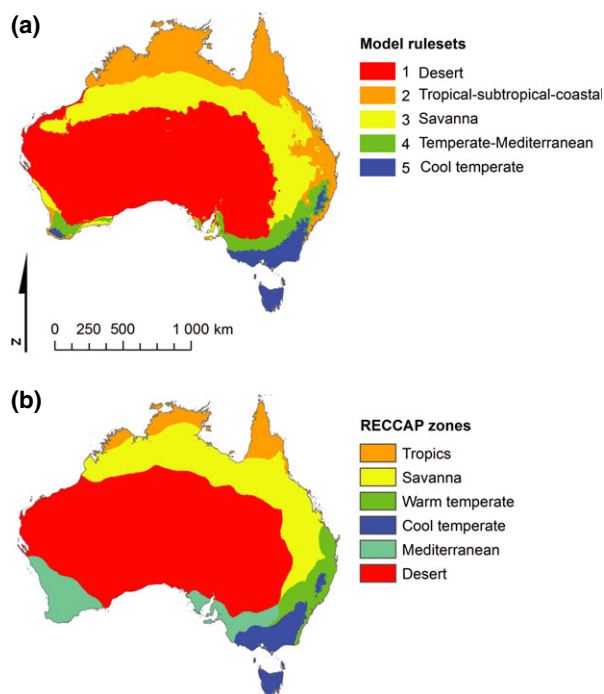


Fig. 3 Maps of (a) the Australian soil organic C stock and (b) its uncertainty expressed in standardized form as the range of the 95% confidence intervals divided by their mean. The insets are examples of the estimates for each State or Territory showing the multi-scale detail achieved by mapping at 90 m. They are the Northern Territory (NT), Western Australia (WA), South Australia (SA), Queensland (Qld), New South Wales (NSW), Victoria (Vic) and Tasmania (Tas).

Table 5 Rule sets of the CUBIST model showing the proxies for the environmental factors (Table 2) that it used in the conditions and in the linear models of the rules. Values in parentheses are the proportions of the predictors used

Rules	Bioclimatic zone	Conditions (top 3 or 100% usage)	Linear models (Top 10 or 100% usage)
1–7	Desert	Rainfall (67%), PC3 (67%), DEM (67%)	Rainfall (100%), Fpar-e (100%), Kaolinite (100%), Prescott (83%), PC3 (83%), max. temp. (67%), min. temp. (67%), PC1 (50%), Relief (33%), Aspect (33%)
13	Tropical, sub-tropical, coastal	Rainfall (100%), min. temp. (100%), Prescott (100%)	PC1 (100%), Smectite (100%), DEM (100%), max. temp. (100%), Solar radiation (100%), Fpar-e (100%), Fpar-r (100%), min. temp. (100%), PC3 (100%), Gravity (100%)
4, 8	Savanna	Rainfall (100%), Prescott (100%), min. temp. (50%)	Smectite (100%), Prescott (100%), max. temp. (100%), Fpar-e (100%), Fpar-r (100%), Rainfall (50%), PC3 (50%), PET (50%), Slope (50%), min. temp. (50%)
9–11	Temperate mediterranean	PET (100%), min. temp. (100%), PC3 (67%)	Kaolinite (100%), DEM (100%), PC3 (100%), PET (100%), max. temp. (100%), Solar radiation (100%), DEM (100%), PC2 (100%), Rainfall (100%), gamma K (100%), PC1 (100%)
12, 14	Cool temperate	Solar radiation (100%), PET (50%)	Kaolinite (100%), PC1 (100%), PC3 (100%), Rainfall (100%), min. temp. (100%), DEM (100%), gamma K (100%), Fpar-e (100%), PC2 (100%), gamma K (100%), gravity (100%)

**Fig. 4** Maps of (a) the model-derived bioclimatic zones and (b) those from the Regional Carbon Cycle Assessment and Process (RECCAP).

environment were well represented by the characteristics of the predictors. In contrast, uncertainties were larger where data were sparser or lacking and where the soil and environmental conditions were poorly represented by the characteristics of the predictors. In future, these rangeland, savanna and desert regions

(Fig. 3b) should be sampled more densely to improve the certainty of predictions and the national carbon accounts.

The CUBIST model used 14 rules to characterize the spatial variation in S_C across Australia. To help with interpretation, we grouped these rules into spatially coherent sets that represent model-derived bioclimatic zones. Their descriptions and the predictors used in the conditions and in the linear models of the rules are given in Table 5. The resulting map of the rules is shown in Fig. 4a.

The amount of organic C in soil and its spatial distribution depend on the environment—its climate, biota, landscape, lithology, soil type, as well as the type of land use and management, all of which interact in time to affect rates of addition and decomposition of C. The conditions in the rules discretized the continent using predictors that account for pedogenetic processes that determine the distribution of organic C in Australian soil. In the linear models, predictors capture regional and smaller scale variations in stock across the landscape.

As might be expected, and found in previous investigations (Wynn *et al.*, 2006; Bui *et al.*, 2009), predictors that represent climate are most frequently used and provide the greatest contribution to the conditions of the rule sets (Table 5). Climate influences the rate of mineralization of organic C in the soil to CO_2 as well as the productivity of vegetation and therefore addition of organic C to the soil. Low temperatures and waterlogging inhibit decomposition and mineralization. The result is that carbon tends to accumulate, and soil in such environments contains much organic C. Soil with

abundant smectite and illite, some goethite and organic matter, represented by the third principal component of Australian visible–near infrared spectra (PC3) (Viscarrá Rossel & Chen, 2011), and the DEM were also used in the conditions that characterize deserts and temperate Mediterranean zones (Table 5). As in the conditions, predictors that represent climate are also frequently used in the linear models. However, other predictors play a role. They include the type of vegetation, which determines to some extent the rate at which organic C is added to the soil; the type of clay and parent material, which determine, also to some extent, the capacity of soil to support plant growth and the ability to protect organic C against decomposition; and the form of the terrain through its effect on drainage (Tables 5 and 2).

We compared our map of the model-derived bioclimatic zones (Fig. 4a) with the RECCAP bioclimatic zones (Fig. 4b) and found good agreement between them. However, the bioregional classification that the model produced to predict soil organic C provides more detailed zoning in south western Western Australia. Elsewhere, like the RECCAP zones, the model-derived bioregional classification reflects the seasonality of climate, patterns in plant growth and composition and prime features of the landscape. They, in turn, reflect the different land uses across the continent.

Baselines of the soil organic C stock in Australia and its states and territories

We estimate that the average stock of organic C in Australian soil in the 0–30 cm layer is 29.7 t ha^{-1} . The uncertainty around this value, calculated by 95% confidence intervals, ranges between 22.6 and 37.9 t ha^{-1} . By aggregating the spatial estimates over the continent we calculate, using Eqn (11), that the total stock of

organic C in Australian soil to a depth of 30 cm is 24.98 Gt. The uncertainty around our estimate ranges between 19.04 and 31.83 Gt.

Table 6 lists our estimates of the mean and total stock of organic C for the Australian states and territories, and their uncertainties. The average stock of the 0–30 cm soil layer in the states and territories follows roughly a mean annual temperature gradient and are listed in decreasing order in Table 6. States and territories in southern Australia where there is more than average rain, and cooler than average mean annual temperatures and at high elevations, where conditions are conducive to more biomass production and slower decomposition, on average, have larger stocks of organic C than those further north with Mediterranean, sub tropical, tropical climates (Fig. 3a; Table 6). The exception is South Australia where the soil contains the least organic C (Table 6), largely because it spans large extents of desert. The total stock of organic C stock in the 0–30-cm layer of soil in the states and territories, however, does not follow the temperature gradient. Instead, it is well correlated with their total areas, with Western Australia having the largest area and total stock, and the Australian Capital Territory and the Jervis Bay Territory the smallest of both (Table 6).

The soil organic carbon stock in bioclimatic regions

Estimates of the mean and total stock for the RECCAP and CUBIST bioclimatic zones (Fig. 4) and their uncertainties are shown in Fig. 5a and b. The largest contents of organic C occur in the cool, temperate and Mediterranean bioclimatic zones (Fig. 5a and b, also see Figs 3a and 4a) with extensive rain forests and many types of eucalypt forests. There, the environment favours vigorous vegetation growth and slow decomposition of organic matter (dry hot summers, cool

Table 6 Estimates of the stocks of soil organic C of Australia and its States and Territories and their uncertainties expressed as 95% confidence intervals

States and Territories	Mean S_C (t ha^{-1})	Lower 95% CI	Upper 95% CI	Total S_C (Gt)	Lower 95% CI	Upper 95% CI	Area (km^2)
Tasmania	133.99	108.44	162.40	1.048	0.848	1.270	64 519
Jervis Bay Territory	95.59	75.56	118.02	0.000 67	0.000 53	0.000 83	72.0
Victoria	66.69	54.66	80.03	1.684	1.381	2.022	227 010
Australian Capital Territory	62.29	48.76	77.55	0.01623	0.01271	0.02021	2358
New South Wales	42.40	34.55	51.12	3.701	3.016	4.462	800 628
Queensland	31.15	24.33	38.92	5.883	4.595	7.350	1 723 936
Western Australia	25.77	18.99	33.66	7.087	5.222	9.259	2 526 786
Northern Territory	22.61	15.85	30.61	3.364	2.358	4.554	1 335 742
South Australia	20.32	14.86	26.76	2.171	1.587	2.858	978 810
Australia	29.712	22.65	37.86	24.977	19.038	31.826	7 659 861

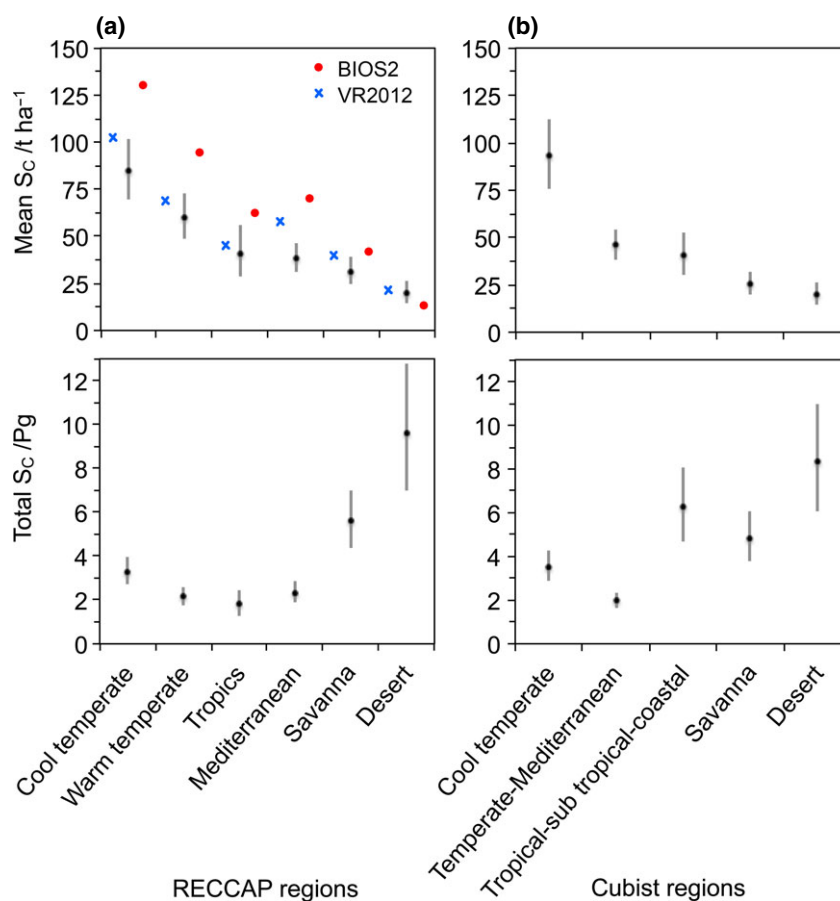


Fig. 5 Soil organic C stocks of (a) the Regional Carbon Cycle Assessment and Process (RECCAP) and (b) the model-derived regions, and their uncertainty expressed as 95% confidence intervals. We provide a comparison to estimates made by Haverd *et al.* (2013) who use an exponential organic C profile to estimate the stock for the 0–10 cm layer. Thus, to convert the 0–10 cm estimate of the stock to the stock in the 0–30 cm layer, we multiplied by $\{\exp(-30k)-1\} / \{\exp(-10k)-1\}$, with $k = 0.0101$.

winter temperatures and abundant rain in the more temperate regions) coincide.

The amounts of carbon per unit area are small in the savanna, desert and northern tropical and subtropical coastal zones where the dominant vegetation is open woodland, acacia shrub- and wood-land, tussock grasslands and arid spinifex grasslands (Fig. 3 and Fig. 5a and b). Nevertheless, because these types of vegetation cover such large areas of Australia, they have largest total stocks of organic C.

The soil organic carbon stock of vegetation groups

The soil of tall and short open eucalypt forests, rain forests, short open forests, heath-lands, open eucalypt forests and tall dense thickets have mean contents of organic C exceeding 50 t ha^{-1} (Table 7). All other major types of vegetation contain less. The mean stock of organic C (and its uncertainty in parenthesis) in the soil under tall open eucalypt forests (with dominant trees

>30 m tall) is 110 t ha^{-1} ($88\text{--}135 \text{ t ha}^{-1}$), which is more than that under any other type of vegetation (Table 7). These are cool temperate evergreen forests, dominated by *Eucalyptus*, *Corymbia* and *Angophora* genera, that have a secondary stratum of rainforest species and often form landscape mosaics with rain forests (DEH, 2006). These forests are home to the world's tallest flowering plant, *Eucalyptus regnans*, and are among the world's most C-dense, with average living above-ground carbon content of 1053 t ha^{-1} (Keith *et al.*, 2009). The rain forests have the next largest mean carbon content of 92 t ha^{-1} ($68\text{--}120 \text{ t ha}^{-1}$) (Table 7). Open eucalypt forests contain somewhat less carbon with a mean of 69 t ha^{-1} ($54\text{--}86 \text{ t ha}^{-1}$), but they occupy the largest area of all forest types and therefore their soil contains the largest total stock of organic C (Table 7).

Although the soil under eucalypt woodlands and arid spinifex grasslands contain small proportions of organic C, they store more organic C than the soil

Table 7 Estimates of the stocks of soil organic C in major vegetation groups and their uncertainties expressed as 95% confidence intervals. Values are derived from the National Vegetation Information System (NVIS)

Major vegetation	Mean S_C (t ha ⁻¹)	Lower 95% CI	Upper 95% CI	Total S_C (Gt)	Lower 95% CI	Upper 95% CI	Area / km ²
Eucalypt tall open forests (>30 m)	109.88	87.75	134.61	0.390	0.311	0.477	35 467
Rainforests	91.94	67.96	119.90	0.324	0.240	0.423	35 283
Eucalypt low open forests (<10 m)	90.06	72.23	109.97	0.036	0.029	0.045	4049
Heathlands	69.82	55.65	85.65	0.056	0.044	0.068	7969
Eucalypt open forests (10–30 m)	68.65	53.74	85.56	1.866	1.461	2.326	271 882
Tall dense thickets	53.08	41.88	65.67	0.085	0.067	0.105	16 056
Mangroves	44.91	31.33	61.08	0.036	0.025	0.049	8077
Regrowth, modified native vegetation	40.05	33.48	47.28	0.114	0.095	0.134	28 429
Swampy grasses and sedges	39.33	29.97	50.06	0.252	0.192	0.321	64 187
Tropical eucalypt woodlands/grasslands	36.84	26.03	49.60	0.423	0.299	0.569	114 763
Eucalypt woodlands	36.59	27.95	46.51	3.258	2.488	4.140	890 181
Callitris forests and woodlands	34.67	28.43	41.57	0.110	0.090	0.132	31 738
Other forests and woodlands	31.34	24.10	39.58	0.225	0.173	0.285	71 903
Melaleuca forests and woodlands	30.68	22.57	40.09	0.307	0.226	0.401	100 136
Mallee woodlands and shrublands	28.38	21.47	36.30	0.757	0.573	0.969	266 956
Other shrublands	27.39	20.74	35.05	0.332	0.251	0.424	121 030
Eucalypt open woodlands	25.40	18.97	32.86	1.154	0.862	1.493	454 395
Acacia forests and woodlands	23.77	18.17	30.19	0.951	0.727	1.208	400 013
Casuarina forests and woodlands	21.84	16.20	28.37	0.319	0.237	0.415	146 180
Tussock grasslands	21.06	16.13	26.73	1.092	0.837	1.386	518 556
Acacia shrublands	20.59	14.87	27.29	1.714	1.238	2.272	832 510
Arid spinifex grasslands	19.46	13.40	26.70	2.616	1.800	3.589	1 344 034
Acacia open woodlands	18.45	13.60	24.06	0.566	0.418	0.739	306 972
Salt bushes and salt marshes	18.23	13.26	24.04	0.780	0.567	1.029	427 807

under other major vegetation groups simply because they are the two most extensive, covering 890 181 and 1 344 034 km² respectively (Table 7). Acacia shrub-lands occupy almost the same area as eucalypt woodlands, but the soil under them holds only half as much organic C (Table 7). Soil under chenopod shrub-lands (salt-bushes and marshes), with small above-ground biomass and organic C content, has a larger total stock than does that under tall eucalypt open forests and rain forests combined; these types

of vegetation occupy an area of 427 807 km², which is approximately six times larger than the combined area of the two forest types.

The soil organic C stock in types of land use

Table 8 lists our estimates of the current, average and total stock of organic C for types of land use, and their uncertainties, arranged in decreasing order of their mean stock. It shows that areas with natural vegetation,

Table 8 Estimates of the stocks of soil organic C by land use and their uncertainties expressed as 95% confidence intervals

Land use	Mean S_C (t ha ⁻¹)	Lower 95% CI	Upper 95% CI	Total S_C (Gt)	Lower 95% CI	Upper 95% CI	Area (km ²)
Nature conservation	83.25	67.41	100.88	1.148	0.930	1.391	137 729
Horticulture	67.14	53.66	82.31	0.022	0.018	0.027	3190
Irrigated horticulture	64.67	50.95	80.18	0.0079	0.0062	0.0098	1187
Forestry	56.86	44.98	70.33	0.029	0.023	0.036	4984
Improved grazing	45.88	38.27	54.26	3.246	2.707	3.839	707 006
Irrigated cropping	44.32	36.80	52.64	0.059	0.049	0.070	13 306
Cropping	35.36	29.58	41.73	0.897	0.750	1.058	253 186
Minimal use	28.98	21.26	37.99	8.204	6.012	10.763	2 860 605
Grazing	24.35	18.28	31.37	8.528	6.402	10.988	3 678 668
Agriculture total				12.760	9.931	15.991	4 656 543

Table 9 Comparison of our estimates of the total stock of organic C in Australian soil with others found in the literature

Source	Depth (cm)	Estimate (Gt)	Lower 95% CI	Upper 95% CI
Gifford <i>et al.</i> (1992)	rooting depth	51.8		
Barrett (2002)	0–20	20.0		
Barrett (2013)*	0–100	26.9	20.3	33.5
Grace <i>et al.</i> (2006)	0–30	18.8		
	0–100	34.2		
Our estimate	0–30	24.97	19.04	31.83

*We approximated the values of the lower and upper 95% CIs as twice the standard deviation of the estimates provided by Barrett (2013).

those that are under horticulture and forestry have the largest average organic C stock. Their total stocks, however, are small and reflect the small areas in which they are present (Table 8).

Improved pastures for grazing and irrigated land have the next largest, and similar average carbon stocks. The total stock of C in the soil under improved pastures is larger than that in soil under irrigation, because its area is larger (Table 8). Soil used for crop production contains on average 35 t ha⁻¹ organic C, (30–42 t ha⁻¹ 95% confidence intervals), while the total carbon stock is 0.90 Gt (0.75–1.1 Gt).

Little-used land (including land with residual native cover, deserts, areas of rehabilitation and that for traditional indigenous use) and grazing on native vegetation hold the smallest average amounts of organic C, with 29 and 24 t ha⁻¹. Nevertheless, these land-use types hold the largest total stocks with 8.2 and 8.5 Gt, because they cover such a large area (Table 8).

Agricultural land, including the large areas of land used for grazing on native vegetation, occurs over approximately 61% of all land in Australia and holds

around 51% of the total soil organic C stock. The total soil organic C stock of agricultural land is 12.76 Gt ha⁻¹ with 95% confidence intervals of 9.93 and 15.99 Gt ha⁻¹ (Table 8).

Discussion

We mapped the stock of organic C in the 0–30 cm layer of soil across Australia at 90 m pixel resolution from fairly sparse data with little bias and with estimates of uncertainty. The data that we used represent all states and territories of Australia, all soil types and all land-use classes (Table 1), and our spatial modelling is based on harmonized data and required few assumptions. Therefore, we believe that our estimates are more reliable than those reported previously in the literature, which are derived from sparse data or from simulation.

Our estimates are somewhat different from those previously reported. Most other estimates were for a 1990 baseline, and most were derived by simulation modelling. They are described below and where possible, summarized in Table 9. Gifford *et al.* (1992) derived estimates of the stock of organic C in the root zone of Australian ecosystems using estimates of the total organic C in live vegetation and assumed ratios of soil C to plant C. They derived the estimates of organic C in live vegetation from a global assessment by Olson *et al.* (1985). In that study, the Earth's surface was divided into 50 m × 50 m cells each of which was heuristically coded with typical values for carbon in live vegetation and net annual primary production as recorded in the literature.

Using a patchwork of surveys of various scales from the states and territories, and sparse and largely biased data, AGO (2002) derived a map of soil organic C for the 0–30 cm layer before the land was cleared. Its result provides only an approximate range in stock, suggesting that there could be less than 10 t ha⁻¹ of organic C

in arid, desert areas and possibly more than 250 t ha⁻¹ in highland areas of southeastern Australia and in Tasmania.

With the Vegetation and Soil-carbon Transfer (VAST) model, Barrett (2002, 2013) estimated the organic C stock of Australian soil in the 0–20 and 0–100 cm layers to be 20 and 26.9 Gt respectively. The 0–100 cm estimates were provided with a standard deviation of 3.3 Gt (Barrett, 2013). Barrett used 341 measurements of soil C content and 50 measurements of soil bulk density, all within the 0–15 cm layer.

Using the SOCRATES terrestrial carbon model, Grace *et al.* (2006) estimated the 1990 baseline stock of organic C in Australian soil in the 0–30 cm layer to be 18.8 Gt. Their simulations used a biogeographical regionalization of Australia, information on soil type and texture and an assumed BD of 1.3 g cm⁻³. Their estimate for the 0–30 cm layer is significantly different from ours as it falls outside our 95% confidence interval (Table 9).

Haverd *et al.* (2013) estimated the stock of organic C for Australia using a modified version of the CSIRO Atmosphere and Biosphere Land Exchange (CABLE) land surface model implemented in BIOS2. They compared their estimates with a map of stock for the 0–10-cm layer derived by the lead author here (which we refer to as the VR2012 map), using data from ASRIS recorded from 1950 to 2010 and spectroscopic estimates of organic C on samples collected for the NGSa (see above).

Their spatial estimates produced with BIOS2 show spatial patterns similar to those of the VR2012 map and ours here [compare Fig. 3a to 10 in Haverd *et al.* (2013)], but with some apparent discontinuities in their map. Both the BIOS2 and VR2012 estimates are generally larger than ours here. A possible reason for the larger estimates in the VR2012 map is that that map was produced with data recorded during six decades (1950–2010) and not from the recent data for the SCaRP.

Haverd *et al.* (2013) compare their estimates of BIOS2 and those of VR2012 with the mean organic C stock for the RECCAP zones. The trends in those estimates are fairly similar to ours; all reveal there to be more carbon in the cool temperate zone than in the warm temperate, the tropics, the Mediterranean, savannas and the desert (Fig. 5a). However, all of the BIOS2 estimates lie outside of the 95% confidence limits of our estimates, and all except for those of the desert zone are significantly larger than ours (Fig. 5a). Their estimates for the desert zone are significantly smaller than ours.

The VR2012 estimates of the mean organic C stock in the RECCAP zones are all larger than ours, but all except for the estimates in the Mediterranean zone, lie

within our 95% confidence limits (Fig. 5a). We have already commented above on the reason for these overestimates. However, we note that they are significant in the Mediterranean zone because, unlike our estimates, the VR2012 estimates do not use the data from the SCaRP, many of which are in this zone. Therefore we believe that our estimates are more reliable.

Our estimate of the total stock of organic C in Australian soil is somewhat less than the global average for the upper 0–30 cm of soil. Foley (1995) and Post *et al.* (1982) report that the average amount of organic C in soil to a depth of 1 m is approximately 104 t ha⁻¹. If we assume that the amount of organic C in the upper 30 cm of soil is 39–70% of the total in top 1 m of soil (Batjes, 1996), then globally the average organic C content for this depth layer is 40–72 t ha⁻¹. The organic C in this layer is most sensitive to interactions with the atmosphere and to change in the environment. The smaller average stock of organic C in Australian soil than the global estimate is due to the large areas of arid and semi-arid land in the continent with soil that contains little organic C (Figs 3 and 4).

Our estimate represents approximately 3.5% of the total stock in the upper 30 cm worldwide. Australia occupies 5.2% of the global land area, and so the total organic C stock of Australian soil makes an important contribution to the global carbon cycle, and it provides significant potential for sequestration and possible source of CO₂ released to the atmosphere.

Conversion of native land for agriculture, and cultivation in particular, has resulted in the loss of between 0.086 and 0.222 Gt organic C in the top 30 cm of soil between 1960 and 2010 (Wang *et al.*, 2013). This is why it is generally thought that agriculture could sequester and store carbon through improvements in management, the use of conservation agricultural practices or conversion to other land use such as grasslands (Luo *et al.*, 2010). The relevance for Australian agriculture lies in that even a small increase in the soil organic C stock across the vast area of agricultural land (Table 8) could sequester a significant amount of organic C and thereby mitigate the emissions of greenhouse gases. Such assessments depend on accurate spatially explicit baseline estimates like the one we report here.

Using data-driven spatial modelling, we have made the most spatially detailed and accurate estimates of the stocks of organic C in Australian soil to date, and these are accompanied by measures of their uncertainty. Our maps of the estimates and their uncertainties have important applications. They could be used to improve future modelling and reduce the uncertainties of their estimates.

The maps could help identify the potential of Australian soil to sequester carbon and guide the formulation

of policy around carbon offset schemes. Although our maps might not be directly used for local carbon trading, they might provide, as a starting point, probable regional baselines of the soil organic C stock.

The maps could serve to direct future soil sampling for inventory. For example, uncertainties in our estimates were larger in the rangelands, savannas and deserts (Fig. 3) where the soil holds large total stocks of organic C. Future sampling efforts should target such areas so that we might better estimate their soil organic C stocks, and thereby improve the national accounts.

The maps could guide the design of networks in which organic C in the soil is repeatedly measured for monitoring. Our estimates are for the year 2010 and could be used as reference against which to monitor and evaluate the impacts of changes in land cover, land management and climate on Australia's organic C stock.

Our spatially explicit estimates and their uncertainties might also support Australia's National Carbon Accounting System; they might improve Australia's terrestrial carbon balances and help develop strategies that will mitigate and adapt to the effects of changing climate.

Acknowledgements

We thank the CSIRO's Sustainable Agriculture Flagship Theme 1179 *Greenhouse Gas Abatement and Carbon Storage* for funding this research (Project R-02622-01–National digital mapping of soil carbon stocks for improved assessment and modelling). We thank also S. Tuomi, P. Leppert, M. Virueda and G. Navarrette for their help with the spectroscopic measurements, C. Chen for preparing the data, P.H. Campbell from the CSIRO's Scientific Computing facility for helping to improve the efficiency of computations, B. Henderson for his advice on some of the statistics, P. de Caritat for the soil samples from the NGS, D. Jacquier for help with the ASRIS database and Tristan Viscarra Rossel for scientific editing. We thank the SCArP for the collection of soil samples. SCArP was funded by the Climate Change Research Program of the Australian Department of Agriculture and the Grains Research and Development Corporation with contributions to the collection of and analysis of soil samples from agricultural regions across Australia by CSIRO, University of Western Australia, Department of Agriculture and Food of Western Australia, Victorian Department of Primary Industries, Murray Catchment Management Authority, Department of Environment and Natural Resources of South Australia, Queensland Government, University of New England, New South Wales Department of Primary Industries, The University of Tasmania, and The Tasmanian Institute of Agricultural Research.

References

- AGO (2002) Pre-clearing Soil Carbon Levels in Australia: Technical report no 12. In: *National Carbon Accounting System*, (ed. Webb A), pp. 2–5. Australian Greenhouse Office, Canberra.
- Baldock JA, Sanderman J, Macdonald D et al. (2013) Australian Soil Carbon Research Program, v1. CSIRO Data Collection. doi: 10.4225/08/5101F31440A36.

- Bartels R, Beatty J, Barsky B (1987) *An Introduction to Splines for Use in Computer Graphics and Geometric Modelling*. Morgan Kaufmann Publishers Inc, Los Altos, California.
- Barrett DJ (2002) Steady state turnover time of carbon in the Australian terrestrial biosphere. *Global Biogeochemical Cycles*, **16**, 55–1–55–21.
- Barrett DJ (2013) NPP Multi-Biome: VAST Calibration Data, 1965–1998, R1. Data set available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. doi: 10.3334/ORNLDAAC/576.
- Batjes NH (1996) Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, **47**, 151–163.
- Bui E, Henderson B, Viergever K (2009) Using knowledge discovery with data mining from the Australian soil resource information system database to inform soil carbon mapping in Australia. *Global Biogeochemical Cycles*, **23**, GB4033.
- Chaplot V, Bouahom B, Valentin C (2010) Soil organic carbon stocks in Laos: spatial variations and controlling factors. *Global Change Biology*, **16**, 1380–1393.
- Cleugh H, Stafford Smith M, Battaglia M et al. (2011) *Climate Change Science and Solutions for Australia*. CSIRO Publishing, Collingwood.
- Cresswell H, Hamilton G (2002) Soil Physical Measurement and Interpretation for Land Evaluation. *Australian Soil and Land Survey Handbook Series*, vol. 5, CSIRO Publishing, Collingwood.
- DAFF (2010) *Land Use of Australia, Version 4, 2005–06*. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra.
- de Caritat P, Lech ME, McPherson AA (2008) Geochemical mapping 'down under': selected results from pilot projects and strategy outline for the National Geochemical Survey of Australia, Geochemistry. *Exploration, Environment, Analysis*, **8**, 301–312.
- DEH (2006) *Major Vegetation Groups in Australia, Version 3.0*. Department of the Environment and Heritage, Canberra.
- Donohue RJ, McVicar TR, Roderick ML (2009) Climate-related trends in Australian vegetation cover as inferred from satellite observations, 1981–2006. *Global Change Biology*, **15**, 1025–1039.
- Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Eswaran H, Van den Berg E, Reich P et al. 1995) Global soil C resources. In: *Soils and global change* (eds Lal R, Kimble JM, Levine EE, Stewart BA), pp. 27–43. Lewis Publishers, Boca Raton, Florida.
- Eswaran H, Reich P, Kimble J (2000) Global carbon stocks. In: *Global Climate Change and Pedogenic Carbonates* (eds Lal R, Kimble J, Eswaran HH, Stewart BA), pp. 15–25. CRC Press, Boca Raton, Florida.
- Foley JA (1995) An equilibrium model of the terrestrial carbon budget. *Tellus Series B—Chemical and Physical Meteorology*, **47**, 310–319.
- GA (2009) *Gravity Grid of Australia and Surrounding Areas (National Geoscience Dataset)*. Geoscience Australia, Symonston.
- Gallant JC, Dowling TI, Read AM et al. (2011) *3 second SRTM Derived Digital Elevation Models User Guide*. Geoscience Australia. www.ga.gov.au/topographic-mapping/digital-elevation-data.html
- Gallant JC, Dowling TI (2003) A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, **39**, 1347.
- Garnaut R (2011) *The Garnaut Review 2011: Australia in the Global Response to Climate Change*. Cambridge University Press, Canberra.
- Gifford RM, Cheney NP, Noble JC et al. 1992) Australian land use, primary production of vegetation and carbon pools in relation to atmospheric carbon dioxide concentration. In: *Australia's Renewable Resources, Sustainability and Global Change* (eds Gifford RM and Barson MM), pp. 151–187. Bureau of Rural Resources Proceedings No 14. Bureau of Rural Resources and CSIRO Division of Plant Industry, Parkes, ACT.
- Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Grace P, Post WM, Hennessy K (2006) The potential impact of climate change on Australia's soil organic carbon resources. *Carbon Balance and Management*, **1**, 14.
- Haverd V, Raupach MR, Briggs PR et al. (2013) Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles. *Biogeosciences*, **10**, 2011–2040.
- Houghton R (2005) The contemporary carbon cycle. In: *Biogeochemistry* (ed. Schlesinger H), pp. 473–513. Elsevier Science, Amsterdam.
- Hutchinson MF, McIntyre S, Hobbs RJ et al. (2005) Integrating a global agro-climatic classification with bioregional boundaries in Australia. *Global Ecology and Biogeography*, **14**, 197–212.
- Isbell R (2002) *The Australian Soil Classification. Revised Edition, Australian Soil and Land Survey Handbooks Series 4*. CSIRO Publishing, Collingwood.

- Jobbágy EG, Jackson RB (2000) The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, **10**, 423–436.
- Johnston RM, Barry SJ, Bleyes E *et al.* (2003) ASRIS: the database. *Australian Journal of Soil Research*, **41**, 1021–1036.
- Keith H, Mackey BG, Lindenmayer DB (2009) Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *Proceedings of the National Academy of Sciences*, **106**, 11635–11640.
- Lal R (2004) Soil carbon sequestration to mitigate climate change. *Geoderma*, **123**, 1–22.
- Landsat (2010) *Landsat Multispectral Imagery*. United States Geological Survey, Reston, VA, USA.
- Lin, LI (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268.
- Lugato E, Panagos P, Bampa F *et al.* (2013) A new baseline of organic carbon stock in European agricultural soils using a modelling approach. *Global Change Biology*, **20**, 313–326.
- Luo Z, Wang E, Sun OJ (2010) Soil carbon change and its responses to agricultural practices in Australian agro-ecosystems: A review and synthesis. *Geoderma*, **155**, 211–223.
- Lymburner L, Tan P, Mueller N, *et al.* (2011) *The National Dynamic Land Cover Dataset – Technical report. Record 2011/031*. Geoscience Australia, Canberra.
- Marchant BP, Viscarra Rossel RA, Webster R (2013) Fluctuations in method-of-moments variograms caused by clustered sampling and their elimination by de-clustering and residual maximum likelihood estimation. *European Journal of Soil Science*, **64**, 401–409.
- Milligan PR, Franklin R, Ravat D (2004) A new generation magnetic anomaly grid database of Australia (MAGDA) – use of independent data increases the accuracy of long wavelength components of continental-scale merges. *Preview*, pp. 25–29. Australian Society of Exploration Geophysicists, Perth.
- Minty B, Franklin R, Milligan P *et al.* (2009) The radiometric map of Australia. Paper presented at 20th International Geophysical Conference and Exhibition, Australian Society of Exploration Geophysicists, Adelaide, 22–25 February 2009.
- Olson JS, Watts JA, Allison, LJ (1985) Major world ecosystem complexes ranked by carbon in live vegetation: a database. Report NDP-017. United States Department of Energy, Oak Ridge, Tennessee.
- Payne RW (2013) *The Guide to GenStat 16th Edition, Part 2 – Statistics*. VSN international, Hemel Hempstead.
- Post WM, Emanuel WR, Zinke PJ *et al.* (1982) Soil carbon pools and world life zones. *Nature*, **298**, 156–159.
- Prescott JA (1950) A climatic index for the leaching factor in soil formation. *Journal of Soil Science*, **1**, 9–19.
- Quinlan J (1992) Learning with continuous classes. In: *AI'92: proceedings of the 5th Australian Joint Conference on Artificial Intelligence* (eds A. Adams ASterling, L), pp. 343–348. World Scientific, Singapore.
- Rayment G, Lyons D (2011) *Soil Chemical Methods – Australasia*. CSIRO Publishing, Collingwood.
- Richards AE, Cook GD, Lynch BT (2011) Optimal fire regimes for soil carbon storage in tropical savannas of northern Australia. *Ecosystems*, **14**, 503–518.
- Sombroek WG, Nachtergaele FO, Hebel A (1993) Amounts, dynamics and sequestering of carbon in tropical and subtropical soils. *Ambio*, **22**, 417–426.
- Stein J (2008) Metadata: Environmental attributes compiled for the continental GDM analysis (technical report). Fenner School of Environment and Society, The Australian National University, Canberra.
- Viscarra Rossel RA (2011) Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *Journal of Geophysical Research*, **116**, F04023.
- Viscarra Rossel RA, Chen C (2011) Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. *Remote Sensing of Environment*, **115**, 1443–1455.
- Viscarra Rossel RA, Webster R (2012) Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. *European Journal of Soil Science*, **63**, 848–860.
- Viscarra Rossel RA, Webster R, Kidd D (2013) 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*, doi: 10.1002/esp.3476 [E publication ahead of print].
- Wang G, Huang Y, Wang E *et al.* (2013) Modelling soil organic carbon change across Australian wheat growing areas, 1960–2010. *PLoS ONE*, **8**, e63324.
- Webster R, Oliver MA (2007) *Geostatistics for Environmental Scientists*, 2nd edn. John Wiley & Sons, Chichester.
- Wynn JG, Bird MI, Vellen L *et al.* (2006) Continental-scale measurement of the soil organic carbon pool with climatic, edaphic, and biotic controls. *Global Biogeochemical Cycles*, **20**, GB1007.
- Xu T, Hutchinson MF (2011) *ANUCLIM Version 6.1. Fenner School of Environment and Society*. The Australian National University, Canberra.
- Zhao G, Bryan BA, King D *et al.* (2013) Impact of agricultural management practices on soil organic carbon: simulation of Australian wheat systems. *Global Change Biology*, **19**, 1585–1597.
- Zhou X-H, Gao S (1997). Confidence intervals for the log-normal mean. *Statistics in Medicine*, **16**, 783–790.