Environmental and human influences on organic carbon fractions down the soil profile

Eleanor U. HOBLEY*¹

Jeff BALDOCK²

Brian WILSON^{1,3}

¹ University of New England, Armidale, New South Wales, Australia

²CSIRO Agriculture, Urrbrae, South Australia, Australia

³ New South Wales Office of Environment and Heritage, Armidale, New South Wales,
Australia

* Corresponding author:

Currently at: Technische Universität München, 85354, Weihenstephan, Germany

nellie.hobley@wzw.tum.de

Tel: +49 8161 71 4186

Keypoints

- Three fractions of soil organic carbon down the whole profile were modelled
- Proportions of fractions are primarily influenced by soil depth
- Land-use and management are not important to the proportions of the fractions

Abstract

We investigated the stability of whole profile soil organic carbon (SOC) based upon three mid-infrared predicted fractions – particulate organic carbon (POC), humus organic carbon (HOC) and resistant organic carbon (ROC) – at 100 sites across eastern Australia. Our aim was to identify the controls on SOC stability down the whole soil profile, in particular relating to climate, site and human influences. To do this we used three data-mining algorithms (randomForests, gradient boosting machines and multiplicative adaptive regression splines) to identify and assess the controls on the relative proportions of the three fractions down the soil profile. Depth was the key influence on all three fractions, with the proportion of POC decreasing, and the proportion of HOC carbon increasing with increasing depth. SOC was strongly linked with POC, suggesting that the soils in the region are input driven. HOC and ROC were controlled additionally by climate and soil physico-chemical properties (e.g. clay content, pH), with SOC being less important to these fractions. Human influences (land-use and management) were not important to the proportion of the fractions, implying that the controls humans can exert on SOC stability in these environments may be limited.

Keywords: land-use, climate, machine learning, randomForests, gradient boosting machines, multiplicative adaptive regression splines

1. Introduction

Soil organic carbon (SOC) represents a large component of the global carbon cycle, with around 1500 Pg of organic carbon stored in the top meter of soil (Jobbagy and Jackson, 2000), and between 2344 (Jobbagy and Jackson, 2000) and 2400 Pg (Batjes, 1996) stored in the top 2 – 3 m of soils worldwide. Due to its origins in plant photosynthesis and loss via heterogenic respiration (predominately via soil microbial activity) it is a highly dynamic part of the global carbon cycle (Batjes, 1996).

Conceptually, SOC is separated into numerous pools with varying turnover rates ranging from hours to millennia (Trumbore, 1997), which are linked with the bioavailability of SOC and therefore its degradability (Baldock and Skjemstad, 2000). Mechanisms such as mineral-association (Oades, 1988), aggregation (and associated occlusion of SOC in aggregate interiors) (Tisdall and Oades, 1982) and physico-chemical transformation (e.g. via pyrolysis, Skjemstad *et al.*, 1996) can reduce the biological accessibility of organic matter, converting it into more stable forms which are resistant to degradation processes and remain in the soil environment for long time periods.

Such pools of organic matter have been successfully isolated using physico-chemical fractionation procedures (e.g. Christensen, 1992; Golchin *et al.*, 1994) which can be used to initialize soil carbon turnover models (Skjemstad *et al.*, 2004), providing a link between measurable SOC fractions and conceptualized SOC pools. Recent advancements have provided the ability to detect different carbon pools without physico-chemical fractionation of soil samples, instead relying on mid-infrared (MIR) spectra. Baldock *et al.* (2013a) applied partial least square regression (PLSR) analyses to MIR spectra to derive predictive algorithms to estimate three SOC fractions of conceptually differing bioavailability.

Particulate organic carbon (POC) was linked with fresh, more easily decomposable organic

material; humus organic carbon (HOC) represented a fraction of SOC which has been altered from its initial state via decomposition, mineral-association and aggregation processes.

Resistant organic carbon (ROC) represented a biologically resistant poly-aryl carbon fraction with long residence time in soils (Baldock *et al.*, 2013b).

These three SOC fractions are therefore conceptually associated with different processes and mechanisms driving SOC dynamics, namely production (POC, via photosynthesis, although turnover will reduce POC), turnover and retention (HOC, via microbiological activity and diagenesis with subsequent sorption on mineral surfaces and/or occlusion in aggregates) and abiotic transformation (ROC, i.e. alteration in fire). Thus POC represents a more labile pool of SOC, whereas HOC and ROC represent SOC pools of greater stability which will remain in soils for longer time periods. Identifying what drives not only the total amount of SOC in soils, but also the relative proportions of these fractions is fundamental to our understanding of SOC dynamics and the processes controlling it. This understanding will not only enable us to manage our soil carbon reserves for the future, but also to interpret SOC response to past environmental changes.

Recent work investigating the factors affecting the stock of these three fractions (POC, HOC, ROC) in eastern Australian soils showed that environmental and site properties exert the greatest influence, with human activities contributing a lesser, but still substantial amount to the variance in the fractions (Rabbi *et al.*, 2014). This is consistent with investigations into the drivers of total SOC storage in eastern Australia, which have been identified as climate, soil texture and bulk density, with the effects of land-use on SOC being negligible at the surface (Rabbi *et al.*, 2015) but becoming more important with increasing soil depth (Hobley *et al.*, 2015).

However, like many studies into SOC dynamics in Australia, these investigations focused on processes within the top 30 cm of soil. Subsoil organic carbon comprises a large (~ 70 % of SOC is located below the top 20 cm, Jobbagy and Jackson, 2000) yet still poorly understood (Rumpel and Kögel-Knabner, 2011) component of the global carbon cycle. With increasing depth from the surface, soil organic matter generally increases in age (e.g. Hobbey *et al.*, 2013; Kaiser *et al.*, 2002), implying enhanced stability, and its utilization as a carbon source by microorganisms is altered (Kramer and Gleixner, 2008) implying different carbon dynamics at depth than at the surface. Understanding soil organic carbon dynamics and its role in the global carbon cycle therefore requires investigation of both topsoils and subsoils. In this study we investigated controls on SOC dynamics and stability down the soil profile to depths of up to 1 m. Several hypotheses were tested relating to the drivers of the SOC fractions:

- That the relative contribution of POC would be driven by similar factors to the drivers of surface SOC, predominately climate variables related to water availability (Hobbey *et al.*, 2015), as this fraction is related to plant productivity. Land-use will negatively affect the relative contribution of POC to SOC via lowered input in agricultural systems (most notably in cropped systems) compared with native vegetation systems (Hobbey *et al.*, 2015; Hobbey and Wilson, 2016; Rabbi *et al.*, 2014);
- In contrast, we hypothesized that the relative contribution of HOC to SOC will be driven more by soil physico-chemical properties such as texture and pH, as this fraction is associated with SOC turnover and retention on soil minerals, which are influenced by the soil physico-chemical environment. As a driver of SOC turnover, temperature will also be important to this fraction.

- ROC will likely be affected by both climate and anthropogenic factors, as temperature, vegetation type and human fire management will dictate the (unknown) fire histories of the sites.
- Lastly, we hypothesized that, associated with the generally greater age (and by implication stability) of subsoil OC than surface OC, the relative contribution of the labile POC fraction will decrease with depth whereas relative contributions of the more stable ROC and HOC fractions will increase with depth.

To test these hypotheses we quantified the concentrations of SOC fractions as predicted from MIR spectra acquired from different soil depths of 100 soil profiles sampled at 100 sites from across New South Wales, eastern Australia and used several data-mining approaches to model and investigate the factors accounting for variations in SOC composition.

2. Methods

2.1 Soils and sites

The State of New South Wales covers a large region of over 800,000 km² in eastern Australia. Although largely within the global temperate zone, temperatures range from winter minima of <0 °C to summer maxima >45 °C depending on altitude and proximity to the coast. Mean annual precipitation (MAP) varies considerably, ranging from <200 mm yr⁻¹ in the west to >1500 mm yr⁻¹ in the North-East (Australian Government Bureau of Meteorology, 2012). For this study, 100 sites across New South Wales (Figure 1) were selected for analysis from an archive of over 1500 sites across the region (Hobley et al. 2015). Whole soil cores were sampled using a 50 mm diameter stainless steel core and subdivided into depth intervals on site. Maximum sampling depth at each was 1 m (or bedrock). The sites were selected to represent a range of MAP, mean annual temperature (MAT), and

land-uses (native systems, grazed/cleared systems, cropped systems), using the matrix described in Hobley and Wilson (2016).

In selecting the sites we attempted to create a balanced dataset with equal representation of land-use across the range of climates, whilst minimizing the variation of other factors (e.g. soil type) important to SOC storage and depth distribution. The final set of sites was, however, not entirely balanced due to the availability of the archived samples and the distribution of land-use within climate zones. Climate, land-use and soil type characteristics of the 100 site dataset are shown in Table 1.

2.2 Mid-infrared analysis

Archived samples were sieved to <2 mm and ground to <100 μm prior to analysis. Mid-infrared spectra were obtained on sample aliquots (~100 mg) in 9 mm auto-sampler cups using a Nicolet 6700 FTIR spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA) equipped with a KBr beam splitter, DTGS detector and an AutoDiff-Automated diffuse reflectance accessory (Pike Technologies, Madison, WI, USA). Spectra were averaged over 60 scans and auto-corrected for background signal intensity against a silicon carbide disk.

Estimation of the mass concentration (%_{mass}) of POC, ROC and HOC fractions was done using truncated (1030-6000 cm^{-1}), baseline-offset and mean-centered spectra according to the methodology reported in Baldock *et al.* (2013a), using an updated dataset which incorporated an additional 230 soil samples to those originally reported. The mass concentration of carbon within each of three fractions from 551 Australian soils were derived from particle-size separation combined with NMR analysis. The soils were dispersed and separated into a coarse (2000-50 μm) and a fine (<50 μm) fraction. NMR was then used to estimate the allocation of SOC to the ROC (poly-aryl carbon in the whole sample), POC (>50 μm non-ROC) and HOC (<50 μm non-ROC)(Baldock *et al.*, 2013b). The unfractionated samples

were analysed via MIR and the spectra were used to develop partial least-square regressions (PLSR) to predict the mass concentration of carbon in each fraction. The new MIR-PLSR prediction algorithms applied in this study had coefficients of determination (calculated via leave-one-out cross-validation) of 0.81, 0.87 and 0.86 for the POC, HOC and ROC fractions respectively. These predicted values for these fraction estimates are henceforth referred to as baseline (BSL) estimates. In addition to the prediction methodology reported above, fractions were predicted using a Savitzky-Golay (SG) first derivative spectral transformation with a 2nd order polynomial and 3 smoothing points. Spectral transformation was applied to calibration and validation spectra to create and validate the predictive algorithms which were then applied to the transformed spectra of the soils under investigation, yielding POC, HOC and ROC contents referred to as Savitzky-Golay (SG) estimates. These SG estimates had (leave-one-out cross-validation) coefficients of determination of 0.82, 0.87 and 0.86 for POC, HOC and ROC fractions respectively. Estimates were also obtained for results averaged (AVE) between the BSL and SG estimates. This yielded in total 9 SOC fraction estimates (three estimates – BSL, SG, AVE - of three fractions – POC, HOC, ROC).

2.3 Modelling the drivers of soil organic carbon fractions

To identify the drivers of the various fraction estimates, several data-mining techniques were applied and the results aggregated, similar to a model averaging approach. The underlying idea of this approach was that individual models have specific limitations, but combining results from several different models helps to overcome these limitations and produce more robust results. The models used for identification (and assessment) of the drivers were: randomForests (Breiman, 2001), gradient boosting machines (gbms, Friedman, 2001) and multivariate adaptive regression splines (MARS, Friedman, 1991). All algorithms are capable of modelling non-linear interactions between response and predictor variables without the

requirements of normality and homoscedasticity and can be used for variable identification purposes. Modelling was done using the R statistical software package (R Core Team, 2014).

The predictor variables used to construct the models included climate, land-use and management, soil physico-chemical properties, soil type, and topographical and geological features (Table 2). Predictors were measured directly on the soil samples, recorded on site at the time of sampling or derived subsequently from GIS datasets.

To overcome potential issues of collinearity between the concentration of the fractions in a given sample and the corresponding SOC concentration (Figure 2), the mass proportion of the individual fractions in a given sample were normalized to the sum of all three fractions (which is a proxy for SOC content) to obtain their relative enrichment (E):

$$E_{Fraction,i} = \frac{Mass\ proportion_{Fraction,i}}{\sum_i Mass\ proportion_{Fraction,i}} \quad (1)$$

This enabled investigation of the whole dataset without the need to compare individual sampling depths which were not consistent across the dataset (due to discrepancies in sampling depths from within the different projects). Furthermore, SOC content and depth (which were correlated both with each other and with the fractions) could then be investigated as potential drivers of the relative abundance of individual fractions because the sum of the fractions at each depth and SOC concentration was then 1 and therefore independent of depth and SOC content, although individual fractions may still be linked with depth and SOC (Figure S.1).

The randomForest model parameters were adopted according to the optimized models reported in (Hobley *et al.*, 2015). For each of these models, 100 randomForest models were constructed using different random seeds (which determine the random selection of response and predictor variables used in each model) so that each response variable was modelled in

total 50,000 times (each randomForest was constructed using 500 trees). All models were created using the party package in R (Strobl *et al.*, 2007). From each model the predictor variable importance was extracted and weighted by dividing by the mean squared error (MSE of prediction) and the weighted predictor variable importance was then averaged over all models. Variables with a relative importance greater than $1/N$ (with N the number of predictors in the model) were defined as being important predictors (Hobley *et al.*, 2015).

The gradient boosting machines were optimized for learning rate, depth of trees, minimum number of observations per node and number of trees in the model using 5-fold cross-validation. The optimized model parameters were then used to create a total of 50,000 individual gradient boosting machines for each fraction estimate. For each individual model, the data were split randomly into a subset for model fitting (75% of data points) and prediction (25% of data points) and the MSE and coefficient of determination ($R^2=1-\text{MSE}/\text{variance}$) calculated for both model fit and prediction. All models were created using the gbm package in R (Ridgeway, 2013). Analogous to the randomForest models, the predictor variable importance was extracted from the individual gbms, weighted by dividing by the MSE (prediction), averaged over all model runs and the predictors important to the models identified via their relative importance.

Multiplicative adaptive regression splines (MARS) of each fraction estimate were then created to assess the predictors identified with the randomForests and gbms as important to the fraction estimates. For a given fraction, important predictor variables were pooled and fed into the MARS algorithm. For highly collinear predictors ($R>0.8$), either the predictor with the greater importance was selected for the MARS algorithm (e.g. vapor pressure deficit (VPD) was selected over evaporation, $R=0.99$) or one predictor was normalized by dividing by the other (e.g. N concentration was divided by SOC concentration, $R=0.86$; and pH at 60-100 cm was divided by pH at 0-5 cm, $R=0.91$). Response variables were standardized (mean-

centered and divided by standard deviation) prior to model fitting. Models were implemented using the earth package in R (Milborrow, 2009).

The MARS models were optimized using 10 runs of 10-fold cross-validation. Models were optimized for number of terms, degrees of interaction (1, 2 or 3, representing simple additive models and interaction models allowing 2 or 3-parameter interactions) and the (minimum) number of data points between knots. The number of terms selected was based upon cross-validated out-of-fold R^2 as a function of number of model terms. If an additional term led to an increase >0.001 in the out-of-fold CV- R^2 , the term was included. This threshold corresponds to the termination threshold for term inclusion in the general (non-cross validated) model algorithm. The span between knots was either set to 1 (i.e. no limitation) or calculated internally by the algorithm as a function of the data. The models of varying degree and number of data-points between knots were then compared for each fraction estimate, based upon the generalized cross-validation R^2 , which gives an estimate of the predictive power of the models and model complexity (degree of model and the number of terms and predictor used in the model).

3. Results

3.1 MIR predictions

SOC contents predicted using BSL, SG and AVE techniques were generally well correlated ($R=0.94$ for BSL, 0.92 for SG, 0.95 for AVE estimates) to SOC contents measured using a dry combustion technique (Hobley and Wilson, 2016) up to $\sim 8\%$ but were poor at higher concentrations (Figure 2). Although this does not necessarily indicate a failure of the predictions of the individual fractions at SOC contents above this value, the reliability of such estimates cannot be assessed without further analyses (e.g. full fractionation). To avoid potentially erroneous results whilst simultaneously retaining maximal data analysis power,

sites with SOC>10% (N=6) were deemed unreliable and excluded from the analysis. The final analyzed dataset comprised a total of 606 data-points.

Overall, the HOC fraction contributed the largest proportion to bulk SOC, accounting for an average of 63 ± 15 % (mean $\pm$ standard deviation) of SOC across all samples, depths and estimates. ROC was the intermediate fraction by proportion, contributing an average of 30 ± 12 % to bulk SOC. POC represented the smallest fraction, contributing on average 7 ± 7 % to bulk SOC.

3.2 randomForests and gradient boosting machines

The randomForest models performed very well at fitting (explained variance 88-92%) and well at predicting (explained variance 74-82%) most fraction estimates (Table S.1). The individual gradient boosting machine models showed a greater variance in explanatory power, ranging from very poor to very good (explained prediction variance between -2 and 85%).

The variables selected as important by the different algorithms differed between (i) fractions, (ii) estimation methods and (iii) algorithms. Averaged across all three MIR estimates, the POC fraction was influenced predominately by sampling depth, SOC and N (analyzed using LECO), containing only a few less important predictors, mainly associated with climate (Figures 3 and 4). In contrast, models of HOC and ROC contained a greater number of important predictors, including more predictors related to climate (Köppen climate class, seasonal rainfall), soil physical properties (bulk density, texture) and soil chemical properties (pH, EC).

3.3 Multivariate adaptive regression models

The MARS model performance was good overall, accounting for 61-89% of variance in the fraction data (Table S.2). Performance improved when allowing second degree interactions between predictors compared with simple additive models, but there was generally no benefit in allowing third degree interactions compared with second degree interactions. There was no trend to better prediction of any estimation method, nor for predictions with or without limitations to the number of data points between knots. As such, results and interpretation are reported for variables which were common to second degree models of all estimates (i.e. limited and unlimited knot-span, BSL, SG and AVE estimates).

Sampling depth was the most important predictor for all fractions identified with the MARS models (Table 3). Climate variables were also important to all fractions, though with differing influence. Depth of regolith (i.e. depth to solid bedrock) and total soil depth were also important predictors in all fractions, above all for POC. In contrast to the HOC and ROC fractions, the soil physical properties (texture, bulk density) were not important to POC, and with the exception of SOC and N, neither were soil chemical properties (Figure 5).

3.4 Predictors of POC proportion

Sampling depth and SOC were primary predictors of the proportion of POC, occurring in all models with and without interaction terms (Table 2). The depth of regolith and mean annual precipitation occurred as primary predictors of POC in some models. The total soil depth and Köppen climate class occurred in interaction with other predictors. Across all models, the proportion of POC: (i) decreased with increasing depth to ~15 cm and remained stable or declined at a lower gradient thereafter; (ii) increased with increasing SOC content until a concentration of ca. ~2% and increased at a lower gradient or remained stable thereafter (Figures 6 and 7); (iii) decreased with increasing depth of regolith to ~19 m and remained stable thereafter (Supplementary Information).

3.5 Predictors of HOC proportion

The most important predictors of the proportion of HOC were sampling depth, surface (0-5 cm) clay content, MAP, the normalized subsurface (60-100 cm) clay content and TN/SOC (Table 3), although not all appeared in non-interactive terms. Additionally, variables related to soil physical (bulk density, ρ , and sand content) and chemical (pH, SOC) properties and the depth of regolith appeared in the models.

Across all models, the proportion of HOC: (i) increased with sampling depth to a depth of around 45 cm and stayed stable or increased slightly thereafter; (ii) decreased sharply with increasing SOC up to a concentration of around 0.8%, and declined with a smaller gradient thereafter; (iii) increased with surface (0-5 cm) clay content up to a maximum of ~23% and declined thereafter; (iv) increased with increasing surface (0-5 cm) pH until a maximum at around 5.5 and declined thereafter (Figure 8). A trend to higher proportions of HOC in regions of higher (above $\sim 800 \text{ mm yr}^{-1}$) rainfall was discernible, especially in shallower ($> \sim 0.5 \text{ m}$) depths (Figure 9).

3.6 Predictors of ROC proportion

The most important predictors of the proportion of ROC were sampling depth, surface (0-5 cm) clay content, MAP, TN/SOC and SOC (Table 3), although not all appeared in non-interactive terms. The variables interacting in the ROC MARS models were predominately associated with soil chemical characteristics (i.e. SOC, pH, ECE).

The proportion of ROC increased with sampling depth to $\sim 13 \text{ cm}$, and decreased thereafter to $\sim 40 \text{ cm}$ (Figure 10). Below 40 cm the trends were inconsistent across models, with some models showing increases and other decreases in ROC below 40 cm. Although the proportion of ROC varied with surface (0-5 cm) clay content, identifying relationships between clay

content and ROC were difficult, as clay content had numerous interactions with other variables (e.g. depth or pH). A consistent trend was the decrease of ROC in soils with increasing surface clay content up to ~22 % (Figure 11).

All models showed an interaction between surface clay content and MAP, with soils in higher rainfall regions with higher clay content having very low proportions of ROC (Figure 11). Clay content also interacted with SOC content in almost all models, with soils low in clay having proportionally more ROC than soils high in clay, especially in soils with low overall SOC content. A large interaction effect was observed between sampling depth and SOC content, with soils of above average SOC content at depth having larger proportions of ROC. The interaction of sampling depth with MAP revealed that areas of high rainfall have overall less ROC than areas of low rainfall.

4. Discussion

4.1 Model performance and interpretation

Although the individual data-mining models identified (at times) different variables of differing importance, all models were informative and their results lead to similar conclusions. The generally good to very good explanation of variance (c.f. Balshi *et al.*, 2009; Hobley *et al.*, 2015; Wiesmeier *et al.*, 2014), indicates the suitability of these modelling approaches for investigating SOC dynamics and fractions. Using a consensus approach based upon multiple models can help to eliminate bias of individual models which yield varied results (Araújo and New, 2007), and we recommend the use of multiple methods to identify and assess the influence of environmental and site factors on SOC stability.

Although the models employed in this study make no assumptions as to linearity or distribution of relationships between the response and the predictor variables, it is noted that

certain assumptions underlie the interpretations of the identified relationships. In particular, the relationships revealed by individual and interactive terms in the MARS models assume that the other predictors not under consideration (i.e. by coefficients or graphing) are held constant (at the median value for continuous variables and in the first coded category for factor variables). Thus, the absolute response values should be considered with caution as the predictions may lie outside of the actual data range (e.g. there were no site with 8% SOC at a depth of 1 m, but the models will predict for these values). In such instances, patterns and trends across the data range should be compared instead.

4.2 Depth is the key influence on soil organic carbon fractions down the profile

The predictors important to the proportions of the three fractions differed and present differing pictures of what influences individual fractions, but a common result was the overarching importance of depth to all three fractions.

The relative proportion of POC decreased with increasing depth, supporting our hypothesis and suggesting that this fraction was related to surface input of organic carbon. Similarly, the positive relationship between proportion of HOC and depth suggests an increasing degree of humification with increasing soil depth. This implies that in subsoils, the input of fresh organic matter (e.g. via roots) is less important to carbon storage than at the surface. Given the generally greater age (and by implication stability) of subsurface SOC (Hobley *et al.*, 2014a; Rethemeyer *et al.*, 2005; Rumpel *et al.*, 2002) and the lower likelihood of microbial activity due to the lower SOC concentrations in subsoils (Don *et al.*, 2013), this result cannot be explained by higher turnover rates of fresh (root) carbon in subsoils compared with surface soils leading to a reduction in POC proportion and increase in HOC proportion.

Instead we attribute the greater contribution of HOC to SOC in subsoils to (1) a greater retention capacity of subsoils due to the lower SOC concentration and therefore lower

saturation of the soil with regards to SOC sorption capacity (Six *et al.*, 2002), and (2) a greater relative influence of DOC translocation to subsurface SOC storage. Future research should aim to quantify the relative contributions of root derived vs. translocated SOC in subsoils.

The relationship between ROC and depth was not as clear as for the other fractions and it would appear that depth interacts with other factors (such as rainfall and clay content) to influence the contribution of ROC to total SOC (see also Section 4.6). However, that soils with an above-average SOC content at depth had higher proportions of ROC – a fraction of SOC consistent with pyrogenic origins – suggests that SOC of pyrogenic origin contributes to enhanced subsurface SOC storage (Hobley *et al.*, 2014b) and is potentially more stable at depth (Czimczik and Masiello, 2007).

4.3 Soil organic carbon concentration and profile SOC stability

Soil organic carbon concentration was highly important to and positively associated with the proportion of POC. This implies that bulk and particulate SOC generally have similar drivers and respond in similar ways to environmental and site factors, despite the fact that POC represents, on average, the smallest fraction of soil organic carbon. Although SOC concentration was also important to both HOC and ROC, its relative importance to these fractions was smaller than to the POC fractions, and there was an inverse relationship between SOC and HOC proportion. As POC represents a labile, freshly produced fraction of SOC (Skjemstad *et al.*, 2001) these results suggests that, on the whole, SOC across the region is more driven by input than by turnover and transformation processes.

The high proportion of HOC in soils with low SOC concentrations (<0.8%) also implies that relatively little SOC remains unaltered from its original state in low-C soils, which are

usually in the subsurface, providing further evidence that subsurface SOC is mainly derived from humification and transport processes.

4.4 Climate and environmental controls on SOC stability

The importance of regolith depth to POC supports the idea that geological and topographical features play a role in determining SOC dynamics (Jenny, 1941), as this variable is determined mainly by lithology, weathering intensity, climate and relief (personal communication, Ross Searle, CSIRO). In this case, the knot-point in the regolith depth at ~19 m in the MARS models is a good indicator of geography in the study region, essentially separating the Centre and Western Plains (generally with regolith depths of ~20-80 m) from the highlands of the Great Dividing Range and Eastern Seaboard, which have considerably shallower regoliths. In the Highlands and Eastern Seaboard, the negative relationship between POC and regolith depth may be an indication of the effects of erosion and deposition on SOC dynamics. However, the geographical boundary identified by the regolith depth of ~19 m can also be viewed as a proxy for climate, with the moister climates of the (cooler) highlands and (warm) Eastern Seaboard climatically distinct from the hotter, drier climates in the center and west of NSW. As such, the importance of regolith depth to POC may actually reflect the influence of rainfall on POC, and not necessarily reflect geographical and topographical influences.

Climate is an important driver of SOC fractions, with increased rainfall associated with greater proportions of HOC. This indicates that water availability is a key factor determining SOC dynamics in these soils. This may reflect a limiting influence of water-availability on SOC turnover thereby reducing production of HOC, but the greater proportion of HOC in higher rainfall regions could also result from greater translocation of DOC and subsequent retention of HOC in subsurface soil depths.

SOC turnover is traditionally thought to be more sensitive to temperature than precipitation (Lützow and Kögel-Knabner, 2009). A previous investigation of surface soils (0-30 cm) in the region found that precipitation was the key driver of total SOC, but suggested that turnover is more controlled by temperature than water availability (Hobley *et al.*, 2015). However, the results of this current study suggest that temperature does not contribute substantially to SOC turnover and dynamics down the entire soil profile but that water availability is the key driving factor in the study region. The influence of temperature may be indirectly implied by the importance of the Köppen climate classes, which are categorical descriptors of seasonal temperature, rainfall and vegetation, but temperature is not a direct influence on SOC dynamics. Therefore, our initial hypothesis that temperature is important to turnover processes was not confirmed. The air temperatures in the study region tend to be mild, with sub-zero temperatures limited to higher altitudes in winter, so that this result may be affected by the lack of colder climates in the region. Our results, although generally reflecting current understanding of SOC dynamics, are therefore not representative of larger regions around the globe which encompass colder climates.

4.5 Soil physico-chemical properties and SOC stability

In contrast to the relatively few controls on POC, the proportion of HOC was determined by more factors, including soil physical and chemical properties such as texture and pH, implying that these factors are controls on humification and/or retention processes and supporting our original hypothesis. Specifically, it would appear that there is an optimum pH (at around 5.5 assuming median values of other predictors), which results in the greatest production of HOC. Soil pH has been associated with the stocks of SOC fractions in the surface (0-30 cm) soils of eastern Australia (Rabbi *et al.*, 2014). Specifically, pH was negatively correlated with POC stocks. Our findings do not corroborate this result but do imply that pH affects humification and SOC retention processes throughout the soil profile.

Low pH has been associated with increased mean residence times of soil organic matter fractions, and substrate availability is likely to be reduced at pH<5 (Leifeld *et al.*, 2011). Our results suggest that weakly acidic soils offer optimal conditions to soil microorganisms but that alteration of the chemical environment (e.g. reduction in pH) affects the sorption of HOC to soil minerals, aggregation and therefore substrate availability.

Similarly, a clay content of ~23% (assuming median values of other predictors) appears to result in the greatest proportion of HOC. Links between soil texture and SOC storage are well established (Hassink, 1997; Schöning *et al.*, 2005) and heavy textured soils generally store larger quantities of SOC than light textured soils as a result of physical-chemical protection of SOC associated with fine mineral surfaces and occluded in aggregates (Baldock and Skjemstad, 2000; Six *et al.*, 2002). Our results support the idea that heavy-textured soils hinder degradation processes, likely a result of these stabilization mechanisms. Furthermore, they suggest that SOC degradation is not a linear function of texture but that an ‘ideal’ texture for microbial activity and retention of the resultant HOC in soils exists, possibly due to the influence of texture on surface area, structure and pore volume (i.e. living space) and water retention capacity. Texture also influences infiltration and percolation of rainwater in soils, and a reduction of rainwater percolation and therefore DOC translocation in heavy-textured soils may be reflected in the proportionally lower HOC contents of the soils at higher clay contents.

Once again, we stress that the optimal values identified are dependent upon the other variables in the model and assume median values of these other variables. As such, they should be treated with caution and do not necessarily represent ‘absolute’ values. They are, however, suitable for assessing trends and patterns in the relationships between covariates.

4.6 Controls on pyrogenic carbon

Interestingly, the greater proportion of ROC associated with higher SOC concentrations in deep soils suggests that poly-aryl organic material (consistent with the structure of charred material) may contribute substantially to subsurface SOC storage (Hobley *et al.*, 2013). This appears to be mainly in areas of low rainfall and in soils with low clay content. This suggests that texture plays an important role in the incorporation and retention of pyrogenic carbon in soil organic matter (Hobley *et al.*, 2014b) and provides support for the translocation of charcoal into subsurface depths in light-textured soils (Bird *et al.*, 1999; Czimczik and Masiello, 2007; Hobley *et al.*, 2013; Skjemstad *et al.*, 1996). In contrast, the low proportion of ROC in heavy-textured soils indicates that vertical translocation of ROC is comparatively reduced in heavy-textured soils. This suggests that the finer pore-structures of these soils may hinder the infiltration of ROC particles into the soil. As the study region included numerous Vertosols, which are characterized by shrink-swell capacities and associated crack-formation, this result indicates that redistribution via physical turbation along cracks is not very important to the vertical redistribution of SOC (Czimczik and Masiello, 2007).

The link between lower rainfall and higher ROC proportion is most likely an indicator of the importance of climate as a driving factor of fire in natural ecosystems (Mooney *et al.*, 2011). Although fire regimes are generally linked with higher ambient temperatures than with precipitation, our results suggest that pyrogenic material in soils was preferentially retained in drier rather than wetter regions. This is interesting because it is conceivable that low-rainfall will reduce translocation of particulate matter due to reduced infiltration. However, it may not be the *absolute amount* of rainfall that affects the advective vertical translocation of particles in soils but instead rainfall *intensity*. Australia is known as the continent of ‘droughts and flooding rains’ and low frequency, high intensity rainfall events may contribute to translocation of particulate SOC in the drier regions. Future research should focus on links

between precipitation and the production, retention and incorporation of pyrogenic material in soils.

4.7 A conceptual model of SOC fractions profile distribution

From these results we present a conceptual model for the distribution of these three fractions down the soil profile (Figure 12). The production of SOC is predominately in the form of POC, via shoots, leaves and roots, although input of root exudates may lead to direct production of HOC if the exudates are retained on mineral surfaces. This production is greatest near the soil surface and declines with depth as rooting intensity and litter input decreases, resulting in the distribution of POC which is very similar to SOC distribution in the soil, namely greatest content at the soil surface which declines with depth. Vertical redistribution of POC may result from bioturbation, cryoturbation, or via cracks in shrink-swelling clays, however the strong decline in proportional content of this fraction with depth indicates that these processes of redistribution are not very important to the vertical distribution of this fraction. This implies that the vertical redistribution of POC, if it occurs, does not result in enhanced protection of the redistributed POC in the subsoil and that this fraction indeed represents a labile fraction of SOC.

Turnover of POC results in the production of HOC, however, this fraction can only be retained in the soil if mineral surfaces are available for sorption and/or aggregation. At the soil surface, where SOC concentrations are highest, the soil is most likely to be saturated or near saturation (Six *et al.*, 2002), so that proportional HOC retention is reduced because mineral surfaces are not available for protecting the fraction from further degradation. However, in subsoils, turnover of the POC results in a relatively greater retention of HOC because the soil is not saturated with respect to SOC retention.

512 Additionally, dissolution of SOC at the surface results in vertical translocation via percolating
513 rainwater as DOC (Kaiser and Kalbitz, 2012; Sanderman *et al.*, 2008). In the subsoil, the
514 percolating DOC can sorb to the unsaturated mineral surfaces and be retained at greater soil
515 depths leading to an increase in the proportion of HOC in subsoils. Although other
516 mechanisms of redistribution (e.g. bioturbation, cryoturbation, or via redistribution along
517 cracks in shrink-swell soils) may also contribute to vertical redistribution, they are not limited
518 to the HOC fraction (i.e. may also redistribute POC and ROC), so that microbial activity,
519 DOC translocation and SOC sorption capacity of the soil minerals can be considered as
520 having the greatest importance to the vertical distribution of HOC. This results in the
521 proportional increase of HOC with depth, though the amount of HOC retained in a given soil
522 is limited by HOC production (i.e. turnover of POC), DOC percolation (limited by soil
523 texture and rainfall) and the type and availability of mineral surfaces for SOC retention
524 (controlled by texture and mineralogy) (Lawrence *et al.*, 2015; Torn *et al.*, 1997).

525 The production of ROC will be determined by fire frequency and intensity which are linked
526 above-all with climate (Mooney *et al.*, 2011). However, as ROC is predominately produced
527 when fire acts on the top few (~3-10) cm of soil (Chafer, 2008; Doerr *et al.*, 2006) as well as
528 above-ground biomass, its presence in subsurface depths is limited by its incorporation into
529 the subsoil via redistribution processes. As this is a particulate fraction, this necessarily
530 implies redistribution of particles (or potentially colloidal material) from the surface to
531 subsurface. Several processes are conceivably responsible for vertical translocation of
532 particles down the soil profile: bioturbation, cryoturbation, physical turbation via cracks in
533 shrink-swelling soils and advection with percolating rainwater. The effect of bioturbation and
534 cryoturbation could not be assessed here due to the lack of biological data and the relatively
535 moderate temperatures in the region. However, our results suggest that redistribution via
536 physical turbation in cracking soils is not important to incorporation of ROC into the soil

profile. In contrast, the presence of ROC in light-textured soils indicates that incorporation of this fraction into the soil may occur via advection with percolating rainwater in the larger pores of sandy soils. As such, the vertical distribution of ROC is related to its incorporation into soils, a function of soil texture and climate (above-all rainfall).

4.8 Land-use and management do not control SOC stability down the profile

Lastly, it is notable that although land-use and management have been found to influence SOC storage (Hobley *et al.*, 2015; Wilson *et al.*, 2008; Wilson *et al.*, 2011) and the stocks of SOC fractions (Rabbi *et al.*, 2014) in the top 30 cm of soil across NSW, neither land-use nor land-management factors influenced the relative proportions of the fractions down the whole soil profile.

Profile SOC depth distribution is influenced strongly by land-use and management, as is subsurface SOC concentration (Hobley and Wilson, 2016), and we hypothesized that human influence would also be noticeable in the relative proportions of the fractions down the soil profile. The lack of importance of these factors to the proportions of the individual SOC fractions therefore implies that the influence of land-use and management on SOC dynamics is not limited to a specific fraction: although individual fractions may be sensitive to land-use change at specific depths (e.g. POC near the surface, HOC at depth), when viewing the whole profile, all fractions are responsive to land-use change.

Many studies have shown that particulate organic matter and light fractions are more sensitive to land-use than other fractions (e.g. Cambardella and Elliott, 1993; Chan *et al.*, 2002; Conant *et al.*, 2004; Skjemstad *et al.*, 2001). However, it has also been suggested other indicators such as total SOC content and its depth distribution are better than specific fractions for assessing SOC response to land-use change (Leifeld and Kögel-Knabner, 2005). Furthermore, such studies have frequently been limited to surface soils (0-30 cm) and there is

a lack of understanding of subsurface SOC response to land-management practices and land-use change. Our results indicate that whole profile SOC stability, assessed by three fractions – POC, HOC and ROC – is not influenced by human activities across the study region but that all fractions respond in a similar fashion to human activities.

We hypothesize that this is because soil organic carbon dynamics are input driven. When we reduce SOC input via agricultural systems (most noticeably through cultivation), we primarily reduce the input of POC, predominately in shallower soil depths. However, this results in a reduction of HOC and ROC production, which are produced from POC (and - in the case of ROC - above-ground biomass), so that these fractions are not incorporated into the soil profile and cannot be retained in the subsoil. The net effect is that, when the whole profile is viewed, all fractions react to land-use change, albeit at different depths. This has implications for the concept of SOC stability and we recommend future investigations into whole profile SOC dynamics.

5. Conclusions

The modern statistical techniques employed here enabled us to investigate the relative allocation of SOC to three of its component fractions – particulate, humus and resistant organic carbon - down the entire soil profile without the requirements of data normality and homoscedasticity which can limit traditional statistical applications, enabling us to simultaneously examine the allocation of SOC to fractions across numerous, at times inconsistent, sampling depths and over a wide range of SOC concentrations. Furthermore, rather than limit our results to one model, we aggregated results from the different approaches to enhance the robustness of the results.

Depth was the key determinant of the allocation of SOC to its component fractions, with enhanced proportions of humus found in increasingly deep soils, concurrent with a depletion

of particulate organic carbon. This implied that subsoil organic carbon consisted largely of humified soil organic matter, retained after POC turnover in subsoils or translocated as DOC from the surface. The particulate organic matter fraction was closely associated with total SOC concentrations, suggesting that the OC in soils of the study region was input driven. Greater proportions of resistant organic carbon (associated with pyrogenic carbon forms) were found in light-textured soils at depth, implying that pyrogenic carbon may have been a substantial stock of organic carbon at depth in sandy soils. The main drivers of SOC fractions were related to climate and site factors, as well as soil physico-chemical properties. Land-use and management did not influence the relative proportions of the fractions down the soil profile. This implies that the manipulation of soil carbon stocks by humans with a view to increase stable carbon forms of high longevity in soils may be limited, as all carbon fractions appear to respond similarly to human influence.

6. Acknowledgements

This research was supported by funding from the Australian Government Department of Agriculture. Thank you to Bruce Hawke, Kristen Clancy and Mano Veeragathipillai for laboratory work and to Arjan Wilkie for the map.

7. Bibliography

- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42-47.
- Baldock, J.A., Hawke, B., Sanderman, J., Macdonald, L.M., 2013a. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. *Soil Research* 51, 577-595.

607 Baldock, J.A., Sanderman, J., Macdonald, L.M., Puccini, A., Hawke, B., Szarvas, S.,
 608 McGowan, J., 2013b. Quantifying the allocation of soil organic carbon to biologically
 609 significant fractions. *Soil Research* 51, 561-576.

610 Baldock, J.A., Skjemstad, J.O., 2000. Role of the soil matrix and minerals in protecting
 611 natural organic materials against biological attack. *Org. Geochem.* 31, 697-710.

612 Balshi, M.S., McGuire, A.D., Duffy, P., Flannigan, M., Walsh, J., Melillo, J., 2009.
 613 Assessing the response of area burned to changing climate in western boreal North America
 614 using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biol.*
 615 15, 578-600.

616 Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. *European Journal of*
 617 *Soil Science* 47, 151-163.

618 Bird, M.I., Moyo, C., Veenendaal, E.M., Lloyd, J., Frost, P., 1999. Stability of elemental
 619 carbon in a savanna soil. *Global Biogeochemical Cycles* 13, 923-932.

620 Breiman, L., 2001. Random forests. *Machine Learning* 45, 5-32.

621 Cambardella, C.A., Elliott, E.T., 1993. Methods for physical separation and characterization
 622 of soil organic matter fractions. *Geoderma* 56, 449-457.

623 Chafer, C.J., 2008. A comparison of fire severity measures: An Australian example and
 624 implications for predicting major areas of soil erosion. *CATENA* 74, 235-245.

625 Chan, K.Y., Heenan, D.P., Oates, A., 2002. Soil carbon fractions and relationship to soil
 626 quality under different tillage and stubble management. *Soil & Tillage Research* 63, 133-139.

627 Christensen, B.T., 1992. Physical fractionation of soil and organic matter in primary particle
 628 size and density separates. *Adv. Soil Sci.* 20, 1-90.

629 Conant, R., Six, J., Paustian, K., 2004. Land use effects on soil carbon fractions in the
 630 southeastern United States. II. changes in soil carbon fractions along a forest to pasture
 631 chronosequence. *Biol. Fertility Soils* 40, 194-200.

632 Czimczik, C.I., Masiello, C.A., 2007. Controls on black carbon storage in soils. *Global*
633 *Biogeochem. Cycles* 21, GB3005.

634 Doerr, S.H., Shakesby, R.A., Blake, W.H., Chafer, C.J., Humphreys, G.S., Wallbrink, P.J.,
635 2006. Effects of differing wildfire severities on soil wettability and implications for
636 hydrological response. *J. Hydrol.* 319, 295-311.

637 Don, A., Rödenbeck, C., Gleixner, G., 2013. Unexpected control of soil carbon turnover by
638 soil carbon concentration. *Env. Chem. Letters* 11, 407-413.

639 Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. 1-67.

640 Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Ann.*
641 *Stat.* 29, 1189-1232.

642 Golchin, A., Oades, J.M., Skjemstad, J.O., Clarke, P., 1994. Study of free and occluded
643 particulate organic matter in soils by solid state ¹³C Cp/MAS NMR spectroscopy and
644 scanning electron microscopy. *Australian Journal of Soil Research* 32, 285-309.

645 Hassink, J., 1997. The capacity of soils to preserve organic C and N by their association with
646 clay and silt particles. *Plant Soil* 191, 77-87.

647 Hobley, E., Willgoose, G.R., Frisia, S., Jacobsen, G., 2013. Environmental and site factors
648 controlling the vertical distribution and radiocarbon ages of organic carbon in a sandy soil.
649 *Biol. Fertility Soils* 49, 1015-1026.

650 Hobley, E., Willgoose, G.R., Frisia, S., Jacobsen, G., 2014a. Stability and storage of soil
651 organic carbon in a heavy-textured Karst soil from south-eastern Australia. *Soil Research* 52,
652 476-482.

653 Hobley, E., Willgoose, G.R., Frisia, S., Jacobsen, G., 2014b. Vertical distribution of charcoal
654 in a sandy soil: evidence from DRIFT spectra and field emission scanning electron
655 microscopy. *European Journal of Soil Science* 65, 751-762.

656 Hobley, E., Wilson, B., Wilkie, A., Gray, J., Koen, T., 2015. Drivers of soil organic carbon
 657 storage and vertical distribution in Eastern Australia. *Plant Soil* 390, 111-127.
 658 Hobley, E.U., Wilson, B., 2016. The depth distribution of organic carbon in the soils of
 659 eastern Australia. *Ecosphere* 7, DOI: 10.1002/ecs1002.1214.
 660 Jenny, H., 1941. Factors of soil formation : a system of quantitative pedology. McGraw-Hill,
 661 New York.
 662 Jobbagy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its
 663 relation to climate and vegetation. *Ecol. Appl.* 10, 423-436.
 664 Kaiser, K., Eusterhues, K., Rumpel, C., Guggenberger, G., Kögel-Knabner, I., 2002.
 665 Stabilization of organic matter by soil minerals - investigations of density and particle-size
 666 fractions from two acid forest soils. *J. Plant Nutr. Soil Sci.* 165, 451-459.
 667 Kaiser, K., Kalbitz, K., 2012. Cycling downwards – dissolved organic matter in soils. *Soil*
 668 *Biol. Biochem.* 52, 29-32.
 669 Kramer, C., Gleixner, G., 2008. Soil organic matter in soil depth profiles: Distinct carbon
 670 preferences of microbial groups during carbon transformation. *Soil Biol. Biochem.* 40, 425-
 671 433.
 672 Lawrence, C.R., Harden, J.W., Xu, X., Schulz, M.S., Trumbore, S.E., 2015. Long-term
 673 controls on soil organic carbon with depth and time: A case study from the Cowlitz River
 674 Chronosequence, WA USA. *Geoderma* 247–248, 73-87.
 675 Leifeld, J., Bassin, S., Conen, F., Hajdas, I., Egli, M., Fuhrer, J., 2011. Control of soil pH on
 676 turnover of belowground organic matter in subalpine grassland. *Biogeochemistry* 112, 59-69.
 677 Leifeld, J., Kögel-Knabner, I., 2005. Soil organic matter fractions as early indicators for
 678 carbon stock changes under different land-use? *Geoderma* 124, 143-155.
 679 Lützow, M., Kögel-Knabner, I., 2009. Temperature sensitivity of soil organic matter
 680 decomposition—what do we know? *Biol. Fertility Soils* 46, 1-15.

681 Australian Government Bureau of Meteorology, 2012. www.bom.gov.au
 682 Milborrow, S., 2009. earth: Multivariate Adaptive Regression Splines (Derived from
 683 mda:mars by Trevor Hastie and Rob Tibshirani).
 684 Mooney, S.D., Harrison, S.P., Bartlein, P.J., Daniau, A.L., Stevenson, J., Brownlie, K.C.,
 685 Buckman, S., Cupper, M., Luly, J., Black, M., Colhoun, E., D'Costa, D., Dodson, J., Haberle,
 686 S., Hope, G.S., Kershaw, P., Kenyon, C., McKenzie, M., Williams, N., 2011. Late
 687 Quaternary fire regimes of Australasia. *Quaternary Sci. Rev.* 30, 28-46.
 688 Oades, J.M., 1988. The Retention of Organic Matter in Soils. *Biogeochemistry* 5, 35-70.
 689 R Core Team, 2014. R: A language for statistical computing. R Foundation for Statistical
 690 Computing, Vienna, Austria.
 691 Rabbi, S.M.F., Tighe, M., Cowie, A., Wilson, B.R., Schwenke, G., Mcleod, M., Badgery, W.,
 692 Baldock, J.A., 2014. The relationships between land uses, soil management practices, and
 693 soil carbon fractions in South Eastern Australia. *Agriculture Ecosyst. Environ.* 197, 41-52.
 694 Rabbi, S.M.F., Tighe, M., Delgado-Baquerizo, M., Cowie, A., Robertson, F., Dalal, R., Page,
 695 K., Crawford, D., Wilson, B.R., Schwenke, G., McLeod, M., Badgery, W., Dang, Y.P., Bell,
 696 M., O'Leary, G., Liu, D.L., Baldock, J., 2015. Climate and soil properties limit the positive
 697 effects of land use reversion on carbon storage in Eastern Australia. *Sci. Rep.* 5, 17866.
 698 Rethemeyer, J., Kramer, C., Gleixner, G., John, B., Yamashita, T., Flessa, H., Andersen, N.,
 699 Nadeau, M.-J., Grootes, P.M., 2005. Transformation of organic matter in agricultural soils:
 700 radiocarbon concentration versus soil depth. *Geoderma* 128, 94-105.
 701 Ridgeway, G., 2013. gbm: Generalized Boosted Regression Models. R package version 2.1.
 702 Rumpel, C., Kögel-Knabner, I., 2011. Deep soil organic matter—a key but poorly understood
 703 component of terrestrial C cycle. *Plant Soil* 338, 143-158.

704 Rumpel, C., Kögel-Knabner, I., Bruhn, F., 2002. Vertical distribution, age, and chemical
 705 composition of organic carbon in two forest soils of different pedogenesis. *Org. Geochem.*
 706 33, 1131-1142.

707 Sanderman, J., Baldock, J., Amundson, R., 2008. Dissolved organic carbon chemistry and
 708 dynamics in contrasting forest and grassland soils. *Biogeochemistry* 89, 181-198.

709 Schöning, I., Morgenroth, G., Kögel-Knabner, I., 2005. O/N-Alkyl and Alkyl C Are
 710 Stabilised in Fine Particle Size Fractions of Forest Soils. *Biogeochemistry* 73, 475-497.

711 Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic
 712 matter: Implications for C-saturation of soils. *Plant Soil* 241, 155-176.

713 Skjemstad, J.O., Clarke, P., Taylor, J.A., Oades, J.M., McClure, S.G., 1996. The chemistry
 714 and nature of protected carbon in soil. *Australian Journal of Soil Research* 34, 251-271.

715 Skjemstad, J.O., Dalal, R.C., Janik, L.J., McGowan, J.A., 2001. Changes in chemical nature
 716 of soil organic carbon in Vertisols under wheat in south-eastern Queensland. *Australian*
 717 *Journal of Soil Research* 39, 343-359.

718 Skjemstad, J.O., Spouncer, L.R., Cowie, B., Swift, R.S., 2004. Calibration of the
 719 Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic
 720 carbon pools *Australian Journal of Soil Research* 42, 79-88.

721 Strobl, C., Boulesteix, A., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable
 722 importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 8, 25.

723 Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. *J. Soil*
 724 *Sci.* 33, 141-163.

725 Torn, M.S., Trumbore, S.E., Chadwick, O.A., Vitousek, P.M., Hendricks, D.M., 1997.
 726 Mineral control of soil organic carbon storage and turnover. *Nature* 389, 170-173.

727 Trumbore, S., 1997. Potential responses of soil organic carbon to global environmental
 728 change. *Proceedings of the National Academy of Sciences of the USA* 94, 8284-8291.

729 Wiesmeier, M., Hübner, R., Spörlein, P., Geuß, U., Hangen, E., Reischl, A., Schilling, B.,
730 von Lützow, M., Kögel-Knabner, I., 2014. Carbon sequestration potential of soils in
731 southeast Germany derived from stable soil organic carbon saturation. *Global Change Biol.*
732 20, 653-665.

733 Wilson, B.R., Grown, I., Lemon, J., 2008. Land-use effects on soil properties on the north-
734 western slopes of New South Wales: Implications for soil condition assessment. *Soil*
735 *Research* 46, 359-367.

736 Wilson, B.R., Koen, T.B., Barnes, P., Ghosh, S., King, D., 2011. Soil carbon and related soil
737 properties along a soil type and land-use intensity gradient, New South Wales, Australia. *Soil*
738 *Use Manag.* 27, 437-447.

739

740

741 List of Figures

742 Figure 1: Location of study sites

743 Figure 2: Relationships between total SOC measured via LECO analysis and predicted using
744 MIR for (a) total soil organic carbon, (b) particulate organic carbon (POC) fraction, (c)
745 humus organic carbon (HOC) fraction, (d) resistant organic carbon (ROC) fraction. The MIR
746 estimates shown are the average of the baseline-offset and Savtizky-Golay calibration
747 estimates. The dashed line in (a) represents the 1:1 fit.

748 Figure 3: Relative variable importance (variable importance of a given variable normalized to
749 the sum of variable importance of all variables in the model) of important predictors (Table
750 2) identified using the randomForest models (averaged of estimates for individual fractions)
751 for (a) POC fraction, (b) HOC fraction and (c) ROC fraction. The vertical blue lines indicate
752 the expected importance in a completely random model.

753 Figure 4: Relative variable importance (variable importance of a given variable normalized to
754 the sum of variable importance of all variables in the model) of important predictors (Table
755 2) identified using the gradient boosting machines (averaged of estimates for individual
756 fractions) for (a) POC fraction, (b) HOC fraction and (c) ROC fraction. The vertical blue
757 lines indicate the expected importance in a completely random model.

758 Figure 5: Relative influence of predictors (Table 2) selected by the MARS models on three
759 SOC fractions: POC, HOC and ROC. Relative influence of a predictor is the decrease in
760 residual sum of squares attained by the variable divided by the maximum residual sum of
761 squares decrease attained by a single variable in the model, and weighted by the proportion of
762 model subsets the variable appeared in.

763 Figure 6: Primary plots between relative proportion of POC and key covariates in MARS
764 models: (a) sampling depth, (b) soil organic carbon content.

765 Figure 7: Interaction plots between relative proportion of POC and key covariates in MARS
766 models: (a) SOC concentration and sampling depth and (b) sampling depth and mean annual
767 precipitation.

768 Figure 8: Primary plots between relative proportion of HOC and key covariates in MARS
769 models: (a) sampling depth, (b) surface clay content, (c) soil organic carbon content (%), (d)
770 surface pH (0-5 cm).

771 Figure 9: Effect of mean annual precipitation and sampling depth on the relative proportion
772 of HOC in the MARS models.

773 Figure 10: Primary plots between the proportion of ROC key covariates in MARS models
774 and: (a) sampling depth, (b) surface clay content.

775 Figure 11: Interaction plots between proportion of ROC and key predictors in MARS models:
776 (a) SOC concentration and sampling depth, (b) SOC concentration and surface clay content,
777 (c) sampling depth and mean annual precipitation, (d) clay content and mean annual
778 precipitation.

779 Figure 12: A conceptual model of the proportions of soil organic carbon fractions (POC –
780 particulate organic carbon, HOC – humus organic carbon, ROC – resistant organic carbon),
781 their vertical distribution and the factors influencing their vertical distribution.

782 Figure S.1: Proportion of soil organic carbon fractions as a function of total soil organic
783 carbon concentration and depth: (a) POC vs. SOC; (b) HOC vs. SOC; (c) ROC vs. SOC; (d)
784 POC vs. sampling depth; (e) HOC vs. sampling depth; (f) ROC vs. sampling depth.

785 Figure S.2: Interaction plots between relative proportion of POC and key covariates in MARS
786 models: (a) regolith depth and sampling depth, and (b) SOC concentration and regolith depth
787 within the Köppen climate class ‘Temperate, no dry season, warm summer’.

Table 1

Table 1: Climate, land-use and soil type characteristics of dataset

Variable	Statistic	Value		
Mean annual precipitation (mm yr ⁻¹)	Range	280 – 1402		
	1 st quartile	484		
	Median	668		
	Mean	695		
	3 rd quartile	849		
Mean annual temperature (°C)		<i>Minimum</i>	<i>Average</i>	<i>Maximum</i>
	Range	3.6 – 13.7	9.2 – 20.7	14.6 – 27.5
	1 st quartile	3.6	14.1	20.3
	Median	10.8	17.7	24.3
	Mean	9.9	16.5	23.1
	3 rd quartile	11.7	18.5	25.6
Land-use class	Native vegetation	35		
	Grazed	34		
	Cropped	31		
Land management class	Natural	36		
	Native pasture	17		
	Introduced pasture	16		
	Crop-pasture rotation	6		
	No / low till cropping	13		
	Traditional cropping	12		
Soil type (Australian Soil Classification, World Reference Base; United States Department of Agriculture in brackets)	Calcarosols	11		
	(Calcisols; Alfisols)			
	Chromosols	13		
	(Acrisols; Alfisols)			
	Dermosols	4		
	(Luvisols; Alfisols or Ultisols)			
	Ferrosols (Nitisol; Oxisols)	9		
	Kandosols (Acrisols; Alfisols)	6		
	Kurosols	20		
	(Acrisols; Alfisols or Ultisols)			
	Rudosols and Tenosols	12		
	(Arenosols/ Regosols; Entisols)			
	Sodosols	6		
	(Solonetz; Alfisols)			
	Vertosols	19		
	(Vertisols; Vertisols)			

Table 2: Predictor variables used in machine learning algorithms

Variable name	Type (Levels)	Explanation	Source and further information
Climate variables			
Climate zone	F (4)	Climate zone defined according to temperature and humidity	A: Australian Bureau of Meteorology (BOM) 2006 Published scale: ~2 km x ~2 km (0.025 degrees) Based on a standard 30-year climatology (1961-1990)
Evaporation	C	Evaporation (mm yr ⁻¹)	B: Australian Department of Environment Published scale: ~1km x ~1km 1970-2012 monthly averages.
Köppen class	F (8)	Detailed classification of Köppen climate zone based on a modified Köppen classification system defined for Australia	A
Köppen group	F (4)	Major Köppen climate classification	A
MAP	C	Mean annual precipitation (mm yr ⁻¹)	B
MARH	C	Mean annual relative humidity (% , 9 am data: MARH _{9am} , 3 pm data: MARH _{3pm})	BOM 2008 Published scale: ~10 km x ~10 km (0.1 degrees) Based on a 30-year climatology (1976-2005)
MAT	C	Average mean annual temperature (°C)	B
VPD	C	Vapour pressure deficit (kPa)	B
Seasonal rainfall	F (9)	Seasonal rainfall zones based on median annual rainfall (November to April and May to October) and seasonal incidence of rainfall.	BOM 2006 Published scale: ~25 km x ~25 km (0.25 degrees) Based on a 100-year period (1900-1999)
Summer daylight	C	Peak number of summer daylight hours derived from site latitude	Authors
Winter daylight	C	Minimum number of winter daylight hours derived from site latitude	Authors
Land-use variables			
LM codes	F (6)	Land management categories based upon original land-use recorded during sampling.	Authors
LU Class	F (3)	Land use classes based upon land-use recorded during sampling and via visual cross-checking using satellite imagery from Bing and Google maps.	Authors
Site and soil variables			

ASC	F (9)	Australian soil class from map of soil types across NSW using the Australian Soils Classification at Order level	NSW Office of Environment and Heritage 2012 Published scale: 1:250,000
AWC*	C	Available water capacity (% , 0-5 cm and 60-100 cm depths)	C: Soil and landscape grid of Australia CSIRO 2014 Published scale 3 arc seconds (approx. 90 m)
Clay*	C	Clay mass fraction (% , 0-5 cm and 60-100 cm depths)	C
ECE*	C	Cations extracted using BaCl ₂ plus exchangeable H + Al (m _{eq} /100 g) in a given soil depth (0-5 cm and 60-100 cm)	C
Elevation	C	Elevation (m above sea level)	Surface Geology of Australia Geoscience Australia 2012 Published scale: 1:1,000,000
N*	C	Sample N concentration (% _{mass} , LECO analysed)	C
P*	C	Total P concentration (% _{mass} , 0-5 cm and 60-100 cm depths)	C
pH*	C	pH in CaCl ₂ (0-5 cm and 60-100 cm depths)	C
ρ*	C	Bulk density of soil (g cm ⁻³) in a given depth	C
Sampling depth	C	Mean of upper and low sampling depths	Authors
Sand*	C	Sand mass fraction (% , 0-5 cm and 60-100 cm depths)	C
Silt*	C	Silt mass fraction (% , 0-5 cm and 60-100 cm depths)	C
Regolith depth	C	Depth of regolith to solid bedrock (m)	C
SOC	C	Sample soil organic carbon concentration (% _{mass} , LECO analysed)	Authors
Total soil depth	C	Depth of soil A and B horizons (m)	C
TWI	C	Topographic wetness index (dimensionless units)	CSIRO 2012 Published scale: one second of arc (approx. 30 m)

*Subscript ₀₋₅ indicates the 0-5 cm depth, ₁₀₀₋₂₀₀ indicates the 100-200 cm depth. Variable type: F-factor, C-continuous.

Table 3

Fraction	Predictor variables
POC	Sampling depth ¹ + SOC ¹ + Regolith Depth ² + Soil Depth + MAP ³ + Köppen class* + TN/SOC ^{3*}
HOC	Sampling depth ¹ + Clay ₀₋₅ ² + MAP + Clay ₆₀₋₁₀₀ ³ + TN/SOC + Sand ₆₀₋₁₀₀ ³ + ρ ₆₀₋₁₀₀ ^{3*} + pH ₀₋₅ ^{3*} + Regolith Depth ² + SOC ² + pH ₆₀₋₁₀₀
ROC	Sampling depth ¹ + Clay ₀₋₅ ² + MAP + N/C ² + SOC ² + EC ₀₋₅ + ρ ₀₋₅ + pH ₀₋₅ ³ + Regolith Depth* + pH ₆₀₋₁₀₀ ³

Table 3: Predictors common to all fraction estimates in the MARS models. Variables appear according to their mean variable importance ranking in the models (relative contribution to residual sum of squares reduction weighted by frequency of occurrence in model subsets). All variables appear interactively with other variables in the models. ¹ the predictor occurs as a primary predictor (i.e. without interaction) in all models. ² the predictor occurs as a primary predictor in most (>66%) models. ³ the predictor occurs as a primary predictor in some (≤50 %) models. *the predictor appeared in most (>66%) models.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11

Figure 12