REGULAR ARTICLE

Drivers of soil organic carbon storage and vertical distribution in Eastern Australia

**Eleanor Hobley · Brian Wilson · Arjan Wilkie ·
Jonathan Gray · Terry Koen**

Abstract

Aims Drivers of soil organic carbon (SOC) storage are likely to vary in importance in different regions and at different depths due to local factors influencing SOC dynamics. This paper explores the factors influencing SOC to a depth of 30 cm in eastern Australia.

Methods We used a machine learning approach to identify the key drivers of SOC storage and vertical distribution at 1401 sites from New South Wales, Australia. We then assessed the influence of the identified factors using traditional statistical approaches.

Results Precipitation was important to and positively associated with SOC content, whereas temperature was important to and negatively associated with SOC vertical distribution. The importance of geology to SOC content increased with increasing soil depth. Land-use was important to both SOC content and its vertical distribution.

Conclusion We attribute these results to the influence of precipitation on primary production controlling SOC content, and the stronger influence of temperature on microbial activity affecting SOC degradation and vertical distribution. Geology affects SOC retention below the surface. Land-use controls SOC via production, removal and vertical mixing. The factors driving SOC storage are not identical to those driving SOC vertical distribution. Changes to these drivers will have differential effects on SOC storage and depth distribution.

Keywords Machine learning · Vertical distribution · Land-use · Climate · Geology

E. Hobley (✉) · B. Wilson · A. Wilkie
University of New England,
Armidale, NSW 2351, Australia
e-mail: ehobley@une.edu.au

B. Wilson
NSW Office of Environment and Heritage,
Armidale, NSW 2351, Australia

J. Gray
NSW Office of Environment and Heritage,
Parramatta, NSW 2124, Australia

T. Koen
NSW Office of Environment and Heritage,
Cowra, NSW 2794, Australia

Introduction

Globally, a significant body of research has focussed on contrasting land-uses and management with a view to quantifying differences in SOC stocks in different agricultural systems (e.g. West and Post 2002; Dawson and Smith 2007). In Australia, the Soil Carbon Research Program (and subsequent National Soil Carbon Program) commenced in 2009 to this aim and has yielded numerous studies attempting to quantify such effects (Baldock et al. 2013; Sanderman et al. 2013; Wilson and Lonergan 2013; Rabbi et al. 2014). Whilst quantifying the effects of land-use and land-management on SOC stocks is an important undertaking for carbon accounting purposes (Richards 2001), it assumes that

the changes in SOC associated with alteration of land-use and management are easily quantified. Unfortunately, detecting changes in SOC between land-use systems can be difficult within the necessary timeframes for C accounting.

This difficulty in successfully distinguishing SOC differences under varying land-uses may be (partially) attributable to the numerous, complex and at times conflicting factors influencing SOC dynamics. These influential factors were described by Jenny (1941) and include time, parent material, topography, climate and organisms (including humans). Although Jenny characterised these factors as being key to soil formation, they are all conceptually linked with SOC dynamics as SOC is intimately linked with soil development.

Human influence on SOC is therefore but one of many driving factors, so that the influence of land-use and management on SOC storage may be masked by the influence of other over-arching factors such as climate. For example, Jackson et al. (2002) reported overall gains in SOC after encroachment of grasslands by woody species in drier climates but losses in wetter climates. However, a greater understanding of the driving factors of SOC dynamics may enable de-trending SOC data from the various influential signals, thereby allowing detection of differences between land-uses and land management schemes previously thought to be undetectable. To this end it was recently noted that to improve future IPCC assessments 'it is critical to understand the drivers of soil carbon dynamics in the models, as well as the real world' (Todd-Brown and Luo 2014).

Globally, SOC storage has been linked with vegetation, climate and physico-chemical soil properties such as texture or soil type (Schlesinger 1977; Post et al. 1982; Burke et al. 1989; Jobbagy and Jackson 2000) and climate also appears most relevant to SOC storage at regional scales (Wiesmeier et al. 2013b). Topography, on the other hand, appears to become relevant at the sub-regional scale (Grimm et al. 2008; Davy and Koen 2013). In their landmark paper, Jobbagy and Jackson (2000) assessed not only SOC storage but also its vertical distribution, and found that the factors influencing total SOC content differed from those influencing SOC vertical distribution. Specifically, climate was most influential in shallow depths (0–20 cm), whereas soil texture (sand and clay content) became more important at greater depths (>20 cm). More recently, similar results were reported by Badgery et al. (2013) who found that the influence of climate is most pronounced near the

surface (0–10 cm), but at 20–30 cm soil texture and mineralogy become more influential to SOC storage. This suggests that an in depth understanding of the drivers of SOC dynamics requires assessment at different depths.

Changes in land-use lead to alteration of SOC content near the soil surface (0–10 cm) (Young et al. 2005; Luo et al. 2010; Wilson et al. 2010, 2011). Theoretically, changes in land-use should also affect SOC stocks below this depth, as land-use change usually implies vegetation change (e.g. reforestation or clearing of land for agricultural use), and vegetation type is strongly linked to SOC storage and its vertical distribution (Jackson et al. 2000). However, despite international evidence of subsoil response to changing land-use and management (Wiesmeier et al. 2013a), response of SOC below the surface due to land-use changes is yet to be clearly demonstrated in Australian soils.

Similarly, changes in land management are generally reported to have the greatest influence on the SOC content nearest the soil surface (0–15 cm) (Cotching 2012; Sanderman et al. 2013; Wilson and Lonergan 2013), with effects being more difficult to detect deeper in the soil (Badgery et al. 2013). This has led to the suggestion that subsurface SOC stocks may better reflect historical rather than current land-management (Wilson and Lonergan 2013). Despite this, altered land-management practices have been reported to influence subsurface carbon stocks (Meersmans et al. 2009) and the stratification of SOC (Yang et al. 2008) on decadal timeframes internationally, and more research is needed to elucidate the effects of both land-management and land-use on SOC stocks at different depths.

A possible reason for the discrepancies between international and Australian findings regarding the effects of land-use and management on SOC storage is the unique nature of Australian soils. These are characterised by low SOC storage compared with soils internationally, which has been attributed to Australian climatic influences (low precipitation, high temperature) leading to low inputs of organic matter into the soil (Hassink 1997). Given the above-mentioned difficulties in detecting anthropogenic-driven SOC changes below the surface 10 cm in Australian soils, it is likely that—in contrast to many international soils—the active soil zone typically referred to as 'topsoil' many only comprise the surface 10 cm in Australian soils.

We hypothesise that the drivers of SOC storage vary with depth, namely that the factors influencing the total amount of SOC stored by a soil vary with depth and also differ from those influencing the vertical distribution of SOC. To test this hypothesis we used two analysis approaches—machine learning and classical parametric statistics—to assess the importance of various possible explanatory factors on SOC storage and vertical distribution to a depth of 30 cm. The explanatory factors were derived from site observations and GIS data. The SOC dataset comprised over 1400 sites from New South Wales (NSW), Australia, drawn from existing data relating to soils derived from the NSW Soil Monitoring Program and the National Soil Carbon Research Program.

Methodology

Study area

New South Wales is a state in eastern Australia, covering an area of 801 600 km^2. Although lying within the global temperate zone, NSW encompasses a wide range of climates, with mean annual precipitation (MAP) ranging from <200 mm year^{-1} in the west to >1500 mm year^{-1} in the north-east of the state (Fig. 1). Rainfall generally increases from the west to east. The temperatures are mild overall but can be very hot in the north-west (>45 °C) and very cold in the alpine regions of the south-east (<0 °C). This variability in climate is reflected in the biodiversity of the state, with over 17 bioregions represented within the State, ranging from sandy deserts to lush rainforests (Sahukar et al. 2003). The State is traversed from north to south by the Great Dividing Range, which has a maximum elevation of >2200 m, and is a geographical boundary between the eastern seaboard and western centre.

The soils of NSW are highly diverse, with all 14 orders of the Australian Soil Classification (Isbell 2002) represented, such as siliceous sands (Tenosols, World Reference Base (WRB) Arenosol), cracking clays (Vertosols, WRB Vertisol) and water-influenced gleys (Hydrosols, WRB Gleysol) (Charman and Murphy 2007). The diversity of the soil landscape reflects the variation in climate, geology, topography and vegetation found within the State. This diversity in soil type and climate enables a range of agricultural land-uses, with the vast majority (74 %) of land used for grazing

and cropping, with other uses being forestry and nature conservation (Australian Bureau of Statistics 2013). Overall the state of the soils of NSW is considered fair, but a noticeable, moderate decline in soil health is observable in the State (State of NSW Environmental Protection Authority 2012).

Soil and site data

The data presented here were merged from two separate datasets. The first was derived from the NSW soil condition Monitoring, Evaluation and Reporting (MER) Program. This dataset contains information on 800 sites and provides a provisional 'baseline' data layer gathered in 2008–09. The second dataset was from the Soil Carbon Research Program (SCRP), initiated across Australia in 2009 with the aim of assessing current carbon stocks across a range of representative agricultural land-use systems, focussing on those land-uses that are believed to hold the most promise for carbon storage. The work reported here was a part of the NSW component of the SCRP program, comprising a total of 741 sites.

Both datasets contained information regarding SOC concentration (%$_{SOC}$), bulk density (ρ, in g cm^{-3}) and gravimetric rock content (RM, dimensionless units), from which the C stocks (t ha^{-1}) in a given sampling depth (cm) were calculated according to:

$$C_{Stock} = \%_{SOC} \times \rho \times (1-RM) \times depth \qquad (1)$$

Further information in the datasets related to the land-use at the time of sampling and the GIS co-ordinates of the site. This information was complimented with data which we believe to potentially influence SOC dynamics, namely variables describing climate, land-use and site such as soil type, elevation and geology (Table 1). The data were acquired from GIS data layers from different sources and where possible numerous descriptors describing the same or similar variables (e.g. maximum, minimum and average MAT, or numerous geological and lithological descriptors) were obtained, so as to reduce potential bias towards a single (set of) variable(s) or a single GIS data source.

Dataset collation and processing

A procedure was developed to merge the two datasets, which due to their different histories contained

inconsistent sampling depth information. The MER project had sampled four depths: 0–5, 5–10, 10–20 and 20–30 cm, whereas within the SCRP project only three depths were sampled (0–10, 10–20, 20–30 cm). To unify the datasets, the MER data from the 0–5 and 5–10 cm were algebraically averaged, to create one depth corresponding to 0–10 cm. The appropriateness of this procedure was tested by comparing the carbon stocks calculated according to Eq. (1) for the 0–10 cm from the averaged values with those calculated as the sum of the C stocks in the individual depths (0–5 and 5–10 cm). The relationship between the two calculations was linear with Pearson's product moment $r > 0.99$, so that the algebraic averaging was deemed sufficient for the purposes of the project.

Incomplete profiles were eliminated, as well as obviously erroneous values (e.g. SOC>100 % or rock content=100 %). The bulk density data were highly heterogeneous and at times implausible, ranging from 0.2 to 2.9 g cm^{-3} (which is greater than the density of granite), so that a statistical procedure to process the data was developed. Extreme bulk density values were identified using the interquartile range (IQR). Values above the

limit set by the 3rd quartile plus 1.5 times the IQR (corresponding to a bulk density of >1.85 g cm^{-3} in the top 10 cm and >2.06 g cm^{-3} in the 20–30 cm depth) as well as values below the 1st quartile minus 3 times the IQR (corresponding to a bulk density of <0.33 g cm^{-3} in the top 10 cm, and <0.46 g/cm^3 in the 20–30 cm depth) were omitted. The greater tolerance for extremely low bulk densities was due to the presence of organic soils (Organosols and Hydrosols, WRB Histosols) in the dataset, which would have been excluded using the same tolerance as for the upper limits. A total of 1401 data points remained in the processed dataset for analysis (Fig. 1).

SOC storage and vertical distribution

SOC storage was assessed using several variables, namely %$_{SOC}$ and C$_{stock}$ in each depth increment (0-10, 10–20, 20–30 cm), as well as the sum of %$_{SOC}$ and C$_{stock}$ over the three depth increments. Although assessed individually, we hereafter refer to %$_{SOC}$ and C$_{stock}$ combined as SOC content. To assess the vertical
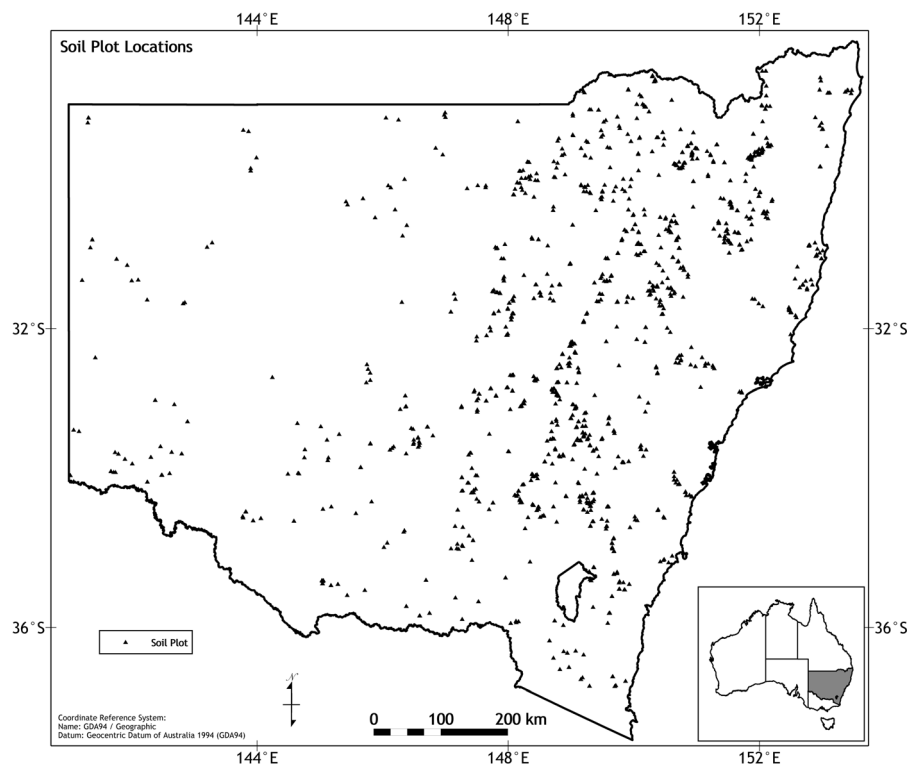

Fig. 1 The study region and location of sampling sites

distribution of SOC, numerous indicators were investigated, namely:

The proportion of C in the top 10 cm to the sum of C at the three depth intervals:

$$C_{proportion} = \frac{SOC_{0-10\ cm}}{\sum SOC_{0-30\ cm}} \qquad (2)$$

the ratio of C in the upper depth to lower depth:

$$C_{ratio} = \frac{SOC_{0-10\ cm}}{SOC_{20-30\ cm}} \qquad (3)$$

and the difference between C in the upper 10 cm and lower 10 cm normalised to the amount of C in the lower depth:

$$\Delta C_{ratio} = \frac{SOC_{0-10\ cm} - SOC_{20-30\ cm}}{SOC_{20-30\ cm}} \qquad (4)$$

The $C_{proportion}$ is best related to SOC in the top 10 cm (which contributes most to SOC in the 0–30 cm depth) and implies that the vertical distribution of SOC is strongly dependent on SOC near the surface (i.e. production). In contrast, the C_{ratio} gives a greater importance to the SOC content below the surface, and implies that SOC ratio is more influenced by C retention in the soil. The ΔC_{ratio} is an intermediate between these two measures. These three indices were calculated for both $\%_{SOC}$ and C_{stocks}, and are hereafter collectively referred to as $C_{gradient}$.

Identifying factors important to SOC storage
and vertical distribution

To identify the key drivers of SOC storage and vertical distribution a machine-learning procedure similar to a random forest (Breiman 2001) was employed. Numerous (500) regression trees were bagged to an ensemble and the variable importance of the predictors extracted to identify the drivers of C storage and vertical distribution. Tree ensembles were grown using the conditional inference forest algorithm (cforest command in the party 1.0–15 package, Hothorn et al. 2006) with the R language for statistical computing (version 3.1.1, R Core Team 2014). In growing a conditional inference tree, the decision to split a node is based upon the outcome of a test of the global null hypothesis of independence between the response variable and the predictor variables selected for splitting a node. If the null hypothesis can be rejected at the specified significance level ($\alpha=0.05$), the node is

split. Otherwise, tree growth is terminated. This approach eliminates the possibility of overgrown trees (and therefore the need for tree pruning), and overcomes the bias of traditional random forests to split upon categorical variables with many factors, or along continuous variables with a broader scale (Strobl et al. 2007).

Tree ensembles were generated using natural and log-transformed data for both the above-mentioned C variables and the available (continuous) predictor variables (Table 1), resulting in three models (natural, log, mixed) for each response variable. Log-transformation of selected variables was undertaken to both stabilise variance in and linearise the relationship between response and (continuous) predictor variables. Log-transformation of a variable was applied where it led to an improvement in the correlation between response and predictor variable.

Model performance was assessed using explained variance defined by the coefficient of determination (R^2):

$$R^2 = 1 - \frac{MSE}{Variance} \qquad (5)$$

with *MSE* the mean square error of the average of individual estimates for each tree in an ensemble, and *Variance* the variance of the modelled response variable. As our aim was not to produce predictive models but merely describe the factors influencing SOC storage and vertical distribution, we assessed model performance using the fitted, not predicted values.

From these models, each predictor's variable importance, VI, (determined both from model accuracy and area under curve) was extracted from each model. Where numerous, highly collinear (Pearson's product moment $r>0.9$) predictor variables describing similar factors (e.g. MAT_{max}, MAT_{ave} and MAT_{min}, or EVAP and VPD) were indicated as important in the models, the lower ranked were eliminated and new ensembles grown to produce final models without high correlation of important variables as this may affect variable importance results (Nicodemus et al. 2010). From these final models, VI was extracted and averaged over the models to generate the final results.

Assessing the influence of important factors influencing
SOC storage and vertical distribution

To assess the influence of the important factors driving SOC storage and vertical distribution, the results of the

Table 1 Predictor variables used to fit tree ensemble models. (C): continuous variable, F (n): categorical variable with n levels

Variable name	Type (levels)	Explanation	Source and further information
Climate variables			
CLMZone	F (4)	Climate zone defined according to temperature and humidity	A: Australian Bureau of Meteorology (BOM) Published scale: ~2×~2 km (0.025 degrees) Based on a standard 30-year climatology (1961–1990)
EVAP	C	Evaporation (mm year^{-1})	B: Australian Department of Environment Published scale: ~1×~1 km 1970–2012 monthly averages.
KoppenCode	F (10)	Detailed classification of Köppen climate zone based on a modified Köppen classification system defined for Australia	A
KoppenGroup	F (4)	Major Köppen Classification Group	A
KoppenMP	F (7)	Revised international Köppen-Geiger climate classes	(Peel et al. 2007) Published scale: ~11×~11 km (0.1×0.1 degrees)
MAP	C	Mean annual precipitation (mm year^{-1})	B
MARH	C	Mean annual relative humidity (%, 9 am data: $MARH_{9am}$, 3 pm data: $MARH_{3pm}$)	BOM Published scale: ~10×~10 km (0.1 degrees) Based on a 30-year climatology (1976–2005)
MAT	C	Mean annual temperature (°C, average: MAT_{ave}, minimum: MAT_{min} and maximum: MAT_{max})	B
T_{diff}	C	MAT_{max}–MAT_{min}	Authors
$T_{diffnorm}$	C	T_{diff} divided by MAT_{ave}	Authors
T_{ratio}	C	Ratio of MAT_{max} to MAT_{min}	Authors
VPD	C	Vapour pressure deficit (kPa)	B
SrnGRP	F (4)	Major Seasonal Rainfall Zones based on median annual rainfall (November to April and May to October) and seasonal incidence of rainfall.	C: BOM Published scale: ~25×~25 km (0.25 degrees) Based on a 100-year period (1900–1999)
SrnAll_CODE	F (10)	Detailed Seasonal Rainfall Zones	C
Land-use variables			
LMcodes	F (8)	Merged categories from the LUOrig due to discrepancies between the two separate databases. Eight categories were defined covering different land-use and management intensities: native vegetation, tree crops, native pasture, introduced pasture, crop-pasture rotation, non-conventional cropping (minimum/no-till), irrigated cropping, conventional cropping.	Authors
LUClass	F (4)	Simplified land use classes based upon LUOrig, OEH.LU.Det, and via visual detection using satellite imagery from Bing and Google maps. Eight categories were defined: native vegetation, tree crops, cleared/grazed sites, cropped sites.	Authors
LUOrig	F (38)	Original land use description in MER/SCRP dataset.	Data obtained from land-owners during sampling
OEH.LU.Det	F (33)	The detailed land use category according to the Australian Land Use and Management (ALUM) Classification	D: NSW Office of Environment and Heritage (OEH) Compiled: June 2000–June 2007 Published scale: 1: 50,000
OEH.LU.Maj	F (22)	The major land use category according to the ALUM Classification	D

Table 1 (continued)

Variable name	Type (levels)	Explanation	Source and further information
Site variables			
ASC	F (14)	Australian soil class from map of soil types across NSW using the Australian Soils Classification at Order level, derived from GSG	E: OEH Published scale: 1:250,000
Elev	C	Elevation (m above sea level)	F: Surface Geology of Australia Geoscience Australia 2012 Published scale: 1:1,000,000
ρ	C	Bulk density of fine soil (g cm^{-3}) in a given sampling depth	Original datasets
GSG	F (33)	Great soil group from map of soil types across NSW using the Great Soil Group Classification.	E
Lith	F (37)	A summary description of the lithological composition of the geologic unit	F
PlotSym	F (47)	Geologic unit displayed on a map	F
PM	F (8)	Parent material classification	Authors
RM	C	Gravimetric rock content (g g^{-1}) in a given sampling depth	Original datasets
sdL	F (49)	Single dominant lithology of geological unit	NSW Geological Dataset Department of Mineral Resources Published scale: 1:250,000
SI	F (16)	Silica index derived from primary mineralogy and soil type	Authors
TWI	C	Topographic wetness index (dimensionless units)	CSIRO Published scale: one second of arc (approx. 30 m)

tree ensemble data-mining exercise were used to inform classical, parametric-based data analyses. The aim hereby was not to develop predictive models, but to assess and compare the influence of those predictor variables indicated by the tree ensembles as important to SOC storage and vertical distribution.

For each response variable, multiple regressions were created using only those predictor variables whose relative importance, $VI_{i, rel} = \frac{VI_i}{\sum_1^k VI_i} \times 100\%$, was greater than that expected from a theoretical model where all predictor variables are equally important (i.e. $VI = \frac{1}{k} \rightarrow 100\%$ for k predictors in the tree ensembles). The models were created with a mixture of natural and log-transformed data, based upon a correlation analysis of the best relationships between response and predictor variables. Log-transformation was not applied to ΔC_{ratios} due to the presence of negative values in the dataset. The models were built using the variable order indicated by the rank of the VI_{rel} from the tree

ensembles. An ANOVA of each regression model was then performed, and non-significant variables ($p > 0.05$) dropped from the models to build the final model. From these final models, the relative contribution of climate, site and land-use variables (Table 1) were calculated as the total sum of squares from the respective predictors divided by the model sum of squares.

The influence of continuous variables on SOC variables was assessed via the coefficients of each model. For categorical variables, a combination of partial regression, ANOVA and the Games-Howell post-hoc test (an extension of Tukey's Honestly Significant Differences which adjusts for unbalanced groups and unequal variance) of the relevant variable was performed after controlling for other important variables. For example, to assess the importance of land-use on SOC, partial regressions of important land-use variables were performed whilst controlling for climate and site variables (i.e. the other important variables indicated by the tree ensembles) and the Games-Howell test used to detect

significant differences between land-uses. To do this, multiple regression models were created for each SOC variable to its important environmental and site factors and the residuals from these models were then regressed to the important land-use factors and an ANOVA performed. The Games-Howell post-hoc analysis was then undertaken on the results to identify significant differences between land-use types. A comparatively conservative significance level of $p < 0.01$ was chosen due to the large number of observations in the dataset.

Results

SOC storage and vertical distribution

SOC concentration and stocks were characterised by large variance and non-normal distribution, exhibiting a positive skew. C_{stock} in the 0–30 cm ranged from 14 to 203 t ha^{-1} with a coefficient of variation (CV) of >50 % for the entire dataset (Table 2). Relative variance was even larger for $\%_{SOC}$ with a CV around 80 %. At the site scale, the average $\%_{SOC}$ CV was 24 % for the three depths (range 0.5–155 %), while the average for C_{stocks} was 29 % (range 0.5–156 %). Generally, SOC concentrations and stocks declined with depth, though not at all sites, indicated by the negative values observed as minima for ΔC_{ratios} (Table 2). Of the $C_{gradients}$, the proportions had the lowest, the ratios the highest relative variance.

Tree ensemble models

The SOC content nearest the surface and integrated over all depths were best fit by the models, which explained 76 % of observed variance (Table 3). Explained model variance was poorer for the C_{ratios} at 67 %. Model performance decreased with increasing soil depth.

Bulk density and MAP were important explanatory variables for $\%_{SOC}$ at all depths, and these variables were top ranked at all depths except 20–30 cm. In the top 10 cm, vapour pressure deficit (VPD) and mean annual relative humidity at 3 pm (MARH$_{3pm}$) were important, but at 10–20 and 20–30 cm Great Soil Group (GSG) and the land-use recorded during sampling (LUorig) emerged as important splitting variables (Fig. 2). At a depth of 20–30 cm, LUorig was the most important variable in the models. Other factors identified as important in some models were the categories describing climate classes and seasonality

(KoppenCode, ClmZone, SrnAll_Code—all depths), as well as the geological/lithological categories PlotSym and sdL (20–30 cm depth).

In contrast, bulk density was not an important splitting variable for C_{stocks} (Fig. 2). The variable importance rankings for C_{stocks} revealed LUorig, MAP, and KoppenCode as important variables for the models at all depths (0–10, 10–20, 20–30, and 0–30 cm). LUorig was the most important splitting variable at all depths, with the exception of the top 10 cm, where VPD was most important. Other important splitting variables were relative humidity (MARH$_{3pm}$, all depths except 20–30 cm), soil type (GSG, all depths, though not in all models of 0–10 cm), seasonal rainfall (SrnAll_Code, all depths except 20–30 cm), geology (PlotSym) and rock content (20–30 cm) and lithology (Lith, some models at 20–30 cm).

Generally, the importance of climate variables and descriptors dominated in the top 10 cm, and with increasing depth land-use and geological descriptors became more important for both $\%_{SOC}$ and C_{stocks}.

In contrast to the differences observable in variable importance rankings between $\%_{SOC}$ and C_{stocks}, all measures of vertical distribution of SOC yielded LUorig as the most important predictor variable, with a normalised importance of up to 10 times greater than expected. Further important variables in all models were temperature, geology, Köppen class, and evaporation (MAT$_{max}$, PlotSym, KoppenCode, EVAP, Fig. 3). In some models, soil type (GSG) and lithology (Lith) were indicated as being of marginal importance (normalised importance less than twice the expected value).

Multiple regression analyses

Model performance and factors contributing to SOC storage, stocks and vertical distribution

The multiple regressions best explained the variance in $\%_{SOC}$, and performance declined for the C_{stocks} and the indicators of C vertical distribution. The amount of variance explained by the models declined with increasing soil depth (Table 4). Between 49 % ($C_{stock\ 20–30}$) and 74 % ($\%_{SOC\ 0–10}$) of total variance in the SOC variables was explained by the multiple regression models. Of the explained variance in the $\%_{SOC}$ models, climate factors contributed the largest amount in the top 0–10 cm, accounting for 51 % of explained variance, with 35 % of attributable to site factors and 14 % to land-use. With increasing depth, the importance of climate factors to

Table 2 SOC concentration, stocks and vertical distribution indicators

	SOC concentration (%)				C_{stock} (t ha^{-1})				SOC concentration (%)			C_{stock} (t ha^{-1})		
	Depth (cm)								Depth distribution indicator					
	0–10	10–20	20–30	∑0–30	0–10	10–20	20–30	0–30	Proportion	Ratio	Δ_{ratio}	Proportion	Ratio	Δ_{ratio}
Minimum	0.43	0.19	0.09	0.92	5	3	1	14	0.21	0.34	−0.45	0.13	0.19	−0.54
Maximum	19.21	10.71	9.87	36.36	106	61	54	203	0.82	16.92	0.76	0.80	17.71	0.75
Range	18.79	10.52	9.78	35.45	102	58	53	190	0.62	16.58	1.20	0.67	17.52	1.29
Median	1.69	0.86	0.66	3.24	22	12	10	44	0.51	2.41	0.30	0.48	2.11	0.25
Mean	2.08	1.10	0.85	4.03	25	15	12	51	0.51	2.93	0.30	0.48	2.56	0.25
SE mean	0.04	0.02	0.02	0.08	0.4	0.2	0.2	0.7	0.00	0.05	0.00	0.00	0.04	0.00
95 % CI mean	0.09	0.05	0.04	0.16	0.7	0.5	0.4	1.5	0.01	0.10	0.01	0.01	0.09	0.01
Variance	2.73	0.75	0.47	8.96	199.9	79.9	58.4	773.5	0.01	3.56	0.03	0.01	2.69	0.03
Standard deviation	1.65	0.86	0.68	2.99	14.1	8.9	7.6	27.8	0.11	1.89	0.18	0.11	1.64	0.18
CV	0.80	0.78	0.81	0.74	0.57	0.60	0.65	0.54	0.21	0.64	0.61	0.22	0.64	0.71

explained variance decreased and the relative contribution of both site and land-use factors increased, so that at 20–30 cm depth, climate factors contributed 8 % of explained variance in %$_{SOC}$ models, with 61 % attributable to site and 30 % attributable to land-use. The amount of variance attributable to the climate, site and land-use for the C_{stocks} differed, but the general pattern was similar, with climate influence dominating in the top 10 cm, and land-use and site becoming more important with increasing soil depth (Table 4).

Table 3 Performance of tree ensembles models

Variable	R^2
%$_{SOC\ 0–10}$	0.75
%$_{SOC\ 10–20}$	0.74
%$_{SOC\ 20–30}$	0.71
∑%$_{SOC\ 0–30}$	0.75
$C_{stock\ 0–10}$	0.74
$C_{stock\ 10–20}$	0.71
$C_{stock\ 20–30}$	0.70
$C_{stock\ 0–30}$	0.76
%$_{SOC\ proportion}$	0.76
%$_{SOC\ ratio}$	0.67
Δ%$_{SOC\ ratio}$	0.75
$C_{stock\ proportion}$	0.75
$C_{stock\ ratio}$	0.67
$\Delta C_{stock\ ratio}$	0.75

All models of the $C_{gradients}$ (i.e., the indices representing the vertical distribution of SOC) yielded similar results, with 56–58 % of variance explained by the multiple regressions (Table 4). Land-use was the most important influence on SOC vertical distribution, accounting for 68±4 % of explained variance, with climate accounting for 22±6 % and site factors accounting for 10±4 % (averages±standard deviation of the individual $C_{gradient}$ results).

Influences on SOC storage and stocks

Of the continuous variables included in the multiple regressions, MAP and MARH$_{3pm}$ were positively associated with %$_{SOC}$ and C_{stocks}, whereas VPD was negatively associated with both %$_{SOC}$ and C_{stocks}. Bulk density was negatively associated with %$_{SOC}$.

Of the categorical variables, only Köppen climate classes (KoppenCode) and the land-use recorded during sampling (LUorig) are considered here, as they were important in all tree ensemble models and significant in the multiple regressions. Although GSG was significant in many models, the Games-Howell post-hoc analysis failed to identify significant differences between soil groups.

After controlling for other variables, the Games-Howell post-hoc analysis of %$_{SOC}$ and C_{stocks} to Köppen climate classes indicated generally greater SOC in temperate than in either subtropical, grassland or desert
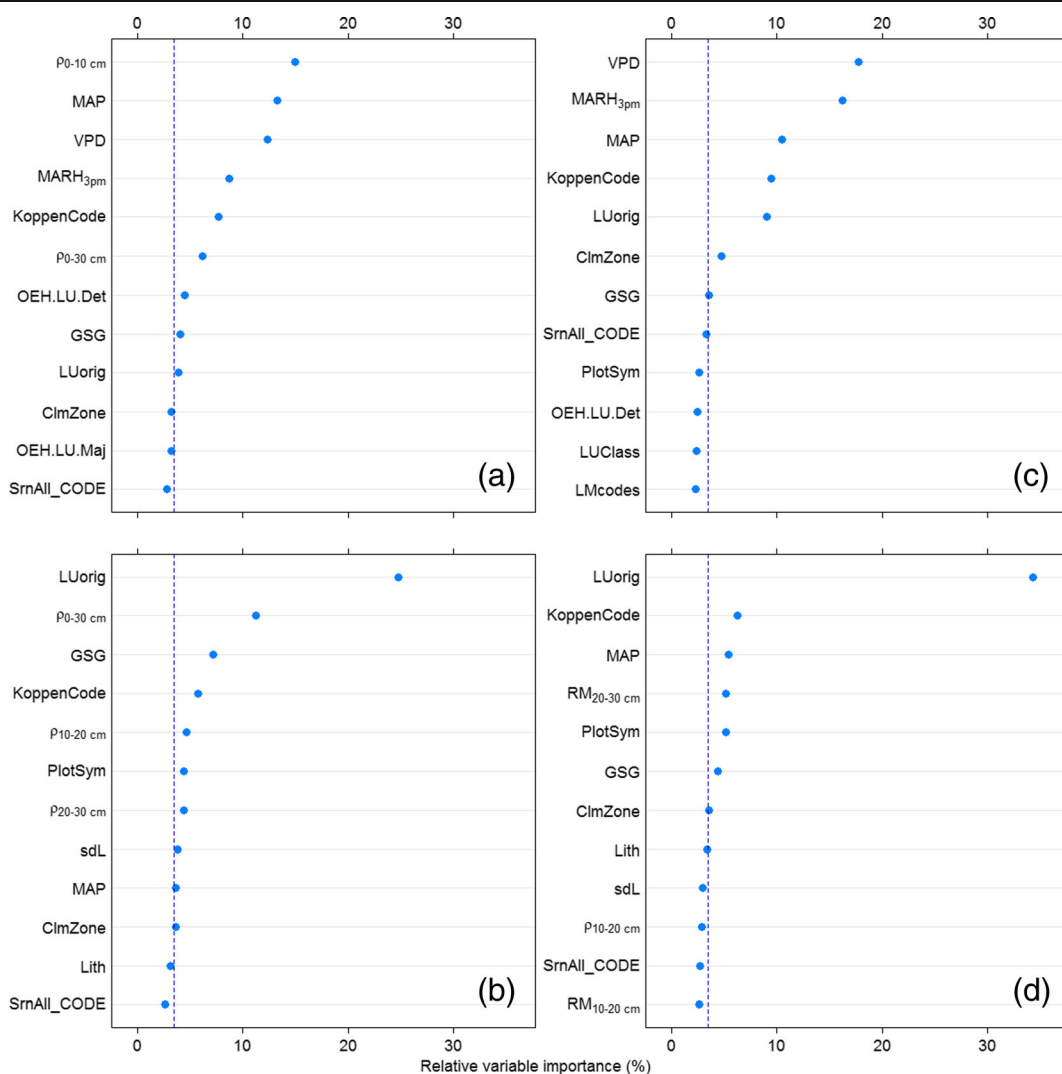
Fig. 2 Relative variable importance from the regression tree models for SOC concentration and stocks. (a) %$_{SOC}$ at 0–10 cm, (b) %$_{SOC}$ at 20–30 cm, (c) C$_{stocks}$ at 0–10 cm and (d) C$_{stocks}$ 20– 30 cm. The *blue dotted line* represents the expected variable importance in a model where all variables are equally ranked. For an explanation of variables, see Table 1

climates (Table 5). Differences were not consistent for all depths, nor when considering both %$_{SOC}$ and C$_{stocks}$.

After controlling for other variables, the Games-Howell post-hoc analysis of %$_{SOC}$ and C$_{stocks}$ to land-use indicated the highest C concentrations and stocks in native systems (unused timber/scrub and native grasses). C concentrations and stocks declined in the order: unused timber/scrub/native grasses > improved pasture ≅ grazing systems (including low-grazed systems, native pastures, set-stocking systems and rotational grazing systems) > softwood plantation ≅ modern farming systems (including rotational crop-pasture systems, minimum till

systems and carbon farming systems) > tillage cropping > irrigated cotton ≅ dryland cropping.

Influences on SOC vertical distribution

The two continuous variables indicated as important in all the tree ensemble models and therefore included in the multiple regressions were temperature and evaporation (MAT$_{max}$ and EVAP), and both were negatively associated with the C$_{gradient}$. Of the categorical variables, only Köppen climate classes (KoppenCode) and land-use at sampling (LUorig) are considered here, as they

were important in all models and significant in the multiple regressions. Although GSG was important in the tree ensemble models for the $C_{gradients}$ of %$_{SOC}$, no significant differences between different soil types were found using the Games-Howell post-hoc test.

After controlling for other variables, the Games-Howell post-hoc analysis of %$_{SOC}$ and C_{stocks} to Köppen climate classes revealed fewer significant differences in $C_{gradients}$ than in %$_{SOC}$ or C_{stocks}. More significant differences were found in vertical distribution of C_{stocks} than in the vertical distribution of %$_{SOC}$ (Table 5).

Similarly, fewer significant differences were found for analyses of land-use in the $C_{gradients}$ than in %$_{SOC}$ and C_{stocks}. Smaller $C_{gradients}$ were found under cropped systems than under most other systems. $C_{gradients}$ declined in order: grazing systems (including improved pasture, low-grazed systems, native pastures, set-stocking systems and rotational grazing systems) $\cong$ modern farming systems (including rotational crop-pasture systems, minimum till systems and carbon farming systems) > cropping.
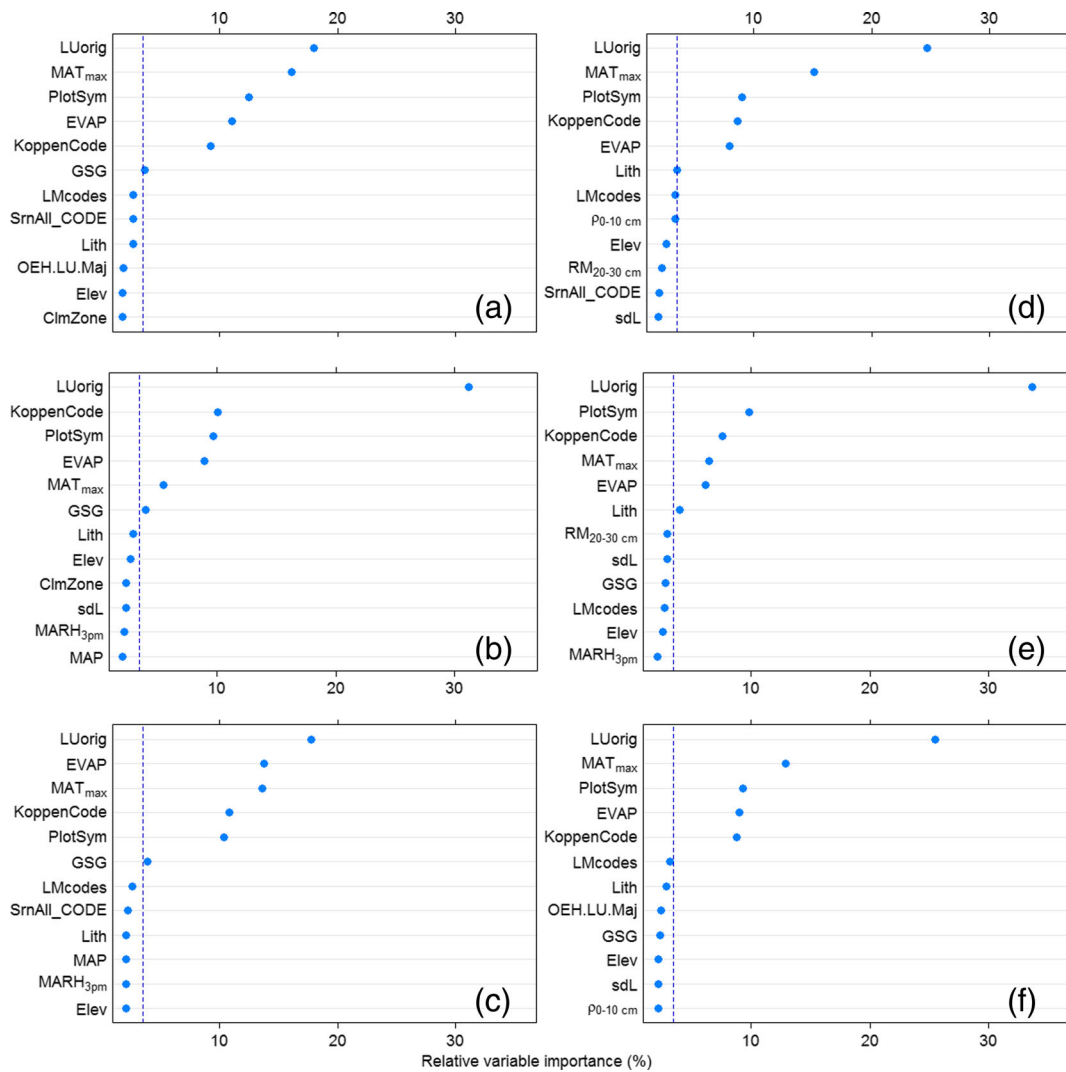

Fig. 3 Relative variable importance from the regression tree models for vertical distribution indicators. (a) %$_{SOC,proportion}$, (b) %$_{SOC,ratio}$, (c) Δ%$_{SOC,ratio}$, (d) $C_{stock\ proportion}$, (e) $C_{stock\ ratio}$, and (f) $\Delta C_{stock,ratio}$. The *blue line* represents the expected variable importance in a model where all variables are equally ranked. For an explanation of variables, see Table 1

Table 4 Multiple regression models

Variable	Final model	Adjusted R^2	Proportion of explained variance		
		All	Climate	Site	Land-use
$\%_{SOC\ 0–10}$	$\log(\%_{SOC\ 0–10}) \sim \log(BD_{0–10}) + \log(MAP) + VPD + MARH_{3pm} +$ KoppenCode + OEH.LU.Det + GSG + LUorig	0.74	0.51	0.35	0.14
$\%_{SOC\ 10–20}$	$\%_{SOC\ 10–20} \sim \log(BD_{0–10}) + MAP + GSG + LUorig + SrnAll\_CODE +$ $\log(BD_{0–10})$ + KoppenCode + ClmZone	0.66	0.25	0.58	0.17
$\%_{SOC\ 20–30}$	$\%_{SOC\ 20–30} \sim LUorig + \log(BD_{0–30}) + GSG + KoppenCode + PlotSym + sdL +$ ClmZone + MAP	0.59	0.08	0.61	0.30
$\sum\%_{SOC\ 0–30}$	$\log(\sum\%_{SOC\ 0–30}) \sim \log(MAP) + \log(BD_{0–10}) + GSG + \log(BD_{0–30}) +$ KoppenCode + LUorig + SrnAll_CODE + MARH_{3pm}	0.69	0.52	0.28	0.20
$C_{stock\ 0–10}$	$\log(C_{stock\ 0–10}) \sim VPD + MARH_{3pm} + \log(MAP) + KoppenCode + LUorig +$ GSG + SrnAll_CODE	0.65	0.80	0.03	0.17
$C_{stock\ 10–20}$	$C_{stock\ 0–10} \sim LUorig + MAP + GSG + KoppenCode + SrnAll\_CODE +$ ClmZone + $\log(BD_{0–10})$ + MARH_{3pm}	0.56	0.43	0.17	0.39
$C_{stock\ 20–30}$	$C_{stock\ 20–30} \sim LUorig + KoppenCode + MAP + RM_{20–30} + PlotSym + GSG$	0.49	0.28	0.17	0.55
$C_{stock\ 0–30}$	$\log(C_{stock\ 0–30}) \sim LUorig + MAP + KoppenCode + GSG + SrnAll\_CODE +$ VPD + MARH_{3pm}	0.60	0.46	0.07	0.47
$\%_{SOC\ proportion}$	$\%_{SOC\ proportion} \sim LUorig + MAT_{max} + PlotSym + EVAP + KoppenCode + GSG$	0.58	0.26	0.10	0.64
$C_{stock\ proportion}$	$C_{stock\ proportion} \sim LUorig + MAT_{max} + PlotSym + KoppenCode + EVAP$	0.55	0.22	0.06	0.72
$\%_{SOC\ ratio}$	$\log(\%_{SOC\ ratio}) \sim LUorig + KoppenCode + PlotSym + EVAP + MAT_{max} + GSG$	0.57	0.25	0.09	0.66
$C_{stock\ ratio}$	$\log(C_{stock\ ratio}) \sim LUorig + PlotSym + KoppenCode + MAT_{max} + EVAP$	0.57	0.10	0.16	0.74
$\Delta\%_{SOC\ ratio}$	$\Delta\%_{SOC\ ratio} \sim LUorig + EVAP + MAT_{max} + KoppenCode + PlotSym + GSG$	0.58	0.28	0.08	0.64
$\Delta C_{stock\ ratio}$	$\Delta C_{stock\ ratio} \sim LUorig + MAT_{max} + PlotSym + EVAP + KoppenCode$	0.56	0.22	0.05	0.72

Discussion

Model performance

For the tree ensembles, the R^2 indicate a good explanation of variance between measured and modelled SOC concentration and stocks (e.g. Wiesmeier et al. 2014), indicating their suitability at describing the observable variance in SOC dependent upon the explanatory variables available from sampling and site (GIS) data. The gap between modelled and total variance (~25–30 %) resembled closely the CV at site level (for 10 replicates per site CV averaged ~25 %, data not shown), so that our models appear not only adequate at explaining variance but do not overfit the data. There was no large difference in fitted model performance between the two measures of SOC content: $\%_{SOC}$ performed slightly better for the individual depths, but for the combined 0–30 cm depth interval, C_{stocks} were described slightly better, indicating that the models perform equally well at describing trends in both SOC concentration and stocks.

With the exception of $\%_{SOC}$ at 0–10 cm, the multiple regression models generally explained less variance than the tree ensemble models. However, given the fact that they were designed to analyse the influence of specific variables on SOC storage and vertical distribution, and not as predictive models their lower performance is not of great import.

Factors influencing SOC storage and vertical distribution

The tree ensemble models indicated similarities and differences in the variables important to SOC concentration and storage. Notably, although bulk density was very important to SOC concentrations, it was not an important factor determining C stocks. Relationships between SOC concentration and bulk density are well established and it is often assumed that SOC concentration affects bulk density (Ruehlmann and Körschens 2009). However, in compacted soils water infiltration, gas exchange, plant growth and root penetration are

limited, inhibiting SOC production so that bulk density also affects SOC. As such, bulk density can be viewed as important indicator of soil physical health with a strong influence on SOC. That bulk density was important to SOC concentrations but not stocks, despite its occurrence in the formula used to calculate C_{stocks}, is likely a result of the inverse relationship between SOC concentrations and bulk density. As SOC concentration increases, bulk density decreases. However, the relationship between SOC concentration and bulk density is not linear, so that the product of these two variables is not clearly delineated rendering bulk density unimportant to predicting C_{stocks}. Due to the nature of the highly diverse soils and fixed sampling depths in the original datasets, we calculated SOC stocks without adjusting for the effects of SOC on bulk density and therefore sampling depth via equivalent soil mass (ESM) equation (Ellert and Bettany 1995). It is possible that using ESM to calculate SOC stocks may further illuminate the importance of bulk density to soil physical health and SOC stocks, and this should be the focus of future investigations.

The fact that the variables identified as important in the tree ensemble models differed (absolutely and in rank) between $\%_{SOC}$ and C_{stocks} accounts for the different contributions to explained variance in the multiple regression models of the different factors (climate, site, land-use) for SOC concentration and stocks. A further explanation for the different contributions of the various factors to model variance is the differences in total variance explained by the models (SOC concentrations are better modelled than C_{stocks}), which leads to a shift in the relative contribution of a given factor to the explained variance. In this dataset, $\%_{SOC}$ appears influenced by bulk density, whereas C_{stocks}, with one exception, are not, with the result that the multiple regression models describe C_{stocks} less effectively than $\%_{SOC}$, and the influence of site factors (which include bulk density) is greater for $\%_{SOC}$ than for C_{stocks}.

Similar to other investigations into SOC storage at different depths (Jobbagy and Jackson 2000; Badgery et al. 2013), climate factors dominated as important explanatory variables for both SOC concentrations and stocks near the surface, with precipitation, vapour

Table 5 Games-Howell post-hoc analysis on the partial regression of Köppen climate classes to C indicator variables, controlled for other explanatory variables in the model

Köppen classes compared	$\%_{SOC}$				C_{stock}				Depth distribution	
	Depth (cm)									
	0–10	10–20	20–30	0–30	0–10	10–20	20–30	0–30	$\%_{SOC}$	C_{stock}
Temperate, no dry season, hot summer: grassland, hot, persistently dry	+	+	0	+	+	+	0	+	0	0
Temperate, no dry season, mild summer: grassland, hot, persistently dry	0	+	0	+	0	+	0	+	0	−
Temperate, no dry season, mild summer: grassland, warm, persistently dry	0	0	0	+	0	0	0	0	−	−
Temperate, no dry season, mild summer: subtropical, moderately dry winter	0	0	0	+	0	0	0	+	0	0
Temperate, no dry season, mild summer: subtropical, no dry season	0	0	0	0	0	0	0	0	0	−
Temperate, no dry season, mild summer: temperate, no dry season, hot summer	0	0	0	0	0	0	0	0	−	−
Temperate, no dry season, warm summer: temperate, no dry season, hot summer	−	0	0	−	−	−	0	−	0	0
Temperate, no dry season, warm summer: temperate, no dry season, mild summer	0	0	0	−	0	0	0	−	0	+

+ indicates a higher value of the C-indicator variable in the first-mentioned Köppen climate class than in the second-mentioned Köppen climate class; − indicates a lower value of the C-indicator variable in the first-mentioned Köppen climate class than in the second-mentioned Köppen climate class; 0 indicates no significant difference ($p < 0.01$) between zones. Only comparisons with a significant difference in either the surface or subsurface depths are reported

pressure deficit and relative humidity highly important in the top 10 cm but less important below the surface. Consistent with our understanding of the relationships between rainfall, net primary production (Michaletz et al. 2014) and SOC storage (Jobbagy and Jackson 2000; Wiesmeier et al. 2013b), MAP was positively associated with SOC content, whereas VPD was negatively associated with SOC content. These results suggest that SOC production near the surface is limited by water availability. The effect of water-limitation will be most noticeable in SOC content at the soil surface, as dryness throughout the soil profile will not only limit below-ground SOC production (i.e. root growth and exudates), but also above-ground plant productivity, e.g. leaf-litter production. Therefore, water-stress is likely to affect the SOC content at the soil surface to a greater degree than below the surface. Compounding this is the fact that soil temperatures are greatest near the surface, resulting in a higher evaporation and drier soils, which will further limit SOC production in shallow-rooting vegetation types (e.g. pasture).

Despite long-standing links between temperature and ecosystem productivity (Michaletz et al. 2014), temperature was not an important variable in the tree ensemble models of SOC content. However, temperature was highly important to the vertical distribution of SOC. Temperature has been shown to have a stronger association with SOC vertical distribution than precipitation (Wang et al. 2004), and we believe that our results indicate that SOC vertical distribution is driven more by degradation than by production processes. Microbial activity and SOC decomposition rates are positively associated with temperature, but are also limited by substrate availability (Kirschbaum 2006). Although water-availability will limit microbial activity (VPD is an important splitting variable in the tree ensembles), precipitation directly limits SOC production, and will therefore limit microbial activity and SOC degradation indirectly by limiting substrate availability. As the SOC content is generally greatest near the surface, higher temperatures will lead to enhanced turnover of SOC near the surface, where substrate limitation is less likely to affect microbial activity, resulting in a lower gradient of SOC from the surface to subsurface and thereby reducing the $C_{gradients}$.

With increasing soil depth, the Köppen climates classes (and other categorical climate descriptors) became more important to the models. This indicates that the absolute amount of rainfall, evaporation and humidity are less important than the climate patterns (i.e. seasonality, temperature regime). This can be explained by reduced water and temperature fluctuations with increasing depth from the soil surface, which dampens the absolute climate signal and results in SOC dynamics below the surface being driven by seasonal trends.

With increasing soil depth, the influence of climate variables diminished and site and land-use factors became more important to SOC storage. That site factors such as soil type, lithology and geology become more important with depth can be explained by the fact that the ability of the soil inorganic matrix to retain SOC is linked with mineralogy and texture (Six et al. 2002), which are derived from bedrock and weathering properties and are inherent to the sampling site. Near the surface, SOC content is highest and so fine minerals are most likely to be saturated with SOC, limiting their retention capacity. Below the surface SOC content generally decreases, so that saturation of fine minerals is less likely to be an issue and mineralogy becomes more relevant to SOC storage. This is consistent with the results of Grimm et al. (2008), who found that subsoil (10–50 cm) SOC variance was best explained by soil textural classes as derived from soil mapping units. These results imply that, for the environment and sites represented in this dataset, the amount of SOC produced (i.e. near the surface) at a site is limited predominantly by climate, above-all precipitation and water-availability, but that C retention (i.e. below the surface) is more closely linked with geological and mineralogical site properties than with climate variables.

Mineralogy and (the closely associated) soil texture have long been recognised as relevant to SOC retention (Christensen 1992), particularly in Australian soils (Hassink 1997). In our study we did not have access to soil textural data or measured mineralogical properties, so can only assess the importance of mineralogy via the (proxy) GIS variables related to geology and lithology. Our results highlight their importance to SOC storage below the surface and we strongly recommend the assessment of textural and mineralogical data in SOC research projects so that the importance of these variables can be better understood in future research.

It is notable that the site factors identified as important in the models were inherent to internal soil properties, i.e. bulk density, soil type, geology and lithology, and not descriptive of external (topographical) site properties such as elevation or TWI. Topography is one of Jenny's soil forming factors (Jenny 1941) and has been

found to be highly important to SOC stocks (Grimm et al. 2008; Davy and Koen 2013). However, those studies assessed considerably smaller areas, characterised by lower climate variance, but highly varied topographies. Minasny et al. (2013) suggest that local terrain attributes (slope, aspect, curvatures) are important at small scales (<100 m) but that position in landscape is more important at larger scales (>100 m). Our results suggest that despite the fact that topography varies greatly across NSW, for the scale of this study, which encompasses a large area around the size of Germany, Poland, the Czech Republic, and Austria combined, topographical features are not important drivers of SOC storage and vertical distribution. Instead, climate, geology and humans have the greatest influence on SOC stocks and vertical distribution.

Traditionally, land-use has been reported to affect SOC content near the surface (0–10 cm), with effects more difficult to detect deeper in the soil (Wilson et al. 2008, 2010; Luo et al. 2010). However, the tree ensemble models all indicated land-use as highly important to C content below the surface and the contribution of land-use to explained variance in the multiple regression models more than doubled from the surface 10 cm to the 20–30 cm depth. Further evidence for the importance of land-use in determining subsurface OC content is the identification of land-use as by far the most important variable to C vertical distribution in the tree ensemble models, and its accounting for around two-thirds of the variance in the multiple regression models of SOC vertical distribution.

That land-use becomes important with depth may reflect stratification of SOC down the soil profile under different land-use systems as a result of differential accumulation or loss or SOC at a different depths down the soil profile, for example accumulation of SOC at the surface of unused timber/scrub areas, indicated by their higher SOC content and gradients. For pasture systems, it appears that grazing may reduce the input of SOC at greater soil depths (via removal of plant matter at the surface, thereby lowering translocation), resulting in lower SOC content below the surface and therefore higher vertical distribution indicators. Furthermore, the deposition of animal faeces onto the soil surface will help to enrich the soil surface in organic matter, enhancing the gradient from surface to subsurface. Alternatively, land-use can affect vertical distribution via mixing of the soil, as indicated by the low $C_{gradient}$ and SOC content in conventionally cropped systems.

Land-use not only affected the vertical distribution of SOC to a depth of 30 cm, but also the absolute SOC storage, with native, unused systems storing significantly larger amounts of C than anthropogenically managed systems. These results confirm the current understanding of the effect of land-use and soil disturbance on soil carbon stores, namely that compared with natural systems, the greatest depletion of SOC occurs in highly disturbed, cropped and tilled systems, with modern farming methods (including rotational crop-pasture systems, minimum till systems and carbon farming systems) and grazed systems not as greatly depleted in SOC.

Finally, the results and discussion related to the vertical distribution of SOC have been drawn on data from the top 30 cm of soil. As outlined in the Introduction, Australian soils are unique and it is likely that the 'topsoil' (i.e. active/reactive soil zone) can adequately be assessed in the 0–10 cm depth. Although tillage depth in the State is generally 10 cm, occasionally deeper tillage to depths of 20–25 cm are used for specific purposes (e.g. the placement of 'deep' P fertiliser), which may affect the indices used for assessing the factors influencing SOC vertical distribution of SOC, as will the rooting depth of grasses. Although many Australian soils are shallow, with depths of <1 m (Soil and Landscape Grid of Australia 2014) and exhibit poor profile development, generalisation of these results to entire soil profiles should be treated with caution. Instead, the results and discussion presented here can be considered a starting point to assess the drivers of the depth distribution of SOC in eastern Australia. Specifically, the hypotheses developed based upon these results, namely that climate is important near the surface but that geological and mineralogical characteristics become more important in subsoils, that precipitation is the largest climatic driver of SOC stocks, and that the depth distribution is driven more by temperature than by precipitation should be tested in future research exploring greater soil depths.

Conclusions

We investigated the drivers of soil organic carbon (SOC) storage at three depths (0–10, 10–20, and 20–30 cm) as well as the gradient of SOC from the surface to subsurface in the soils of New South Wales, Australia using a combination of machine-learning and classical statistics.

Our results indicate that the storage of SOC near the surface is driven predominantly by climate, most specifically water availability, indicated by the positive association of SOC with precipitation and relative humidity, but negative association with vapour pressure deficit. With increasing soil depth the influence of climate waned and inherent site factors (bulk density, soil type, geology and lithology) and land-use became more important to SOC content. Below the surface, seasonality and climate regimes appear to be more important than absolute values of precipitation and vapour pressure deficit. This we attribute to the dampening of the absolute climate signal (e.g. temperature fluctuations, wetting–drying cycles) at greater soil depths. Temperature was not important to SOC content, but highly important to SOC vertical distribution, and we hypothesise that this is a result of the temperature dependence of microbial activity and SOC degradation. In warmer climates, microbial activity is enhanced, and is greatest near the soil surface, where SOC content is highest. Thus, warmer temperatures lead to a comparatively greater turnover of surface SOC and a correspondingly lower gradient of SOC from surface to sub-soil than in colder climates. Land-use affects the absolute storage of SOC, with natural systems containing the greatest amounts of carbon and conventionally cropped systems the least. Grazed systems and modern farming systems (e.g. rotational cropping-grazing or carbon farming systems) had intermediate SOC contents between these end-members. Importantly, land-use is the most important predictor of the vertical distribution of SOC in the investigated depths, which can be attributed to preferential accumulation at the soil surface (unused timber/scrub systems), reduced input into the sub-soil (grazed systems) or mixing of the soil profile (cropping). Lastly, our study assessed SOC storage and vertical distribution in the top 30 cm of soil. Although this is a good starting point for investigating SOC storage and vertical distribution, future research efforts should be focussed on testing the hypotheses developed here regarding the relative importance of the numerous drivers of SOC at multiple depths as well as the depth distribution of SOC in deeper soil profiles.

References

Badgery WB, Simmons AT, Murphy BM, Rawson A, Andersson KO, Lonergan VE, van de Ven R (2013) Relationship between environmental and land-use variables on soil carbon levels at the regional scale in central New South Wales, Australia. Soil Res 51:645–656

Baldock JA, Sanderman J, Macdonald LM, Puccini A, Hawke B, Szarvas S, McGowan J (2013) Quantifying the allocation of soil organic carbon to biologically significant fractions. Soil Res 51:561–576

Breiman L (2001) Random forests. Mach Learn 45:5–32

Burke IC, Yonker CM, Parton WJ, Cole CV, Schimel DS, Flach K (1989) Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils. Soil Sci Soc Am J 53:800–805

Charman PEV, Murphy BW (eds) (2007) Soils their properties and management, 3rd edn. Oxford University Press, South Melbourne

Christensen BT (1992) Physical fractionation of soil and organic matter in primary particle size and density separates. Adv Soil Sci 20:1–90

Cotching WE (2012) Carbon stocks in Tasmanian soils. Soil Res 50:83–90

Davy MC, Koen TB (2013) Variations in soil organic carbon for two soil types and six land uses in the Murray Catchment, New South Wales, Australia. Soil Res 51:631–644

Dawson JJC, Smith P (2007) Carbon losses from soil and its consequences for land-use management. Sci Total Environ 382:165–190

Ellert BH, Bettany JR (1995) Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Can J Soil Sci 75:529–538

Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis. Geoderma 146:102–113

Hassink J (1997) The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil 191:77–87

Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 15:651–674

Isbell R (2002) The Australian soil classification. CSIRO Publishing, Melbourne

Jackson RB et al (2000) Belowground consequences of vegetation change and their treatment in models. Ecol Appl 10:470–483

Jackson RB, Banner JL, Jobbagy EG, Pockman WT, Wall DH (2002) Ecosystem carbon loss with woody plant invasion of grasslands. Nature 418:623–626

Jenny H (1941) Factors of soil formation: a system of quantitative pedology. McGraw-Hill, New York

Jobbagy EG, Jackson RB (2000) The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol Appl 10:423–436

Kirschbaum MUF (2006) The temperature dependence of organic-matter decomposition—still a topic of debate. Soil Biol Biochem 38:2510–2518

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

Meersmans J, Van Wesemael B, De Ridder F, Fallas Dotti M, De Baets S, Van Molle M (2009) Changes in organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. Glob Chang Biol 15:2739–2750

Michaletz ST, Cheng D, Kerkhoff AJ, Enquist BJ (2014) Convergence of terrestrial plant production across global climate gradients. Nature. doi:10.1038/nature13470

Minasny B, McBratney AB, Malone BP, Wheeler I (2013) Digital mapping of soil carbon. Adv Agron 118:1–47

Nicodemus KK, Malley JD, Strobl C, Ziegler A (2010) The behaviour of random forest permutation-based variable importance measures under predictor correlation. BMC Bioinforma 11:110

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci Discuss 4:439–473

Post WM, Emanuel WR, Zinke PJ, Stangenberger AG (1982) Soil carbon pools and world life zones. Nature 298:156–159

R Core Team (2014) R: A language for statistical computing 3.1.1. R Foundation for Statistical Computing, Vienna

Rabbi SMF et al (2014) The relationships between land uses, soil management practices, and soil carbon fractions in South Eastern Australia. Agric Ecosyst Environ 197:41–52

Richards G (2001) The FullCAM Carbon Accounting Model: Development, Calibration and Implementation. Carbon accounting and emissions trading related to bioenergy, wood products and carbon sequestration: 1

Ruehlmann J, Körschens M (2009) Calculating the effect of soil organic matter concentration on soil bulk density. Soil Sci Soc Am J 73:876–885

Sahukar R, Gallery C, Smart J, Mitchell P (2003) The bioregions of New South Wales. NSW National Parks and Wildlife Service, Hurstville

Sanderman J et al (2013) Carbon sequestration under subtropical perennial pastures II: carbon dynamics. Soil Res 51:771–780

Schlesinger WH (1977) Carbon balance in terrestrial detritus. Annu Rev Ecol Syst 8:51–81

Six J, Conant RT, Paul EA, Paustian K (2002) Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant Soil 241:155–176

Soil and Landscape Grid National Soil Attribute Maps - Soil Depth (3″ resolution) - Release 1. v2. (2014) CSIRO. http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html

State of NSW Environmental Protection Authority (2012) New South Wales state of the environment. Office of Environment and Heritage

Statistics ABo (2013) 7121.0—Agricultural Commodities, Australia, 2012–2013

Strobl C, Boulesteix A, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinforma. doi:10.1186/1471-2105-8-25

Todd-Brown K, Luo Y (2014) Future research directions for soil carbon modeling. Eos 95:371

Wang S, Huang M, Shao X, Mickler RA, Li K, Ji J (2004) Vertical distribution of soil organic carbon in China. Environ Manag 33:S200–S209

West TO, Post WM (2002) Soil organic carbon sequestration rates by tillage and crop rotation. Soil Sci Soc Am J 66:1930–1946

Wiesmeier M et al (2013a) Amount, distribution and driving factors of soil organic carbon and nitrogen in cropland and grassland soils of southeast Germany (Bavaria). Agric Ecosyst Environ 176:39–52

Wiesmeier M et al (2013b) Storage and drivers of organic carbon in forest soils of southeast Germany (Bavaria)—implications for carbon sequestration. For Ecol Manag 295:162–172

Wiesmeier M et al (2014) Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany). Geoderma Reg 1:67–78

Wilson BR, Lonergan VE (2013) Land-use and historical management effects on soil organic carbon in grazing systems on the Northern Tablelands of New South Wales. Soil Res 51:668–679

Wilson BR, Growns I, Lemon J (2008) Land-use effects on soil properties on the north-western slopes of New South Wales: implications for soil condition assessment. Soil Res 46:359–367

Wilson BR, Barnes P, Koen TB, Ghosh S, King D (2010) Measurement and estimation of land-use effects on soil carbon and related properties for soil monitoring: a study on a basalt landscape of northern New South Wales, Australia. Soil Res 48:421–433

Wilson BR, Koen TB, Barnes P, Ghosh S, King D (2011) Soil carbon and related soil properties along a soil type and land-use intensity gradient, New South Wales, Australia. Soil Use Manag 27:437–447

Yang XM, Drury CF, Reynolds WD, Tan CS (2008) Impacts of long-term and recently imposed tillage practices on the vertical distribution of soil organic carbon. Soil Tillage Res 100:120–124

Young R, Wilson BR, McLeod M, Alston C (2005) Carbon storage in the soils and vegetation of contrasting land uses in northern New South Wales, Australia. Aust J Soil Res 43:21–31