Predicting carbon stocks following reforestation of pastures: a sampling scenario-based approach for testing the utility of field-measured and remotely-derived variables.

[a,*]Cavagnaro Timothy R. and [b,c]Cunningham Shaun C.

Affiliations

[a] School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, PMB1 Glen Osmond, SA, 5064, Australia.

[b] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Burwood, Vic 3125, Australia.

[c] Institute for Applied Ecology, University of Canberra, Bruce, ACT, 2617, Australia.

***Corresponding author:**

Email: timothy.cavagnaro@adelaide.edu.au

Phone: +61 8 8313 2770

Key Words: Bayesian GLM, carbon (C) sequestration, leaf litter, reforestation, soil C, soil ecology

ABSTRACT

Reforestation of agricultural lands is an important means of restoring land and sequestering C. At large scales, the labour and costs of direct measurement of ecosystem responses can be prohibitive, making the development of models valuable. Here, we develop a new sampling scenario-based modelling approach coupled with Bayesian model averaging (BMA) to build predictive models for absolute values in mixed-species woody plantings, and differences from their adjacent pasture, for litter stocks, soil C stocks, and soil C:N ratios. Modelling scenarios of increasing data availability and effort were tested. These included variables that could be derived without a site visit (e.g. location, climate, management), that were sampled in the adjacent pasture (e.g. soil C and nutrients) or were sampled in the environmental planting (e.g. vegetation, litter properties, soil C and nutrients). The predictive power of models varied considerably among C variables (litter stocks, soil C stocks and soil C:N ratios in tree plantings, and their differences to their adjacent pastures) and the model scenarios used. The use of a sampling scenario-based approach to building predictive models shows promise for monitoring changes in tree plantings following reforestation. The approach could also be readily adapted to other contexts where sampling effort for predictor variables in models is a major potential limitation to model utilization. This study demonstrates the benefit of exploring scenarios of data availability during modelling, and will be especially valuable where the sampling effort differs greatly among variables.

INTRODUCTION

Carbon trading schemes are developing around the world as a means of mitigating global climate change. While the nature of individual schemes differs, they all involve the making of payments for activities that sequester C. Therefore, monitoring and evaluating C sequestration is essential to the success of these schemes. Soils contain one of the largest pools of C on the planet, and are a major focus for sequestration activities (Parras-Alcántara & Lozano-García, 2014; Muñoz-Rojas *et al.*, 2015; Parras-Alcántara *et al.*, 2015; Novara *et al.*, 2016). This is because the global soil C pool has been reduced substantially (e.g. up to 60% decrease in temperate regions) following conversion of native vegetation to agricultural production (Lal, 2004). There is potential to reverse this decline in soil C in a number of ways, including reducing soil disturbance (Minoshima *et al.*, 2007; Parras-Alcántara & Lozano-García, 2014), adding amendments to the soil (Ng *et al.*, 2014; Srinivasarao *et al.*, 2014; Cavagnaro, 2015), and planting tress on degraded and marginal agricultural lands (Paul *et al.*, 2003; Cunningham *et al.*, 2015a).

Reforestation of agricultural lands is an important means of sequestering C (Paul *et al.*, 2002; Cunningham *et al.*, 2015a; Cavagnaro *et al.*, 2016), as well as providing habitat for native plants and animals, reducing erosion and improving water quality (Cunningham *et al.*, 2015b). Tree plantings contain three major C stocks: plant biomass (both above- and below-ground), plant litter and soil C. Stocks of C in these pools can be substantial (Novara *et al.*, 2015; Cavagnaro, 2016) with typical stocks of C in plant biomass, litter and soil being 150, 25 and 120 Mg ha^{-1} respectively in mature forests compared with 5, 1 and 90 Mg ha^{-1} respectively in agricultural fields (Cunningham *et al.*, 2015b).

Above-ground biomass of forests can be readily quantified or modelled (Paul *et al.*, 2013; Paul *et al.*, 2015). While litter stocks can be modelled (Paul *et al.*, 2003), this can be difficult for mixed-species plantings (compared with single-species plantations) due to variation in the plant species present, and their litter inputs and decomposition rates. Accurately estimating changes in soil C stocks following reforestation is challenging, and typically requires intensive field sampling (Allen *et al.*, 2010; Chaudhuri *et al.*, 2011; Cunningham *et al.*, 2012) and access to specialised analytical techniques (Baldock *et al.*, 2014).

In addition to sequestering C above- and below-ground, reforestation can change soil chemistry (Berthrong *et al.*, 2009; Jiao *et al.*, 2012; MacKay *et al.*, 2016), with potentially important consequences for C cycling and sequestration (Cunningham *et al.*, 2015b). For example, soil C:N ratios can increase with reforestation, which can affect microbially-mediated decomposition of soil organic matter (Paul, 2006; Fierer *et al.*, 2009; Harrison & Bardgett, 2010) and thence, C sequestration (Hoogmoed *et al.*, 2014b). Similarly, a shift towards a less labile C pool has been observed following reforestation (Cunningham *et al.*, 2015a).

While direct measurement is clearly the most accurate way to monitor C stocks and sequestration, it requires intensive sampling for accurate estimates and, therefore, is labour intensive and cost prohibitive over large areas. This is especially true of soil C and other soil properties (Baldock *et al.*, 2014), and for other C stocks such as leaf litter. In contrast, modelling C stocks and dynamics provides a pragmatic solution for evaluating C sequestration, as it does not require intensive field work after development. For example, models of C accounting, forest growth, and litter decomposition can be linked to predict changes in tree biomass, plant litter and soil C after reforestation (Paul *et al.*, 2003). Similar process-based models have been developed for C dynamics in agroecosystems (Parton *et al.*, 1994; Coleman & Jenkinson, 1996). Although such models are used widely in making predictions of C stocks and dynamics, as with all models, they require validation and potentially recalibration prior to use in new areas and land-use contexts.

An alternative to process-based modelling is the development of correlative models. This approach is especially useful where the response variable is difficult and/or expensive to measure, but might be predicted using easily-measured environmental variables (Jardine & Siikamäki, 2014). This approach has been used to develop predictive models for soil C on a national scale (McNeill *et* *al.*, 2014). Such modelling approaches generally consider a suite of potential predictor variables and use a step-wise reduction process to select the simplest and best performing model. However, this does not consider the relative effort required to measure different predictor variables. For example, if the predictor variables in a model require as much or more effort to collect than the response variable, then a modelling-based approach is less practical than a field survey. To address this problem, we developed and presented, a modelling approach that tested a series of sampling scenarios using combinations of variables that require different amounts of effort to obtain. Modelling scenarios could include using potential predictors that require data collected on-site, using predictors that can be derived remotely, or using both on-site and remote predictors. Ultimately, the most useful model will use easily-collected variables while performing as well as more complex models that use difficult and expensive to measure variables. The development of such predictive models of C stocks will require regional data from multiple sites, and the use of a flexible but robust modelling approach.

One of the challenges in building predictive models is that focusing on the outcome of a single model is prone to statistical bias and underestimation of model uncertainty (Wöhling & Vrugt, 2008). Bayesian model averaging (BMA) overcomes this problem by accounting for the model uncertainty inherent in the variable selection problem. This is achieved by fitting a large number of models to the data, and then averaging over the best models in the model class according to approximate posterior model probability, to provide an averaged model (Raftery *et al.*, 2009). This approach has been used successfully to deal with multivariate data sets in a range of research contexts, including soil science. For example, BMA has been used to make predictions of total soil C concentrations using loss on ignition data (Leon & Gonzalez, 2009), identify links between previous land-use, climate and soil carbon (England *et al.*, 2016), and to model vadose zone hydrology (Wöhling & Vrugt, 2008). Recently, it has been used to develop predictive models of C in mangrove soils on a global scale (Jardine & Siikamäki, 2014).

Here, we develop a sampling scenario-based modelling approach coupled with BMA. The context for this work is the development of predictive models for absolute values in mixed-species woody plantings and differences from their adjacent pasture, litter stocks, soil C stocks, and soil C:N ratios. To build these models, we used an existing survey (Cunningham *et al.*, 2015a) of 36 environmental plantings and their adjacent pastures in south-eastern Australia. Our Bayesian sampling scenario-based approach was used to identify models that predicted the C variables using a suite of environmental variables. The sampling scenario-based approach involved building models using data sets ranging from variables that required intensive field work (e.g. vegetation properties, litter stocks, and local soils information) through to variables that could be extracted from existing data bases and maps (e.g. location, climate, soil type and age). The aim was to develop models that could predict the variables of interest using variables that can be collected with minimal effort and specialised analysis, in an effort to a) build predictive models for important C stocks and soil properties, and b) to explore a sampling scenario-based modelling approach.

MATERIALS AND METHODS

Data sources

Predictive models for litter stocks, soil C stocks, soil C:N ratios in reforested pastures, and differences in soil C stocks and C:N ratios between tree plantings and their adjacent reference pastures, were built using data collected in the field and derived from a number of other sources (see below). Differences in C stocks, as well as absolute values, were included here as C accounting is based on the additionality that the difference measures provide, whereas the absolute values demonstrate actual stocks in the plantings. We have focused on litter stocks rather than litter C, because it is a relatively easily measured (e.g. by farmers) potential predictor for soil C stocks and C:N ratios. In addition, the available data set had litter stocks not C contents, and litter stocks could not be reliably converted to litter C stocks because of the high variability in the C concentration of litter among different tree species and developmental stages.

The predictor variables used here were divided into three groups: derived predictors, planting data and pasture data (Table 1). Derived predictors included latitude, longitude, soil type, soil texture, landscape position (riparian or dryland), age of tree planting, mean annual rainfall, lifetime rainfall (of the planting) and maximum temperature of the growing season. The derived predictors included data from data bases (e.g. climate variables) and GIS layers (i.e. could be obtained without visiting a site), or site information provided by the land owners (e.g. planting age). Planting data included tree canopy extent, basal area of the trees and litter stock. Pasture data included: pasture (adjacent to the planting) soil C content and pasture (adjacent to planting) soil C:N. Both planting data and pasture data predictors were calculated from samples collected in the field during site visits.

Field data

Tree plantings were on grazing properties in northern Victorian, Australia (36.5 ^{o}S 146.0 ^{o}E). The climate in this region is temperate with seasonal changes in mean monthly maximum temperature (12.6–30.8 °C) and minimum temperature (2.9–16.5 °C), and a winter-dominant annual precipitation (570-715 mm yr^{-1}, BOM, 2014). There were 36 tree plantings; 1-9 ha, 5-45 years post planting in 2010 when the soils were collected (see Cunningham *et al.*, 2015a for a full description of field sites and sampling). Ten of the plantings were in riparian and 26 in upland (i.e. non-riparian) positions. The sites were planted with a mixture of 2–15 regionally endemic trees and shrubs from the genera *Acacia* Mill., *Allocasaurina* L.A.S. Johnson, *Callistemon* R. Br., *Eucalyptus* L'Hér and *Melaleuca* L. The soils at the plantings were predominantly sodosols except for three of the riparian plantings that were on chromosols, according to the Australian Soils Classification (ABARES, 2004).

Vegetation was surveyed (austral spring to summer, 2010) using three randomly-placed plots of 900 m^2 at each planting. Stem diameter was measured at breast height (1.3 m high) for trees and at the base of shrubs (10 cm high) due to the multi-stemmed form of most shrub species. For each planting, total basal area was calculated from these diameter measurements.

In the austral winter of 2010, soil C and litter stocks were estimated at each site of the chronosequence from a single plot (400 m^2) within the tree planting and another plot in the adjacent pasture. The pastures were situated approx. 50 m from the planting, and were originally part of the same field as the tree plantings and along the same contour to minimize differences in soil type (see Cunningham *et al.*, 2015a). The adjacent pastures, which continued to be grazed by stock, were sampled to determine differences in soil organic C between land uses, and to standardize for potential differences in soil characteristics and disturbance histories among the farms. At each sampling point, plant litter was collected destructively within a 25 cm × 25 cm quadrat. Plant litter samples were air-dried for two weeks, oven-dried at 60 °C for 48 h and weighed. After litter was removed, soils were sampled at the centre of the quadrat from upper (0–5 cm) soil layer, with five independent samples collected. Additional samples were taken from three of the sampling points to measure bulk density (following, Minoshima *et al.*, 2007). Gravimetric moisture was determined after drying approx. 20 g subsample of moist soil at 105 ^{o}C for 48 h. The remainder of each sample was air dried, sieved to < 2 mm and roots ≥ 1 mm diameter were removed by manual dry picking (Damsma *et al.*, 2015). These soil samples were then ground to a fine powder and C and N content determined by dry combustion (vario MICRO cube, Elementar Analysensysteme GmbH, Hanau, Germany). Values of C concentration for each soil sample were converted to content (Mg ha^{-1}) using the mean bulk density from the appropriate site. Digital photographs of the tree canopy were taken above soil cores in the tree planting. These images were then used to estimate canopy extent by placing a 25-cell grid laid over the image and counting the number of cells dominated by canopy (Cunningham *et al.*, 2012).

Derived data

Derived variables were collated as follows. Latitude and longitude were recorded on site using a handheld GPS unit (GPS 60, Garmin). Soil type from the digital version of the Atlas of Australian Soils (ABARES, 2004) and soil texture from a GIS layer based on the Atlas of Australian Soils (McKenzie *et al.*, 2000). Historical monthly climate data (1971-2000) were obtained for each planting (Queensland Government, 2014) to estimate mean maximum temperature of the warmest six months, mean maximum temperature over the growing season, mean annual rainfall and lifetime rainfall (cumulative rainfall over time since reforestation), as potential indicators of growth rates. Landscape position (*Riparian* or *Upland*) was included to test for potential differences in C sequestration between the tree plantings established at different landscape positions, for example due to potential differences in productivity because of additional water in riparian zones.

Data analysis

Generalized linear modeling with Bayesian model averaging (GLM with BMA) was used to determine strong predictors of C variables (see Hoeting *et al.*, 1999, for a detailed description of, and background on, GLM with BMA). A key feature of GLM with BMA is that it accounts for the model uncertainty inherent in the variable selection problem by averaging over the best models in the model class according to approximate posterior model probability. Further, GLM with BMA differs from traditional (frequentist) GLM in that a probability is assigned to a given hypothesis, rather than a hypothesis being tested without a probability being assigned. BMA combines the predictions from multiple models by calculating a weighted average of those predictions. The weights used are the posterior model probabilities or the relative strengths of evidence in favour of each model (Raftery *et al.*, 1997). The important outcome is that the averaged model is likely to be substantially more generalizable beyond the build data set than a model of best fit (e.g. maximum likelihood). An environmental variable with a probability of inclusion (Pr(inc)) > 0.75 generally is considered a strong candidate for inclusion in the model and a 'key predictor' for the response variable (Thomson *et al.*, 2007). The posterior mean coefficient for a predictor is a measure of the magnitude and direction of its relationship with the response variable. Prior to undertaking BMA with GLM, the data were checked for highly-correlated variables ($r > 0.7$), as they can influence the ability of this approach to resolve coefficients values for each predictor (Thomson *et al.*, 2007) but none were identified. The 'bic.glm' function in the 'BMA' package of R was used for the analyses (Raftery *et al.*, 2008). A Gaussian error distribution and link function were used in all GLMs to allow for non-linear relationships between the response and predictors. Results from the GLMs with BMA are presented as probabilities of coefficients for predictors being non-zero, estimates of these coefficients and the strength (R^2) of relationships between the observed and predicted values of the variables of interest. We also present Bayesian Information Criterion (BIC) for each model, where models with a lower BIC value are deemed to be 'better' (Schwarz, 1978). The following scenarios of decreasing data availability (i.e. sets of Predictor Variables – see above and Table 1) were used to develop the predictive models:

<u>Model Scenario 1.</u> Potential predictor variables used in model construction were: derived predictors, planting data and pasture data.

<u>Model Scenario 2.</u> Potential predictor variables used in model construction were: derived predictors and planting data.

<u>Model Scenario 3:</u> Potential predictor variables used in model construction were: derived predictors, and pasture data.

<u>Model Scenario 4:</u> Potential predictor variables used in model construction were: derived

predictors.

These model scenarios represent different levels of effort in data collection and the need to visit field sites. For example, Model Scenario 1 requires the measurement of structural properties in the planting, sampling soils from the adjacent pasture, which may be collected as part of routine soil testing (e.g. for fertilizer decision-making) and derived data. In contrast, Model Scenario 4 only requires derived data that could be compiled without a site visit. Thus, a model based on Scenario 4 that has similar predictive power as a model based on Scenario 1, may be deemed to be a more efficient option for making predictions.

RESULTS

Predicting litter stocks.

The best model for predicting litter stock (Model Scenario 1) was based on tree basal area (Figure 1a, Table 2). When data from the adjacent pasture soils was removed (Model Scenario 2), basal area remained the only predictor of litter stocks and the model showed negligible reduction in predictive power (Figure 1b, Table 2). Model Scenario 3 (derived data and adjacent pasture soils data) and Model Scenario 4 (derived data only) predicted litter stocks on the basis of the age of the tree planting (Table 2), but performed worse (4-13% reduction in R^2 values) than Model Scenarios 1 and 2 (Figures 1 c, d).

Predicting soil C stock

Soil C stocks in the tree plantings were best predicted by Model Scenario 1, which included the predictor variables latitude, growing season maximum temperature, litter stock, and pasture soil C stock (Figure 1e, Table 3). In Model Scenario 2, soil C stock was predicted with the same remaining variables, but with less predictive power (Figure 1f, Table 3). Model Scenario 3 predicted soil C stock with planting age better than Model Scenario 2, but less well than Model Scenario 1 (Figure 1g, Table 3), using the predictor variables growing season maximum temperature, and pasture soil C stock. Model Scenario 4 predicted soil C stock using latitude, planting age and growing season maximum temperature (Figure 1h, Table 3), but performed less well than Model Scenarios 1-3.

Predicting differences in soil C stock

The difference in soil C stocks in the tree plantings was predicted by Model Scenario 1 with moderate ($0.5 < R^2 < 0.75$) predictive power (Figure 1i, Table 4), and included the predictor variables latitude, growing season maximum temperature, litter stock, and pasture soil C stock. Model Scenario 2 predicted the difference in soil C stock with longitude (Figure 1j, Table 4) but with less predictive power than Model Scenario 1. The difference in soil C stock was predicted in Model Scenario 3 by the age of the tree planting, growing season maximum temperature, and pasture soil C stock (Figure 1k, Table 4). Although Model Scenario 3 outperformed Model Scenario 2, it did not perform as well as Model Scenario 1 (compare Figures 1i-k). Model Scenario 4 predicted the difference in soil C stock based on the age of the tree planting (Figure 1l, Table 4), but had little predictive power ($R^2 < 0.5$). Importantly, the regressions of the observed versus predicted difference in soil C stocks were dominated by two data points with high leverage (see Figure 1i-l). Repeating the BMA with GLM analysis with these two sites excluded, Model Scenarios 1 and 3 predicted the difference in soil C stock using Pasture C Content as the predictor (Model Scenario 1: Pr(inc) = 1.0, Coefficient ± S.D.= -0.34 ± 0.01; Model Scenario 3: Pr(inc) = 1.0, Coefficient ± S.D.= -0.38 ± 0.01) and yielded observed versus predicted plots with R^2 values of 0.63 for both models (data not shown); these R^2 values are 3-7% lower than those for models using the full data set (i.e. compare to Figures 1i, k). Model Scenarios 2 and 4 failed to predict the difference in soil C stock when the outliers were omitted from the data set (data not shown).

Predicting soil C:N

The C:N ratio of soil from the tree planting was predicted by Model Scenario 1 with strong predictive power ($R^2 > 0.75$) by including basal area of the tree planting and the C:N ratio of the adjacent pasture soil in the model (Figure 1m, Table 5). Simplifying the model (Model Scenario 2) yielded a model that included the basal area of the trees but had a lower predictive power (Figure 1n, Table 5). In contrast, Model Scenario 3 predicted the C:N ratio of soil from the tree planting almost as well as Model Scenario 1 (5% reduction in R^2) using the age of the tree planting, and the C:N ratio of the adjacent pasture as predictor variables (Figure 1o, Table 5). Model Scenario 4 predicted the C:N ratio of soils from the tree planting on the basis of age of the tree planting but with little predictive power as indicated by a low R^2 value (Figure 1p, Table 5).

Predicting differences in soil C:N

Moderate predictions (R^2 = 0.63) of the difference in C:N ratio of soil between the tree planting and the adjacent pasture were provided by Model Scenario 1 (Figure 1q, Table 6), using basal area of the tree planting as the predictor. Model Scenario 2 performed equally well as Model Scenario 1 (same R^2 value) and included the same predictor variable (Figure 1r, Table 6). Model Scenarios 3 and 4 both performed moderately well (R^2 = 0.56 and 0.57, respectively), with both Model Scenarios including the age of the tree planting (Model Scenario 3: Figure 1s, Table 6; Model Scenario 4: Figure 1t, Table 6) as predictor variables.

DISCUSSION

Here, we present a new sampling scenario-based modelling approach for testing the utility of field-measured and remotely-derived variables. A key feature of this approach is that the best models is selected based on both the strength of prediction and the relative effort required to sample the predictor variables. For example, if two models have similar predictive power but one is based on more easily collected data, then it may present a more practical option for end-users. Using this approach, we were able to predict litter stocks, soil C stocks and soil C:N ratios in tree plantings, and differences in soil C stocks and soil C:N ratios between tree plantings and their adjacent pastures. The predictive power of the models varied considerably between variables of interest and model scenarios. For example, the models with the highest and lowest predictive power were for soil C:N ratios in tree plantings with Model Scenarios 1 and 4, respectively (Model Scenario 1 R^2 = 0.8; Model Scenario 4 R^2 = 0.3; see Figure 1). Generally, models that included all the field-measured and remotely-derived variables performed best, but this was not always the case. For some variables (litter stock, difference in soil C stock, soil C stock and difference in C:N), models built using simpler data sets had the same or similar predictive power to those built using the full set of potential predictor variables (e.g. compare observed versus predicted plots for Model Scenarios depicted in Figures 1 a-c, i & k, m & o, and q & r). The use of a scenario-based approach to build predictive models and their potential implications for monitoring soil C sequestration, are now discussed, as is the potential to use this approach to develop predictive models for other variables of interest.

Litter stocks.

In addition to being a significant C stock in tree plantings (Cunningham *et al.*, 2015b), the litter layer is an intermediary between above-ground biomass C and soil C (Attiwill & Adams, 1993; Maguire, 1994). Litter stocks in the tree plantings were predicted with moderate accuracy by a simple model based on the basal area of a tree planting (Model Scenario 2). Including information on the adjacent pasture soil did little to improve the performance of this model. The inclusion of basal area in the model likely reflects the general relationship that the amount of litter inputs increases with tree size and density (Lehtonen *et al.*, 2004). In support of this, model scenarios that did not include field estimates of vegetation in the tree plantings performed worse than those that did. In general, there was an approximate 5-10% reduction in the strength of prediction (R^2 values in observed versus predicted plots) for Model Scenarios 3 and 4 compared with Model Scenarios 1 and 2 (see Figure 1). Although the weaker models included the age of the planting as a predictor variable, which reflects the positive relationship between tree age, size and litter inputs, it is clear that age alone is not equal to information on the development of vegetation structure and species mix. Given that tree growth is influenced by factors including rainfall, temperature and soil fertility (Landsberg & Waring, 1994; Paul *et al.*, 2015), all of which varied among farms and planting lifetimes (Cunningham *et al.*, 2015a), this was not unexpected.

Although measurement of basal area requires field work, it can be collected readily without specialized equipment or training. Advances in remote sensing suggest it may be possible to remotely estimate basal area with sufficient accuracy (Lefsky *et al.*, 1999), thereby removing the need for site visits. However, the reliability of such an approach needs to be directly tested and validated, especially in evergreen forests where interference from the canopy signal may affect the accuracy of basal area estimates. Finally, it may be possible to predict basal area of trees using simple forest growth models (see Torres & Lovett, 2012). However, this would be complicated in mixed-species woody plantings where a range of species often are planted in unknown proportions, and whose relative abundance will change with stand development. Nevertheless, the development of allometric equations for a range of species commonly found in environmental plantings may help overcome this challenge (Paul *et al.*, 2013).

Soil C stocks

Both soil C stocks in the tree plantings, and the difference in soil C stocks between the tree plantings and their adjacent pastures, were best predicted by models that included data from the tree plantings, adjacent pastures and derived variables. The better performing models for soil C stock and the difference in soil C stock included information on the location, climate, litter stock in the tree planting, and pasture soil C stock. This was not unexpected given that soil C stocks are affected by the amount and nature of biomass inputs, climate (especially mean maximum temperature) and soil chemistry (Lal, 2004). Interestingly, tree canopy extent was not identified as a predictor of soil C in any of our models, as it was in earlier work on a single site in the same region (Smith *et al.*, 2012). However, this earlier study was at one farm with a strong contrast between plots with and without tree canopy, whereas the present study included a range of canopy cover among tree plantings.

Models of soil C stocks and differences in soil C stocks that included data from the adjacent pasture soil data had higher predictive power than those using vegetation data from the planting (compare Figures 1f & g and Figures 1j & k). That soil C stocks were better predicted by models that included data from the adjacent pasture suggests the importance of differences in initial C stocks before reforestation to potential C sequestration (Paul et al. 2002). They may also reflect relationships between soil C and other underlying soil properties that may affects tree growth and C accrual. Models of soil C stocks and differences in soil C stocks based only on derived data performed poorly (see Figures 1h & l). This suggests that site-specific factors, such as species choice and past or present land management, are important determinants of soil C stocks following reforestation and, therefore, when building models for a region. With inclusion of additional site-specific information and a larger data set, it may be possible to build a stronger predictive model using only derived variables, or using process-based models (Parton *et al.*, 1994; Coleman & Jenkinson, 1996; Paul *et al.*, 2003; Viaud *et al.*, 2010).

Soil C:N ratios

Of all the C variables, the C:N ratio of soils from the tree plantings was the most readily predicted. The best performing model (Model Scenario 1) included the age of the tree plantings and the C:N ratio of the adjacent pasture soils. The inclusion of planting age is consistent with earlier work showing that soil C:N ratios tend to increase with time since reforestation due to a concomitant increase in soil C and decrease in soil N with time since reforestation (Cunningham *et al.*, 2012). The inclusion of adjacent soil C:N in the models is likely to have improved model performance by accounting for site-specific variation in soil C:N ratios, which may be due to inherent differences in soil properties and/or land management, as is the case for soil C (see above). The predictive power of the models was only marginally improved by including vegetation data from the tree planting (compare Model Scenarios 1 and 3, Figure 1 m & o), but not enough to justify the additional effort and expense required to collect vegetation data. However, model predictions may be improved by including other types of vegetation data, such as relative abundance of N-fixing trees, which can affect both soil N and C levels (Ussiri *et al.*, 2006; Hoogmoed *et al.*, 2014a, b). For example, while planting trees generally increases soil C, N-fixing trees can increase soil N while non-N-fixing trees can deplete soil N to support their growth. Consequently, we would expect soil C:N ratios to decrease under N-fixing trees relative to where tree plantings are dominated by non-N-fixing trees. This requires further investigation.

Although typically not considered in C sequestration schemes, soil C:N ratios are important determinants of terrestrial C cycling via their impacts on soil biota involved in soil C cycling (Paul *et al.*, 2002; Bardgett & Wardle, 2010). For example, a higher soil C:N ratio is often associated with a more stable soil C stock (Cunningham *et al.*, 2015b). Thus, building robust models for soil C:N ratios in tree plantings is important. This may be especially true for efforts seeking to develop process models for soil C cycling at both the local and landscape scales (Viaud *et al.*, 2010).

A sampling scenario-based modelling approach.

The value of the sampling scenario-based modelling approach developed here is highlighted by the models for predicting soil C:N ratios. Model Scenario 1 (for soil C:N in the tree plantings), which was based on data derived from existing GIS layers, and collected from the tree planting and adjacent pasture, had strong predictive power, but only marginally outperformed Model Scenario 3, which did not include the field-collected vegetation data. Thus, predictions of soil C:N can be made using derived data, and adjacent pasture data, which farmers may already collect as part of the fertiliser use decision making process, negating the need for a specific site visit. However, this does assume that soil sampling strategies for making fertiliser decisions are sufficiently rigorous to account for spatial variation in soil C, which may not always be the case (Cunningham *et al.*, 2012). Although we caution here that the models developed here should not be extrapolated beyond the region from which they were developed (i.e. northern Victoria), many of the variables found to be strong predictors here should be considered as candidate predictor variables in other regions.

The sampling scenario-based modelling approach has the potential to be adapted to a wide range of other contexts where data collection represents a major limitation, such as where sites are spread over large geographic areas, difficult to access locations, or where the cost of field sampling and analysis is prohibitively expensive. This is first attempt to use a sampling scenario-based modelling approach to predict important C stocks following reforestation of former pastures. The best performing models were for stocks of soil C, plant litter stocks and soil C:N ratios. These predictions, and those for differences in soil C stocks and soil C:N ratios, are likely to be improved by a more comprehensive survey with more field sites and/or additional predictor variables. Nevertheless, this study demonstrates that with refinement, the sampling scenario-based approach has the potential to develop models for these important C stocks and soil properties. This approach may reveal that the necessary accuracy of predictions can be achieved with remotely-derived data and avoiding intensive field work.

Conclusions

Here, a robust modelling technique (BMA with GLM) was combined with a sampling scenario-based approach. The rational for this approach was to build predictive models for variables of interest with not only the least number predictor variables, but more importantly, with easily and ideally remotely-collected data. Although applied to the context of C sequestration, the sampling scenario-based approach developed here could be easily used to build predictive models for other variables in a range of different contexts, and would be especially valuable where the ease with which potential predictors variables can be collected varies.

ACKNOWLEDGEMENTS

Thanks to Dr Jim Thomson for being on hand to answer questions when troubleshooting the R code. We wish to thank the land-holders for access to their farms, and Scott MacDonald for his tireless assistance in the field and laboratory. We thank the editor and an anonymous reviewer for valuable feedback that helped to improve the manuscript. This research was funded by the Australian Research Council Linkage Program (LP0990038), Victorian Department of Sustainability and Environment, Goulburn-Broken Catchment Management Authority (CMA), Victorian EPA, Kilter Pty Ltd and the North Central CMA. TRC gratefully acknowledges the ARC for supporting his research via the award of a Future Fellowship (FT120100463).

REFERENCES

ABARES, 2004. Digital Atlas of Australian Soils. Australian Bureau of Agricultural and Resource Economics and Sciences, Department of Agriculture Fisheries and Forestry, Canberra.

Allen, DE, Pringle, MJ, Page, KL, Dalal, RC, 2010. A review of sampling designs for the measurement of soil organic carbon in Australian grazing lands. Rangeland Journal **32**: 227-246. DOI: 10.1071/RJ09043.

Attiwill, PM, Adams, MA, 1993. Tansley review No. 50. Nutrient cycling in forests. New Phytol. **124**: 561-582.

Baldock, JA, Hawke, B, Sanderman, J, L.M., M, 2014. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Research **51**: 577-595. DOI: 10.1071/SR13077.

Bardgett, RD, Wardle, DA, 2010. Aboveground-belowground linkages: biotic interactions, ecosystem processes, and global change. Oxford University Press, Oxford. DOI: 10.1111/j.1442-9993.2012.02405.x.

Berthrong, ST, Jobbagy, EG, Jackson, RB, 2009. A global meta-analysis of soil exchangeable cations, pH, carbon, and nitrogen with afforestation. Ecological Applications **19**: 2228-2241.

BOM, 2014. Climate Data Online. Australian Government Bureau of Meteorology. http://www.bom.gov.au/climate/averages/.

Cavagnaro, TR, 2015. Biologically regulated nutrient supply systems: compost and arbuscular mycorrhizas: a review. Advances in Agronomy **129**: 293-321. DOI: 10.1016/bs.agron.2014.09.005.

Cavagnaro, TR, 2016. Life at the interface: above- and below-ground responses of a grazed pasture soil to reforestation. Appl. Soil Ecol. **100**: 27-37.

Cavagnaro, TR, Cunningham, SC, Fitzpatrick, S, 2016. Pastures to woodlands: changes in soil microbial communities and carbon following reforestation. Applied Soil Ecology **107**: 24-32. DOI: 10.1016/j.apsoil.2016.05.003.

Chaudhuri, S, Pena-Yewtukhiw, EM, McDonald, LM, Skousen, J, Sperow, M, 2011. Land use effects on sample size requirements for soil organic carbon stock estimations. Soil Science **176**: 110-114. DOI: 10.1097/SS.0b013e31820a0fe2.

Coleman, K, Jenkinson, DS, 1996. RothC-26.3 – A model for the turnover of carbon in soil. In: Powlson, D.S., Smith, P., Smith, J.U. (Eds.), Evaluation of soil organic matter models – using existing long-term datasets. NATO ASI Subseries I. Springer-Verlag, Heidelberg, Germany, pp. 237–246. DOI: 10.1007/978-3-642-61094-3.

Cunningham, SC, Cavagnaro, TR, Mac Nally, R, Paul, KI, Baker, PJ, Beringer, J, Thomson, J, Thompson, RM, 2015a. Reforestation with native mixed-species plantings in a temperate continental climate effectively sequesters and stabilizes carbon within decades. Global Change Biology **21**: 1552-1566. DOI: 10.1111/gcb.12746.

Cunningham, SC, Mac Nally, R, Baker, PJ, Cavagnaro, TR, Beringer, J, Thomson, JR, Thompson, RM, 2015b. Balancing the environmental benefits of reforestation in agricultural regions. Perspectives in Plant Ecology, Evolution and Systematics **17**: 301-317. DOI: 10.1007/978-3-642-61094-3.

Cunningham, SC, Metzeling, KJ, Mac Nally, R, Thomson, JR, Cavagnaro, TR, 2012. Changes in soil carbon of pastures after afforestation with mixed species: sampling, heterogeneity and surrogates. Agriculture Ecosystems and Environment **158**: 58-65. DOI: 10.1016/j.agee.2012.05.019.

Damsma, KM, Rose, MT, Cavagnaro, TR, 2015. Landscape scale survey of indicators of soil health in grazing systems. Soil Research **53**: 154-167. DOI: 10.1071/SR14147.

England, JR, Paul, KI, Cunningham, SC, Madhacan, DB, KBaker, TG, Read, Z, Wilson, BR, Cavagnaro, TR, Lewis, T, Perring, MP, Herrmann, T, Polglase, PJ, 2016. Previous land use and climate influence differences in soil organic carbon following reforestation of agricultural land with mixed-species plantings. Agriculture, Ecosystems and Environment **277**: 61-72. DOI: 10.1016/j.agee.2016.04.026.

Fierer, N, Strickland, MS, Liptzin, D, Bradford, MA, Cleveland, CE, 2009. Global patterns in belowground communities. Ecol. Lett. **12**: 1-12.

Harrison, KA, Bardgett, RD, 2010. Influence of plant species and soil conditions on plant-soil feedback in mixed grassland communities. Journal of Ecology **98**: 384-395. DOI: 10.1111/j.1365-2745.2009.01614.x.

Hoeting, JA, Madigan, D, Raftery, AE, Volinsky, CT, 1999. Bayesian model averaging: a tutorial. Stat. Sci. **14**: 382-417.

Hoogmoed, M, Cunningham, SC, Baker, PJ, Beringer, J, Cavagnaro, TR, 2014a. Is there more soil carbon under nitrogen-fixing trees than under non-nitrogen-fixing trees in mixed-species restoration plantings? Soil Biology and Biochemistry **188**: 80-84. DOI: 10.1016/j.agee.2014.02.013.

Hoogmoed, M, Cunningham, SC, Baker, PJ, Beringer, J, Cavagnaro, TR, 2014b. N-fixing trees in restoration plantings: effects on nitrogen supply and soil microbial communities. Soil Biology and Biochemistry **77**: 203-212. DOI: 10.1016/j.soilbio.2014.06.008.

Jardine, SL, Siikamäki, JV, 2014. A global predictive model of carbon in mangrove soils. Environmental Research Letters **9**: 104013. DOI: 10.1088/1748-9326/9/10/104013.

Jiao, J, Zhang, Z, Bai, W, Jia, Y, Wang, N, 2012. Assessing the ecological success of restoration by afforestation on the Chinese Loess Plateau. Restor. Ecol. **20**: 240-249.

Lal, R, 2004. Carbon sequestration in dryland ecosystems. Environmental Management **33**: 528–544. DOI: 10.1007/s00267-003-9110-9.

Landsberg, JJ, Waring, RH, 1994. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and Management **95**: 209-228. DOI: 10.1016/S0378-1127(97)00026-1.

Lefsky, MA, Harding, D, Cohen, WB, Parker, G, Shugart, HH, 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. Remote Sensing of Environment **67**: 83-98. DOI: 10.1016/S0034-4257(98)00071-6.

Lehtonen, A, Sievänen, R, A., Mk, Mäkipää, R, Korhonen, KT, Hokkanen, T, 2004. Potential litterfall of Scots pine branches in southern Finland. Ecological Modelling **180**: 305-315. DOI: 10.1016/j.ecolmodel.2004.04.024.

Leon, A, Gonzalez, RL, 2009. Predicting soil organic carbon percentage from loss-on-ignition using Bayesian Model Averaging. Australian Journal of Soil Research **47**: 763-769. DOI: 10.1071/SR08119.

MacKay, JE, Cunningham, SC, Cavagnaro, TR, 2016. Riparian reforestation: are there changes in soil carbon and soil microbial communities? Science of the Total Environment **566-567**: 960-967. DOI: 10.1016/j.scitotenv.2016.05.045.

Maguire, DA, 1994. 1994. Branch mortality and potential Litter fall from Douglas-fir trees in stands of varying density. Forest Ecology and Management **70**: 41-53. DOI: 10.1016/0378-1127(94)90073-6.

McKenzie, NJ, Jacquier, DW, Ashton, LJ, Cresswell, HP, 2000. Estimation of Soil Properties Using the Atlas of Australian Soils. CSIRO Land and Water, Canberra, Australia.

McNeill, SJE, Golubiewski, N, Barringer, J, 2014. Development and calibration of a soil carbon inventory model for New Zealand. Soil Research **52**: 789-804. DOI: 1029/2002WR001619.

Minoshima, H, Jackson, LE, Cavagnaro, TR, Sánchez-Moreno, S, Ferris, H, Temple, SR, Mitchell, JP, 2007. Soil food webs and carbon dynamics in response to conservation tillage in legume rotations in California. Soil Science Society of America Journal **71**: 952-963.

Muñoz-Rojas, M, Jordán, A, Zavala, LM, De la Rosa, D, Abd-Elmabod, SK, Anaya-Romero, M, 2015. Impact of land use and land cover changes on organic carbon stocks in Mediterranean soils (1956-2007). Land Degradation and Development **26**: 168-179. DOI: 10.1002/ldr.2194.

Ng, E, Patti, AF, Rose, MT, Schefe, CR, Wilkinson, K, Cavagnaro, TR, 2014. Functional stoichiometry of soil microbial communities after amendment with stabilised organic matter. Soil Biology and Biochemistry **76**: 170-178. DOI: 10.1016/j.soilbio.2014.05.016.

Novara, A, Poma, I, Sarno, M, Venezia, G, Gristina, L, 2016. 2016. Long-term durum wheat-based cropping systems result in the rapid saturation of soil carbon in the Mediterranean semi-arid environment. Land Degradation and Development **27**: 612-619. DOI: 10.1002/ldr.2468

Novara, A, Rühl, J, La Mantia, T, Gristina, L, La Bella, S, Tuttolomondo, T, 2015. Litter contribution to soil organic carbon in the processes of agriculture abandon. Soild Earth **6**: 425-432. DOI: 10.5194/se-6-425-2015.

Parras-Alcántara, L, Lozano-García, B, 2014. Conventional tillage versus organic farming in relation to soil organic carbon stock in olive groves in Mediterranean rangelands (southern Spain). Solid Earth **5**: 299-311. DOI: 10.5194/se-5-299-2014.

Parras-Alcántara, L, Lozano-García, B, Brevik, EC, Cerdá, A, 2015. Soil organic carbon stocks assessment in Mediterranean natural areas: a comparison of entire soil profiles and soil control sections. Journal of Environmental Management **155**: 219-228. DOI: 10.1016/j.jenvman.2015.03.039.

Parton, WJ, Ojima, DS, Cole, CV, Schimel, DS, 1994. A general model for soil organic matter dynamics: sensitivity to litter chemistry, texture and management. In: Bryant, R.B., Arnold, R.W. (Eds.), Quantitative modeling of soil forming processes. SSSA Special Publication 39. Soil Science Society of America, Madison, WI, USA, pp. 147–167. DOI: 10.2136/sssaspecpub39.c9.

Paul, EA, 2006. Soil microbiology, ecology and biochemistry. Academic Press. DOI: 10.1111/j.1365-2389.2008.01052_2.x.

Paul, KI, Polglase, PJ, Nyakuengama, JG, Khanna, PK, 2002. Changes in soil carbon following afforestation. Forest Ecology and Management **168**: 241-257. DOI: 10.3334/CDIAC/tcm.008.

Paul, KI, Polglase, PJ, Richards, GP, 2003. Predicted change in soil carbon following afforestation or reforestation, and analysis of controlling factors by linking a C accounting model (CAMFor) to models of forest growth (3PG), litter decomposition (GENDEC) and soil C turnover (RothC). Forest Ecology and Management **177**: 485-501. DOI: 10.1016/S0378-1127(02)00454-1.

Paul, KI, Roxburgh, SH, England, JR, de Ligt, R, Larmour, JS, Brooksbank, K, Murphy, S, Ritson, P, Hobbs, T, Lewis, T, Preece, ND, Cunningham, SC, Read, Z, Clifford, D, Raison, RJ, 2015. Improved models for estimating temporal changes in carbon sequestration in above-ground biomass of mixed-species environmental plantings. Forest Ecology and Management **338**: 208-218. DOI: 10.1016/j.foreco.2014.11.025.

Paul, KI, Roxburgh, SH, England, JR, Ritson, P, Hobbs, T, Brooksbank, K, Raison, RJ, Larmour, JS, Murphy, S, Norris, J, Neumann, C, Lewis, T, Jonson, J, Carter, JL, McArthur, G, Barton, C, Rose, B, 2013. Development and testing of generic allometric equations for estimating above-ground biomass in mixed-species environmental plantings. Forest Ecol. Manage. **310**: 483-494.

Queensland Government, 2014. Data Drill. Queensland Department of Science, Information Technology, Innovation and the Arts. http://www.longpaddock.qld.gov.au/silo/datadrill/index.php.

Raftery, A, Hoeting, J, Volinsky, C, Painter, I, Yeung, KY, 2008. The BMA package: version 3.05., http://cran.r-project.org/web/packages/BMA/BMA.pdf

Raftery, AE, Hoeting, JA, Volinsky, CT, Painter, I, Yeung, KY, 2009. Package 'BMA'. Technical report. Software manual.

Raftery, AE, Madigan, D, Hoeting, JA, 1997. Bayesian model averaging for linear regression models. J. Am. Stat. Soc. **92**: 179-191.

Schwarz, GE, 1978. Estimating the dimension of a model. Annals of Statistics **6**: 461–464.

Smith, M, Conte, P, Berns, AE, Thomson, J, Cavagnaro, TR, 2012. Spatial patterns of, and environmental controls on, soil properties at a riparian-paddock interface. Soil Biology and Biochemistry **49**: 39-45. DOI: 10.1016/j.soilbio.2012.02.007.

Srinivasarao, C, Venkateswarlu, B, Lal, R, Singh, AK, Kundu, S, Vittal, KPR, Patel, JJ, Patel, MM, 2014. Long-term manuring and fertilizer effects on depletion of soil organic carbon stocks under pearl millet-cluster bean-castor rotation in Western India. Land Degradation and Development **25**: 173-183. DOI: 1002/ldr.1158.

Thomson, JR, Mac Nally, R, Fleishman, E, Horrocks, G, 2007. Predicting bird species distributions in reconstructed landscapes. Conservation Biology **21**: 752-766. DOI: 10.1111/j.1523-1739.2007.00687.x.

Torres, AB, Lovett, JC, 2012. Using basal area to estimate aboveground carbon stocks in forests: La Primavera Biosphere's Reserve, Mexico. Forestry **86**: 1-15. DOI: 10.1093/forestry/cps084.

Ussiri, DAN, Lal, R, Jacinthe, PA, 2006. Soil properties and carbon sequestrationof afforested pastures in reclaimed minesoils of Ohio. Soil Science Society of America Journal **70**: 1797–1806. DOI: 10.1111/10.2136/sssaj2005.0352.

Viaud, V, Angers, DA, Walter, C, 2010. Towards landscape-scale modeling of soil organic matter dynamics in agroecosystems. Soil Science Society of America Journal **74**: 1-14. DOI: 10.2136/sssaj2009.0412.

Wöhling, T, Vrugt, JA, 2008. Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models. Water Resource Ressearch **44**: W12432. DOI: 10.1029/2008WR007154.

Table 1. Potential predictor variables used in model development.

Variables
Response variables
Soil C stock, tree planting (Mg ha^{-1} in 0-5 cm soil layer)
Litter stock, tree planting (Mg ha^{-1})
Soil C:N ratio, tree planting (0-5 cm soil layer)
Difference[a] in soil C stock (Mg ha^{-1} in 0-5 cm soil layer)
Difference[a] in soil C:N ratio
Derived predictors
Site Age (years since reforestation)
Latitude (^{o}S)
Longitude (^{o}E)
Landscape position (riparian/upland)
Soil Type (Australian Soil Classification)
Soil texture
Annual Rainfall (mm)
Lifetime rainfall (cumulative rainfall over time since reforestation, mm)
Mean maximum temperature over the growing season (^{o}C)
Planting data
Basal area of trees (m^2 ha^{-1})
Canopy extent (%)
Litter stock[b], tree planting (Mg ha^{-1})
Pasture data
Pasture soil C stock (Mg ha^{-1} in 0-5 cm soil layer)
Pasture soil C:N ratio (0-5 cm soil layer)

[a] Difference refers to different in variable of interest between tree planting and adjacent pasture (see Materials and Methods). [b] Litter stock was only included as a potential predictor variable for soil C stocks and soil C:N ratios.

Table 2. Probability of a non-zero coefficient in the predictor model [Pr(inc)] and Bayesian Information Criterion (BIC), for litter stock in the tree plantings, determined by BMA for Model Scenarios 1-4 (see text and Figure 1). See Table 1 for explanations of the predictor variables.

	Model Scenario 1		**Model Scenario 2**		**Model Scenario 3**		**Model Scenario 4**	
Predictor	**Pr(inc)[e]**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**
Intercept	1.00	-4.41 ± 79.2	1.00	-8.29 ± 92.19	1.00	-2.56 ± 146.1	1.00	-2.89 ± 148
Derived predictors								
Planting age	0.62		0.62		1.00	0.34 ± 0.06	1.00	0.34 ± 0.06
Latitude	0.08		0.09		0.29		0.30	
Longitude	0.09		0.12		0.19		0.20	
Landscape position	0.37		0.43		0.19		0.22	
Soil type	0.06		0.06		0.09		0.10	
Soil texture	0.10		0.12		0.07		0.90	
Annual rainfall	0.07		0.07		0.08		0.10	
Lifetime rainfall[a]	0.64		0.59		0.35		0.36	
Mean max temp[b]	0.10		0.11		0.28		0.29	
Planting data								
Basal area	0.96	0.29 ± 0.12	0.94	0.28 ± 0.12	-		-	-
Canopy extent	0.06		0.08		-		-	-
Pasture data								
Pasture soil C[c]	0.26		-	-	0.10		-	-
Pasture soil C:N	0.05		-	-	0.08		-	-
Model BIC[d]	-86.5		-86.1		-90.1		-89.2	

[a]Cumilative rainfall over time since revegetation; [b]Mean Max temperature over the growing season; [c]Pasture soil C stock; [d]BIC = Bayesian Information Criterion; [e]Pr(inc) = probability of inclusion.

Table 3. Probability of a non-zero coefficient in the predictor model [Pr(inc)] and Bayesian Information Criterion (BIC), for soil C stock in the tree plantings, determined by BMA for Model Scenarios 1-4 (see text and Figure 1). See Table 1 for explanations of the predictor variables.

	Model Scenario 1		**Model Scenario 2**		**Model Scenario 3**		**Model Scenario 4**	
Predictor	**Pr(inc)[e]**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**
Intercept	1.00	-388.4 ± 675.8	1.00	-1115 ± 326	1.00	-293.8 ± 734.5	1.00	-1140 ± 361.0
Derived predictors								
Planting age	0.26		0.10		1.00		1.00	0.41 ± 0.12
Latitude	0.78	-13.99 ± 9.56	1.00	-26.66 ± 6.27	0.70		1.00	-26.81 ± 6.99
Longitude	0.33		0.08		0.35		0.10	
Landscape position	0.15		0.08		0.21		0.09	
Soil Type	0.08		0.07		0.18		0.11	
Soil texture	0.07		0.07		0.14		0.09	
Annual rainfall	0.33		0.35		0.22		0.33	
Lifetime rainfall[a]	0.09		0.08		0.24		0.09	
Mean max temp[b]	0.79	2.81 ± 2.00	1.00	5.16 ± 1.57	0.85	3.50 ± 2.22	1.00	6.20 ± 1.87
Planting data								
Basal area	0.08		0.09		-	-	-	-
Canopy extent	0.11		0.07		-	-	-	-
Litter stock	0.99	0.83 ± 0.27	1.00	1.01 ± 0.25	-	-	-	-
Pasture data								
Pasture soil C[c]	1.00	0.51 ± 0.14	-	-	1.00	0.59 ± 0.15	-	-
Pasture soil C:N	0.13		-	-	0.16		-	-
Model BIC[d]	-81.1		-85.9		-80.2		-84.7	

[a]Cumilative rainfall over time since revegetation; [b]Mean Max temperature over the growing season; [c]Pasture soil C stock; [d]BIC = Bayesian Information Criterion; [e]Pr(inc) = probability of inclusion.

Table 4. Probability of a non-zero coefficient in the predictor model [Pr(inc)] and Bayesian Information Criterion (BIC), for difference in soil C stock in the tree plantings, determined by BMA for Model Scenarios 1-4 (see text and Figure 1). See Table 1 for explanations of the predictor variables.

	Model Scenario 1		**Model Scenario 2**		**Model Scenario 3**		**Model Scenario 4**	
Predictor	**Pr(inc)[e]**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**
Intercept	1.00	-388.4 ± 675.8	1.00	617.2 ± 518.0	1.00	-261.2 ± 749.9	1.00	487.8 ± 560.9
Derived predictors								
Planting age	0.26		0.35		1.00	0.42 ± 0.11	1.00	0.43 ± 0.11
Latitude	0.78	-13.99 ± 9.58	0.19		0.68		0.18	
Longitude	0.33		0.79	-4.42 ± 3.21	0.36		0.63	
Landscape position	0.15		0.26		0.22		0.42	
Soil type	0.07		0.09		0.18		0.18	
Soil texture	0.07		0.07		0.13		0.11	
Annual rainfall	0.33		0.19		0.22		0.12	
Lifetime rainfall[a]	0.09		0.36		0.24		0.47	
Mean max temp[b]	0.79	2.81 ± 2.00	0.32		0.84	3.46 ± 2.21	0.45	
Planting data								
Basal area	0.08		0.54		-	-	-	-
Canopy extent	0.11		0.15		-	-	-	-
Litter stock	0.99	0.83 ± 0.27	0.40		-	-	-	-
Pasture data								
Pasture soil C[c]	1.00	-0.50 ± 0.14	-	-	0.97	-0.40 ± 0.16	-	-
Pasture soil C:N	0.13		-	-	0.16		-	-
Model BIC[d]	-81.4		-84.8		-80.2		-84.7	

[a]Cumilative rainfall over time since revegetation; [b]Mean Max temperature over the growing season; [c]Pasture soil C stock; [d]BIC = Bayesian Information Criterion; [e]Pr(inc) = probability of inclusion.

Table 5. Probability of a non-zero coefficient in the predictor model [Pr(inc)] and Bayesian Information Criterion (BIC), for soil C:N ratio in the tree plantings, determined by BMA for Model Scenarios 1-4 (see text and Figure 1). See Table 1 for explanations of the predictor variables.

	Model Scenario 1		Model Scenario 2		Model Scenario 3		Model Scenario 4	
Predictor	**Pr(inc)[e]**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**
Intercept	1.00	-4.30 ±36.70	1.00	29.22 ± 66.67	1.00	-28.16 ± 74.62	1.00	12.99 ± 26.12
Derived predictors								
Planting age	0.15		0.21		1.00	0.11 ± 0.03	1.00	0.13 ± 0.04
Latitude	0.09		0.11		0.16		0.07	
Longitude	0.08		0.16		0.28		0.06	
Landscape position	0.08		0.12		0.10		0.07	
Soil type	0.06		0.06		0.08		0.09	
Soil texture	0.35		0.20		0.10		0.08	
Annual rainfall	0.07		0.26		0.21		0.09	
Lifetime rainfall[a]	0.38		0.08		0.57		0.07	
Mean max temp[b]	0.23		0.07		0.15		0.11	
Planting data								
Basal area	0.98	0.13 ± 0.04	0.88	0.14 ± 0.07	-		-	-
Canopy extent	0.18		0.16		-		-	-
Litter stock	0.11		0.10		-		-	-
Pasture data								
Pasture soil C[c]	0.08		-	-	0.07		-	-
Pasture soil C:N	1.00	0.88 ± 0.13	-	-	1.00	0.92 ± 0.14	-	-
Model BIC[d]	-88.1		-86.8		-87.1		-90.7	

[a]Cumilative rainfall over time since revegetation; [b]Mean Max temperature over the growing season; [c]Pasture soil C stock; [d]BIC = Bayesian Information Criterion; [e]Pr(inc) = probability of inclusion.

Table 6. Probability of a non-zero coefficient in the predictor model [Pr(inc)] and Bayesian Information Criterion (BIC), for difference in soil C:N ratio in the tree plantings, determined by BMA for Model Scenarios 1-4 (see text and Figure 1). See Table 1 for explanations of the predictor variables.

	Model Scenario 1		**Model Scenario 2**		**Model Scenario 3**		**Model Scenario 4**	
Predictor	**Pr(inc)[e]**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**	**Pr(inc)**	**Coefficient ± SD**
Intercept	1.00	-9.52 ± 44.93	1.00	-12.43 ± 51.44	1.00	-33.30 ± 79.19	1.00	-0.35 ± 82.26
Derived predictors								
Planting age	0.15		0.16		1.00	0.11 ± 0.03	1.00	0.11 ± 0.02
Latitude	0.08		0.10		0.17		0.18	
Longitude	0.09		0.11		0.30		0.30	
Landscape position	0.09		0.11		0.09		0.11	
Soil type	0.07		0.08		0.06		0.07	
Soil texture	0.29		0.30		0.09		0.10	
Annual rainfall	0.09		0.10		0.23		0.26	
Lifetime rainfall[a]	0.57		0.63		0.66		0.70	
Mean max temp[c]	0.11		0.12		0.17		0.18	
Planting data								
Basal area	0.93	0.12 ± 0.05	0.93	0.12 ± 0.05	-		-	-
Canopy extent	0.15		0.17		-		-	-
Litter stock	0.16		0.17		-		-	-
Pasture data								
Pasture soil C[c]	0.08		-	-	0.06		-	-
Pasture soil C:N	0.12		-	-	0.08		-	-
Model BIC[d]	-90.6		-89.2		-90.5		-88.3	

[a]Cumilative rainfall over time since revegetation; [b]Mean Max temperature over the growing season; [c]Pasture soil C stock; [d]BIC = Bayesian Information Criterion; [e]Pr(inc) = probability of inclusion

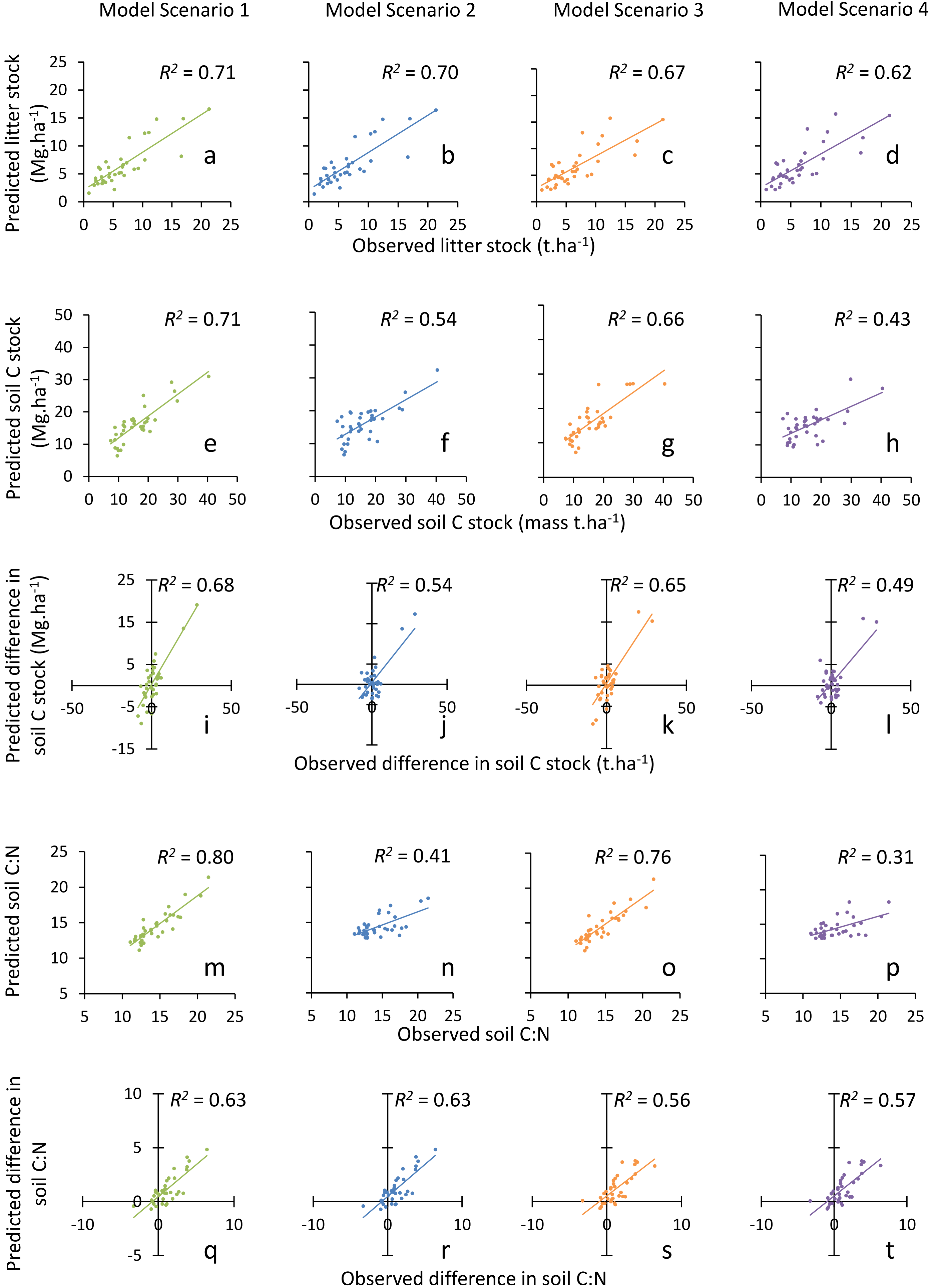

Figure 1. Relationships between observed and predicted values of C variables (see text and Tables 3-7), for Model Scenarios 14 for (a-d) litter mass, (e-h) soil C stock, (i-l) difference in soil C stock, (m-p) soil C:N and (q-t) difference in soil C:N. For graphs i-l there are two points in each Figure with strong leverage; see text for regression output with these values removed.