Int J Appl Earth Obs Geoinformation 73 (2018) 397–406

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Multi-sensor airborne and satellite data for upscaling tree number information in a structurally complex eucalypt forest

Iurii Shendryk[a,b,*], Mark Broich[a], Mirela G. Tulbure[a,*]

[a] School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, NSW, 2052, Australia
[b] CSIRO, Agriculture and Food, St. Lucia, QLD, 4067, Australia

ARTICLE INFO

Keywords:
WorldView-2
SAR
ALS
Tree numbers
Random forest
Upscaling
Object-oriented classification
Australia
River Red Gum
Eucalypt

ABSTRACT

Detailed information on the number and density of trees is important for conservation and sustainable use of forest resources. In this respect, remote sensing technology is a reliable tool for deriving timely and fine-scale information on forest inventory attributes. However, to better predict and understand the functioning of the forest, fine-scale measures of tree number and density must be extrapolated to the forest plot or stand levels through upscaling. In this study, we compared and combined three sources of remotely sensed data, including low point density airborne laser scans (ALS), synthetic aperture radar (SAR) and very-high resolution WorldView-2 imagery to upscale the total number of trees to the plot level in a structurally complex eucalypt forest using random forest regression. We used information on number of trees previously derived from high point density ALS as training data for a random forest regressor and field inventory data for validation. Overall, our modelled estimates resulted in significant fits ($p < 0.05$) with goodness-of-fit (R^2) of 0.61, but systematically underestimated tree numbers. The ALS predictor variables (e.g. canopy cover and height) were the best for estimating tree numbers ($R^2 = 0.48$, $nRMSE = 61\%$), as compared to WorldView-2 and SAR predictor variables ($R^2 < 0.35$). Overall, the combined use of WorldView-2, ALS and SAR predictors for estimating tree numbers showed substantial improvement in R^2 of up to 0.13 as compared to their individual use. Our findings demonstrate the potential of using low point density ALS, SAR and WorldView-2 imagery to upscale high point density ALS derived tree numbers at the plot level in a structurally complex eucalypt forest.

1. Introduction

A systematic derivation of information on the number and density of trees is required for conservation and sustainable use of forest resources, and could assist with understanding the causes of forest decline (Hudak et al., 2006; Bowen et al., 2011). Therefore, a spatial representation of this attribute is required over large areas to assist in the planning of management actions. Conventional methods for estimating tree numbers through field surveys and aerial photograph interpretation are costly, time consuming and not pertinent to large areas. Furthermore, it is not always feasible to conduct an extensive field survey due to logistical constraints. In recent years, remote sensing technology became an operational method for management and monitoring of forest resources as part of state inventory programs (Næsset, 2014; Magnussen et al., 2018) as well as by commercial timber companies (Rombouts et al., 2014; Interpine Innovation, 2018). Moreover, semi-automated individual tree delineation methods have been proposed for direct retrieval of information on number and density of trees from high point density airborne laser scans (ALS) (Reitberger et al., 2009; Ferraz et al., 2012; Shendryk et al., 2016a) and airborne imagery (Culvenor, 2002; Jing et al., 2012). Although high resolution airborne remote sensing data could cover larger areas at lower cost per unit area as compared to field surveys, they are still costly and not feasible on a regional scale. Therefore, there is a need for testing the applicability of low-cost airborne and satellite data for upscaling tree number and density information to the plot or stand levels.

Previous studies have shown that satellite remote sensing as well as low point density ALS can upscale field-derived information on forest composition across large spatial extents. For example, Næsset and Bjerknes (2001) and Lindberg and Hollaus (2012) used field data collected within 39 and 68 plots, respectively to estimate total number of trees from ALS data in boreal forests dominated by Norway spruce and Scots pine with normalized RMSE (*nRMSE*) as low as 27.9% (Næsset and Bjerknes, 2001) and 37.3% (Lindberg and Hollaus, 2012). Similarly, Humagain et al. (2017) and Pflugmacher et al. (2012) used field-derived data in combination with Landsat imagery to upscale tree

* Corresponding authors.
E-mail addresses: yuri.shendryk@csiro.au (I. Shendryk), mirela.tulbure@unsw.edu.au (M.G. Tulbure).

https://doi.org/10.1016/j.jag.2018.07.011
Received 5 April 2018; Received in revised form 5 July 2018; Accepted 6 July 2018
Available online 18 July 2018

density and attributes that are related to tree density (e.g. basal area) at the plot level. However, they found that mixing of understory and open soil signals in 30 m, medium resolution pixels containing forested areas led to suboptimal model performance. In this respect, with the launch of the first very-high resolution (VHR) satellite sensor in 1999 (Dial et al., 2003), it is now possible to account for the mixing effect of understory and open soil signal in pixels representing forested areas. Finally SAR imagery, although lacking the high spatial resolution necessary for individual tree detection was previously used for large area forest inventory (Tomppo et al., 1994). While SAR data is commonly reported to be reliable for estimating forest biomass, especially the L- and P-band cross-polarized backscatter (*HV* and *VH*) (Le Toan et al., 1992; Mitchard et al., 2009), there is little research using SAR data to quantify tree number or density.

As different sources of remotely sensed data (e.g. ALS, SAR and VHR multispectral satellite imagery) become increasingly accessible and freely available, it is crucial to test their individual and integrated applicability in the derivation of forest attributes important for forest management. The common technique for estimating tree numbers at the plot level is through regression analysis of field and ALS data (Næsset and Bjerknes, 2001; Maltamo et al., 2004), while the predictive ability of VHR satellite multispectral and SAR imagery in this respect is commonly neglected in the literature. Therefore, the objective of this study was to quantify and compare the suitability of low point density ALS and SAR imagery, providing detailed information on vegetation structure, with VHR WorldView-2 (WV2) imagery providing information on spectral characteristics of trees for estimating tree numbers across the largest River Red Gum (RRG) forest in the world – Barmah-Millewa Forest (BMF). Our approach relies on low point density ALS data to mask understory and open soil signals in VHR satellite imagery, and unlike other studies upscaling field survey data, we used high point density ALS derived information on total number of trees as training data for a random forest regressor. The final map of total tree numbers is presented at the plot level, which is an important forest management unit for many forestry applications (Wu et al., 2014), and validated against independent field data.

2. Methods

Our methodology consisted of pre-processing three sources of remote sensing data (i.e. low point density ALS, SAR and VHR WV2 imagery) and their integration with the information on number of trees for 2014 that was derived using high point density ALS (further referred to as ALS_{2014}) data (Shendryk et al., 2016a, b) using random forest regression (RF) (Breiman, 2001). Our final RF models were validated using data on number of trees collected in the field in 36 (30 m × 30 m) plots in 2014 and 50 (50 m × 50 m) plots in 2013. RF – one of the most popular and most powerful machine learning algorithms – was used in this study twice. Firstly, it was used to generate cloud- and shadow-free WV2 imagery, and secondly to estimate total tree numbers from low point density ALS, SAR and WV2 data. The flow chart of methodological steps is outlined in Fig. 1.

2.1. Study area

Barmah-Millewa Forest (BMF) (Fig. 2), which occupies 737 km^2, is the largest contiguous area of RRGs (*Eucalyptus camaldulensis*) in the world. This mono-specific, uneven-aged forest complex consists of the RRG forest (71%), RRG woodland (23%) and mixed box eucalypt woodland (6%) (MDBA, 2010). BMF is structurally complex with highly variable tree densities ranging from more than 700 trees (ha^{-1}) in forests to less than 50 trees (ha^{-1}) in open woodlands (Kerle, 2005; OEH, 2012).

Previous remote sensing studies were valuable in estimating tree density of the BMF at the plot (Cunningham et al., 2010) and forest stand (Bowen et al., 2011) levels, but relied on extensive field data covering 175 plots (Cunningham et al., 2010) or aerial photo interpretation (Bowen et al., 2011).

2.2. Data

In order to produce a seamless map of total tree numbers at the plot level covering the whole extent of BMF, we used VHR WV2 imagery acquired in 2014 (further referred to as $WV2_{2014}$), Sentinel-1 SAR data collected in 2014 (further referred to as $S1_{2014}$), ALOS PALSAR-2 SAR mosaic generated using imagery collected in 2015 (further referred to as PA_{2015}) and low point density (< 0.5 pts/m^2) discrete return airborne ALS collected in 2001 (further referred to as ALS_{2001}) covering the whole extent of BMF. In addition, the information on number of trees derived using ALS_{2014} covering ~14% of BMF was used to train RF models, while two sets of data on the number of trees collected in the field were used to validate RF models.

2.2.1. Field data for RF model validation

In June 2014, detailed measurements of trees were performed within 36 (30 m × 30 m) plots that overlapped with ALS_{2014} data. This field survey was denoted as *Inventory A*, with a total of 1051 trees (Table 1). The delineation of plots and positioning of every tree ≥ 13 cm in diameter at breast height (DBH) in *Inventory A* was done using differential GPS (dGPS) rover operating on the SmartNet Australia DGNSS RTK CORS network (Leica, 2012). A similar field survey denoted as *Inventory B,* with a total of 3703 trees, consisted of 50 (50 m × 50 m) plots surveyed as part of the five Icon Sites of the Living Murray program (MDBC, 2005) by NSW National Parks and Wildlife service in 2013. The delineation of plots in *Inventory B* was done using a standard GPS unit with every tree ≥ 10 cm DBH within the defined sampling plot being assessed (Cunningham, 2016). The data in both *Inventory A* and *B* were further used exclusively for validation of RF models.

2.2.2. ALS_{2014} derived data for RF model training

The methods used to extract information on number of trees from the ALS_{2014} were published in Shendryk et al. (2016a,b), and their reliability to retrieve individual tree characteristics was already assessed. As a result, they are only briefly described here, and our main effort is directed at investigating the relationship between high point density ALS_{2014} derived tree numbers and three sources of remotely sensed data (i.e. low point density ALS, SAR and VHR WV2 imagery).

The overall accuracy of individual tree detections in ALS_{2014} data (Fig. 3B) was 64% based on dGPS measurements of trees with DBH ≥ 13 cm as part of *Inventory A* (Shendryk et al., 2016a, b). To have the same resolution for both training and validation datasets, the ALS_{2014} derived data was further aggregated in grid cells of 30 m × 30 m and 50 m × 50 m representing individual plots and matching the resolution of the field data collected in *Inventory A* and *B*, respectively (Fig. 3C).

There were 77,923 (30 m × 30 m) and 28,336 (50 m × 50 m) grid cells that were completely within the ALS_{2014} extent, respectively (Table 2). The numbers of trees (as derived from ALS_{2014}) in grid cells were further used exclusively for training of RF models.

Given that the data for RF model training was derived using ALS_{2014} in this study, it is expected that our models will underestimate actual tree numbers, but will be sufficient to consistently and in relative terms investigate the relationship between tree numbers and the three sources of remotely sensed data (i.e. low point density ALS, SAR and VHR WV2).

2.2.3. Low point density ALS data

The low point density discrete return ALS_{2001} data were gathered in July 2001 and acquired by the MDB Commission using Optech ALTM 1225 airborne laser scanning system (Miura and Jones, 2008). Table 3 details the specifications of the ALS_{2001} sensor.