



OPEN

DATA DESCRIPTOR

Baseline high-resolution maps of organic carbon content in Australian soils

Alexandre M. J.-C. Wadoux¹✉, Mercedes Román Dobarco¹, Brendan Malone², Budiman Minasny¹, Alex B. McBratney¹ & Ross Searle³

We introduce a new dataset of high-resolution gridded total soil organic carbon content data produced at 30 m × 30 m and 90 m × 90 m resolutions across Australia. For each product resolution, the dataset consists of six maps of soil organic carbon content along with an estimate of the uncertainty represented by the 90% prediction interval. Soil organic carbon maps were produced up to a depth of 200 cm, for six intervals: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm and 100–200 cm. The maps were obtained through interpolation of 90,025 depth-harmonized organic carbon measurements using quantile regression forest and a large set of environmental covariates. Validation with 10-fold cross-validation showed that all six maps had relatively small errors and that prediction uncertainty was adequately estimated. The soil carbon maps provide a new baseline from which change in future carbon stocks can be monitored and the influence of climate change, land management, and greenhouse gas offset can be assessed.

Background & Summary

In the last two decades, there has been a growing interest in estimating soil organic carbon (SOC) content and stocks for management (e.g. carbon sequestration), economic (e.g. greenhouse gas emission trading schemes, commercial incentives for Net Zero targets)¹ and scientific (e.g. dynamic of carbon cycle) purposes. Soils are an essential component of the ecosystem carbon cycle, storing approximately two-thirds of the total terrestrial organic carbon pool². Organic carbon is also a key indicator of the overall soil functioning. It is the main constituent of soil organic matter and is related to most soil functions such as water and nutrient cycling, the productivity of plants, carbon storage and climate mitigation, among others³.

The total SOC concentration is conventionally measured at a point using laboratory techniques (e.g. the Walkley-Black method⁴ or high-temperature combustion), the values of which can be used in models of soil C dynamics models such as RothC⁵. Organic carbon, however, is a continuum of compounds with different origins, multiple stages of decomposition and decay and chemical composition⁶. The processes that control organic carbon composition vary spatially and with depth, depending on soil, climate, land use and management practices and are controlled by a myriad of biotic and abiotic factors⁷. Several recent lines of work have therefore been developed to model the spatial distribution of SOC at regional⁸, national⁹ and continental¹⁰ scales. The mapping of SOC also reflects an increasing demand for spatially explicit organic carbon assessment to be used. For example, to prioritize local actions in soil carbon sequestration and to monitor SOC change over time. SOC maps are also used as input into Earth System Models and are relevant to calibrate and initiate mechanistic simulation models of the terrestrial carbon cycle¹¹.

SOC maps are usually made using statistical or non-statistical models that exploit the quantitative relationships between point-measured values of SOC and a set of environmental covariates that control SOC spatial distribution^{11,12}. Various methods can be used for this purpose, including geostatistical methods that rely on the variogram and kriging and recently more complex algorithmic tools from machine learning¹³. These models are used to predict the SOC at unobserved locations using the fitted relationship obtained at observation points and the spatially explicit covariates, such as terrain attributes and remote sensing imagery.

¹Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Camperdown, Australia. ²CSIRO Agriculture and Food, Black Mountain, ACT, Australia. ³CSIRO Agriculture and Food, 306 Carmody Road, St Lucia, QLD, Australia. ✉e-mail: alexandre.wadoux@sydney.edu.au

Laboratory methods	Description	Reference
6A1, 6A1_UC	Organic C. Walkley and Black wet Oxidation by dichromate-sulfuric acid. TOC approximate due to incomplete chemical reaction	4
6B2	Total C. Dumas high-temperature combustion, volumetric CO ₂ measurement (no soil pretreatment). Measures TOC in absence of carbonates/bicarbonates	32
6B2b	Total organic C. Dumas high-temperature oxidative combustion, infrared/thermal conductivity detection (no soil pretreatment).	32
6B3, 6B3a	Total organic C. Dumas high-temperature oxidative combustion furnace, infrared/thermal conductivity detection. The results are corrected for the percentage of carbonate/bicarbonate determined separately.	33

Table 1. Datasets used for spatial modelling along with the laboratory method and reference.

In Australia, digital maps of SOC have been produced extensively for nearly two decades. The Soil and Landscape Grid of Australia (SLGA)¹⁴, for example, used a large soil inventory composed of more than 27,000 sites where SOC is recorded, either measured or inferred with spectroscopic techniques. These SOC maps have been extensively used for national carbon accounting and to monitor soil carbon change. However, data collection and processing techniques have progressed rapidly in the last 8 years. It is worthwhile to include these recent developments to produce new high-resolution maps of SOC in Australia.

In this paper, we present updated continental-scale maps of SOC. These maps are part of the new version of the SLGA and will be the new baseline maps of Australia from which carbon stock can be estimated and changes in SOC can be monitored. The resolution at which the maps are produced (90 m × 90 m and 30 m × 30 m) enables applications from regional assessment to local-scale soil management. The maps are produced for six depth intervals, following the specifications of the GlobalSoilMap project¹⁵ and we also produced maps of the prediction uncertainty. The maps are based on measured total organic carbon (TOC) content in soils compiled from various sources which represent the most comprehensive dataset on SOC currently available in Australia.

Methods

Organic carbon data. Data on total organic carbon (TOC) concentration (%) was extracted with the Soil Data Federator developed by CSIRO with support from the Terrestrial Ecosystem Research Network (TERN). The Soil Data Federator is a web API that compiles soil data from different institutions and government agencies throughout Australia. The SOC data used in this study are publicly available through the Soil Data Federator (<https://esoil.io/TERNLandscapes/Public/Pages/SoilDataFederator/SoilDataFederatorHelp.html>) managed by CSIRO¹⁶. The laboratory methods for total organic carbon included in the study are presented in Table 1. We selected TOC data from the period 1970–2020 to get a compromise between the representativity of current TOC concentration and spatial coverage. The data was error checked and processed to harmonize units, excluding duplicates and potentially wrong data entries (e.g. missing upper or lower horizon depths, extreme TOC values, unknown sampling date). Additional TOC measurements from the Biome of Australian Soil Environments (BASE) contextual data¹⁷ were also included in the analyses. TOC concentration for BASE samples was determined by the Walkley-Black method⁴ (method 6A1 in Table 1). Upper limits for TOC concentration by biome and land cover classes were set according to published literature, and consistent datasets (Australian national Soil Carbon Research Program (SCaRP) and BASE), see refs. ^{17,18} and data exploration to exclude unrealistic TOC values (e.g. maximum TOC = 30% in temperate forests, maximum TOC = 14% in temperate rainforests). Since TOC concentration in Australian ecosystems has been underestimated by previous SOC maps¹⁹, we did not set conservative TOC upper limits, knowing that the machine learning model would likely underestimate high SOC values.

Data for TOC concentration come from bulk soil samples taken at various depth intervals in the soil profile. To standardize the depth intervals, we built a mass-preserving depth function using the equal-area quadratic spline. The mathematical description of this function and its application to build continuous depth intervals over the soil profile have been extensively described in the literature^{20,21}. The equal-area quadratic spline function was fitted to the whole collection of pre-processed TOC data, and then values extracted for the 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm depth intervals, following GlobalSoilMap specifications¹⁵. Boxplots with TOC values by biome and land cover after data cleaning and depth standardization are shown in Fig. 1. A total of 90,025 measurements of TOC are used hereafter for mapping.

Environmental covariates. We collected a set of 57 spatially exhaustive environmental covariates made available by the Terrestrial Ecosystems Research Network (TERN), covering Australia and representing proxies for factors influencing SOC formation and spatial distribution: soil properties, climate, organisms/vegetation, relief and parent material/age. The covariates were reprojected to the WGS84 (EPSG:4326) projection and cropped to the same spatial extent. All covariates were resampled using bilinear interpolation or aggregated to conform with a spatial resolution with a grid cell of 90 m × 90 m and 30 m × 30 m. The list of covariates along with their unit and reference is provided in Table 2. The covariates used in this study are freely available through the link <https://data.tern.org.au/landscapes/slga/NationalMaps/SoilAndLandscapeGrid/>. Instructions for accessing the covariate rasters as Cloud Optimised GeoTiffs are provided at <https://esoil.io/TERNLandscapes/Public/Pages/SLGA/GetData-COGSDataStore.html>.

Mapping. The spatial distribution of soil TOC concentration is driven by the combined influence of climate, vegetation, relief and parent materials²². We thus modelled TOC concentration as a function of environmental

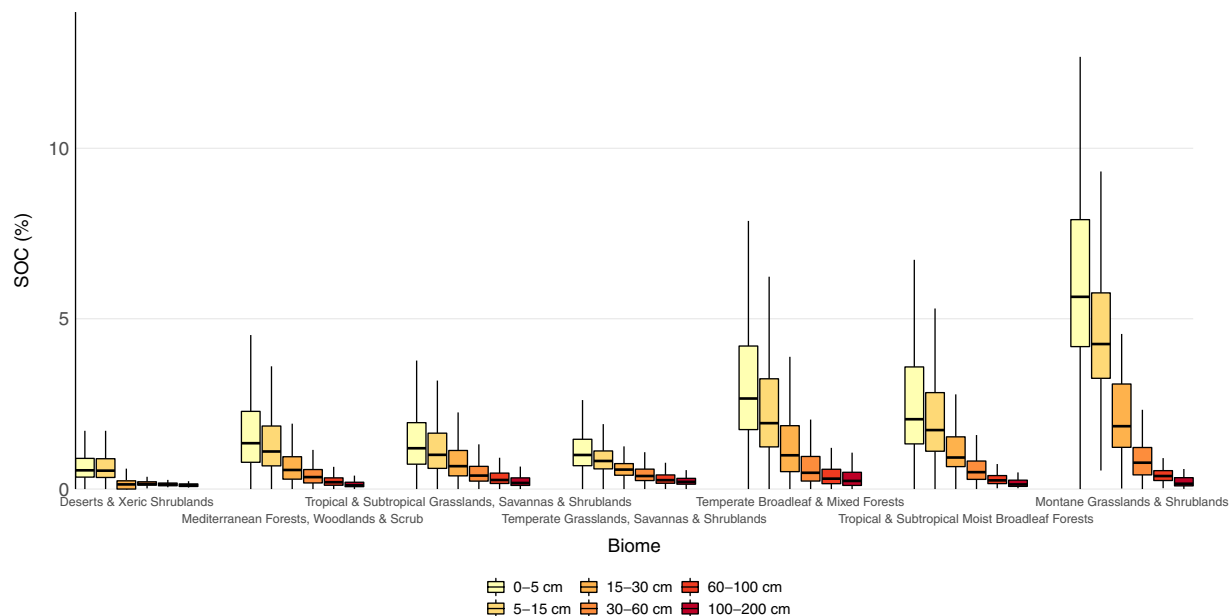


Fig. 1 Values of total organic carbon concentration by biome and land cover after data cleaning and depth standardization.

covariates representing biotic and abiotic control of TOC. The measurements of SOC and their corresponding values of environmental covariates from Table 2 at the same measurement locations were used to fit the mapping model. For the mapping, we used a machine learning algorithm called quantile regression forest.

Quantile regression forest. Quantile regression forest²³ is an ensemble of decision trees. A decision tree is built by partitioning the covariate dataset from the calibration dataset. A number of partitions are evaluated and a splitting metric, the variance, is used to evaluate the partitions. The partition with the smallest splitting metric is selected and undergoes the same procedure until a user-defined parameter, the minimum node size, is reached. For a single tree, the prediction is taken as the average prediction of the values at the end of the node of the tree.

The decision tree is extended by the process of bagging (i.e. bootstrap and aggregating)²⁴, which aims to build an ensemble of decision trees called a random forest. In a random forest, a large number of decision trees is built on bootstrap samples of the original calibration data. For each tree, a random perturbation (i.e. bagging) is introduced during partition where only a subset of size m_{try} from the original number of covariates in the calibration data is used for partition. The final prediction from the random forest is simply the aggregation through averaging of all the decision tree predictions. Extending the standard random forest to QRF is straightforward. Instead of obtaining a single statistic, that is the mean prediction from the decision trees in the random forest, we report all the target values of the leaf node of the decision trees. With QRF, the prediction is thus not a single value but a cumulative distribution of the TOC prediction at each location, which can be used to compute empirical quantile estimates.

Fitting a QRF model is thus based on three user-defined parameters: the partition subset size m_{try} , the number of trees n_{tree} and the stopping criterion for the tree splitting $nodesize$. We fitted a QRF model for each depth interval with parameters m_{try} and $nodesize$ held to their default values. m_{try} is rounded down to the square root of the total number of covariates and $nodesize$ was set to 5. To compromise between computational load and accuracy we fixed the number of trees to 250. We tested parameter tuning using a random grid-search procedure for the three QRF parameters, using 1000 parameter set combinations and the mean of the square error as criterion obtained from a 10-fold cross-validation strategy. We found that parameter tuning was very computationally demanding for a negligible improvement in prediction accuracy. Thus, we did not proceed any further with QRF parameter tuning. We used the R programming language and the *ranger*²⁵ package for model fitting and prediction.

Evaluation of the prediction and uncertainty quantification. Model prediction. Each depth-specific model of TOC was validated based on the results of a K -fold cross-validation. The whole dataset was randomly split into $K = 10$ approximately same-size folds. Each fold was kept apart for the validation and the remaining $K - 1$ folds were used as a calibration dataset. Models were compared using the mean error (ME), the root mean square error (RMSE), the squared Pearson's r correlation coefficient (r^2), and the modelling efficiency coefficient (MEC), defined by the following equations:

Mean error:

$$ME = \frac{1}{n} \sum_{i=1}^n z_i - \hat{z}_i, \quad (1)$$

Factor	Predictor variable	Unit	Reference
Soil	Depth-specific soil clay content	percent	34
	Depth-specific soil sand content	percent	34
Climate	Mean annual aridity index (annual precipitation/annual potential evaporation)	index	35
	Annual potential evaporation	mm	35
	Minimum monthly potential evaporation	mm	35
	Maximum monthly potential evaporation	mm	35
	Prescott Index generated by using Prescott index = 0.445 P/E0.75	index	—
	Annual precipitation	mm	35
	Minimum monthly precipitation	mm	35
	Precipitation seasonality 1- solstice seasonality composite factor ratio	ratio	35
	Precipitation seasonality 2- equinox seasonality composite factor ratio	ratio	35
	Maximum monthly precipitation	mm	35
	Short-wave solar radiation - annual mean (SRAD data)	MJ/m ² /day	35
	Minimum temperature - Annual mean	°C	35
	Annual temperature range	°C	35
	Maximum temperature - Annual mean	°C	35
	Annual atmospheric water deficit (annual precipitation - annual potential evaporation)	mm	35
Organisms/vegetation	Long-term average NDVI Q1	unitless	
	Long-term average NDVI Q2	unitless	
	Long-term average NDVI Q3	unitless	
	Long-term average NDVI Q4	unitless	
	Landsat Fractional cover - Bare Soil - Maximum of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Non Photosynthetic Vegetation - Maximum of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Photosynthetic Vegetation - Maximum of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Bare Soil - Mean of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Non Photosynthetic Vegetation - Mean of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Photosynthetic Vegetation - Mean of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Bare Soil Minimum of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Non Photosynthetic Vegetation - Minimum of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Photosynthetic Vegetation - Minimum of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Bare Soil - Standard deviation of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Non Photosynthetic Vegetation - Standard deviation of the timeseries - 1987–2019	percent	36
	Landsat Fractional cover - Bare Soil - Standard deviation of the timeseries - 1987–2019	percent	36
	Fraction of Photosynthetically Active Radiation (FPAR) - AVHRR - Maximum Value in Timeseries	percent	37
	Fraction of Photosynthetically Active Radiation (FPAR) - AVHRR - Mean Value in Timeseries	percent	37
	Fraction of Photosynthetically Active Radiation (FPAR) - AVHRR - Median Value in Timeseries	percent	37
	Fraction of Photosynthetically Active Radiation (FPAR) - AVHRR - Minimum Value in Timeseries	percent	37
National Dynamic Land Cover Dataset Mean of the timeseries 2000 - 2008	unitless	38	
Landsat 2000–2010 Persistent Green-Vegetation Fraction	unitless	39	
Relief	Elevation 3 Second - Shuttle Radar Topography Mission	metre	40
	Multi-resolution Ridgetop Flatness	unitless	41
	Multiresolution Index of Valley Bottom Flatness (MRVBF)	unitless	42
	Plan curvature	unitless	43
	Profile curvature	unitless	43
	relief roughness	unitless	43
	Slope	percent	44
	Topographic wetness index	unitless	43
Continued			

Factor	Predictor variable	Unit	Reference
Parent material/age	Total Magnetic Intensity (TMI) Gravity Grid of Australia	unitless	45
	Radiometric grid of Australia (Radmap) v4 2019 - Filtered dose	unitless	46
	Radiometric grid of Australia (Radmap) v4 2019 - Potassium	percent	46
	Radiometric grid of Australia (Radmap) v4 2019 - Thorium	ppm	46
	Radiometric grid of Australia (Radmap) v4 2019 - Uranium	ppm	46
	Radiometric grid of Australia (Radmap) v4 2019 - Thorium Potassium ratio	ratio	46
	Radiometric grid of Australia (Radmap) v4 2019 - Uranium Thorium ratio	ratio	46
	Radiometric grid of Australia (Radmap) v4 2019 - Uranium Potassium ratio	ratio	46
	Radiometric grid of Australia (Radmap) v4 2019 - Uranium Thorium ratio	ratio	46
	Weathering index	unitless	47

Table 2. List of environmental covariates with unit and associated reference when applicable. All covariates are in geographic coordinates with 3 arc second grid cell (about 90 m) or 1 arc second grid cell (about 30 m) resolution with coordinate system WGS84 (EPSG:4326) and extent: 112.99958°E - 153.99958°E; 10.0004°S - 44.00042°S.

where z and \hat{z} denote the measured and predicted values of TOC, respectively, and n is the total number of measured values. The ME represent the bias, i.e. the systematic over- or under-prediction of the model.

Root mean square error:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2}. \quad (2)$$

The RMSE represents the magnitude of the error, its optimal value is 0 and is expressed in the unit of TOC (i.e. in %).

Squared Pearson's r correlation coefficient:

$$r^2 = \left(\frac{\sum_{i=1}^n (z_i - \bar{z})(\hat{z}_i - \bar{\hat{z}})}{\sqrt{\sum_{i=1}^n (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^n (\hat{z}_i - \bar{\hat{z}})^2}} \right)^2, \quad (3)$$

where \bar{z} is the mean of the measured values and $\bar{\hat{z}}$ is the mean of the predicted values. The r^2 describes the linear correlation between measured and predicted values and ranges between 0 (no linear correlation) to 1 (perfect linear correlation).

Modelling efficiency coefficient:

$$\text{MEC} = 1 - \frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2}. \quad (4)$$

The MEC optimal value is 1 but it can be negative if the mean of the measured values is a better predictor than the model. Positive MEC values can be interpreted as an amount of variance explained by the model.

We use a solar diagram²⁶ to exploit the relationship between statistical indices and compare the maps. In a Cartesian coordinate system, the x-axis represents the ME (Eq. 1), and the y-axis the standard deviation of the error (SDE). The distance from the origin to any point in the diagram is expressed in terms of RMSE (Eq. 2), which allows to see the individual contribution of ME and SDE to the RMSE. The ME, SDE and RMSE are standardized by the standard deviation of the observation. The solar diagram further includes information on the correlation (Eq. 3) and MEC (Eq. 4). In a solar diagram, the closer the point to the origin of the diagram, the better the map. More information on the statistical representation of the validation statistics is provided in ref.²⁶.

Uncertainty quantification. We report the depth-specific lower ($q_{0.05}$) and upper ($q_{0.95}$) limits of the 90% prediction interval with two maps. Validation of the uncertainty estimates are obtained through a so-called accuracy plot. In an accuracy plot, the proportion of cross-validation observations contained in each q prediction interval is calculated. Ideally, the proportion of observations covered by a q interval is approximately equal to the value of q . If the proportion of observations in q is greater than q , it suggests that the uncertainty is under-estimated, whereas a substantially smaller proportion of observations compared to the nominal value of q suggests an under-estimation of the uncertainty. The process is repeated for all q , and the values of q are plotted against the actual proportion of values covered by q in a scattergram. Ideally, all values should be close to the 1:1 line, which would mean that the uncertainty is adequately estimated. Note that we evaluate uncertainty for all q against cross-validation observations, but report the maps of the 90% prediction intervals only.

Data Records

Prediction results. The depth-specific maps of TOC are distributed through the CSIRO Data Access Portal. Each product contains a map of TOC and its upper and lower prediction intervals. The products are available for grid size cells of 30 m × 30 m²⁷ and 90 m × 90 m²⁸ and for the size depth intervals: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm and 100–200 cm. The unit of TOC is in percent mass (%). The projection system

	ME		RMSE		r^2		MEC	
	30 m	90 m	30 m	90 m	30 m	90 m	30 m	90 m
0–5 m	0.03	0.03	1.25	1.39	0.53	0.53	0.53	0.53
5–15 cm	0.03	0.03	1.07	1.11	0.50	0.54	0.50	0.53
15–30 cm	0.02	0.02	0.90	0.93	0.41	0.44	0.41	0.44
30–60 cm	0.02	0.02	0.74	0.75	0.22	0.22	0.22	0.22
60–100 cm	0.02	0.02	0.50	0.56	0.12	0.13	0.11	0.12
100–200 cm	0.02	0.01	0.38	0.41	0.18	0.18	0.18	0.17

Table 3. Depth-specific validation statistics for the maps. The statistics are obtained by 10-fold cross-validation. Note that 30 m and 90 m refers to the map made at 30 m × 30 m and 90 m × 90 m resolutions, respectively.

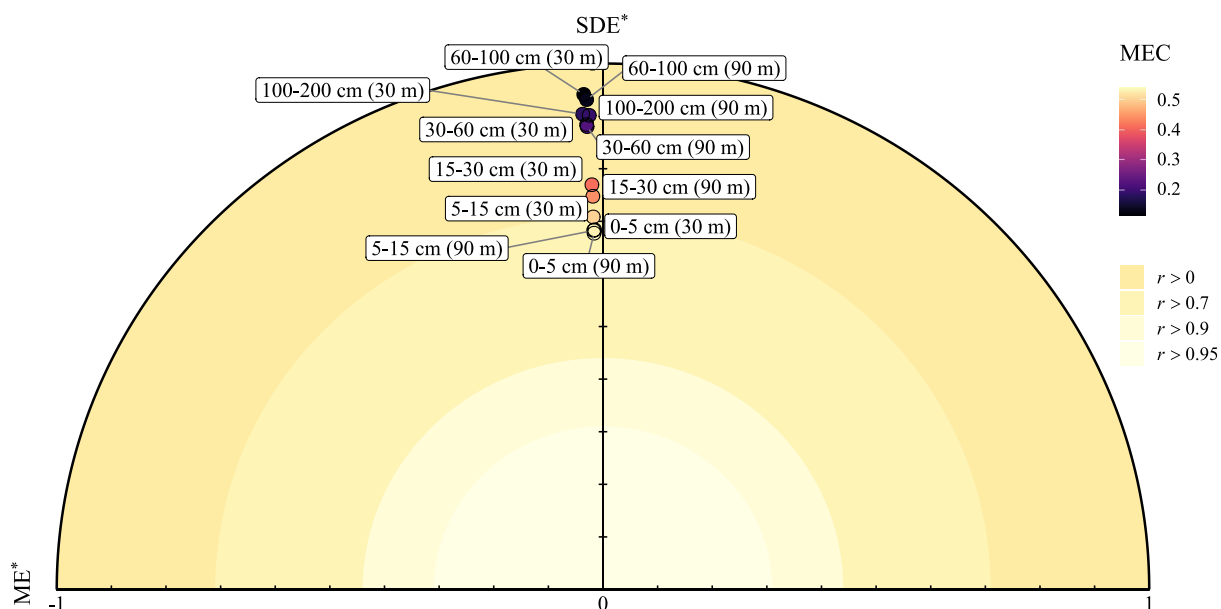


Fig. 2 Summary diagram (solar diagram) of the validation statistics. The statistics are obtained by 10-fold cross-validation. The x-axis represent the mean error (ME), the y-axis is the standard deviation of the error (SDE) which are both standardized by the standard deviation of the observations (denoted by *). Any point in the diagram has a distance to the origin equal to the RMSE. Color scale indicate the Pearson's r correlation coefficient and the modelling efficiency (MEC). Note that 30 m and 90 m refers to the map made at 30 m × 30 m and 90 m × 90 m resolutions, respectively.

is WGS84 (EPSG:4326) and maps have extent: 112.54449°E - 153.38239°E; 9.59539°S - 43.38329°S for the 30 m and 112.99958°E - 153.99958°E; 10.0004°S - 44.00042°S for the 90 m. There are 36 maps in total: 6 depths, each with two maps of lower and upper intervals, for two spatial resolutions. An example of prediction along with the lower and upper intervals (0.05 and 0.95 percentiles) is shown in Fig. 6 for the 0–5 cm depth interval and the 90 m × 90 m spatial resolution. The maps of TOC prediction for all depth intervals and the two spatial resolutions are shown in Fig. 4. For the 90 m × 90 m resolution maps only, they can be used and accessed directly as Cloud Optimised GeoTIFF files through the TERN Data Store²⁸.

Technical Validation

Prediction accuracy. The validation statistics for the QRF model prediction, two spatial resolutions and the six depth intervals are shown in Table 3 and visualized in Fig. 2. All maps have a negligible bias (the ME is close to 0 in all cases). The RMSE decreased with depth, which is expected because it is reported in the units of SOC (i.e. %) and the SOC content decreases with depth (see also Fig. 1). The solar diagram in Fig. 2 confirms that the RMSE is mostly composed of the SDE because all points are close to the a value of ME = 0. The r^2 and MEC suggest that the models are accurate, but that prediction accuracy decreases dramatically for deeper depth intervals. This is an expected result reported in previous studies¹⁴. There is no substantial difference in terms of ME, r^2 and MEC between the two spatial resolutions, and a slight improvement in RMSE between the prediction made with covariates at 30 m and that using 90 m resolution covariates.

Validation of the uncertainty. The validation of the predicted prediction intervals is shown in Fig. 3. Note that while Fig. 2 shows all prediction intervals, we only report the maps of the 90% prediction interval, that is, the

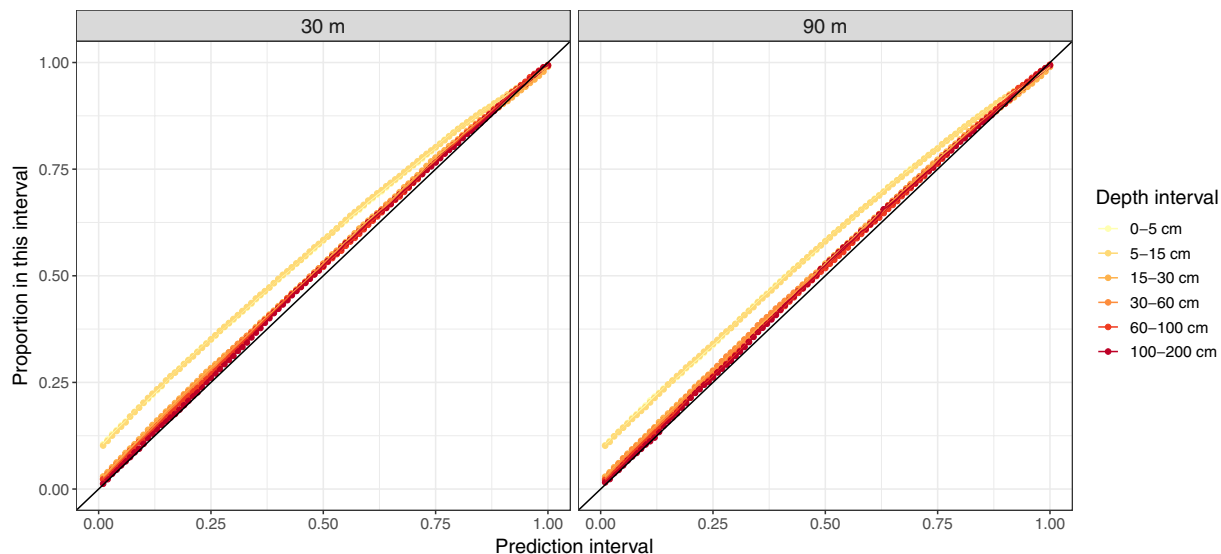


Fig. 3 Accuracy plots showing the probability interval and the proportion falling within this interval, for the six depth intervals and two spatial resolutions.

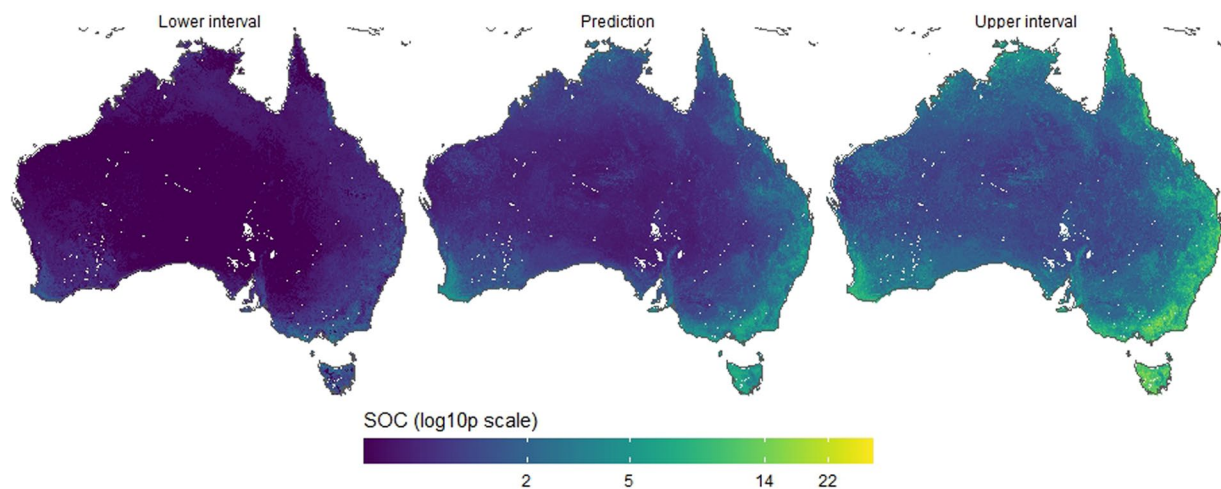


Fig. 4 Map of prediction of TOC for the 0–5 cm depth interval and 90 m × 90 m spatial resolution (center) along with the lower (left) and upper (right) intervals (0.05 and 0.95 percentiles).

maps of the 0.05th and the 0.95 percentiles (see Figs. 4 and 5). Overall, the uncertainty was adequately estimated because most points are close to the line of equality. The uncertainty of SOC for the 0–5 cm depth intervals is slightly overestimated. For example, for the 50% interval, nearly 55% of the observations from the validation dataset fall within this interval.

Comparison with existing maps. We compared the topsoil (0–5 cm) TOC prediction maps at 30 m × 30 m (Fig. 6a) with existing maps available for Australia, the previous version of the SLGA v1 (Fig. 6b²⁹) global map of SoilGrids 2.0 (Fig. 6c³⁰). Note that in Fig. 6 the SLGA v1 map is at 90 m × 90 m whereas the SoilGrids map is at 250 m × 250 m spatial resolution. Three small areas in the West, East and North of Australia are shown. Overall, the prediction results for the three maps have similar patterns and ranges of values for the three small areas. However, the new map reveals much more detailed information than previous maps. The map of SoilGrids has smooth variation while the map of SLGV v1 captures further variation, but missed the detailed variation caused by fields and river beds.

Usage Notes

With its high spatial resolution and national coverage, this dataset should be useful for a range of stakeholders, including policymakers, scientists and land users alike. Potential uses include setting up a benchmark for Australia to estimate the change in soil organic carbon resulting from a change in land use, land cover, soil management practices and greenhouse gas offset activities. They can be also used for researchers to obtain insights

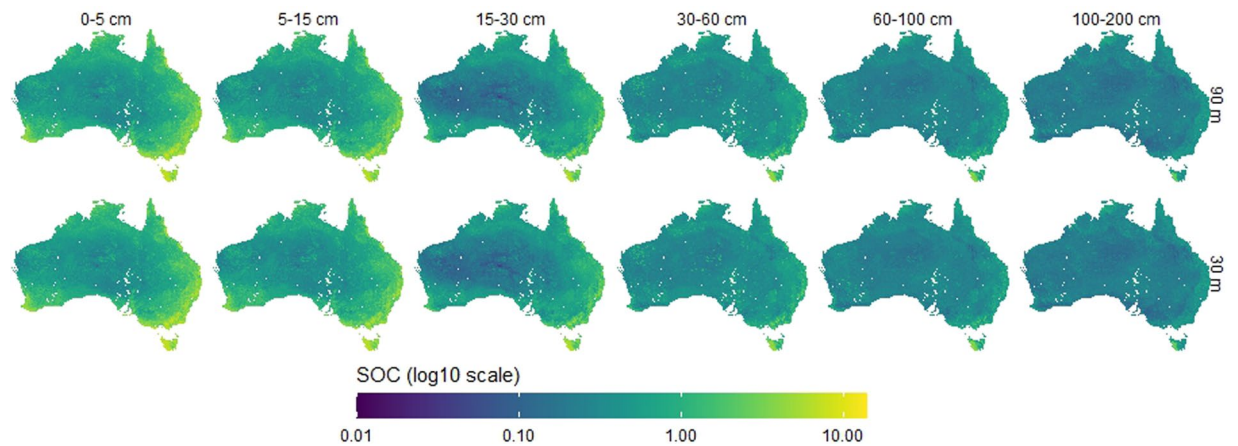


Fig. 5 Maps of total organic carbon prediction for the size depth intervals and two spatial resolution (i.e. 90 m × 90 m and 30 m × 30 m).

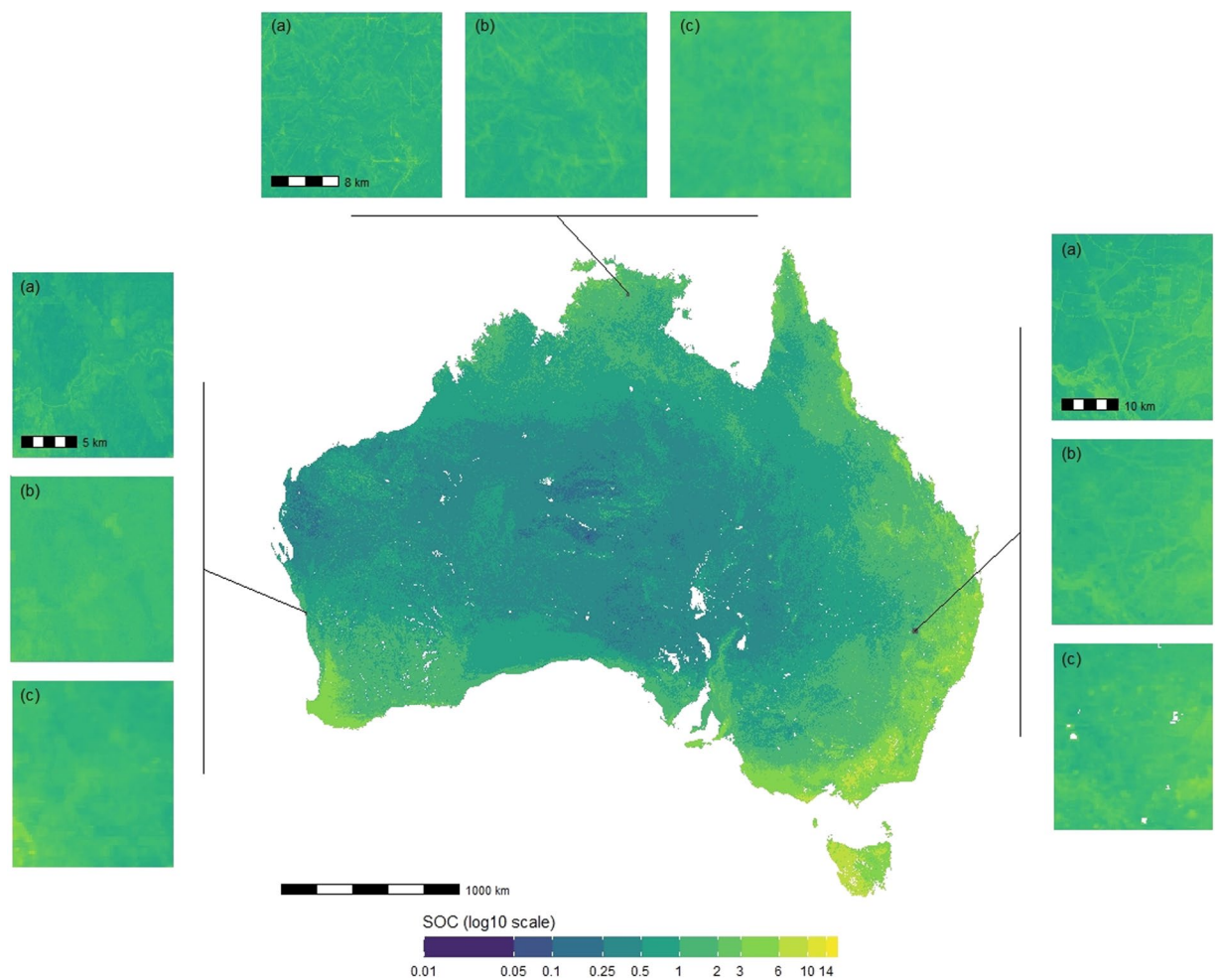


Fig. 6 Topsoil (0–5 cm) (a) TOC prediction maps at 30 m × 30 m spatial resolution compared with (b) the previous version of the Soil and Landscape Grid of Australia and (c) the global map of SoilGrids 2.0³⁰ for three small areas in the West, East and North of Australia.

into the large-scale and local-scale drivers of soil organic carbon in relation to biophysical factors and the environment. National-scale maps of soil organic carbon also constitute an input to guide the design of national soil monitoring networks and soil organic carbon accounting projects. The uncertainty of the maps reported in this

study can be used to guide sampling to refine the existing maps in areas of large uncertainty. At the international level, we envision that this map will help policymakers to report on the national soil carbon budgeting and might help assist with strategies to mitigate climate change through carbon storage in soils.

Code availability

Specific functions for data pre-processing and spline fitting are freely available in the R package *ithir*³¹. Codes associated to the model fitting, cross-validation and mapping are freely available from <https://github.com/AusSoilsDSM/SLGA/tree/main/Production/DSM/SoilOrganicCarbon>. All analyses were performed in the R programming language (version 4.1.0).

Received: 5 October 2022; Accepted: 8 March 2023;

Published online: 31 March 2023

References

- Keenor, S. G. *et al.* Capturing a soil carbon economy. *Royal Society Open Science* **8**, 202305 (2021).
- Batjes, N. H. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science* **47**, 151–163 (1996).
- Amelung, W. *et al.* Towards a global-scale soil climate mitigation strategy. *Nature Communications* **11**, 1–10 (2020).
- Walkley, A. & Black, I. A. 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 (1934).
- Coleman, K. & Jenkinson, D. S. Rothc-26.3-a model for the turnover of carbon in soil. In *Evaluation of soil organic matter models*, 237–246 (Springer, 1996).
- Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).
- Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications* **10**, 423–436 (2000).
- Ellili-Bargaoui, Y., Walter, C., Michot, D. & Lemercier, B. Mapping soil properties at multiple depths from disaggregated legacy soil maps in the Brittany region. *France. Geoderma Regional* **23**, e00342 (2020).
- Adhikari, K. *et al.* Digital mapping of soil organic carbon contents and stocks in denmark. *PLoS ONE* **9**, e105519 (2014).
- de Brogniez, D. *et al.* A map of the topsoil organic carbon content of europe generated by a generalized additive model. *European Journal of Soil Science* **66**, 121–134 (2015).
- Minasny, B., McBratney, A. B., Malone, B. P. & Wheeler, I. Digital mapping of soil carbon. *Advances in Agronomy* **118**, 1–47 (2013).
- McBratney, A. B., Santos, M. L. M. & Minasny, B. On digital soil mapping. *Geoderma* **117**, 3–52 (2003).
- Wadoux, A. M. J.-C., Minasny, B. & McBratney, A. B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews* **210**, 103359 (2020).
- Grundy, M. J. *et al.* Soil and landscape grid of australia. *Soil Research* **53**, 835–844 (2015).
- Arrouays, D. *et al.* The GlobalSoilMap project specifications. In Arrouays, D., McKenzie, N., Hempel, J., de Forges, A. R. & McBratney, A. B. (eds.) *GlobalSoilMap: Basis of the Global Spatial Soil Information System* (CRC Press, London, 2014).
- Searle, R. *et al.* Soil data, united, will never be defeated – The SoilDataFedorator (2021).
- Bissett, A. *et al.* Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *Gigascience* **5**, s13742–016 (2016).
- Baldock, J. A. *et al.* Quantifying the allocation of soil organic carbon to biologically significant fractions. *Soil Research* **51**, 561–576 (2013).
- Bennett, L. T. *et al.* Refining benchmarks for soil organic carbon in australia's temperate forests. *Geoderma* **368**, 114246 (2020).
- Bishop, T., McBratney, A. B. & Laslett, G. M. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27–45 (1999).
- Malone, B. P., McBratney, A. B., Minasny, B. & Laslett, G. M. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **154**, 138–152 (2009).
- Wiesmeier, M. *et al.* Soil organic carbon storage as a key function of soils—a review of drivers and indicators at various scales. *Geoderma* **333**, 149–162 (2019).
- Meinshausen, N. Quantile regression forests. *Journal of Machine Learning Research* **7** (2006).
- Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).
- Wright, M. N. & Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* **77**, 1–17 (2017).
- Wadoux, A. M. J.-C., Walvoort, D. J. J. & Brus, D. J. An integrated approach for the evaluation of quantitative soil maps through taylor and solar diagrams. *Geoderma* **405**, 115332 (2022).
- Wadoux, A. M. J.-C. *et al.* Soil and Landscape Grid National Soil Attribute Maps - Organic Carbon (1" resolution). Release 1. v1. CSIRO <https://doi.org/10.25919/07c3-2n73> (2022).
- Wadoux, A. M. J.-C. *et al.* Soil and Landscape Grid National Soil Attribute Maps - Organic Carbon (3" resolution). Release 2. v1. CSIRO <https://doi.org/10.25919/ejhm-c070> (2022).
- Viscarra Rossel, R. A. *et al.* The australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Research* **53**, 845–864 (2015).
- Poggio, L. *et al.* Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240 (2021).
- Malone, B. Package "ithir". R package version 1.0 [Accessed 10.01.2023] (2021).
- Rayment, G. E. & Higginson, F. R. *Australian Laboratory Handbook of Soil and Water Chemical Methods* (Inkata Press Pty Ltd, 1992).
- Rayment, G. E. & Lyons, D. J. *Soil chemical methods: Australasia*, vol. 3 (CSIRO publishing, 2011).
- Malone, B. & Searle, R. Updating the australian digital soil texture mapping (Part 2*): spatial modelling of merged field and lab measurements. *Soil Research* **59** (2021).
- Harwood, T. 9s climatology for continental Australia 1976–2005: BIOCLIM variable suite. V1. CSIRO. Data Collection (2019).
- Joint Remote Sensing Research Program. Seasonal fractional cover - Landsat, JRSRP algorithm, Australia coverage. Version 1.0.0 (2021).
- Donohue, R., McVicar, T. & Roderick, M. Fraction of Photosynthetically Active Radiation (FPAR) - AVHRR. Version 5. Terrestrial Ecosystem Research Network (TERN) (2021).
- Lymburner, L. *et al.* The National Dynamic Land Cover Dataset (2010).
- Johansen, K. *et al.* Landsat based Persistent Green-Vegetation Fraction for Australia. In *Proceedings of the 16th Australasian Remote Sensing and Photogrammetry Conference* (2012).
- Farr, T. G. *et al.* The shuttle radar topography mission. *Reviews of Geophysics* **45** (2007).
- Gallant, J. C., Dowling, T. & Austin, J. Multi-resolution Ridge Top Flatness (MrRTF). Version v2. (2016).
- Gallant, J. C. & Dowling, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* **39** (2003).

43. Wilson, J. P. & Gallant, J. C. Secondary topographic attributes. *Terrain Analysis: Principles and Applications* 87–131 (2000).
44. Zevenbergen, L. W. & Thorne, C. R. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* **12**, 47–56 (1987).
45. Minty, B. R. S. & Poudjom Djomani, Y. Total Magnetic Intensity (TMI) Grid of Australia. *Seventh edition - 40 m cell size* (2019).
46. Poudjom Djomani, Y. & Minty, B. R. S. Radiometric Grid of Australia (Radmap) v4 2019 filtered pct potassium grid (2019).
47. Wilford, J. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma* **183**, 124–142 (2012).

Acknowledgements

This work was financially supported by Terrestrial Ecosystem Research Network (TERN) and by the use of TERN infrastructure, which is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). We acknowledge support from the Research Portfolio at the University of Sydney. We also acknowledge the high-performance computing facilities from CSIRO that were used for mapping at 30 m resolution.

Author contributions

A.M.J.C.W. and M.R.D. conceived and performed the analysis, B.r.M. preprocessed the point data, A.M.J.C.W. and M.R.D. analysed the results. All authors contributed to the development of the methodology and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.M.J.-C.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023