



Modelling LiDAR derived tree canopy height from Landsat TM, ETM+ and OLI satellite imagery—A machine learning approach[☆]

Grant Staben^{a,b,*}, Arko Lucieer^a, Peter Scarth^{c,d}

^a Discipline of Geography and Spatial Science, School of Technology, Environments and Design, College of Sciences and Engineering, University of Tasmania, Private Bag 76, TAS 7001, Hobart, Tasmania, Australia

^b Department of Environment and Natural Resources, Northern Territory Government, PO Box 496, NT 0831, Palmerston, Northern Territory, Australia

^c Joint Remote Sensing Research Program, School of Geography, Planning and Environmental Management, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia

^d Remote Sensing Centre, Science Delivery, Department of Environment and Science, Queensland Government, 41 Boggo Road, Brisbane, Queensland 4102, Australia



ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Random forest
LiDAR
Canopy height model
Vegetation structure
Landsat

ABSTRACT

Understanding ecological changes in native vegetation communities often requires information over long time periods (multiple decades). Tropical cyclones can have a major impact on woody vegetation structure across northern Australia, however understanding the impacts on woody vegetation structure is limited. Woody vegetation structural attributes such as height are used in ecological studies to identify long term changes and trends. LiDAR has been used to measure woody vegetation structure, however LiDAR datasets cover relatively small areas and historical coverage is restricted, limiting the use of this technology for monitoring long-term change. The Landsat archive spans multiple decades and is suitable for regional/continental assessment. Advances in predictive modelling using machine learning algorithms have enabled complex relationships between dependent and independent variables to be identified. The aim of this study is to develop a predictive model to estimate woody vegetation height from Landsat imagery to assist in understanding change through space and time. A LiDAR canopy height model was produced covering a range of vegetation communities in northern Australia (Darwin region) for use as the dependent variable. A random forest regression model was developed to predict mean LiDAR canopy height (30 m spatial resolution) from Landsat-5 Thematic Mapper (TM). Validation of the random forest model was undertaken on independent data ($n = 30,500$) resulting in an overall $R^2 = 0.53$, RMSE of 2.8 m. Assessment of the RMSE within four broad vegetation communities ranged from 2.5 to 3.7 m with the two dominant communities in the study area Mangrove forests and Eucalyptus communities recording an RMSE value of 2.9 m and 2.5 m respectively. The model was also applied to Landsat-7 Enhanced Thematic Mapper Plus (ETM+) resulting in an R^2 of 0.49, RMSE of 2.8 m. The model was then applied to all cloud free Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 Operational Land Imager (OLI) imagery (106/69 path/row) available between the months April, May and June for 1987 to 2016 to produce annual estimates (29 years) of canopy height. A number of time traces were produced to illustrate tree canopy height through time in the Darwin region which was severely impacted by cyclone (hurricane) Tracy on the 25th December 1974.

1. Introduction

The value of remote sensing in ecological studies has been well recognised (Roughgarden et al., 1991; Wang et al., 2010; Pettorelli et al., 2014). Landsat satellites have been capturing multispectral imagery of the earth surface since 1972 representing the longest record of temporal space-borne land observations (Roy et al., 2010). Landsat data has been used for a variety of applications, such as natural hazard

assessment (Barlow et al., 2003; Joyce et al., 2009), fire scar mapping (Gill et al., 2000; Goodwin and Collett, 2014), coral reef mapping (Joyce et al., 2004), rangeland monitoring (Wallace et al., 2004; Scarth et al., 2010), temperate and tropical forest mapping (Brown et al., 2000; Renó et al., 2011), and many others. The characteristics of the Landsat sensors have been identified as valuable for regional monitoring applications (Cohen and Goward, 2004). The spectral and spatial resolution of the Landsat imagery combined with its temporal

[☆] Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.

E-mail address: Grant.Staben@utas.edu.au (G. Staben).

record make it valuable for monitoring woody cover change across large regions (Woodcock et al., 2001; Danaher et al., 2004; Staben et al., 2016; Gill et al., 2017). Amongst its many applications Landsat imagery has been utilised to detect severe forest damage (Ekstrand, 1996) including damage as a result of cyclonic (hurricane) winds (Preston, 1987; Paling et al., 2008; Staben and Evans, 2008).

Tropical cyclones occur on a frequent basis across the coastline of the Australian Northern Territory. The destructive winds associated with these cyclones can have a major impact on both the man-made and natural environments. The impact of cyclonic winds are greatest on the coastal regions, however they also have the potential to cause significant disturbance further inland (e.g. Cyclone Monica) (Staben and Evans, 2008). The impact on native vegetation can be significant, resulting in major structural changes to vegetation communities. A number of studies have reported on the impact of cyclones on vegetation in the Northern Territory (Stocker, 1976; Fox, 1980; Cameron et al., 1983; Bowman and Panton, 1994; Cook and Goyens, 2004; Staben and Evans, 2008; Williamson et al., 2011; Hutley et al., 2013). These studies have used a number of methods ranging from collection of field data, aerial photography and satellite imagery. Although cyclones are frequent and have the potential to be a major disturbance agent in ecosystems across the Northern Territory (Murphy, 1984), very few studies have been undertaken to quantify the impact and potential role they play in driving the structure of these communities (Cook and Goyens, 2004). While it is well recognised that fire and the stress of the seasonal drought (a characteristic of the wet-dry tropics of northern Australia) are frequent disturbance factors on vegetation communities, very little focus has been given to the impact cyclones have on these ecosystems (Cook and Goyens, 2004; Hutley et al., 2013).

While severe damage to woody vegetation can be relatively easy to identify by comparing satellite imagery captured directly before and after the change event (e.g. cyclones), accurate assessment of the subtle changes through time is enhanced by relating biophysical variables to satellite remote sensing observations. To obtain quantitative information from optical satellite data relationships between biophysical variables need to be established (Moulin et al., 1998). Numerous studies have derived empirical relationships between Landsat imagery and field based measurements such as leaf area index (Coops et al., 1997; Eriksson et al., 2006), above ground biomass of woody vegetation (Foody et al., 2003; Powell et al., 2010; Avitabile et al., 2012), fractional cover (Scarth et al., 2010) and woody vegetation foliage projective cover (Danaher et al., 2004; Armston et al., 2009). A variety of statistical methods have been used to develop these relationships including, linear and non-linear regression models based on single or multiple predictor variables (Cohen et al., 2003), while others have used machine learning algorithms such as neural networks, tree-based models, K-nearest neighbours and support vector machines (Labrecque et al., 2006; Li et al., 2010; Avitabile et al., 2012).

Vegetation height has been identified as a key parameter for inferring long term trends in biomass and carbon stock (Skidmore et al., 2015; Cook et al., 2015). Combined with species and site quality information vegetation height helps to inform estimates of stand age and successional stages (Stojanova et al., 2010). Light detection and ranging (LiDAR) data has been used extensively to measure woody vegetation structure, and while LiDAR is an efficient way to map and measure woody vegetation structure (Lim et al., 2003; Wulder et al., 2012; Goldbergs et al., 2018), the use of these data at a regional level can be prohibitive due to financial constraints (Pascual et al., 2010). Furthermore, the availability of LiDAR for long-term studies (multiple decades) is limited due to the paucity of data. Ecological processes can occur over long time frames, and understanding these processes often requires information recorded over multiple decades, captured at an appropriate spatial, spectral and temporal resolution. Numerous studies have used structural information obtained from LiDAR data to develop predictive models using Landsat sensors with an aim to enhance the spatial and temporal coverage (Hudak et al., 2002; Pascual et al., 2010; Hill et al.,

2011; Ota et al., 2014; Ahmed et al., 2015). These studies have been undertaken across a variety of vegetation communities ranging from conifer forests (Ahmed et al., 2015) to tropical evergreen and deciduous forests (Ota et al., 2014; Hill et al., 2011; Wilkes et al., 2015). In southern Australia Wilkes et al. (2015) predicted canopy height over a 2.9 million ha area of heterogeneous temperate forests by developing a relationship between LiDAR derived canopy height and a combination of satellite imagery (Landsat and Moderate Resolution Imaging Spectroradiometer) using the random forest algorithm. Machine learning techniques based on ensemble models such as random forest have been used successfully for a variety of remote sensing classification and regression modelling applications (Pal, 2005; Avitabile et al., 2012; Mellor et al., 2013, 2015; Mascaro et al., 2014; Karlson et al., 2015; Wilkes et al., 2015). These studies demonstrate the advantages of random forest algorithm such as its robustness to outliers in the training data, ability to handle non-parametric data, its ability to uncover complicated non-linear relationships between variables and the ease in tuning the models parameters.

In this study, we investigate the application of Landsat satellite sensors to predict woody vegetation canopy height and develop a model predicting canopy height across a range of vegetation communities in the wet-dry tropics of Northern Australia. While previous studies have demonstrated a fusion of different sensors and LiDAR to derive predictive models of canopy height in Australia (Wilkes et al., 2015), this study investigates the use of Landsat sensors only for the estimation of canopy height over a long time series of multiple decades. To our knowledge this is the first study to look at predicting LiDAR derived canopy height from Landsat sensors in the wet-dry tropics of northern Australia. A canopy height model (1 m spatial resolution) was produced from a LiDAR dataset captured in 2009 for use as the dependent variable. Random forest regression was used to produce a model to predict LiDAR derived canopy height from a single Landsat-5 Thematic Mapper (TM) image captured in 2009 (30 m spatial resolution). We developed a three-stage approach to identify the important independent variables and optimise the parameters used in the random forest model, which was applied to Landsat-5 TM, Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) sensors.

2. Data and methods

2.1. Study area

This study was undertaken in the Darwin region, located in northern Australia's wet dry tropics (Fig. 1). The average annual temperature for the Darwin region is 32 °C with average annual rainfall of 1729 mm, with the majority of the precipitation occurring during October and April. The study site covers an area of approximately 1800 km² consisting of urban, peri-urban development and native vegetation. The dominant native vegetation communities occurring in the study area include Mangrove forests and Eucalyptus open forest to woodlands. Eucalyptus communities are dominated by *E. tetrodonta* and *E. miniatia* woodlands (average height 15 m) to open forests containing a mid-stratum of semi-deciduous to deciduous trees and shrubs such as *Corymbia polycarpa*, *E. porreota*, *Livistona humifus*, *Terminalia ferdinandiana* and *Xanthostemon paradoxus* and grasses (Brock, 1995; Williams et al., 1997; O'Grady et al., 1999). Mangrove communities consist of a variety of species, the two dominate species mapped by Brocklehurst and Edmeades (1996) included *Ceriops tagal* closed to open forests (heights 1–10 m) and *Rhizophora stylosa* closed forest, heights between 6 to 18 m tall. Stands of Melaleuca species are located on the floodplain and riparian zones across the study area (Brock, 1995). Patches of wet and dry Monsoon forests are also found across the study area, containing a mix of species. Canopy height in wet monsoon forest rarely exceeds 25 m while dry monsoon forest canopies range from 10 m to 17 m (Bach, 2002; Brock, 1995).

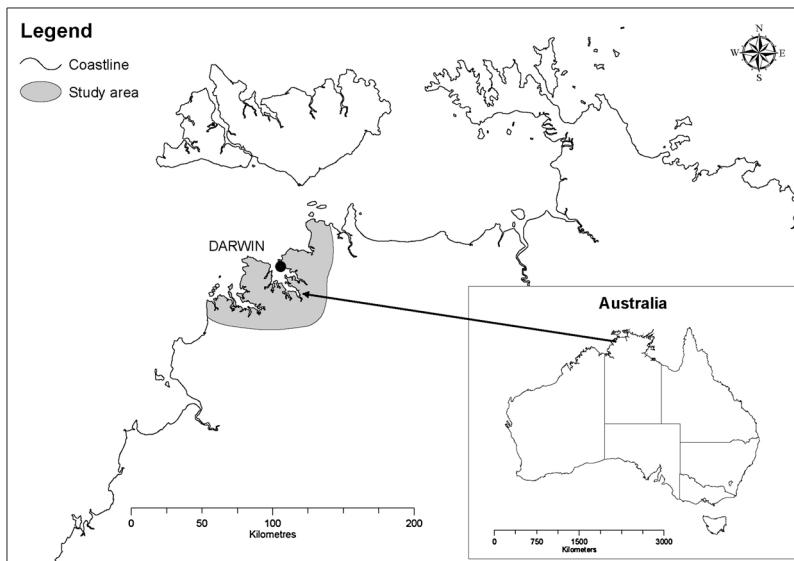


Fig. 1. Location of the study area in Northern Territory of Australia.

2.2. LiDAR canopy height model

The airborne LiDAR used in this study was captured between the 3rd and 5th of July 2009 using a LEICA ALS50-II laser scanner, on-board a Cessna 404 aircraft. LiDAR data were captured within two hours of low tide at a nominal flying height of 2012 m AGL. The LiDAR's foot print was 0.45 m in diameter with an average point density of 1.1 m². These data were pre-processed by contractors and supplied as las files containing 4 discrete returns. A pit filled canopy height model (CHM) was produced at a spatial resolution of 1 m using the suite of LAStools (Khorsravipour et al., 2014) (Fig. 2(a)). The las files were first classified as either ground or non-ground using the lasground tool. These classified las files were then used to normalise the z coordinates of the non-ground returns to represent relative height at ground level (LAStools, 2017). The final step was to produce the raster CHM using the lasgrid tool. As the LiDAR was captured with an average point density of 1.1 m² the pixel size for the output raster layer was set to 1 m. The parameter “-highest” was used as it ensures that only the highest z coordinate for each 1 m² pixel is used to produce the CHM. The parameter “-subcircle” which converts each point to a disc shape was used to remove pits within the tree canopies; for this study the subcircle parameter was set to 0.3 m.¹

At the time of the LiDAR capture (July, mid dry season) the native grasses in the study area were senescent. While the grass understory is senescent it would in some instances represent the highest z coordinate for a given pixel in the development of the CHM. As native grasses in the NT savannas are generally 0.5 m in height (Setterfield et al., 2010) pixels \leq 0.5 m were removed, resulting in a CHM with values ranging from 0.51 to 36.5 m used to produce the dependent variable, mean canopy height.

2.3. Satellite imagery

Details of the Landsat satellite imagery used to develop and validate the canopy height model in this project are shown in Table 1. The canopy height model was developed using a Landsat-5 TM image captured on the 26th of May 2009. This image date was chosen as it was the first cloud free Landsat-5 TM image captured between the transition from the monsoon wet season to the dry season. A number of images

capture in 2009, 2013 and 2016 were used to assess how well the final canopy height model generalised between sensors and years (Table 1). Time series analysis was also undertaken to produce canopy height estimates from Landsat imagery captured between the years 1987 and 2016 to further evaluate the generalisation error. Details of the imagery used to produce the time series are provided as supplementary material.

Imagery captured in the early dry season was chosen to minimise the impact of fire scars in the imagery, as large areas of northern Australia are burnt each year with prescribed burning undertaken by land managers at the onset of the dry season as the grass understory starts to become senescent (Edwards et al., 2013). While imagery captured later into the dry season (June–August) has the potential to provide greater spectral separability between woody and non-woody vegetation the impact of fire scars in the imagery limits the use of these data for producing woody vegetation structure maps. The Landsat image used in this study was atmospherically corrected using 6S radiative transfer code and a bi-directional reflectance distribution function (BRDF) model was applied to the imagery. This BRDF model takes into account topographic illumination effects and produces surface reflectance values to a standard view geometry (Flood, 2014). For a detailed description of the image pre-processing applied to the Landsat imagery used in this study see Flood et al. (2013). To reduce noise introduced by differing atmospheric conditions between image dates, the Blue spectral band which is sensitive to aerosol conditions (Flood, 2014) was not used in the development of the canopy height model. In addition to the five multispectral bands (Green, Red, Near-Infrared (NIR), Short Wave Infrared 1 (SWIR1), Short Wave Infrared 2 (SWIR2)) a number of band ratios and vegetation indices were derived to investigate the relationship between mean canopy height and Landsat-5 TM derived predictor variables.

2.3.1. Vegetation indices and band ratios

The vegetation indices investigated in this study included the more common NIR and Red band indices such as Normalised Difference Vegetation Index (NDVI) and Modified Simple Ratio (MSR), and several indices based on the SWIR and Green spectral bands. A combination of the ratios of all the five surface reflectance bands was also investigated. A total of 12 vegetation indices detailed in Table 2 and 10 band ratios was calculated for this study.

2.4. Training and validation data

The training and validation data were obtained from separate areas

¹ <http://rapidlasso.com/2014/11/04/rasterizing-perfect-canopy-height-models-from-lidar/>

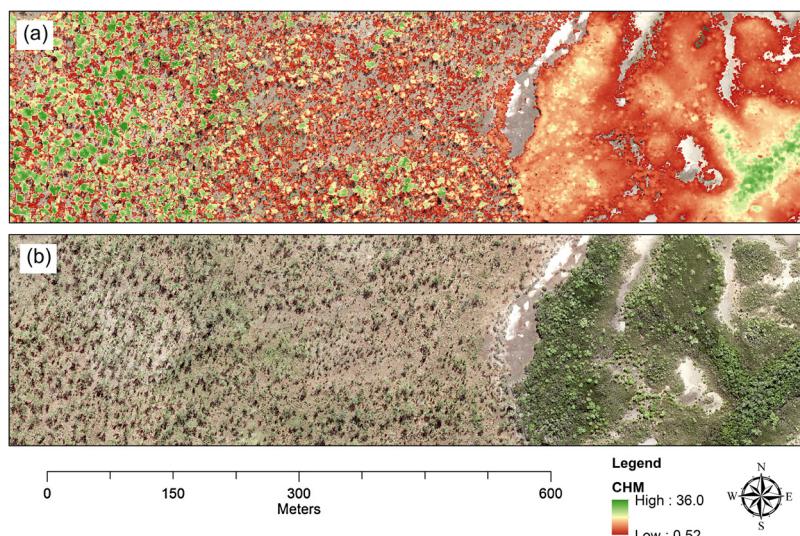


Fig. 2. Example of the canopy height model produced from the LiDAR data captured in 2009 (a) over Eucalyptus woodland (West) and transition to Mangrove forest (in the Eastern side of the figure), 15 cm digital aerial photography (b) shown for clarity.

Table 1
Landsat-5 TM and 7 ETM+ imagery used in the development* and validation of the canopy height model.

Sensor	WRS-2 path/row	Capture date
Landsat-5 TM*	106/69	26/05/2009
Landsat-5 TM	106/68	26/05/2009
Landsat-7 ETM+	106/68	19/06/2009
Landsat-7 ETM+	106/69	27/04/2013
Landsat-8 OLI	106/69	19/04/2013
Landsat-7 ETM+	106/69	21/05/2016
Landsat-8 OLI	106/69	13/05/2016

dataset the mean value for the LiDAR CHM was calculated for each 30×30 m pixel (excluding urban areas), resulting in CHM values ranging from 0 m to 29.1 m and a mean value of 10.5 m. While there is likely to be some misregistration between the location of the LiDAR canopy height model and pixels in the Landsat imagery, we chose a 30×30 window to ensure that the maximum mean LiDAR canopy height range was achieved in the training data. Increasing the window size (e.g. 90×90 m) to account for any misregistration between the two datasets overly reduces the maximum height values (≈ 4 m) in the training dataset. As the random forest algorithm will only predict within the bounds of the training data it was important to maintain as

Table 2
Vegetation indices and band ratios used in this study.

Spectral index	Formula	Reference
Normalized Difference Vegetation Index	$NDVI = \frac{NIR - Red}{NIR + Red}$	Tucker (1979)
Green Soil Adjusted Vegetation Index	$GSAVI = \frac{NIR - Green}{(NIR + Green + 0.5) * (1 + 0.5)}$	Sripada et al. (2006)
Green Normalised Vegetation Index	$GNDVI = \frac{NIR - Green}{NIR + Green}$	Buschmann and Nagel (1993)
Chlorophyll Vegetation Index	$CVI = \frac{NIR * Red}{Green * Green}$	Vincini et al. (2008)
Normalized Difference Greenness Index NDGI	$NDGI = \frac{Green - Red}{Green + Red}$	Bannari et al. (1995)
Normalized Burn Ratio SWIR2 (Band 7)	$NBR = \frac{NIR - SWIR2}{NIR + SWIR2}$	Ji et al. (2011)
Normalized Burn Ratio SWIR1 (Band 5)	$NDII = \frac{NIR - SWIR1}{NIR + SWIR1}$	Ji et al. (2011)
Green Difference Vegetation Index	$GDVI = NIR - Green$	Sripada et al. (2006)
Modified Soil Adjusted Vegetation Index	$MSAVI = \frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - Red)}}{2}$	Qi et al. (1994)
Difference Vegetation Index	$DVI = NIR - Red$	Tucker (1979)
Soil adjusted Vegetation index	$SAVI = \frac{NIR - Red}{(NIR + Red + 0.5) * (1 + 0.5)}$	Huete (1988)
Modified Simple Ratio	$MSR = \frac{\frac{NIR}{Red} - 1}{\left(\frac{\sqrt{NIR}}{\sqrt{Red}}\right) + 1}$	Chen (1996)

within the extent of the LiDAR CHM shown in Fig. 3.

2.4.1. Training data

The training data were located within the extent of Landsat scene 106/69 (path/row) while the validation data were located within the extent of the scene 106/68. Validation data were obtained from areas outside of the extent of the training data to reduce autocorrelation between the training and validation data. To produce the training

much of the height range as possible. It also ensures that as much of the variance in the spectral values of the Landsat image is seen by the model. A stratified random sampling approach for height ranges between 0 and 18 m was taken to ensure that the values of the dependent variable (mean canopy height) were evenly distributed across the height range in the study area (Fig. 3). For canopy heights ranging between 19 m and 20 m only 405 pixels were available, while for heights between 20 m and 29.1 m, only 927 pixels were present. In an

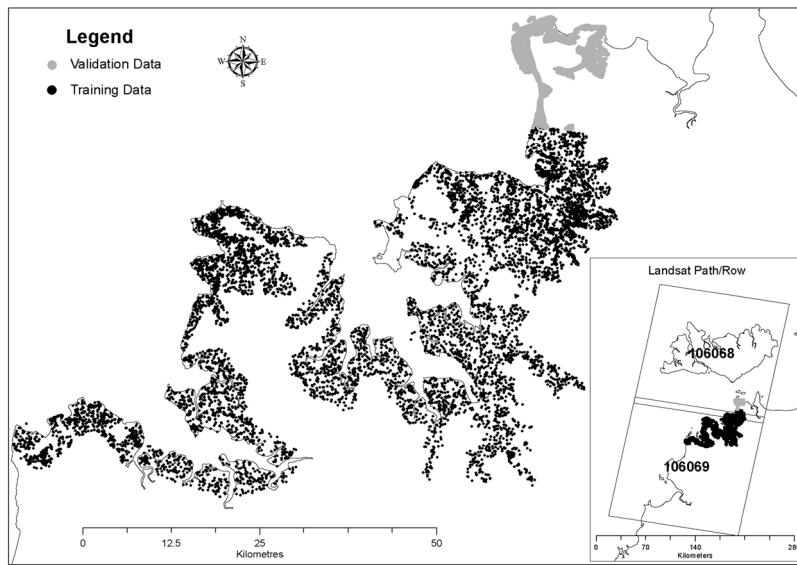


Fig. 3. Location of training and validation data within the extent of the LiDAR data.

attempt to create a more balanced training dataset and retain as much variance as possible 500 pixels were randomly selected at 1 m intervals between 0.51 m and 18 m. An additional 500 pixels were randomly selected from pixels where no canopy height was recorded, resulting in a training dataset with 11,322 pixels.

2.4.2. Validation data

Canopy height estimates predicted from Landsat-5 TM and Landsat-7 ETM+ (Table 1) imagery for 106/68 (path/row) were obtained and compared with the mean LiDAR CHM values. Pixels in the Landsat-5 TM corresponding with the pixels affected by the scan-line corrector error in the Landsat-7 ETM+ imagery were removed from further analysis. Any pixels impacted by cloud, water or fire in both the Landsat-5 TM and Landsat-7 ETM+ imagery were also removed, resulting in a total of 30,500 pixels available to assess the accuracy of the final canopy height model. The CHM values for the validation pixels ranged from 0 m to 25.1 m and a mean value of 7.9 m. Additional assessment was also undertaken to produce accuracy statistics in four vegetation community groups in the region using 1:25,000 scale remnant vegetation mapping (Brock, 1995). To further evaluate the generalisation error the canopy height model was applied to a time series of Landsat imagery captured between 1987 and 2016 for sites located in the Darwin region scene 106/69 (path/row).

2.5. Random forest model development

Random forest is an ensemble learning algorithm (Breiman, 2001) consisting of many decision trees built from a bootstrap sample of the training data. These individual decision trees are then combined to produce a more accurate model (Cutler et al., 2007). Random forest is able to handle thousands of input variables with complex non-linear relationships (Breiman, 2001; Cutler et al., 2007). The random forest algorithm used in this study is implemented in the open source Python module Scikit-learn (Pedregosa et al., 2011). At a minimum there are two parameters that need to be set to produce a random forest model (Rodriguez-Galiano et al., 2012). These parameters define the number of trees (*n_estimator*) used and the number of prediction variables (*max_features*) used to grow the tree. One advantage of random forest is that it also calculates the relative feature importance for each predictor variable for the model (Strobl et al., 2008). In this study we undertook a number of experiments to identify the number of decision trees, important predictor variables and the number of features to grow the trees in the random forest model. This was done to produce a model that

would generalise through time and be computationally practical when applied to an archive of time series imagery (Landsat sensors TM, ETM+, OLI) used in this study.

2.5.1. Model Development Stage One: optimising number of trees

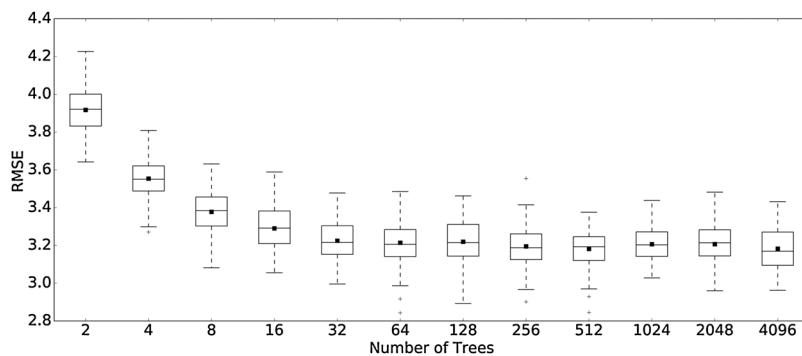
In the first stage, the optimal number of decision trees (*n_estimator* parameter) used to build the model was identified. Studies have shown that predictive accuracy of the random forest algorithm converges with the increase in the number of decision trees (Oshiro et al., 2012; Rodriguez-Galiano et al., 2012; Gao et al., 2015). Similar to the study undertaken by Oshiro et al. (2012) we assessed a range of values (2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, and 4096) to identify the optimal number of decision trees. For each of the *n_estimator* parameters assessed the random forest model was run 100 times and for each iteration 90% of the training data were randomly selected to train the model, while the remaining 10% of these data (data unseen by the model) were used to assess the predictive accuracy of the random forest model using the root mean squared error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where *n* is the number of observations, *y_i* are the observed and *x_i* the predicted mean canopy height respectively. For each test of the random forest model, box plot statistics were produced for the 10% test data's RMSE scores. The candidate *n_estimator* parameter was selected based on the consistency of the predictive accuracy for the 100 iterations of the model. The predictor variables feature importance scores were also calculated for each run of the model and the mean feature importance score for each *n_estimator* was recorded for the 27 parameters tested. The predictor variables feature importance scores (highest to lowest) for the candidate *n_estimator* were then used in the second stage of the models development. For the first stage the *max_features* parameter was set to the default “auto” in Skit-learn which uses all features.

2.5.2. Model Development Stage Two: variable selection

In the second stage highly collinear predictor variables were identified and removed from further analysis. A correlation matrix was used to assess the linear and non-linear relationships between all predictor variables. Visual assessment of the correlation matrix identified a number of the band ratios and vegetation indices that were transformations of each other. Variables that were transformations of each other were identified and the least important variable based on the

**Table 3**

Results for the predictor variable importance scores (Model Development Stage One) for 512 decision trees (n_estimator).

Predictor variable	Importance score	Predictor variable	Importance score
SWIR1	17.9480	SWIR2/SWIR1	1.4282
NIR/Green	14.5789	SWIR2/Green	1.3867
GNDVI	14.4570	SWIR2/NIR	1.3272
GSAVI	10.5619	NBR	1.3054
SWIR2	5.7975	CVI	1.2069
SAVI	4.7715	NIR	1.1859
SWIR1/Red	2.8598	SWIR2/Red	1.1787
SWIR1/Green	2.3242	Red	1.1585
MSR	2.2335	GDVI	0.8464
NIR/Red	2.2216	DVI	0.8196
NDVI	2.2197	MSAVI	0.8149
Green	1.9765	Red/Green	0.7292
NDII	1.9681	NDGI	0.7291
SWIR1/NIR	1.9652		

feature importance score highest to lowest output from stage one were removed from further analysis. For the remaining variables the Pearson's correlation coefficient was calculated and any predictor variables with a score ≥ 0.95 were assessed and the least important of the two variables based on the feature importance score were removed from further analysis. Predictor variables were assessed from highest to lowest based on the feature importance score output in stage one.

2.5.3. Model Development Stage Three: optimising number of variables

In the third stage, a backward elimination process was undertaken to identify the number of predictor variables and the number of features used to grow the trees (max_features). In this analysis the number of decision trees (n_estimators) and predictor variables identified in stage one and two were used as inputs for the random forest model. Two max_features parameters (auto, log2) were evaluated to optimise the number of features used to grow the trees in the random forest model. The setting “auto” is the default setting which equates to using all predictor variables, while the “log2” is the logarithm base 2 of the number of predictor variables. For each suite of predictor variables the mean feature importance score was calculated, the least important variable was removed and the process repeated until only one variable remained. Again for each suite of predictor variables 100 iterations of the random forest were run (as described above) and the mean accuracy statistics based on the independent test data (randomly selected each permutation) were calculated. This approach was taken to expose the random forest model to as much variance in the training data as possible while still assessing the predictive accuracy of the model using an independent set of data. The final candidate model was selected using the accuracy statistics (test data R^2 and RMSE) derived from the 100 independent runs of the random forest model for each suite of predictor variables.

Fig. 4. Box plots showing the RMSE scores for the number of trees (n_estimator) in the random forest model; each box plot represents 100 iterations of the model where 10% of the training data is randomly selected to independently test each iteration of the model (dot = mean, box = 25th and 75th percentile, line = median, whiskers = show the range of the data).

2.5.4. Final model validation

The optimal suite of predictor variables max features and n_estimators identified in stages 1, 2 and 3 were used to produce a final model from all the training data ($n = 11,322$). The final model was then applied to Landsat-5 TM and Landsat-7 ETM+ imagery to predict mean canopy height for path/row 106/68. Accuracy of the model was evaluated using a number of statistics, including the coefficient of determination (Zar, 1984), RMSE, variance and the bias. Variance was used to assess the precision of the predicted observations;

$$\text{variance} = \frac{1}{n-1} \sum_{i=1}^n ((x_i - \bar{x}) - \bar{e})^2 \quad (2)$$

where \bar{e} is the mean of the error. Bias was used to assess the average difference between the predicted and observed mean canopy height values;

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - x_i}{n} \quad (3)$$

Validation statistics were produced from Landsat-5 TM imagery captured on the same date as the models training data, while statistics produced from Landsat-7 ETM+ were used to assess how transferable the model was to other Landsat sensors and image capture dates. The final model was also applied to Landsat-8 OLI imagery to enable comparisons between Landsat-7 ETM+ and produce examples of time series plots for a number of sites. Canopy heights predicted from the three Landsat sensors for the years 2009, 2013 and 2016 were also compared to investigate any bias between the different sensors. This was done by producing scatter plots and using the slope of the regression line forced through the origin (using ridge regression) to identify any systematic bias between sensors (Flood, 2014). The final canopy height model was also applied to all available Landsat imagery for scene 106/69 between the years 1987 and 2016 and a number of time traces were produced to show canopy height estimates for this period in sites that were severely impacted by cyclone Tracy in 1974.

3. Results and discussion

3.1. Model Development Stage One: optimising number of trees

To reduce the computational burden of the random forest model we undertook an experiment to identify the optimal number of decision trees, the results are presented as box plots in Fig. 4. Each box plot represents the RMSE values for the number of trees in the random forest model (based on 100 using independent test data) with mean RMSE values ranging between 3.18 m and 3.92 m. The lowest mean RMSE score was recorded for n_estimator values 512 and 4096. These results are consistent with other studies which have found that as the number of decision trees increase the overall accuracy converges (Oshiro et al., 2012; Gao et al., 2015; Rodriguez-Galiano et al., 2012; Belgiu and Drăgu, 2016). A recent review of remote sensing applications using random forest reported the number of decision trees being used ranged

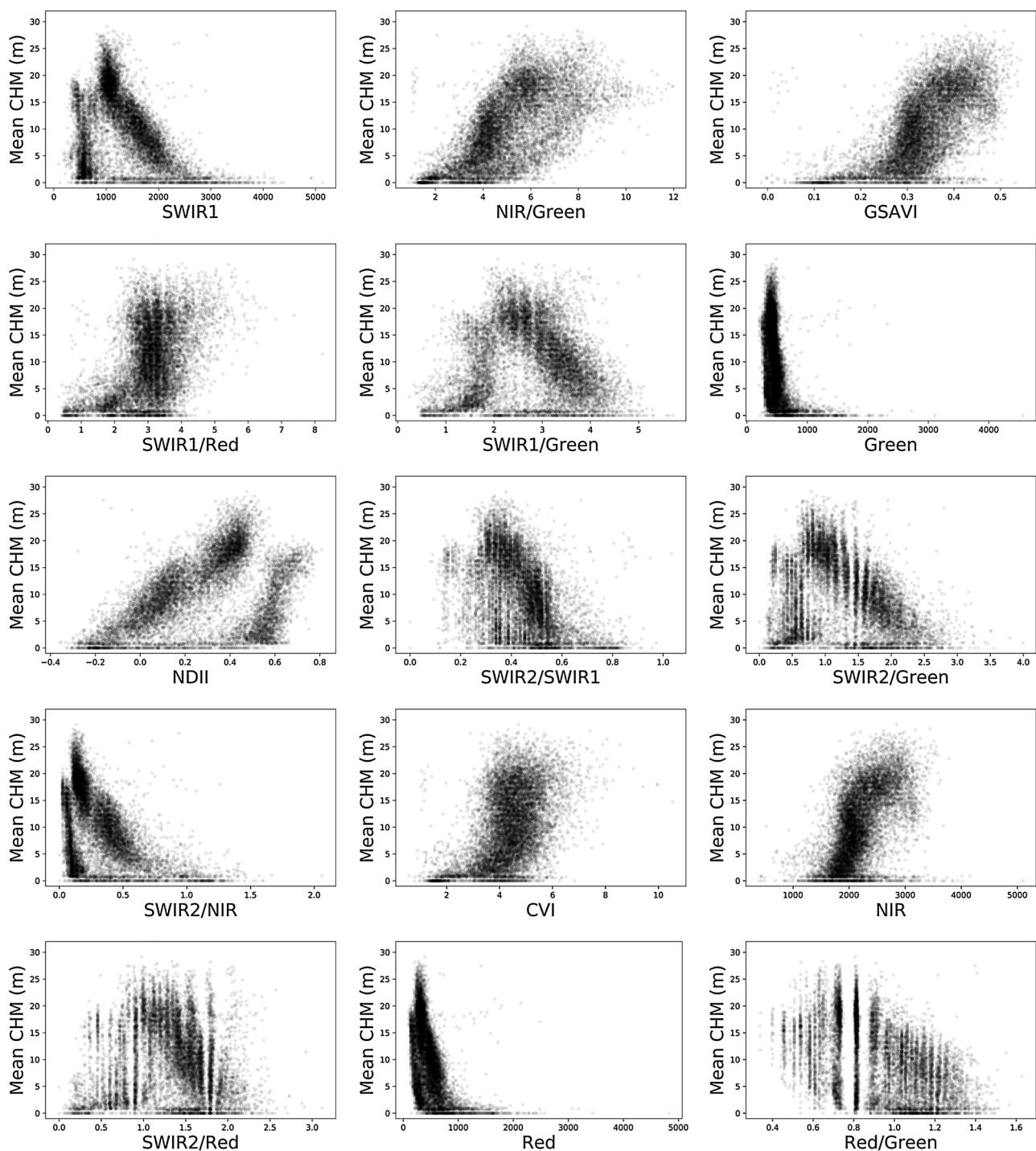


Fig. 5. Scatter plots showing the relationship between mean CHM and the 15 predictor variables identified in stage two of the development of the random forest model.

between 70 and 5000, with a majority using 500 to build their models (Belgiu and Drăgu, 2016). Based on these studies, Belgiu and Drăgu (2016) recommended that 500 represents a default value for remote sensing data. Based on the empirical results of this study we selected 512 decision trees for further use in the development of the random forest model.

3.2. Model Development Stage Two: variable selection

While random forest is a non-parametric model that can handle complex interactions and high dimensional datasets (Breiman, 2001;

Strobl et al., 2008), to further increase the computational efficiency and obtain a more parsimonious model (Wilkes et al., 2015) we identified and removed predictor variables that were highly correlated. Visual assessment of a correlation matrix for the 27 predictor variables assessed in this study identified a number of the band ratios and vegetation indices that were transformations of each other. The predictor variable importance scores output for 512 decision trees (stage one, Table 3) were used to guide the decision on which of the correlated variables were removed. From the original 27 predictor variables 12 were removed. Out of the 15 candidate variables eight were combinations of band ratios, three were vegetation indices and the remaining

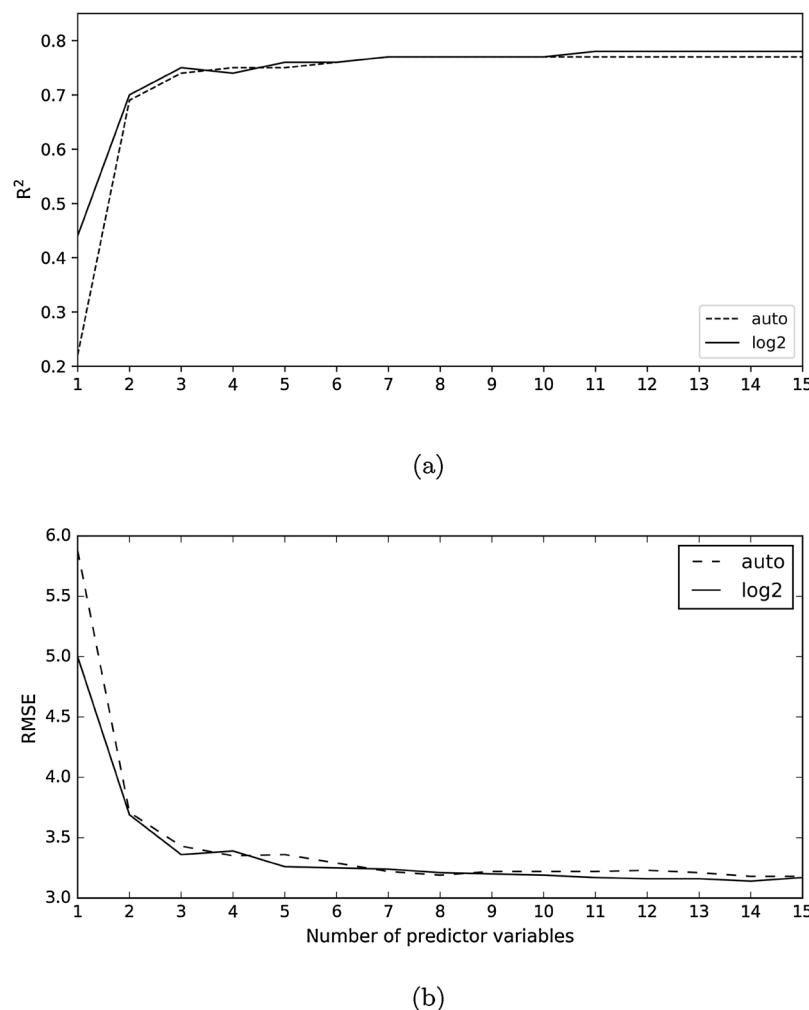


Fig. 6. Mean R^2 (a) and RMSE (b) values (based on 100 iterations) used to identify the number of predictor variables and the number of features used to grow the trees in the random forest model.

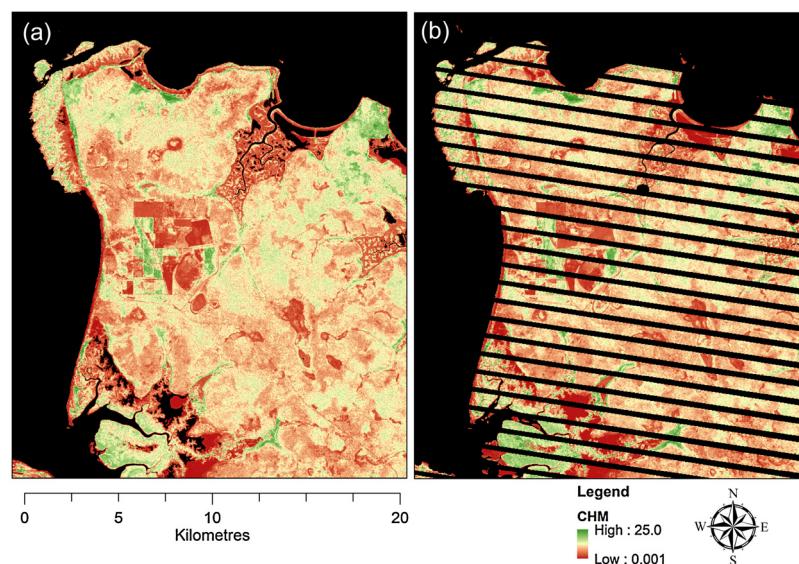


Fig. 7. Random forest canopy height model applied to both (a) Landsat-5 TM (acquired 26/05/2009) and (b) Landsat-7 ETM+ (acquired 19/06/2009) imagery for path/row 106/68 to independently validate the accuracy of the model; black areas denote no data values. Note, the region shown in this figure is the grey area shown in Fig. 3.

were the single surface reflectance bands. The relationship between mean CHM and the remaining 15 candidate predictor variables are shown in scatter plots (Fig. 5). Interestingly none of the variables alone suggest a particularly strong relationship with mean LiDAR canopy

height, and several variables clearly display a bimodal relationship.

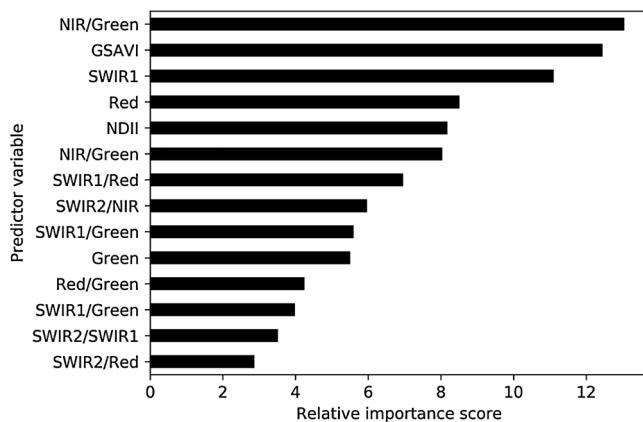


Fig. 8. Relative importance score for the 14 predictor variables used in the final random forest model.

3.3. Model Development Stage Three: optimising number of variables

To reduce over-fitting and further optimise the random forest model we undertook an experiment to assess different values of the max_features and the number of predictor variables used to build the model. The max_features parameter (in the scikit-learn implementation of random forest regressor) enables the user to determine the number of predictor variables randomly selected to split each node of the individual decision trees (Müller and Guido, 2016). Reducing the max_features parameter increases the variance and conversely decreases the bias and produces weaker individual decision trees, however when combined in the ensemble of decision trees it increases the models overall accuracy (Breiman, 2001; Rodriguez-Galiano et al., 2012). Accuracy statistics for the variable reduction and max_features parameters analysis are shown in Fig. 6. RMSE values ranged from 3.14 to 5.88 m

while R^2 values ranged from 22% to 78%. Overall the accuracy statistics were similar for each of the max_features and number of predictor variables, with a decline in accuracy starting to occur around eight predictor variables. The highest R^2 value (78%) was recorded for max_features parameters “log2” and predictor variables between 11 and 15 (78%), while the lowest RMSE value was recorded for 14 predictor variables. Reducing the number of predictor variables randomly selected to split the nodes increased the overall accuracy of the model slightly; based on these results we selected the max_features parameter “log2” and the 14 predictor variables shown in Fig. 8 to produce the canopy height model.

3.4. Canopy height model

The final canopy height model was developed using the following parameters; 512 decision trees, “log2” max_features parameter and 14 predictor variables based on all training data (11,322 data points). The random forest model was serialised to enable it to be applied to Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 OLI imagery. To assess the accuracy of the model on independent data captured outside of the extent of the training data, the model was applied to Landsat-5 TM and Landsat-7 ETM+ imagery captured over path/row 106/68. Canopy height estimates derived from Landsat TM and ETM+ for the area used to validate the final random forest model is shown in Fig. 7.

3.4.1. Predictor variable importance

The importance scores for the 14 predictor variables for the random forest model are shown in Fig. 8. The most important predictor variable was the NIR/Green ratio followed by the vegetation index GSAVI and the SWIR1 band. A combination of band ratios were found to be important, representing eight of the 14 predictor variables.

The green band featured six times, while the Red, NIR and SWIR1 bands occurred five times in either vegetation indices, band ratios or as individual bands. Hill et al. (2011) investigated the relationships

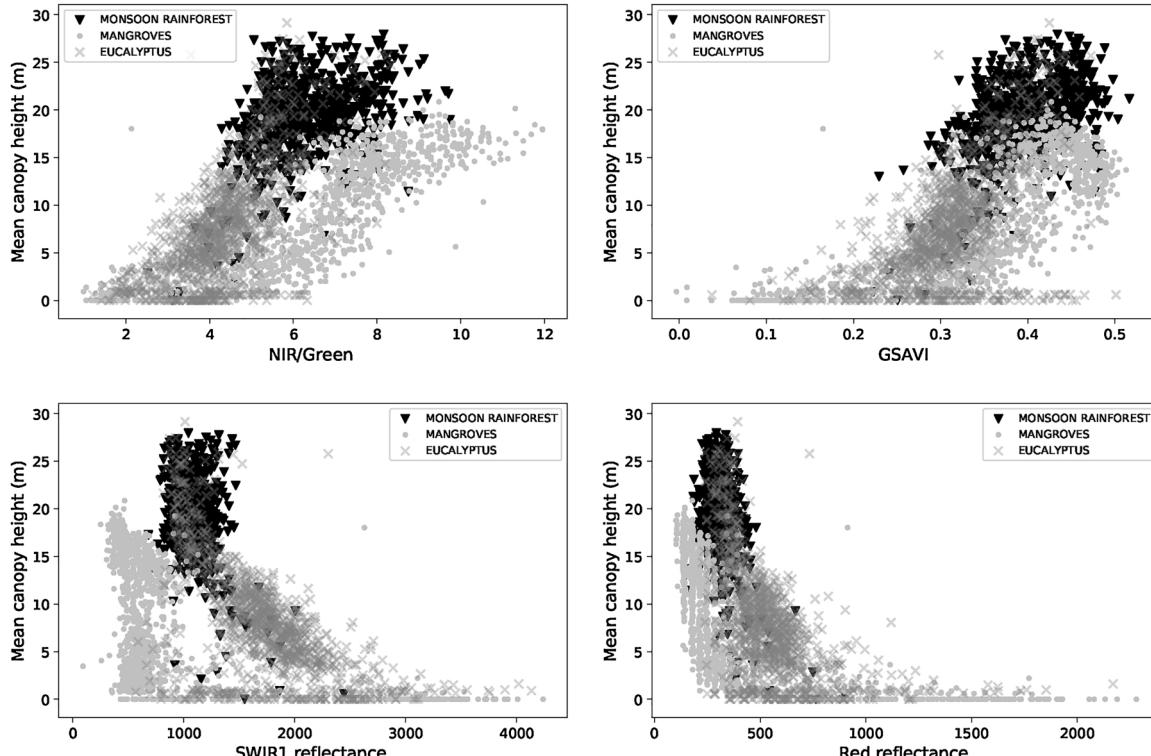


Fig. 9. Scatter plot showing the relationship between mean canopy height and the top four predictor variables, ratio of NIR and Green, GSAVI, SWIR1 reflectance and Red reflectance for three vegetation communities (Monsoon rainforest $n = 759$, Mangroves $n = 1182$, Eucalyptus $n = 1159$). Data points were subset from the training dataset within the extent of the 1:25,000 scale remnant vegetation mapping (Brock, 1995).

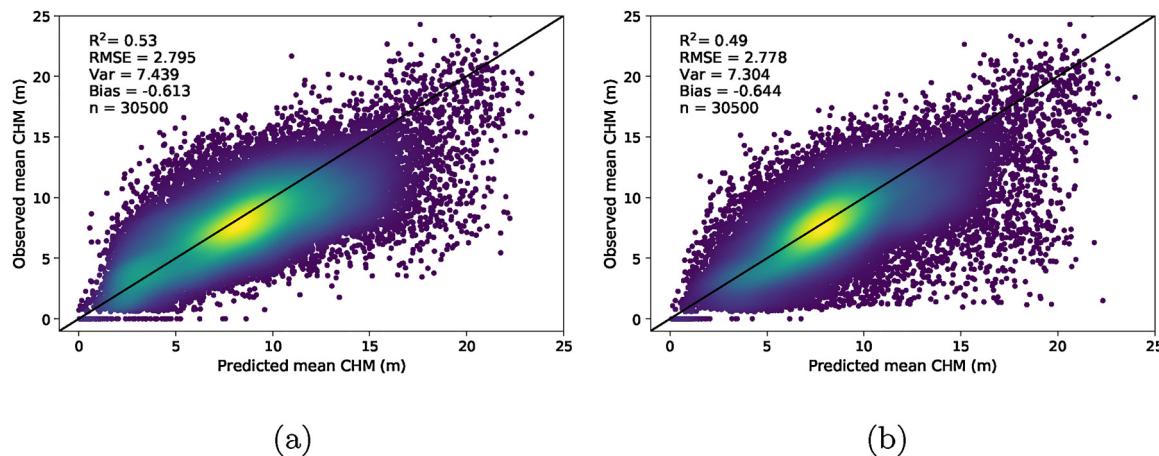


Fig. 10. Scatter plot showing the predicted and observed mean CHM for the validation region: (a) Landsat-5 TM acquired 26/05/2009 and (b) Landsat-7 ETM+ acquired 19/06/2009; solid line = 1:1 line.

between LiDAR derived canopy height and Landsat-7 ETM+ data for four tropical rainforest types. Their study showed that the NIR band was the most sensitive variable when related to mean and maximum canopy height followed by the green normalised vegetation index. They concluded that their study supported the hypothesis that canopy height distribution and shadow effects due to canopy complexity and emergent trees significantly influences spectral response for tropical rainforests (Hill et al., 2011). Hansen et al. (2016) mapped tree height distributions in Sub-Saharan Africa covering a wide range of vegetation community types using Landsat-7 ETM+ and Landsat-8 OLI. They used a regression tree model and found that seasonal derivatives (inter-percentile ranges) from the Red band were the top two important variables, while the SWIR2 band was the third most important (Hansen et al., 2016). These studies highlight the complex relationships between woody vegetation height and spectral reflectance of the Landsat bands. In this study, the woody vegetation communities are structurally variable, ranging from closed forests, woodlands, open woodlands and grasslands with sparse trees. The variability in the relationship between the top four predictor variables and canopy height for three major vegetation communities in this study are shown in Fig. 9. The distinct relationships between canopy height and ratio NIR/Green is evident for each of the three broad vegetation communities. As with Hill et al. (2011) it is likely that the shadow fraction contributed to the spectral reflectance in the Landsat pixels in many of the vegetation communities in this study. The SWIR bands, which are sensitive to leaf water content (Hunt and Rock, 1989) featured eight times in the 14 predictor variables. Typically, the SWIR reflectance values will increase as leaf water content decreases (Ji et al., 2011). The bimodal relationship between SWIR1 reflectance and canopy height for three of the major vegetation communities in the study area can be seen in Fig. 9. SWIR1 reflectance values for the Mangrove communities were low across the range of canopy heights, while mangrove canopy cover is dense and the leaf water content is likely to contribute to the low values, soil moisture and water would also be influencing the SWIR1 values. The monsoon rainforests in the study area generally have dense canopy cover, resulting in low SWIR1 reflectance values clustered above the 12 m canopy height, while for Eucalyptus communities SWIR1 reflectance increases as canopy height declines. While the SWIR1 band was the third most important predictor variable in the model the relationship between canopy height and Mangrove and Monsoon rainforest communities is low. It is clear from the individual relationships between the canopy height and the predictor variables in this study that no one variable has a particularly strong relationship (Figs. 5 and 9). One of the benefits and strengths of the random forest algorithm is its ability to handle non-parametric data and learn from the complex interactions between variables (Grömping, 2009; Strobl et al., 2008; Avitabile et al.,

2012).

3.4.2. Model validation

Accuracy statistics and the predicted and observed mean canopy height for Landsat-5 TM and Landsat-7 ETM+ (path/row 106/68) are shown in Fig. 10. The model performance for the two sensors and image dates were very similar with only slight differences in the accuracy metrics. Canopy height predicted from Landsat-5 TM recorded a R^2 of 0.53 and RMSE value of 2.795 m, while Landsat-7 ETM+ recorded R^2 of 0.49 and RMSE 2.778 m, the variance and bias values were also very similar for both sensors. Even though the validation data were obtained from outside the extent of the training data, it would be expected that the Landsat-5 TM image would have achieved a higher accuracy as it was captured on the same path (106) and date of the image used to develop the model, which is essentially the same image. The similar results between sensors suggest that the random forest model has transferred well to the Landsat-7 ETM+ sensor and has low generalisation error.

To further evaluate the models performance and generalisation error, accuracy statistics were obtained for vegetation communities using existing mapping undertaken by Brock (1995). The predicted and observed mean canopy height and accuracy metrics for the four vegetation communities (Mangrove, Eucalyptus, Monsoon rainforest and Melaleuca) are shown in scatter plots (Fig. 11). There were variations in the models performance between sensors and vegetation communities. The lowest RMSE value 2.332 m was recorded for Eucalyptus communities predicted from Landsat-7 ETM+, while for Landsat-5 TM the lowest RMSE value (2.356 m) was recorded for the Eucalyptus communities, followed by the Mangrove communities (2.864 m). Bias was lowest for the Eucalyptus and Mangrove communities for both sensors with the predicted and observed data reasonably distributed around the 1:1 line. Bias in the model for both the Monsoon Rainforest and Melaleuca communities is clearly seen in Fig. 11 with both sensors overestimating mean canopy height. Landsat-7 ETM+ recorded the lowest RMSE (3.037 m) for the Monsoon Rainforests, while Landsat-5 TM recorded the lowest RMSE for the Melaleuca communities (3.674 m).

The Monsoon Rainforest in the study area is described as either wet or dry (Bach, 2002). Wet monsoon rainforests are predominately evergreen with canopies rarely greater than 25 m, while dry monsoon forest are more diverse and contain deciduous species with a maximum canopy height of 12 m (Bach, 2002). The overestimation of canopy height clustered around the 12–17 m range predicted from Landsat-5 TM are predominately located in dry monsoon rainforest patches. There is a slight improvement in the predictions from dry rainforests from Landsat-7 ETM+ captured later in the year (June), which is likely due to a reduction of the overall canopy greenness as a result of leaf fall in

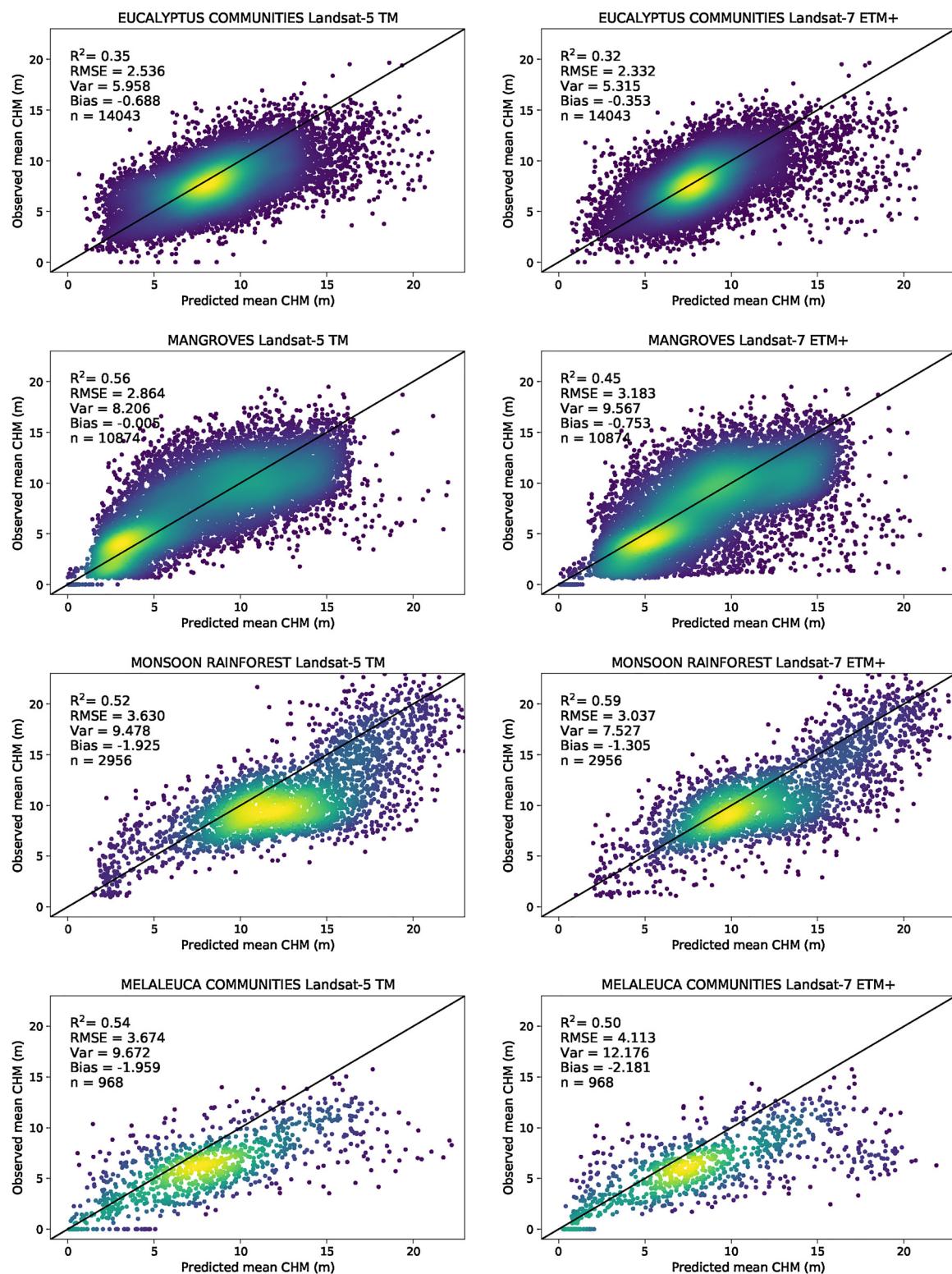


Fig. 11. Scatter plots showing the predicted CHM from Landsat-5 TM (26/05/2009) and Landsat-7 ETM + (19/06/2009) for path/row p106r068 and the observed mean CHM (LiDAR, 03/07/2009) for broad vegetation communities located in the validation region; mapping by [Brock \(1995\)](#), solid line is the 1:1 line.

dry rainforests at this time of the year (Bach, 2002). The bias observed in the dry monsoon rainforest and Melaleuca communities may be also due to the lower number of data points available to train the model. Avitabile et al. (2012) mapped woody biomass using random forest and stressed the importance of obtaining sufficient training data to capture the spectral variance within each classes. While the training data (Section 2) in this study was stratified to obtain an even distribution

across the height ranges, it is likely that there were insufficient samples from both the Dry Monsoon Rainforest and Melaleuca communities, as they represent a small percentage of the overall study area.

3.4.3. Model generalisation/time series

Pre-processing the imagery to standardise the reflectance values between image dates and sensors is important for time series analysis

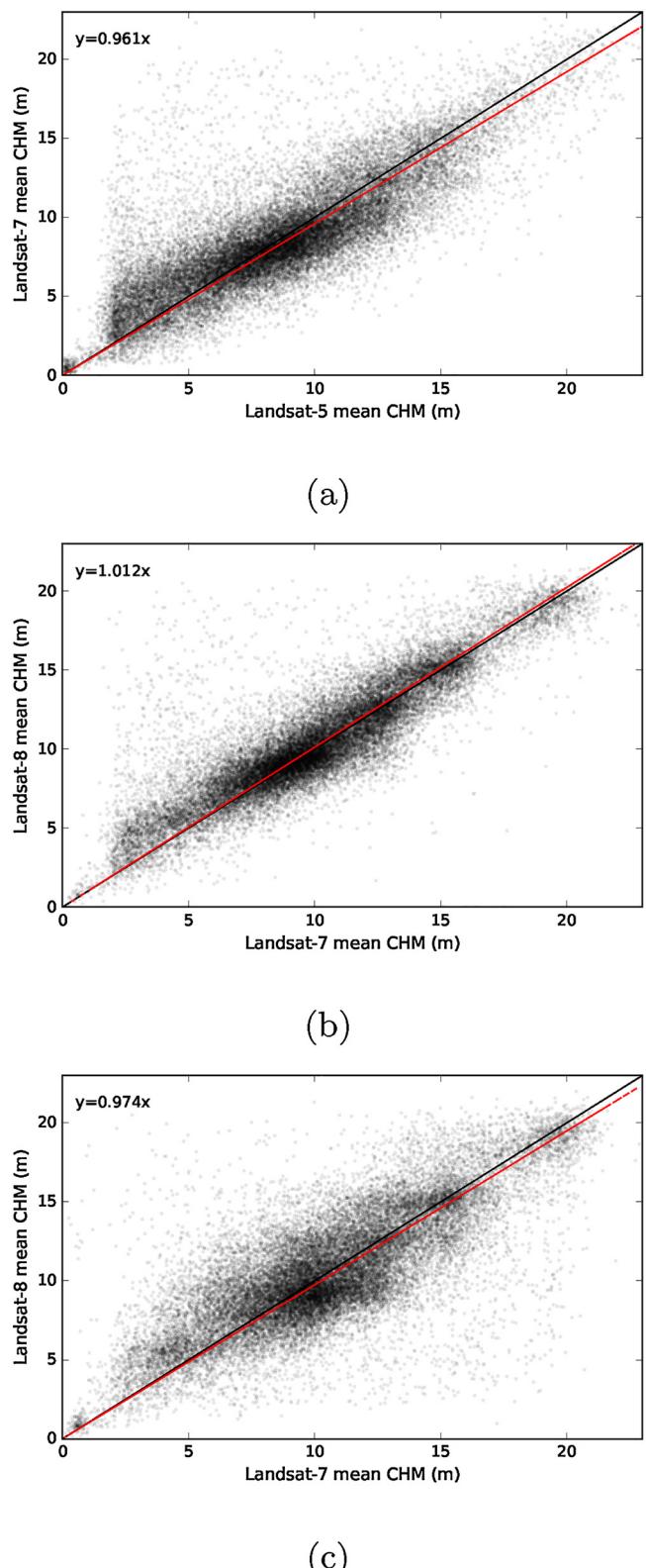


Fig. 12. Scatter plots comparing mean canopy height estimates predicted from Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 OLI in (a) 2009, (b) 2013 and (c) 2016; solid line is the 1:1 line and the dashed red line is the ridge regression. Statistics were only obtained from regions corresponding with valid Landsat-7 ETM+ pixels from the validation region for each year.

(Roy et al., 2016; Vicente-Serrano et al., 2008). In this study, the Landsat images were pre-processed (atmospheric, BRDF and topographic corrected) to a standardised solar zenith angle of 45° (Flood et al., 2013; Flood, 2014) which enabled the model developed from Landsat-5 TM to be directly applied to both Landsat-7 ETM+ and Landsat-8 OLI. Scatter plots comparing canopy heights predicted from Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 OLI from the validation sites are shown in Fig. 12. There was reasonable agreement between sensors, with the ridge regression slope close to 1 for all three years assessed. The results of this study are similar to Flood (2014) who reported similar regression values when comparing NDVI and a fractional cover model predicted from Landsat-7 ETM+ and Landsat-8 OLI imagery.

Canopy height estimates for sites located in the Mangrove community, wet Monsoon Rainforest, and Eucalyptus community predicted from Landsat imagery (captured between April and June) over a period of 29 years (1987–2016) are shown in Fig. 13. These sites were all severely impacted by cyclone Tracy on 25th December 1974 with large numbers of trees windthrown and damaged (Cameron et al., 1983; Stocker, 1976; Fox, 1980). In each time trace the individual canopy height estimates for each Landsat sensor are shown along with smoothed fitted line (dashed line) using a rolling median of three years. Variability in canopy height estimates for individual dates within a year is evident. It is likely that a number of factors are contributing to this variability, including differences in sensor characteristics, atmospheric conditions and seasonal variation in phenology between image capture dates. While pre-processing of imagery has been undertaken to standardise the reflectance values, it is likely that the 6S radiative transfer code used is unable to remove all the atmospheric effects resulting in variability in the spectral values recorded by the sensor. Differences in canopy height values could also be due to variability in spectral response due to differences in timing of leaf fall. The predictor variables used in the random forest model are all sensitive to changes in the level of photosynthetic vegetation and plant vigour, while the two dominant Eucalyptus species in the study region are evergreen and canopy foliage projective cover remains reasonably high through the dry season (Williams et al., 1997). Many of the mid story tree species in the Eucalyptus communities in the study area are semi deciduous or fully deciduous and timing of leaf fall varies between and within species and years (Myers et al., 1997; Williams et al., 1997). This variation in phenology is likely to be more pronounced in Eucalyptus communities where the upper canopy is more open and the mid-story vegetation is more visible to the sensor.

The level of damage to Eucalyptus communities in the region (along McMillan's Road, Darwin) of the time trace (Fig. 13d) six months after cyclone Tracy is evident (Fig. 14a). For the Eucalyptus community sites shown in Fig. 13c and d canopy height appears to be increasing and starts to reach an asymptote around the mid to late 1990s for these sites. Wilson and Bowman (1987) undertook field work in the Howard Peninsula (located in the study area) in 1985 and reported dense regrowth and crown damage in the over story in the Eucalyptus communities and attributed it to the impact of cyclone Tracy in 1974. Likewise Brock (1995) also undertook field work during 1994 and 1995 in the study area and reported that within the Eucalyptus communities there were stands of vigorous uniform regrowth of *E. tetrodonta* and *E. miniatia* to 10 m tall, and attributed this to recovery as a result of extensive damage from cyclone Tracy in 1974. The recovery observed in the 1994–1995 field surveys appears to be evident in the two Eucalyptus time trace examples with canopy heights around 8 m around that time period. The sudden drop in canopy height shown in Fig. 13c around 2013 is an example of land clearing for residential development in the area. In contrast to the Eucalyptus communities the wet Monsoon Rainforests example shows that canopy height remains consistent over the 29 years (Fig. 13b). This Monsoon Rainforest site was also severely impacted by cyclone Tracy with a large number of trees uprooted or having major branch or trunk damage (Stocker, 1976).

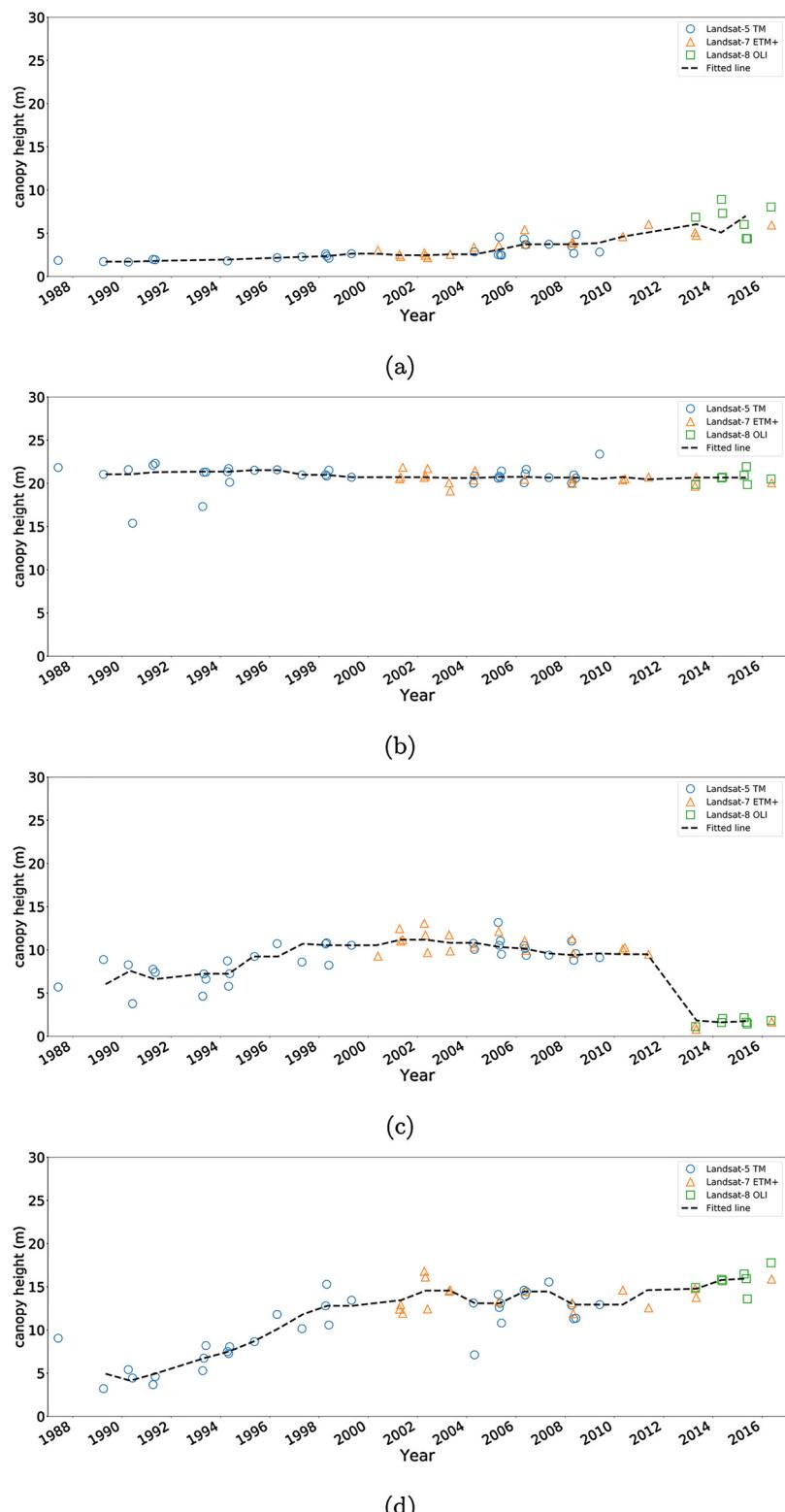


Fig. 13. Canopy height estimates predicted from Landsat-5 TM, Landsat-7 ETM + and Landsat-8 OLI for sites impacted by cyclone Tracy from 1987 to 2016 for a (a) mangrove forest, (b) monsoon rain forest, (c) Eucalyptus woodland with clearing and (d) Eucalyptus woodland.

Regeneration months after cyclone Tracy in the Monsoon forest in the Darwin region was reported to be profuse (Fox, 1980). Wilson and Bowman (1987) reported that Monsoon rainforest in the Howard Peninsula (located 15 km west of the time trace, Fig. 13b) 10 years after cyclone Tracy appeared to have no apparent evidence of damage, other than an absence of canopy emergents. The canopy height for the wet

Monsoon rainforest shown in Fig. 13b is consistent with recovery (from cyclone Tracy) having occurred prior to the capture of the first Landsat image (1987) used in this study. The wet Monsoon rainforest time trace (Fig. 13b) shows less variability in canopy height between sensors; this may be due to the evergreen species in wet Monsoon rainforests resulting in more consistent foliage cover through time (Bach, 2002).



Fig. 14. Example of the damage to Eucalyptus communities (along McMillan's Road, Darwin) due to cyclone Tracy 25th December 1974 and subsequent recovery and regrowth: (a) photograph taken (by Roy Beames) 5th July 1975 precise location on McMillan's Road unknown; (b) photograph taken 27th December 2016 at the location of time trace (Fig. 13(d)).

There is a clear spike in the rainforest canopy height recorded in 2009. This point represents data used to fit the random forest model and indicates that for this wet Monsoon forest it is likely that canopy height is being underestimated. This is in contrast with the dry Monsoon forest (discussed earlier) where canopy height is being overestimated. Damage to the Mangrove forests in the study area due to cyclone Tracy ranged from minor to severe (Stocker, 1976) with mangrove death evident decades later in some areas (Rogers et al., 2017). Mangrove forests in the vicinity of the time trace (Fig. 13a) were severely damaged during cyclone Tracy with dead trees still evident 26 years after the cyclone (Ferwerda et al., 2007). It appears that mangroves at this site have taken decades to recover from the cyclone damage with an increase in canopy height, starting to occur early 2000.

4. Conclusions

In this study we implemented a random forest regression model to predict canopy height from a single date Landsat-5 TM scene, across a variety of natural vegetation communities in the Northern Territory, Australia. The model was trained with a LiDAR-derived canopy height model (CHM) ($R^2 = 0.53$, RMSE = 2.8 m). A three-stage approach was undertaken to tune the random forest model and select the predictor variables used in the final model. Despite none of the individual independent predictor variables derived from Landsat-5 TM having a strong relationship with the dependent variable (LiDAR derived canopy height), the non-parametric random forest algorithm was able to account for the complex relationships between dependent and independent variables. The ability of the model to be applied to both Landsat-7 ETM+ and Landsat-8 OLI imagery was also assessed. The final canopy height model was applied to the first cloud-free Landsat-7 ETM+ image acquired after the LiDAR capture date resulting in a $R^2 = 0.49$ and RMSE = 2.8 m. The model was then applied to the three Landsat sensors to produce time series plots for a period of 27 years (1987–2016). The results showed that the model could be transferred to Landsat-7 ETM+ and Landsat-8 OLI, however, there was variability between canopy height estimates likely to be due to seasonal variations in the image capture dates and sensors characteristics. The four time series plots used as example were all impacted by severe tropical cyclone Tracy in 1974. The canopy height estimates for the Monsoon forest example showed that it had recovered to the current height by 1987 (13 years post cyclone) while the two Eucalyptus community site were still in a state of recovery. The canopy height estimates for the Mangrove community indicate that this site has taken decades to recover and has only recently recovered to its pre-cyclone condition. It should be noted that the random forest regression model will not extrapolate beyond the range of canopy height values in the training dataset. In this study, the minimum and maximum heights in the validation dataset did not exceed the values in the training dataset, and the vegetation communities were similar in both regions. Further work is required to assess how well the random forest model transfers beyond the current study area. This study demonstrates that canopy height can

be predicted from optical Landsat imagery at moderate to high levels of accuracy. The robustness of the model across a range of vegetation communities and three different Landsat sensors illustrate that our approach can be successfully used to explore changes in woody vegetation canopy height through time.

Acknowledgements

This study would not have been possible without the support of the Northern Territory Government and the collaborative partnership between the Northern Territory Government's Department of Environment and Natural Resources, Rangelands Division and Queensland Government's Department of Environment and Science, Remote Sensing Centre. Also thanks to Neil Flood for assistance and advice in the development of the python code used in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jag.2018.08.013>.

References

- Ahmed, O.S., Franklin, S.E., Wulder, M.A., White, J.C., 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS J. Photogramm. Remote Sens.* 101, 89–101. <https://doi.org/10.1016/j.isprsjprs.2014.11.007>. <http://www.sciencedirect.com/science/article/pii/S0924271614002755>.
- Armston, J.D., Denham, R.J., Danaher, T.J., Scarth, P.F., Moffiet, T.N., 2009. Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. *J. Appl. Remote Sens.* 3, 1–28. <https://doi.org/10.1117/1.3216031>.
- Avitabile, V., Baccini, A., Friedl, M.a., Schmullius, C., 2012. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sens. Environ.* 117, 366–380. <https://doi.org/10.1016/j.rse.2011.10.012>. <http://linkinghub.elsevier.com/retrieve/pii/S0034425711003609>.
- Bach, C.S., 2002. Phenological patterns in monsoon rainforests in the Northern Territory, Australia. *Austral Ecol.* 27, 477–489. <https://doi.org/10.1046/j.1442-9993.2002.01209.x>.
- Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. *Remote Sens. Rev.* 13, 95–120. <https://doi.org/10.1080/02757259509532298>.
- Barlow, J., Martin, Y., Franklin, S.E., 2003. Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Can. J. Remote Sens.* 29, 510–517. <https://doi.org/10.5589/m03-018>.
- Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Bowman, D., Panton, W., 1994. Fire and cyclone damage to woody vegetation on the north coast of the Northern Territory, Australia. *Aust. Geogr.* 25, 32–35. <https://doi.org/10.1080/00049189408703096>. URL: <http://www.tandfonline.com/doi/abs/10.6701080/00049189408703096>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brock, J., 1995. Remnant Vegetation Survey: Darwin to Palmerston Region. Technical Report Greening Australia N.T. Darwin. <http://hdl.handle.net/10070/239792>.
- Brocklehurst, P., Edmeades, B., 1996. Mangrove survey of Darwin Harbour Northern Territory (N.T.). Technical Report Dept. of Lands, Planning and Environment Darwin. <http://www.territorystories.nt.gov.au/handle/10070/213495>.
- Brown, L., Chen, J.M., Leblanc, S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. *Remote Sens. Environ.* 71, 16–25. [https://doi.org/10.1016/S0034-4257\(99\)00035-8](https://doi.org/10.1016/S0034-4257(99)00035-8).

- Buschmann, C., Nagel, E., 1993. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *Int. J. Remote Sens.* 14, 711–722. <https://doi.org/10.1080/01431169308904370>.
- Cameron, D., Rance, S., Lukitsch, P., 1983. Tree damage in Darwin parks and gardens during cyclones Tracy and Max. *Landscape Plan.* 10, 89–108. [https://doi.org/10.1016/0304-3924\(83\)90054-0](https://doi.org/10.1016/0304-3924(83)90054-0).
- Chen, J.M., 1996. Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agric. Forest Meteorol.* 80, 135–163. [https://doi.org/10.1016/0168-1923\(95\)02291-0](https://doi.org/10.1016/0168-1923(95)02291-0).
- Cohen, W.B., Goward, S.N., 2004. Landsat's role in ecological applications of remote sensing. *BioScience* 54, 535–545. [https://doi.org/10.1641/0006-3568\(2004\)054\[535:lrleao\]2.0.co;2](https://doi.org/10.1641/0006-3568(2004)054[535:lrleao]2.0.co;2). <https://academic.oup.com/bioscience/article/54/6/535/294251> [Landsat-s Role-in-Ecological-Applications-of].
- Cohen, W.B., Maiersperger, T.K., Gower, S.T., Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sens. Environ.* 84, 561–571.
- Cook, G.D., Goyens, C.M.A.C., 2008. The impact of wind on trees in Australian tropical savannas: lessons from Cyclone Monica. *Austral Ecol.* 33, 462–470. <https://doi.org/10.1111/j.1442-9993.2008.01901.x>.
- Cook, G.D., Liedloff, A.C., Cuff, N.J., Brocklehurst, P.S., Williams, R.J., 2015. Stocks and dynamics of carbon in trees across a rainfall gradient in a tropical savanna. *Austral Ecol.* 40, 845–856. <https://doi.org/10.1111/aec.12262>.
- Coops, N., Delahaye, A., Pook, E., 1997. Estimation of eucalypt forest leaf area index on the South Coast of New South Wales using Landsat MSS data. *Aust. J. Bot.* 45, 757. <https://doi.org/10.1071/BT96021>. <http://www.publish.csiro.au/?paper=BT96021>.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792. <https://doi.org/10.1890/07-0539.1>. <http://www.ncbi.nlm.nih.gov/pubmed/18051647>.
- Danaher, T., Armston, J., Collett, L., 2004. A regression model approach for mapping woody foliage projective cover using Landsat Imagery in Queensland, Australia. In: International Geoscience and Remote Sensing Symposium (IGARSS). Anchorage. pp. 523–527.
- Edwards, A.C., Maier, S.W., Hutley, L.B., Williams, R.J., Russell-Smith, J., 2013. Remote sensing of environment spectral analysis of fire severity in north Australian tropical savannas. *Remote Sens. Environ.* 136, 56–65. <https://doi.org/10.1016/j.rse.2013.04.013>.
- Ekstrand, S., 1996. Landsat TM-based forest damage assessment: correction for topographic effects. *Photogramm. Eng. Remote Sens.* 62, 151–161.
- Eriksson, H.M., Eklundh, L., Kuusk, A., Nilson, T., 2006. Impact of understory vegetation on forest canopy reflectance and remotely sensed LAI estimates. *Remote Sens. Environ.* 103, 408–418. <https://doi.org/10.1016/j.rse.2006.04.005>.
- Ferwerda, J.G., Ketner, P., McGuinness, K.A., 2007. Differences in regeneration between hurricane damaged and clear-cut mangrove stands 25 years after clearing. *Hydrobiologia*. <https://doi.org/10.1007/s10750-007-0782-7>.
- Flood, N., 2014. Continuity of reflectance data between Landsat-7 ETM+ and Landsat-8 OLI, for both top-of-atmosphere and surface reflectance: a study in the Australian landscape. *Remote Sens.* 6, 7952–7970. <https://doi.org/10.3390/rs6097952>.
- Flood, N., Danaher, T., Gill, T., Gillingham, S., 2013. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM+ and SPOT HRG imagery for Eastern Australia. *Remote Sens.* 5, 83–109. <https://doi.org/10.3390/rs5010083>. <http://www.mdpi.com/2072-4292/5/1/83>.
- Foody, G.M., Boyd, D.S., Cutler, M.E., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* 85, 463–474. [https://doi.org/10.1016/S0034-4257\(03\)00039-7](https://doi.org/10.1016/S0034-4257(03)00039-7).
- Fox, E.R., 1980. Deciduous vine thickets of the Darwin area and effects of cyclone 'Tracy' 25 December 1974. Conservation Commission of the Northern Territory, Darwin, NT Technical Report Parks and Wildlife, Technical Bulletin Number Two.
- Gao, T., Zhu, J., Zheng, X., Shang, G., Huang, L., Wu, S., 2015. Mapping spatial distribution of larch plantations from multi-seasonal landsat-8 OLI imagery and multi-scale textures using random forests. *Remote Sens.* 7, 1702–1720. <https://doi.org/10.3390/rs70201702>.
- Gill, A.M., Ryan, P.G., Moore, P.H.R., Gibson, M., 2000. Fire regimes of World Heritage Kakadu National Park, Australia. *Austral Ecol.* 25, 616–625. <https://doi.org/10.1111/j.1442-9993.2000.tb0067.x>.
- Gill, T., Johansen, K., Phinn, S., Trevithick, R., Scarth, P., Armston, J., 2017. A method for mapping Australian woody vegetation cover by linking continental-scale field data and long-term Landsat time series. *Int. J. Remote Sens.* 38, 679–705. <https://doi.org/10.1080/01431161.2016.1266112>.
- Goldbergs, G., Levick, S.R., Lawes, M., Edwards, A., 2018. Hierarchical integration of individual tree and area-based approaches for savanna biomass uncertainty estimation from airborne LiDAR. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2017.11.010>.
- Goodwin, N.R., Collett, L.J., 2014. Development of an automated method for mapping fire history captured in Landsat TM and ETM+ time series across Queensland, Australia. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2014.03.021>.
- Grömping, U., 2009. Variable importance assessment in regression: linear regression versus random forest. *Am. Stat.* 63, 308–319. <https://doi.org/10.1198/tast.2009.08199>.
- Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. *Remote Sens. Environ.* 185, 221–232. <https://doi.org/10.1016/J.RSE.2016.02.023>. <http://www.sciencedirect.com/science/article/pii/S0034429716300530>.
- Hill, R., Boyd, D.S., Hopkinson, C., 2011. Relationship between canopy height and Landsat ETM+ response in lowland Amazonian rainforest. *Remote Sens. Lett.* 2, 203–212. <https://doi.org/10.1080/01431161.2010.510810>. <http://eprints.bournemouth.ac.uk/17133/1/licence.txt>.
- Hudak, A.T., Lefsky, M.A., Cohen, W.B., Berterretche, M., 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sens. Environ.* 82, 397–416. [https://doi.org/10.1016/S0034-4257\(02\)00056-1](https://doi.org/10.1016/S0034-4257(02)00056-1).
- Huet, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309.
- Hunt, E.R., Rock, B.N., 1989. Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sens. Environ.* 54, 43–54.
- Hutley, L.B., Evans, B.J., Beringer, J., Cook, G.D., Maier, S.W., Razon, E., 2013. Impacts of an extreme cyclone event on landscape-scale savanna fire, productivity and greenhouse gas emissions. *Environ. Res. Lett.* 8, 045023. <https://doi.org/10.1088/1748-9326/8/4/045023>. <http://stacks.iop.org/1748-9326/8/i=4/a=045023?key=crossref7c62700d9c92699b3563a9bdc32f8f>.
- Ji, L., Zhang, L., Wyllie, B.K., Rover, J., 2011. On the terminology of the spectral vegetation index (NIRSWIR)/(NIR + SWIR). *Int. J. Remote Sens.* 32, 6901–6909. <https://doi.org/10.1080/01431161.2010.510811>.
- Joyce, K.E., Belliss, S.E., Samsonov, S.V., McNeill, S.J., Glassey, P.J., 2009. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Prog. Phys. Geogr.* 33, 183–207. <https://doi.org/10.1177/030913309339563>.
- Joyce, K.E., Phinn, S.R., Roelfsema, C.M., Neil, D.T., Dennison, W.C., 2004. Combining Landsat ETM+ and Reef Check classifications for mapping coral reefs: a critical assessment from the southern Great Barrier Reef, Australia. *Coral Reefs* 23, 21–25. <https://doi.org/10.1007/s00338-003-0357-7>.
- Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., Mattsson, E., 2015. Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian Woodlands using Landsat 8 and random forest. *Remote Sens.* 7, 10017–10041. <https://doi.org/10.3390/rs70810017>. <http://www.mdpi.com/2072-4292/7/8/10017.htm>.
- Khosrovipour, A., Skidmore, A.K., Isenburg, M., Wang, T., Hussin, Y.A., 2014. Generating pit-free canopy height models from airborne lidar. *Photogramm. Eng. Remote Sens.* 80, 863–872. <https://doi.org/10.14358/PERS.80.9.863>.
- Labrecque, S., Fournier, R.A., Luther, J.E., Piercy, D., 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. *Forest Ecol. Manage.* 226, 129–144. <https://doi.org/10.1016/j.foreco.2006.01.030>.
- LAStools, 2017. LAStools, Efficient LiDAR Processing Software (version 170822, academic), obtained from <http://rapidlasso.com/LAStools>.
- Li, H., Mausel, P., Brondizio, E., Deardorff, D., 2010. A framework for creating and validating a non-linear spectrum-biomass model to estimate the secondary succession biomass in moist tropical forests. *ISPRS J. Photogramm. Remote Sens.* 65, 241–254. <https://doi.org/10.1016/j.isprsjprs.2010.01.002>.
- Lim, K., Treitz, P., Wulder, M.A., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. *Prog. Phys. Geogr.* 27, 88–106. <https://doi.org/10.1191/030913303pp360ra>.
- Mascaro, J., Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Martin, R.E., Anderson, C., Higgins, M., Chadwick, K.D., 2014. A tale of two "Forests": Random Forest machine learning aids tropical forest carbon mapping. *PLoS ONE* 9, 12–16. <https://doi.org/10.1371/journal.pone.0085993>.
- Mellor, A., Boukir, S., Haywood, A., Jones, S., 2015. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS J. Photogramm. Remote Sens.* 105, <https://doi.org/10.1016/j.isprsjprs.2015.03.014>.
- Mellor, A., Haywood, A., Stone, C., Jones, S., 2013. The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sens.* 5, 2838–2856. <https://doi.org/10.3390/rs5062838>. <http://www.mdpi.com/2072-4292/5/6/2838/>.
- Moulin, S., Bondeau, A., Delecolle, R., 1998. Combining agricultural crop models and satellite observations: from field to regional scales. *Int. J. Remote Sens.* 19, 1021–1036. <https://doi.org/10.1080/014311698215586>.
- Müller, A.C., Guido, S., 2016. Introduction to Machine Learning with Python: A Guide for Data Scientists, 1st ed. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, United States of America.
- Murphy, K., 1984. Big Blow up North: A History of Tropical Cyclones in Australia's Northern Territory. University Planning Authority, Darwin.
- Myers, B.A., Duff, G.A., Eamus, D., Fordyce, I.R., O'Grady, A., Williams, R.J., 1997. Seasonal variation in water relations of trees of differing leaf phenology in a wet-dry tropical savanna near Darwin, Northern Australia. *Aust. J. Bot.* 45, 225. <https://doi.org/10.1071/BT96015>.
- O'Grady, A.P., Eamus, D., Hutley, L.B., 1999. Transpiration increases during the dry season: patterns of tree water use in eucalypt open-forests of northern Australia. *Tree Physiol.* 19, 591–597. <https://doi.org/10.1093/treephys/19.4.591>.
- Oshiro, T.M., Perez, P.S., Baranaukas, J.A., 2012. How many trees in a random forest? *LNCS* 7376, 154–168. https://doi.org/10.1007/978-3-642-31537-4_13. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84864950596&partnerID=tZotx3y1>.
- Ota, T., Ahmed, O., Franklin, S., Wulder, M., Kajisa, T., Mizoue, N., Yoshida, S., Takao, G., Hirata, Y., Furuya, N., Sano, T., Heng, S., Vuthy, M., 2014. Estimation of airborne lidar-derived tropical forest canopy height using Landsat time series in Cambodia. *Remote Sens.* 6, 10750–10772. <https://doi.org/10.3390/rs61110750>. <http://www.mdpi.com/2072-4292/6/11/10750/>.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26, 217–222. <https://doi.org/10.1080/01431160412331269698>.
- Paling, E., Kobryn, H., Humphreys, G., 2008. Assessing the extent of mangrove change caused by Cyclone Vance in the eastern Exmouth Gulf, northwestern Australia. *Estuar. Coast. Shelf Sci.* 77, 603–613. <https://doi.org/10.1016/j.ecss.2007.10.019>.

- <http://linkinghub.elsevier.com/retrieve/pii/S0272771407004866>.
- Pascual, C., Garcia-Abril, A., Cohen, W.B., Martin-Fernandez, S., 2010. Relationship between LiDAR-derived forest canopy height and Landsat images. *Int. J. Remote Sens.* 31, 1261–1280. <https://doi.org/10.1080/01431160903380656>.
- Pedregosa, F., Grisel, O., Weiss, R., Passos, A., Brucher, M., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. <https://doi.org/10.1007/s13398-014-0173-7.2>.
- Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: opportunities and challenges. *J. Appl. Ecol.* 51, 839–848. <https://doi.org/10.1111/1365-2664.12261>.
- Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. *Remote Sens. Environ.* 114, 1053–1068. <https://doi.org/10.1016/j.rse.2009.12.018>. <http://linkinghub.elsevier.com/retrieve/pii/S0034425709003745>.
- Preston, R.A., 1987. Use of Landsat 5 MSS imagery to map cyclone damage to rainforest in North Queensland. In: 4th Australasian Remote Sensing Conference. Adelaide.
- Qi, J., Chehbouni, A., Huete, A., Kerr, Y., Sorooshian, S., 1994. A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48, 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1).
- Renó, V.F., Novo, E.M., Suemitsu, C., Rennó, C.D., Silva, T.S., 2011. Assessment of deforestation in the Lower Amazon floodplain using historical Landsat MSS/TM imagery. *Remote Sens. Environ.* 115, 3446–3456. <https://doi.org/10.1016/j.rse.2011.08.008>. <http://linkinghub.elsevier.com/retrieve/pii/S0034425711002902>.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-olmo, M., Rigol-sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for landcover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
- Rogers, K., Lymburner, L., Salum, R., Brooke, B.P., Woodroffe, C.D., 2017. Mapping of mangrove extent and zonation using high and low tide composites of Landsat data. *Hydrobiologia* 1–20. <https://doi.org/10.1007/s10750-017-3257-5>.
- Roughgarden, J., Running, S.W., Matson, P.A., 1991. What does remote sensing do for ecology? *Ecology* 72, 1918–1922. <https://doi.org/10.2307/1941546>.
- Roy, D., Zhang, H., Ju, J., Gomez-Dans, J., Lewis, P., Schaaf, C., Sun, Q., Li, J., Huang, H., Kovalsky, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* 176, 255–271. <https://doi.org/10.1016/j.rse.2016.01.023>.
- Roy, D.P., Ju, J., Mbow, C., Frost, P., Loveland, T., 2010. Accessing free Landsat data via the Internet: Africa's challenge. *Remote Sens. Lett.* 1, 111–117. <https://doi.org/10.1080/01431160903486693>.
- Scarth, P., Roder, A., Schmidt, M., 2010. Tracking grazing pressure and climate interaction? The role of Landsat fractional cover in time series analysis. In: 15th Australasian Remote Sensing and Photogrammetry Conference. 13th–17th September 2010.
- Setterfield, S.A., Rossiter-Rachor, N.A., Hutley, L.B., Douglas, M.M., Williams, R.J., 2010. Turning up the heat: the impacts of *Andropogon gayanus* (gamba grass) invasion on fire behaviour in northern Australian savannas. *Divers. Distrib.* <https://doi.org/10.1111/j.1472-4642.2010.00688.x>.
- Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R., Mücher, C.A., O'Connor, B., Paganini, M., Pereira, H.M., Schaepman, M.E., Turner, W., Wang, T., Wegmann, M., 2015. Environmental science: agree on biodiversity metrics to track from space. *Nature* 523, 403–405. <https://doi.org/10.1038/523403a>.
- Sripada, R.P., Heiniger, R.W., White, J.G., Meijer, A.D., 2006. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agron. J.* <https://doi.org/10.2134/agronj2005.0200>.
- Staben, G., Lucieer, A., Evans, K., Scarth, P., Cook, G., 2016. Obtaining biophysical measurements of woody vegetation from high resolution digital aerial photography in tropical and arid environments: Northern Territory, Australia. *Int. J. Appl. Earth Observ. Geoinform.* 52, 204–220. <https://doi.org/10.1016/j.jag.2016.06.011>. <http://linkinghub.elsevier.com/retrieve/pii/S0303243416300939>.
- Staben, G.W., Evans, K.G., 2008. Estimates of tree canopy loss as a result of Cyclone Monica, in the Magela Creek catchment northern Australia. *Austral Ecol.* 33, 562–569. <https://doi.org/10.1111/j.1442-9993.2008.01911.x>.
- Stocker, G.C., 1976. Report on cyclone damage to natural vegetation in the Darwin area after cyclone Tracy 25 December 1974. Technical Report Forestry and Timber Bureau, Leaflet No. 127 Canberra.
- Stojanova, D., Panov, P., Gjorgioski, V., Kobler, A., Džeroski, S., 2010. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecol. Inform.* 5, 256–266. <https://doi.org/10.1016/j.ecoinf.2010.03.004>.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinform.* 9, 307. <https://doi.org/10.1186/1471-2105-9-307>.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetati. *Remote Sens. Environ.* 8, 127–150.
- Vicente-Serrano, S.M., Pérez-Cabello, F., Lasanta, T., 2008. Assessment of radiometric correction techniques in analyzing vegetation variability and change using time series of Landsat images. *Remote Sens. Environ.* 112, 3916–3934. <https://doi.org/10.1016/j.rse.2008.06.011>.
- Vincini, M., Frazzi, E., D'Alessio, P., 2008. A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precis. Agric.* 9, 303–319. <https://doi.org/10.1007/s11119-008-9075-z>.
- Wallace, J.F., Caccetta, P.A., Kiiveri, H.T., 2004. Recent developments in analysis of spatial and temporal data for landscape qualities and monitoring. *Austral Ecol.* 29, 100–107. <https://doi.org/10.1111/j.1442-9993.2004.01356.x>.
- Wang, K., Franklin, S.E., Guo, X., Cattet, M., 2010. Remote sensing of ecology, biodiversity and conservation: a review from the perspective of remote sensing specialists. *Sensors* 10, 9647–9667. <https://doi.org/10.3390/s101109647>. <http://www.mdpi.com/1424-8220/10/11/9647>.
- Wilkes, P., Jones, S.D., Suarez, L., Mellor, A., Woodgate, W., Soto-Berelov, M., Haywood, A., Skidmore, A.K., 2015. Mapping forest canopy height across large areas by up-scaling ALS estimates with freely available satellite data. *Remote Sens.* 7, 12563–12587. <https://doi.org/10.3390/rs70912563>.
- Williams, R.J., Myers, B.A., Muller, W.J., Duff, G.A., Eamus, D., 1997. Leaf phenology of woody species in a north Australian tropical savanna. *Ecology* 78, 2542–2558. [https://doi.org/10.1890/0012-9658\(1997\)078\[2542:LPOWSI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1997)078[2542:LPOWSI]2.0.CO;2).
- Williamson, G.J., Boggs, G.S., Bowman, D.M.J.S., 2011. Late 20th century mangrove encroachment in the coastal Australian monsoon tropics parallels the regional increase in woody biomass. *Regional Environ. Change* 11, 19–27. <https://doi.org/10.1007/s10113-010-0109-5>.
- Wilson, B.A., Bowman, D.M.J.S., 1987. Fire, storm, flood and drought: the vegetation ecology of Howards Peninsula, Northern Territory, Australia. *Aust. J. Ecol.* 12, 165–174.
- Woodcock, C.E., Macomber, S.A., Pax-lenney, M., Cohen, W.B., 2001. Monitoring large areas for forest change using Landsat: generalization across space, time and Landsat sensors. *Remote Sens. Environ.* 78, 194–203.
- Wulder, M.A., White, J.C., Nelson, R.F., Naesset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. Lidar sampling for large-area forest characterization: a review. *Remote Sens. Environ.* 121, 196–209. <https://doi.org/10.1016/j.rse.2012.02.001>.
- Zar, J.H., 1984. Biostatistical Analysis, 2nd ed. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.