Forest Ecology and Management 432 (2019) 264–275

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

A revised above-ground maximum biomass layer for the Australian continent

Stephen H. Roxburgh[a,*], Senani B. Karunaratne[b], Keryn I. Paul[a], Richard M. Lucas[c,g], John D. Armston[d,e,f], Jingyi Sun[g]

[a] CSIRO Land and Water, GPO Box 1700, Canberra, ACT 2601, Australia
[b] Department of the Environment and Energy, Parkes, ACT, Australia
[c] Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DB, United Kingdom
[d] Department of Environment and Science Remote Sensing Centre, Ecosciences Precinct, Brisbane 4001, Queensland, Australia
[e] Joint Remote Sensing Research Program, School of Earth and Environmental Sciences, University of Queensland, St. Lucia, Brisbane 4067, Queensland, Australia
[f] Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA
[g] School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Keywords:
Forest biomass
Random forest
Carbon accounting
National greenhouse gas inventory

ABSTRACT

The carbon accounting model FullCAM is used in Australia's National Greenhous Gas Inventory to provide estimates of carbon stock changes and emissions in response to deforestation and afforestation/reforestation. FullCAM-predicted above-ground woody biomass is heavily influenced by the parameter *M*, which defines the maximum upper limit to biomass accumulation for any location within the Australian continent. In this study we update FullCAM's *M* spatial input layer through combining an extensive database of 5739 site-based records of above-ground biomass (AGB) with the Random Forest ensemble machine learning algorithm, with model predictions of AGB based on 23 environmental predictor covariates. A Monte-Carlo approach was used, allowing estimates of uncertainty to be calculated. Overall, the new biomass predictions for woodlands, with 20–50% canopy cover, were on average 49.5 ± 1.3 (s.d.) t DM ha^{-1}, and very similar to existing model predictions of 48.5 t DM ha^{-1}. This validates the original FullCAM model calibrations, which had a particular focus on accounting for greenhouse gas emissions in Australian woodlands. In contrast, the prediction of biomass of forests with a canopy cover > 50% increased significantly, from 172.1 t DM ha^{-1}, to 234.4 ± 5.1 t DM ha^{-1}. The change in forest biomass was most pronounced at sub-continental scales, with the largest increases in the states of Tasmania (166 to 351 ± 22 t DM ha^{-1}), Victoria (201 to 333 ± 14 t DM ha^{-1}), New South Wales (210 to 287 ± 9 t DM ha^{-1}), and Western Australia (103 to 264 ± 14 s.d. t DM ha^{-1}). Testing of model predictions against independent data from the savanna woodlands of northern Australia, and from the high biomass *Eucalyptus regnans* forests of Victoria, provided confidence in the predictions across a wide range of forest types and standing biomass. When applied to the Australian Government's National Inventory land clearing accounts there was an overall increase of 6% in continental emissions over the period 1970–2016. Greater changes were seen at sub-continental scales calculated within 6° × 4° analysis tiles, with differences in emissions varying from −21% to +35%. Further testing is required to assess the impacts on other land management activities covered by the National Inventory, such as reforestation; and at more local scales for sequestration projects that utilise FullCAM for determining abatement credits.

1. Introduction

FullCAM (Full Carbon Accounting Model) is a freely available software system for tracking greenhouse gas emissions and changes in carbon stocks associated with land use and management in Australian agricultural and forest systems (Richards, 2001; Richards and Brack, 2004; Richards and Evans, 2004; Brack et al., 2006; Waterworth et al., 2007). It is applied at the national scale for land sector greenhouse gas emissions accounting (Australian Government, 2018), and at the local scale for monitoring and reporting carbon sequestration projects, such as revegetation and the management of regrowth (Paul et al., 2015a,b).

FullCAM predicts the accumulation of above-ground biomass (AGB)

* Corresponding author.
E-mail address: stephen.roxburgh@csiro.au (S.H. Roxburgh).

https://doi.org/10.1016/j.foreco.2018.09.011
Received 8 August 2018; Received in revised form 7 September 2018; Accepted 9 September 2018

in woody vegetation using a hybrid of empirical and process-based modelling via the implementation of the Tree Yield Formula (TYF; Waterworth et al., 2007). The process-based modelling component utilises the forest growth model 3-PG (Landsberg and Waring, 1997) to derive a dimensionless index (the Forest Productivity Index, or FPI) that summarises potential site productivity for any given location based on the Normalised Difference Vegetation Index (NDVI), soil fertility, vapour pressure deficit, soil water content, and temperature (Kesteven and Landsburg, 2004). The empirical component is a statistical relationship between field-based observations of AGB (from minimally disturbed stands) and the FPI (Richards and Brack, 2004). This relationship is used to calculate the parameter *M* (the predicted maximum AGB for a given FPI), and is given by

$$M = (6.011 \times \sqrt{FPI} - 5.291)^2 \tag{1}$$

Parameter *M* is constant for any location in Australia, and is embedded within the FullCAM database as a spatial input layer with a resolution of 0.0025° (or approximately 250 m). Computationally, *M* exerts a strong influence on forest growth, affecting the rate of AGB accumulation, as well as defining the upper maximum biomass limit. *M* is also an important ecosystem property, with links to environmental productivity as well as a being a key indicator of ecosystem structure.

Over recent years evidence has accumulated that predictions of *M* for some vegetation types were biased, particularly for higher-biomass temperate forests, with lower *M* than observations would suggest (Montagu et al., 2003; Waterworth et al., 2007; Wood et al., 2008; Lowson, 2008; Keith et al., 2010; Roxburgh et al., 2010; Fensham et al., 2012; Preece et al., 2012). The presence of such bias may be due to the initial focus during FullCAM development on estimating carbon emissions and sequestration within Australia's woodland ecosystems, due to their ongoing active management. The forest types represented in the original field-based biomass estimates used in the relationship to predict *M* (Eq. (1)) had a strong representation of woodlands, but with < 10% of observations from higher-biomass (> 250 t DM ha^{-1}) temperate forests.

Since the development of FullCAM there has been a large increase in the availability of forest biomass data from across Australia, including from relatively undisturbed high biomass temperate forests. It was therefore timely to explore how these new data can be used to improve the estimation of *M*. The aim of this study was to use these new datasets to update FullCAM's *M* layer, and thus improve the accuracy of predictions of woody biomass growth for Australian woodlands and forests, and hence, Australia's National Greenhouse Gas Inventory.

2. Methods

Whilst it is possible to create *de novo* a new replacement biomass layer, by e.g. re-fitting the existing FPI vs observed biomass relationship on which the existing estimates of *M* are based (Eq. (1)), the approach adopted here was to update rather than replace the current *M* layer. This was to maintain continuity and consistency with the existing FullCAM modelling environment, and to allow new data to be applied only to regions with adequate data representation.

The detailed analysis steps are shown in Fig. 1, and can be summarised as follows:

1. Identify site biomass records that fulfil the criteria of being minimally disturbed, consistent with the definition of maximum biomass, *M*.
2. For each record *i*, calculate the ratio λ_i

$$\lambda_i = \frac{M_i}{O_i} \tag{2}$$

where M_i is the current prediction of maximum biomass (Equation (1)), and O_i is the field observation.

3. Use the Random Forest machine learning algorithm (Brieman 2001) to statistically model and predict λ across the continent, using a range of climatic and edaphic variables.
4. Update the existing *M* layer to *M′* by multiplying by the model-predicted λ

$$M^{'} = \lambda M \tag{3}$$

2.1. Database of above-ground biomass observations

The primary source of AGB observation data was the TERN/Auscover National Biomass Library (NBL), available at http://www.auscover.org.au/purl/biomass-plot-library. This library is a collation of stem inventory and biomass estimates compiled from federal, state and local government departments, universities, private companies and other agencies. The biomass library contains (as of December 2017) 14,453 sites, 887,639 individual tree diameter measurements (> 5 cm), and 1467 species.

For inclusion in the analysis, the AGB estimates were required to represent predominantly mature and undisturbed vegetation (i.e. vegetation that has been minimally impacted by anthropogenic disturbances, and has not had a recent natural disturbance such as a wildfire or cyclone). Because not all sites within the NBL were located in vegetation that could be considered 'mature', it was first necessary to filter the database and exclude those observations that were most likely collected from disturbed vegetation. This was achieved by collating ancillary spatial datasets at both a national and state level that identified areas within which forests were more likely to be undisturbed (such as conservation lands), and also to identify areas where disturbance was more likely, for example areas subject to multiple use, including timber harvesting (Supplementary Data: Appendix A). Information was also gathered from the custodians of the NBL data where this indicated a measurement was located in disturbed or undisturbed (often referred to as remnant) vegetation. Records were also excluded if the observations were non-representative of the broader landscape, such as a number of Tasmanian records that specifically targeted forested areas with higher than average biomass (labelled 'LIMA' and 'LIMI' in the database; D. Mannes pers. comm.). A total of 5739 site records remained following this filtering (Table 1). To provide an additional check of the temporal continuity of forest cover, spatial forest cover mapping (> 20% cover) based on 25 Landsat images extending back to the 1970s were used to confirm woody vegetation cover over the period, thus indicating the absence of major disturbance (Australian Government 2018). Forest cover was defined as the mode within a 3 × 3 pixel window (approximately 75 m × 75 m) centred on the observation.

Preliminary analyses suggested improved empirical model performance could be obtained by stratifying the data and running separate statistical models based on two broad vegetation types corresponding to 'Forests' (with canopy cover > 50%) and 'Woodlands' (with canopy covers between 20 and 50%). The classification of sites within the database was based on forest and woodland cover as defined by the Australian National Forest Inventory (ABARES, 2014).

2.2. Vegetation classification for model prediction

Because *M* represents biomass at forest maturity, the spatial interpolation of the statistical models should represent the potential vegetation that an area could support, not the current vegetation distribution which reflects past land management, such as clearing of woody vegetation. The spatial interpolation was therefore based on the NVIS v4.2 1750 Major Vegetation Subgroups (MVS) classification (NVIS, 2016), which maps the extent of Australia's major vegetation types prior to extensive land clearing, at a 100 m resolution.

The NVIS subgroup for each of the 5739 records was extracted, and any subgroup that was represented by 50 observations or more was