The background of the slide is a stylized, layered representation of a forest canopy. It consists of multiple horizontal bands of green, each containing a dense row of small tree silhouettes. The bands are separated by thin, wavy brown lines, suggesting a cross-section or a multi-scale view of the forest structure. The overall effect is a textured, green landscape that fills the entire frame.

ASSESSMENT OF FOREST
CANOPY VERTICAL
STRUCTURE WITH MULTI-
SCALE REMOTE SENSING:
FROM THE PLOT TO THE
LARGE AREA

PHIL WILKES

**Assessment of forest canopy vertical structure
with multi-scale remote sensing: from the plot
to the large area**

Phil Wilkes

Examining committee:

Prof. dr. M. George Vosselman	University of Twente, ITC
Prof. dr. Wouter Verhoef	University of Twente, ITC
Prof. dr. Martin Herold	Wageningen University
Prof. dr. Mark Danson	University of Salford

ITC dissertation number ...

ITC, P.O. Box 217, 7500 AA Enschede, The Netherlands

ISBN 978-90-365-

DOI 10.3990/1.97890365

Cover designed by

Printed by ITC Printing Department

Copyright © 2015 by Phil Wilkes



**Assessment of forest canopy vertical structure
with multi-scale remote sensing: from the plot
to the large area**

DISSERTATION

to obtain the Double-Badged Degree of Doctor
at the University of Twente, Enschede, The Netherlands; and
RMIT University, Melbourne, Australia,

on the authority of the rector magnificus, prof.dr. H. Brinksma, for
University of Twente

on the authority of the Chancellor, Dr Ziggy Switkowski AO; and
the Vice-Chancellor and President of RMIT University, Mr Martin Bean CBE,
for RMIT University

on account of the decision of the Joint Graduation Committee,
to be publicly defended on 13th day of January 2016 at 16:45 hrs

by

Phil Wilkes

born on 30th October, 1981
in Truro, Cornwall, UK

This thesis is approved by

Prof. dr. Simon D. Jones, promoter

Prof. dr. Andrew K. Skidmore, promoter

Dr. Lola Suarez, co-promoter

Dr. Andrew Haywood, co-promoter

'A mouse took a stroll through the deep dark wood...'

Julia Donaldson, *The Gruffalo*

Acknowledgements

Firstly I would like to give my sincere thanks to my supervisory panel, Prof. Simon Jones and Dr. Lola Suarez from RMIT, Prof. Andrew Skidmore from the University of Twente and Dr. Andrew Haywood from the Victorian Department of Environment, Land, Water and Planning (DELWP). Simon, thank you for accepting me onto this PhD three and a half years ago and for your on-going guidance since then, our regular meetings have given focus to my research but have often covered subjects as diverse as our favourite ska bands to global politics. Lola, thank you for dedicating so much time to me (I'm sure you will miss the constant 'come and look at this!') and providing invaluable advice that has enabled me to get this far. Andrew (Skidmore), thank you for accepting me on to the double-badged programme and making me feel welcome in the Netherlands. Andrew (Haywood) thank you for providing the all-important industry perspective at the genesis of the project. I would like to thank Andrew Mellor from DELWP for guidance with random forest and acting as a sounding board for ideas. I would also like to thank Dr. Peter Woodgate, Dr. Graeme Kernich and the rest of the team at the Cooperative Research Centre for Spatial Information (CRC-SI). From the Commonwealth Scientific and Industrial Research Organisation (CSIRO) I would like to thank Dr. Eva van Gorsel and Dr. Jenny Lovell. I would also like to thank Esther Hondebrink for translating the Summary chapter into Dutch.

I would like to thank the CRC-SI, DELWP and other project partners for providing the funding that made the *Australian Woody Vegetation Landscape Feature Generation from Multi-Source Airborne and Space-Borne Imaging and Ranging Data* project possible, the School of Graduate Research for awarding me an Australian Postgraduate Award, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for awarding me an Office of the Chief Executive postgraduate scholarship and the CRC-SI for awarding me a top-up scholarship. I would like to acknowledge the TERN/AusCover facility, DELWP (Andrew Mellor, Andrew Clark and Rick Frasina) and the Joint Remote Sensing Research Program (Neil Flood) for providing access to data that made this research possible.

I would like to thank my fellow PhD students, post docs, research staff and field crews. From RMIT I would like to thank Mariela Soto-Bereylo, Will

Woodgate, Vaibhav Gupta, Rob Hewson, Bill Langford, Tapasya Arya and Silvana Acevedo. From the University of Twente I would like to thank Maria Fernanda, Xi Zhu, Elnaz Neinavaz, Festus Ihwagi, Anahita Khosravipour, Zihui Wang and all the other students who made me feel so welcome at ITC.

Lastly I would like to thank my family; Mum (Sara), Dad (Glenn), Derek, Ruth, Robbo, Emma, Jack and Zelig, as well as Anne and Brian FitzGerald. Although you are all so far away, you have offered me much support and encouragement throughout my candidature. Most importantly I would like to thank my wife, Laura, and my two kids, Iris and Jonah; the reasons I get out of bed in the morning! You have shown me nothing but the utmost love and support over the past 3 and a half years, you've have had strength, resilience and humour during the months I have been away and provided many hours of much needed distraction when I needed it most.

Table of Contents

Acknowledgements	i
Table of Contents	iii
List of figures	vi
List of tables	xiii

Chapter 1: Introduction	1
1.1. General introduction.....	2
1.2. Problem statement.....	7
1.3. Research questions.....	8
1.4. Thesis structure	9

Chapter 2: Metrics of canopy vertical structure suitable for large area forest attribution	10
2.1. Introduction.....	11
2.1.1. Canopy height	12
2.1.2. Canopy vertical structure.....	14
2.1.3. Aims and objectives.....	16
2.2. Materials and methods.....	16
2.2.1. Study area.....	16
2.2.2. Forest inventory data.....	17
2.2.3. Airborne laser scanning data.....	20
2.3. Data processing.....	21
2.3.1. Canopy height	21
2.3.2. Canopy vertical structure.....	22
2.4. Results.....	24
2.4.1. Canopy height	24
2.4.2. Canopy height profiles.....	27
2.5. Discussion.....	30
2.6. Conclusion	33

Chapter 3: Using discrete-return ALS to quantify number of canopy strata across diverse forest types	35
3.1. Introduction.....	36
3.2. Attributing canopy vertical structure	38
A. Gap probability derivation.....	38

B.	Probability Density Function (PDF) from projected cover (L).....	38
C.	Bootstrapping and P_{gap} filtering.....	40
3.3.	Application across a diverse forested landscape	41
3.3.1.	ALS acquisition and pre-processing	41
3.3.2.	P_{gap} from ALS	43
3.3.3.	Derivation of smoothing coefficient (α)	43
3.3.4.	Bootstrapping simulated point clouds.....	44
3.3.5.	Validation with field inventory.....	46
3.4.	Results and Discussion.....	47
3.4.1.	Methodology evaluation.....	47
3.4.2.	Validation results.....	54
3.4.3.	Canopy vertical structure as an independent metric	56
3.5.	Conclusion	58

Chapter 4: Understanding the effects of ALS pulse density for metric retrieval across diverse forest types 60

4.1.	Introduction.....	61
4.2.	Method.....	63
4.2.1.	Study area and data capture	63
4.2.2.	Data processing	66
4.2.3.	Metrics.....	70
4.3.	Results.....	70
4.3.1.	Canopy height	71
4.3.2.	Canopy cover	75
4.3.3.	Canopy vertical structure.....	75
4.3.4.	Characteristics of thinned point clouds.....	76
4.4.	Discussion.....	77
4.5.	Conclusion	80

Chapter 5: Mapping forest canopy height across large areas by upscaling ALS estimates with freely available satellite data 82

5.1.	Introduction.....	83
5.2.	Materials and methods.....	86
5.2.1.	Study area	86
5.2.2.	Data collection	88
5.2.2.1.	Forest inventory.....	88
5.2.2.2.	Airborne laser scanning data.....	88

5.2.2.3.	Satellite imagery and ancillary data.....	90
5.2.3.	Canopy height estimation with random forest	92
5.2.4.	Random forest implementation.....	93
5.2.5.	Selecting predictor variables	94
5.2.6.	Systematic error in model output.....	95
5.3.	Results.....	96
5.3.1.	Canopy height estimation	96
5.3.2.	Validation with inventory data	99
5.3.3.	Training and validation of random forest using smaller geographic areas	103
5.3.4.	Simulating disparate ALS capture for training a random forest.	103
5.4.	Discussion.....	104
5.5.	Conclusions	108
Chapter 6: Summary and synthesis		110
6.1.	Summary of results.....	111
6.2.	Identifying trends in large-area forest structure	117
6.3.	Remote sensing in sustainable forest management: a future perspective.....	120
7.	References	125
8.	Summary	156
9.	Samenvatting	159
10.	ITC Dissertation List	162

List of figures

- Figure 1. ALS derived point cloud captured over a forest stand. Points are classified into either ground or non-ground, where ground points are coloured by altitude and non-ground points are coloured by low (<2 m) or high vegetation (≥ 2 m). The gap running through the middle of the plot is a road. See Table 3 for acquisition specifications7
- Figure 2. Representative forest structure at the three study areas, derived using ALS point cloud data. For each study area, three panes illustrate: a 1500 m x 30 m transect highlighting topography and structural heterogeneity [top], a downward looking view showing canopy cover (green is canopy and brown is ground) [left] and a side view showing canopy vertical structure [right]. The location of the downward and side ward panes are displayed in the transect plot. Refer to key pane for pane dimensions..... 18
- Figure 3. Boxplots of residual error when using ALS metrics to estimate inventory (INV) measured canopy height. Within each pane, the four groups represent all plots and plots from the *Low Open Woodland*, *Mixed Species Forest* and *Tall Closed canopy Forest* (left to right). Within each group, different point cloud components are displayed.....25
- Figure 4. Scatter plots comparing different techniques for inventory measured and ALS derived canopy height at 27 plots. Also displayed are the metrics derived using three different point cloud components and linear regression using ‘all’ returns (-) including 95th percentile confidence intervals (--). RMSE values are calculated with the outliers (95th percentile of residuals) removed.....26
- Figure 5. (previous page) Canopy height profiles characterised using mixture models where k was iterated from 1 - 3. Values are $wAIC$ calculated using EQ. 9. Function abbreviations are: *GG* generalised gamma, *BCT* Box-Cox- t , *GU* Gumbel, *WEI* Weibull, *GT* generalised y , *WEI3* Weibull (μ the mean), *LO* logistic, *GIG* generalised inverse Gaussian, *TF* t Family, *RG* reverse Gumbel, *IG* inverse Gaussian, *GA* gamma, *LGN* log normal, *ST4* skew t type 4, *ZAIG* zero adjusted inverse

Gaussian and <i>NOF</i> normal family. The order of functions listed for each plot is not the order they were necessarily applied.....	29
Figure 6. Mixture models fitted to 25 m x 25 m ALS plots along a 200 m x 25 m transect (dashed lines identifies plot boundaries). The three forest types represented are; <i>LOW</i> [top], <i>MSF</i> [middle] and <i>TCF</i> [bottom], where k is assumed to be 1, 2 and 3 respectively. ALS data (grey points) has been ground normalised.	30
Figure 7. Derivation of canopy vertical structure from ALS derived gap probability (P_{gap}).....	39
Figure 8. Sensitivity of NoS to increasing α [A] and “apparent” Canopy Height Profiles (CHP) derived by increasing α [B – E]. Also included is the canopy height profile histogram (2 m bins) used in the supervised classification [F] (see Section 3.3.3). Results from a supervised classification suggested this plot is characterised by two canopy strata (equivalent to pane D) and therefore an α_s of 0.04 – 0.38 [A]......	42
Figure 9. Cross validation to determine a suboptimal smoothing coefficient. For this 75% of plots were subset as training data [A], where the horizontal bars represent α_s as determined by a supervised classification (NoSSC). A suboptimal coefficient $\tilde{\alpha}$ was selected as the median value from the range of most frequently observed α values [B]. $\tilde{\alpha}$ was subsequently substituted into EQ. 11 and applied to the test plots [C].	45
Figure 10. Canopy vertical structure from ALS derived P_{gap} for four structurally diverse forest plots [A – D] using forest type (green), study area (purple) and universal (blue) α . P_{gap} from a single bootstrap sample is transformed with a nonparametric cubic spline regression (P_s) [1]. “Apparent” canopy height profile (Lovell et al., 2003) from $f(x)'$, where $x = 1 - P_s$ [2]. ALS return density maxima identified as strata (S_i) from the positive zero-crossings of $f(x)''$ [3]. A comparison of $f(x)'$ and a histogram of weighted ALS return ($wALS$) frequency [2] illustrates the generalisation of the canopy height profile possible with this technique.....	48

Figure 11. <i>NoS</i> and canopy height computed across three diversely forested study areas (Table 2) at a 30 m resolution. Map coordinates are Map Grid of Australia Zone 55.....	49
Figure 12. Results from the supervised classification of 82 plots. Frequency histogram of mean classification response (<i>NoS_{sc}</i>) classified by mode response (i.e. consensus on number of strata) [A] and confusion matrix comparing supervised classification and model output [B].....	51
Figure 13. Linear regression of per plot mean height and mean number of spline polynomials. Mean values were derived from the bootstrap simulations.....	52
Figure 14. A comparison of canopy height profiles derived using a point cloud with a pulse density of 10 pl m ⁻² (black) and a bootstrap (<i>N</i> =50) of a point cloud thinned to 1 pl m ⁻² (see Chapter 4). <i>NoS</i> equals 3 for the original point cloud, 4 for the thinned point cloud and 3.12 for the bootstrap mean of the thinned dataset. Also included as a comparison is a histogram of weighted ALS returns (<i>wALS</i>).....	53
Figure 15. A comparison of 24 canopy height profiles (CHP) generated with GCVM and ALS. Light green and dark green crowns represent eucalypt and non-eucalypt species respectively. Trees are randomly placed within the scene as individual stem location was not recorded. Grey histograms were generated using GCVM and green and blue CHPs were derived from ALS plots with a radius of 11.8 m and 15 m respectively.....	55
Figure 16. Boxplots summarising the statistical association between <i>NoS</i> (rounded to integer values) and canopy height [A], canopy cover [B] and return height coefficient of variation [C]. Boxplot whiskers represent the 95 th percentile.....	57
Figure 17. Comparison of coefficient of variation of return height (<i>C_v</i>) and <i>NoS</i> . Cumulative histogram of <i>C_v</i> where <i>NoS</i> are classified into integers for 239 plots [A]. Canopy height profiles for plots with increasing <i>NoS</i> and similar <i>C_v</i> values [B – E] where canopy height profiles are represented by weighted ALS returns (<i>wALS</i>) and pseudo-waveform generated from $f(x)'$ where $x = 1 - P_s$	58

- Figure 18. Map displaying the location of the six TERN sites used in this investigation, shaded areas indicate forest extent (Montreal Process Implementation Group for Australia, 2013). 63
- Figure 19. Descriptive statistics for 4 metrics of canopy structure across 6 study areas. 95th percentile of canopy height [A]; proportion of cover calculated as $1 - P_{\text{gap}}(z)$ where z equals 1 m [B]; coefficient of variation (C_v) for non-ground return height[C] and NoS [D]. 65
- Figure 20. Visualisation describing the point cloud thinning technique using a target pulse density of 0.5 pl m^{-2} as an example. [A] A single first return is selected for each grid point (+) where the ALS return with the lowest Cartesian distance is selected. Selection is restricted to a search-window around each grid point where the search window dimensions are determined by the desired pulse density (d). Selection of X further returns, determined by the *number of returns* metadata field (in this case $X = 4$), are again chosen by their proximity to the grid point. A search-window restricts the maximum distance of “other” returns where the extent is determined by an estimate of the maximum height of the forest ($z_{\text{max}} = 40 \text{ m}$) and an assumed scan angle $\leq 5^\circ$, in this way a nadir acquisition is simulated. NB. Points have been removed from outside the plot boundary to enhance visualisation [B], when point clouds were thinned points from outside the plot boundary could be selected..... 69
- Figure 21. Difference between the plotwise mean value calculated at 6 pulse densities ($0.05 - 4 \text{ pl m}^{-2}$) and the value calculated at 10 pl m^{-2} . Four metrics are used to summarise canopy structure; [A] canopy height, [B] canopy cover, [C] coefficient of variation (C_v) of return height and [D] NoS . Additionally difference in ground height is shown [E]. Error bars represent the 95th percentile range. 72
- Figure 22. Coefficient of determination where the independent variable was [A] canopy height, [B] canopy cover, [C] return height coefficient of variation (C_v) and [D] NoS at 6 different pulse densities ($0.05 - 4 \text{ pl m}^{-2}$). The dependent variable was the metric derived from a pulse density of 10 pl m^{-2} 73
- Figure 23. Mean intra-plot variance for vegetation structure metrics where variance is calculated as the standard deviation of 9 realisations

drawn systematically from the original dataset. Four metrics were computed; [A] canopy height, [C] canopy cover, [D] coefficient of variation (C_v) of return height and [E] *NoS*. Additionally, [B] C_v of canopy height was included, which normalises for canopy height when calculating variance. Error bars represent standard deviation of mean variance. 74

Figure 24. Example point clouds for each forest type displayed at 4 different pulse densities. Points are classified into either ground (black) or non-ground (grey) i.e. vegetation. Heights are relative to the Australian Height Datum. Point size is shown to increase with decreasing density for visualisation purposes only..... 77

Figure 25. Study area in east Victoria, Australia. A mosaic of 5 Landsat TM false colour composite images covering the study area (outlined in white) and location of the study area within Australia (inset) [A]. Canopy height derived from ALS capture where canopy height values are aggregated into 10 x 10 km cells (grey indicates no data) [B]. The extent of the ALS capture [C]. Forest extent (Mellor et al., 2012) and location of Victorian Forest Monitoring Programme forest inventory plots (VFMP) [D]. Map coordinate system is the projected Map Grid of Australia (MGA) Zone 55..... 87

Figure 26. Relative importance of the 18 variables selected for the final random forest model. Confidence intervals (95th percentile) for variable importance were calculated in a bootstrap ($N=50$). Increase mean square error (MSE) is the mean of the squared prediction error when the variable is permuted for a random variable (Cutler et al., 2007). Numbers in brackets indicate the kernel size..... 95

Figure 27. ALS derived canopy height (H_{95}) compared to model residual error for random forest models (H_{RF}); constructed using 5,000 ALS plots of untransformed response data [A], 5,000 ALS plots where the response variable was resampled to a uniform distribution [B], and subtraction of systematic error from modelled canopy height [C]. The coefficient of determination values (r^2) were calculated from a linear regression of measured canopy height and model residuals..... 97

Figure 28. A comparison of the response variable distribution [A], range of Tasseled Cap wetness values (3 x 3 pixels) [B] and the distributions

of modelled canopy height (random forest [C] and random forest - systematic error [D]) for different height classes. The solid arrows indicate the direction in which the random forest model output was “squeezed” by inequity in response variable distribution; the dashed arrow indicates the direction canopy height values were rescaled after correcting for systematic error..... 98

Figure 29. Canopy height at a 30 m resolution (clipped to forest extent) generated using random forest - systematic error [A]. Model output when compared to ALS derived canopy height (10 km x 10 km resolution) where error is represented as height difference [B] and percentage of height [C]. Coordinate system is the projected Map Grid of Australia (MGA) Zone 55..... 100

Figure 30. A map of canopy height (30 m resolution) for an area of 16.5 km x 16.5 km, highlighting the land use history of an area of mixed use forest [left]. Three logging coupes that were clear-felled between 2002 and 2009 [A - C] and an area that has never been logged [D] are identified (coupe extents outlined in black) [centre column]. Transects of canopy height across each coupe indicating regrowth since logging, coupe boundaries are also identified [right]. Coordinate system is the projected Map Grid of Australia (MGA) zone 55 101

Figure 31. A comparison of inventory measured canopy height (H_{inv}) with random forest (H_{RF}) [A] and random forest corrected for systematic error (H_{RF-SE}) [B] estimated canopy height, at 108 forest inventory plots. 102

Figure 32. A comparison of inventory (H_{inv}) measured canopy height and ALS (H_{ALS}) and random forest - systematic error (H_{RF-SE}) estimated canopy height, for 22 plots from within the ALS capture area. Vertical dotted lines link the same plot estimated with ALS or random forest..... 102

Figure 33. Error in canopy height estimates when constructing random forest models from ALS data selected to represent a combination of a number of disparate (non-random) acquisitions. Model output was validated with ALS plots from outside the training area. For comparison (see boxplot), the results from bootstrapping ($N=50$) random forest trained with a random stratified (by IBRA bioregion)

sample from across the whole study area (~18% of forested area) is included.	104
Figure 34. Canopy height profiles, generated with parametric mixture models and pseudo-waveforms from the derivative of $1 - P_s$, fitted to 25 m x 25 m ALS plots along a 200 m x 25 m transect (dashed lines identify plot boundaries). <i>NoS</i> values have been used as an <i>a priori</i> estimate of the number of model components. Compare with Figure 6.	114
Figure 35. Landscape forest structure captured with ALS transects for a total area of 2.9 million ha. Green, red and blue markers indicate the height of the dominant, shade tolerant and intermediate strata that were calculated using the pseudo-waveform described in Chapter 3. Also illustrated is the suggested limit of the shade tolerant strata and the presence of 3 distinct forest structure types (labelled A - C).....	118
Figure 36. Classification of forest type according to the three types identified in Figure 35.	119
Figure 37. A conceptual model to illustrate the paradigm shift in large-area forest assessment utilising remote sensing.	121

List of tables

Table 1.	Applications of canopy height and canopy vertical structure (CVS).	12
Table 2.	Description of study areas.....	19
Table 3.	Flight and sensor specifications for the ALS acquisition at the three study areas.....	21
Table 4.	Forest type, study area and universal α . <i>Riparian forest</i> is found in both the <i>MSF</i> and <i>TCF</i> study areas. Too few plots were located in <i>Montane Damp Forest</i> to compute α	50
Table 5.	Comparison of GCVM and ALS derived estimates of <i>NoS</i>	56
Table 6.	Flight and sensor specifications for the ALS acquisitions.....	64
Table 7.	ALS capture and instrument specifications	89
Table 8.	List of predictor variables with original image resolution in brackets (*kernel sizes: 3, 5, 15, 31, 65, 99 pixels).....	91
Table 9	Airborne laser scanning datasets utilised.....	112

1. Chapter 1

Introduction

1.1. General introduction

The correlation between healthy forests and a productive wider ecosystem is irrefutable (Costanza et al., 1997; Millennium Ecosystem Assessment, 2005; Jacob et al., 2014). Forests provide ecosystem services, such as fresh water, food and building materials, as well as less tangible services including regulation of climate and maintenance of biodiversity, that make life on earth tenable. In 2010 approximately 40 million km² or 31% of global land area was forested (FAO, 2010); this compares to a preindustrial forested area of 55 million km² (Ramankutty and Foley, 1999). Pressure from forestry and subsequent conversion to agricultural land has increased deforestation rates significantly in recent years, for example 1.5 million km² of forest cover were lost between 2000 – 2012 (Hansen et al., 2013). Furthermore, as forest degradation is not manifested in the complete removal of the canopy, the impact of degradation on the provision of ecosystem services is even harder to quantify (Foley et al., 2005). This is exemplified by the negative impact on forests associated with a rapidly warming climate, including drought stress and increased susceptibility to pest infestation and wildfires (Semple et al., 2010; Park Williams et al., 2012; Sugden et al., 2015).

Forests can be characterised by three key primary attributes; *composition*, *function* and *structure* (Franklin et al., 1981). *Composition* is the species, both floral and faunal, that comprise the biomass, *function* is the “work” a forest does e.g. carbon fixation through photosynthesis and nutrient cycling through decomposition, and *structure* is the three-dimensional architecture of trees and plants (Franklin and Spies, 1991). These three attributes are highly inter-dependent, this can be exemplified using the tall temperate rainforests of south eastern Australia. Examples of this inter-dependency include; *composition* and *function* where tree species composition changes significantly with succession, in particular during senescence, where the tall *Eucalyptus* canopy (e.g. *E. regnans* and *E. delegatensis*) is replaced with *Northofagus* and *Acacia* species (Lindenmayer et al., 2000); *structure* and *composition* where cavities in very tall trees provide homes for arboreal marsupials (Lindenmayer, 2000); and *structure* and *function* where non-catastrophic events have led to the formation of multi-strata canopies and consequently one of the most biomass dense ecosystems on the planet (Keith et al., 2009). As is discussed in more detail below and with regard to

sustainable forest management, assessment of *structure* is the primary focus for land managers as it is key to achieving management objectives (Franklin and Spies, 1991).

Although no globally accepted definition of forest exists, definitions tend to differentiate forests from other land cover types by their *structure*, for example, as having a minimum geographical area and populated with trees that have potential to reach a minimum height and canopy cover (Sasaki and Putz, 2009). For example, the UN Food and Agriculture Organisation (FAO) define forest as having a minimum area of 0.5 ha and a potential to reach >10% canopy cover and >5 m in height (FAO, 2000). The United Nations (UN) Land Cover Classification System (LCCS) classifies forests from other land cover types based on vegetation height and cover (Di Gregorio and Jansen, 1998), then further subdivides forest type on attributes including existence of dominant and shade tolerant strata and floristic composition.

The pivotal role that forests play in addressing key environmental, social and economic challenges has been recognised in international policy and legislation that promotes sustainable forest management. A key framework for assessing the effectiveness of management practices is the use of Criteria and Indicators, these are designed to provide guidance and assess progress towards sustainable forest management (FAO, 2015). For example, following the 1992 United Nations Conference on Environment and Development (Earth Summit) in Rio de Janeiro, the Montreal Process Working Group (MPWG) was established. In February 1995, the MPWG (consisting of 12 member countries) endorsed the non-legally binding Santiago Declaration that set out seven criteria and sixty-seven indicators for national level reporting on sustainable forest management (Montréal Process Working Group, 2009). Of the indicators outlined in the Santiago Declaration, 28 are defined as 'biological' and therefore require measurement of *structure*, *composition* or *function* (Miles, 2002). This therefore requires national and state level coordination, cooperation and standardisation of forest inventory methodologies. For example in Australia, national and state level reporting of forest condition and sustainable management practices is done on a 5-yearly cycle and reported in the *State of the Forest Report* (Montreal Process Implementation Group for Australia, 2013).

A need for a systematic approach to national forest inventory was first identified in the early twentieth century (Smith, 2002; Tomppo et al., 2008, 2010), where forest inventory was designed to assess merchantable timber stocks. As recognition of the wider role forests play has increased, inventory protocols have expanded to incorporate additional “ecosystem indicators” such as species diversity (including non-commercial species) and soil properties (Trumbore et al., 2015). Assessment is achieved by installing small plots within the forest, typically circular with a radius of 10 – 12 m (Tomppo et al., 2010), where different quantitative (e.g. diameter at breast height, tree height) and qualitative (e.g. species composition) measurements are recorded. In most industrialised countries, forest inventory programmes have grown into systematic National Forest Inventories, where large scale programmes have been established (Tomppo et al., 2010). For example, in the US the Forest Inventory and Analysis programme has a network of 1 plot for every 6,000 acres of forest (Bechtold and Patterson, 2005) and in Sweden and Finland combined over 10,000 plots that are visited annually (Tomppo et al., 2008).

The on-going requirement for knowledge on forests across national scales in a cost effective and timely manner has led to the pursuit of techniques that identify the spatial configuration of forests over large areas, such as remote sensing (McRoberts and Tomppo, 2007). Remote sensing comprises a wide range of measurement techniques, in a forest assessment context these can include; hemispherical photography to measure local canopy structure, *in situ* passive sensor networks to measure spatial variation of solar radiation, airborne hyperspectral or ranging sensors to create high resolution spectral or spatial datasets, or space borne multispectral and radar platforms offering high temporal resolution imaging.

Earth Observation (EO) remote sensing, which are typically passive spectrometers located on spaceborne satellite platforms, have significantly broadened the horizons of possibility for large-area forest assessment (Wulder, 1998). These techniques allow for a synoptic capture of the forest canopy, not only incorporating an inventory plot, but also the plots context within the wider landscape e.g. forest extent and connectivity (Wulder, 1998). Since the launch of the first Landsat mission in 1972 and with successive EO missions (e.g. subsequent Landsat missions, MODIS etc.), the

possibility of large area land cover mapping has been realised. This has led to a profusion of techniques to measure forest attributes including extent (Hansen et al., 2013) and canopy height (Cohen and Spies, 1992).

An inherent limitation of passive remote sensing instruments is the inability to penetrate the upper canopy layer to measure the physical three dimensional structure of the forest canopy (Hudak et al., 2002). Active instruments such as Light Detection and Ranging (LiDAR) sensors are capable of measuring three dimensional structure and have been applied successfully to forest inventory challenges (Lim et al., 2003; Wulder et al., 2012b). There are two main types of LiDAR instrument; continuous-wave and pulse-laser, the latter is most commonly used for operational forest inventory (Lim et al., 2003) and will be the focus of this section. Pulse-laser LiDAR instruments measure the range (R) to a target by recording the time (t) of flight for an emitted laser pulse to be reflected off a target and received back at the sensor:

$$R = c \frac{t}{2}$$

EQ. 1

where c is the speed of light ($\sim 3 \times 10^8 \text{ m s}^{-1}$). LiDAR instruments digitise the backscattered signal either, as the full-waveform of backscattered energy where signal intensity is recorded over discretised time, or as discrete “returns”. Discrete returns are recorded when the intensity of the backscattered pulse is greater than a predefined threshold. Interpretation of backscatter intensity can be useful for distinguishing the properties of the intercepted surface (Korpela, 2008; Morsdorf et al., 2010). However, this is complicated by the effects of multiple backscattering (i.e. the outgoing pulse is reflected off a number of targets), differing reflective properties of multiple targets, distance from a target and proprietary methods for interpretation (Ni-Meister et al., 2001; Armston et al., 2013). If a sensors position and attitude is recorded concurrently with the laser measurement, for example with a Global Position System and additionally an Inertial Measurement Unit (IMU) for an instrument in motion (e.g. on board an aircraft), the Cartesian position of the target can be computed (Baltsavias, 1999). Post-processing of LiDAR data produces a binary file (.las format), which in the case of discrete return data contains the X, Y, Z

location of the target and backscatter return intensity along with other metadata (American Society of Photogrammetry and Remote Sensing, 2008). Therefore, discrete return LiDAR contains a rich information source with which to perform analysis, such as classification of returns into ground or non-ground or feature extraction.

The potential for LiDAR to measure forest structure has long been recognised (Aldred and Bonnor, 1985; Nilsson, 1996). For vegetation measurement and forestry applications, LiDAR instruments have been utilised *in situ* (i.e. static) with a Terrestrial Laser Scanner (TLS), as well as onboard airborne and spaceborne platforms. Within an operational forest assessment context, LiDAR instruments tend to be aircraft mounted, with the instrument scanning across a swath (typically $\pm 30^\circ$ off nadir); a technique known as Airborne Laser Scanning (ALS) (Baltsavias, 1999; Wulder et al., 2012b). The instantaneous laser footprint cross-sectional diameter for ALS sensors is usually < 3 m which is referred to as “small-footprint”. “Large footprint” instruments can have a laser footprint diameter > 70 m, for example the GLAS sensor on board the ICESat satellite (Brenner et al., 2003). Typically, sensors were configured to record discrete returns, for example first-and-last (e.g. from the canopy and ground) or as technology improved, up to 4 returns. More recently, full-waveform small footprint instruments have been developed, which can identify up to 7 returns per outgoing pulse. Sampling frequency is determined by the Pulse Repetition Frequency (PRF), the rate at which laser pulses are emitted from the instrument. This has improved from 2 KHz in early experiments (Næsset, 1997) to upwards of 800 KHz in modern instruments (Hansen et al., 2015). Depending on aircraft height and speed this can result in on-ground pulse densities (i.e. the number of pulses that interact with a unit area of open ground) of > 25 pulses m^{-2} (d’Oliveira et al., 2012). Figure 1 illustrates an example of ALS data captured over a forest.

Airborne laser scanning has been successfully incorporated within National Forest Inventories across the world (Tomppo et al., 2010; Wulder et al., 2012b), from Norway (Gobakken et al., 2012) to New Zealand (Beets et al., 2010). Utilisation of ALS in national forest inventories has been focused on the attribution of canopy height and the subsequent inference of biomass. However, a number of additional forest structure attributes have been

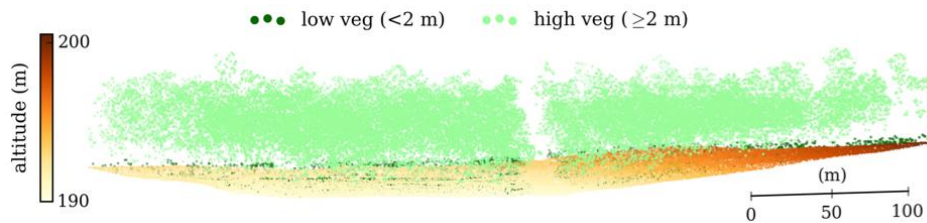


Figure 1. ALS derived point cloud captured over a forest stand. Points are classified into either ground or non-ground, where ground points are coloured by altitude and non-ground points are coloured by low (<2 m) or high vegetation (≥ 2 m). The gap running through the middle of the plot is a road. See Table 3 for acquisition specifications.

derived from ALS, these include canopy cover (Armston et al., 2009; Hopkinson and Chasmer, 2009), canopy vertical structure (Kane et al., 2010b; Jaskierniak et al., 2011) and quantification of coarse woody debris (Miura and Jones, 2012) to highlight but a few.

1.2. Problem statement

Assessment of forests over large, remote and inaccessible areas is an ongoing challenge for land managers (McRoberts et al., 2010; Wulder et al., 2012a, 2012b). There is therefore a requirement for techniques that are applicable across large areas, where traditional inventory is too costly to implement. In this thesis, “large-area” is defined as an area of forest greater than a single forest patch or stand i.e. an area that comprises at least two or more forest types. Technically therefore there is no upper area limit to this definition i.e. large-area can be interpreted as global (Hansen et al., 2013). Applying this definition therefore requires techniques to be transferable between different forest types and not tailored to a particular locality, albeit, they should also be sensitive to local forest conditions and features. This paradox was introduced by Hansen et al. (2013) who state the aim of large area forest assessment is to be globally consistent and yet locally relevant.

Remote sensing has been identified as a key tool in the attribution of forest over large areas (McRoberts and Tomppo, 2007; Wulder et al., 2012a, 2012b), yet there remain unanswered questions regarding the upscaling of

remote sensing measurements from the plot scale to large area attribution. This is particularly relevant for native forests with limited active management where forest structure is not uniform and there is often limited prior knowledge of forest structure.

Owing to the importance of forest *structure*, for example to define forest, as well as the ability to measure *structure* with traditional and remote sensing techniques, and the importance of *structure* to land management and forest scientists (Axelsson et al., 2012), forest structure will be the focus of this thesis. In particular, metrics that are used to estimate canopy height and canopy vertical structure. The importance of measuring canopy height is already well recognised, for example, to estimate biomass volume (Sessa, 2009). Canopy vertical structure, defined as the quantity and arrangement of the canopy along the vertical axis (Brokaw and Lent, 1999), is not currently routinely measured by land management agencies. However, canopy vertical structure has been shown to offer insight into the provision of ecosystem services, such as habitat suitability (Zellweger et al., 2013b) and the hydrological cycle (Jaskierniak et al., 2011).

1.3. Research questions

To address the gaps identified in Section 1.2, three research questions are outlined below:

Question 1. Which metrics of canopy height and canopy vertical structure are suitable for application across forested landscapes?

A set of ALS metrics, that are transferable across and robust to different forest types (Hopkinson et al., 2006), is imperative before up-scaling to the large area can be achieved. This Question aims to test the applicability of existing metrics, and develop new techniques where required, for attributing canopy height and canopy vertical structure.

Question 2. What is an appropriate ALS sampling frequency for attribution of forest structure for different forest types?

Sampling frequency is considered the most important acquisition variable when capturing ALS data for characterisation of vegetation structure (Goodwin et al., 2006). When capturing LiDAR data over a large area, it is likely that different forest types and terrains will be encountered. This

Question aims to test the sensitivity of sampling frequency when attributing vegetation structure and to identify an optimal sampling frequency that is applicable across a range of forest types.

Question 3. How can plot level estimates of canopy structure be upscaled to generate continuous regional maps?

Regional scale wall-to-wall capture of ALS remains prohibitively expensive. An alternative approach is to couple the ability of ALS to measure three dimensional forest structure with the synoptic coverage of passive spaceborne EO sensors. Techniques developed to answer this Question are applied over a forested area where forest structure is highly heterogeneous and assumed unknown *a priori*.

1.4. Thesis structure

Presented in this thesis are four research chapters that aim to answer the questions stated in Section 1.3, this is followed by a synthesis of research outcomes. Chapter 2 tests currently applied techniques to attribute canopy height and canopy vertical structure using inventory and ALS data captured at three structurally diverse study areas. Chapter 3 presents a new technique for attributing canopy vertical structure using ALS that is suitable for application across diverse forested landscapes. Chapter 4 focuses on Question 2 where sampling frequency is investigated across a wide range of forest types. Chapter 5 presents a method to upscale ALS derived plot estimates of canopy height to an area of ~3 million ha using spaceborne passive imagery as predictor data. Finally the thesis concludes with a synthesis of the research presented and a look to the future of remote sensing for large area forest attribution.

2. Chapter 2

Metrics of canopy vertical structure suitable for large area forest attribution*

* This chapter is based on:

Wilkes, P., Jones, S.D., Suarez, L., Haywood, A., Soto-Berelov, M., Mellor, A., Axelsson, C., Woodgate, W., 2012. Deriving metrics of vertical structure at the plot level for use in regional characterisation of S.E. Australian forests. *Proceedings of the GSR_2 Research Symposium*, Melbourne, Australia, 10-12 December 2012.

and

Jones, S., Haywood, A., Suárez, L., Wilkes, P., Woodgate, W., Soto-Berelov, M., Mellor, A., Axelsson, C., (2013). *Literature review for determining optimal data primitives for characterising Australian woody vegetation and scalable for landscape-level woody vegetation feature generation*. CRCSI Project Report, 53 pages, January, 2013. (available online at: <http://goo.gl/knst81>).

2.1. Introduction

Canopy structure metrics such as canopy height and the canopy vertical structure, although useful in themselves, can also be utilised as analogues for metrics that cannot be measured empirically e.g. habitat suitability, or as the basis of empirical or physical modelling to estimate hard to quantify attributes. Biomass, for example, is an Essential Climate Variable and is used to estimate potential sources and sinks of atmospheric carbon dioxide (Sessa, 2009). However, biomass is difficult to measure directly (e.g. with destructive harvesting), and therefore the allometric association with canopy height is routinely used to extrapolate biomass estimates across vegetated landscapes (Lefsky et al., 2002a).

Traditionally, metrics of canopy structure are measured at the plot scale, these are then upscaled to larger areas (e.g. a forest stand) using an expansion factor (McRoberts and Tomppo, 2007). However, a sparse network of plot values may not capture the structural variability apparent across a large area, particularly over native forest where variability in forest structure can be high. Advances in remote sensing techniques have offered an opportunity to synoptically attribute forests (Wulder, 1998; McRoberts and Tomppo, 2007). For example, passive EO instruments and airborne and spaceborne LiDAR instruments have allowed wall-to-wall attribution of forest structure (Lim et al., 2003; Wulder et al., 2012b).

The aim of the majority of studies is to determine an accurate empirical statistical association between an independent variable (or set of variables e.g. ALS return height percentiles or reflectance values) and a dependent target attribute, such as canopy height. The logical technique is therefore to develop a function (model) that minimises the estimation error between the independent and dependent variables. For example, Ordinary Least Squares (OLS) regression fits a linear function that minimises the residual error between predictor variables and observations. The model and its coefficients (the slope and intercept coefficients in the case of OLS regression) can then be used to estimate unknown dependent values. Empirical models have been shown to yield good results in the domain they were developed for (e.g. Næsset (2007)), for example, with the same sensor (i.e. independent variables) and within the range of

Metrics of canopy vertical structure suitable for large area forest attribution

Table 1. Applications of canopy height and canopy vertical structure (CVS).

Application	Metric	Reference
Forest Biomass	Height	Lefsky et al. (2001); Drake et al. (2002); Hurtt et al. (2004); Patenaude et al. (2004); Asner et al. (2010); Latifi et al. (2010); Swatantran et al. (2011); Hudak et al. (2012)
	CVS	Lefsky et al. (1999b, 2001); Drake et al. (2002); Næsset (2004); Zhao et al. (2011)
Habitat mapping	Height	Hill and Thomson (2005); Hyde et al. (2006); Goetz et al. (2007); Hinsley et al. (2009); Haywood and Stone (2011)
	CVS	MacArthur and MacArthur (1961); Goetz et al. (2007); Zellweger et al. (2013)
Species/Floristics /cover	Height	Burgman (1996); Hill and Thomson (2005); Tickle et al. (2006); Mellor et al. (2012); Zhang and Liu (2012)
	CVS	Lucas et al. (2008); Miura and Jones (2010); Zhang and Liu (2012)
Resource management and forest inventory	Height	Næsset (1997, 2007); Brack (2007); Turner (2007); Wulder et al. (2008); Lim et al. (2011)
	CVS	Korpela et al. (2012)

canopy height observations. However, beyond this empirical models can become unreliable (Hopkinson et al., 2006). A more robust approach is to use a simple metric that is not tailored to a specific study area or forest type e.g. standard deviation of return height (Hopkinson et al., 2006), where overall accuracy is sacrificed to improve universality. For example a metric that is transferable between forests of differing species composition, terrain, land use or disturbance history.

2.1.1. *Canopy height*

Tree height is generally defined as the vertical distance from the base point (i.e. the intersection between the ground and the stem) to the highest point of the live crown. Canopy height is then the average height of all, or a subset, of tree heights within a delineated area e.g. an inventory plot or forest stand (Gschwantner et al., 2009). Commonly used metrics of canopy height include dominant height, defined as the mean height of all trees that are not overtopped and whose crowns are not shaded by adjacent trees

(Lefsky *et al.* 1999a), and “Lorey’s height” (h_L) (EQ. 2) which weights the height of all trees in a plot by basal area (Næsset, 1997; Lim *et al.*, 2003).

$$h_L = \frac{\sum_{i=1}^n h_i a_i}{a}$$

EQ. 2

where h_i and a_i are the height and basal area of individual trees. Measurement of tree and canopy height is a core component of a forest inventory (Tomppo *et al.*, 2010), where individual trees are measured using a clinometer, electronic total station or a hypsometer (Andersen *et al.*, 2006; Hollaus *et al.*, 2006).

Owing to the ability to measure three dimensional forest structure, laser ranging techniques, such as LiDAR, lend themselves to the assessment of canopy height, (Lim *et al.*, 2003) and LiDAR derived canopy height is now considered by some as more accurate than inventory techniques (Holmgren and Jonsson, 2004; Maltamo *et al.*, 2006; Magnusson *et al.*, 2007). Canopy height has been estimated using full-waveform and discrete return airborne (Næsset, 1997; Lefsky *et al.*, 1999a; Wulder *et al.*, 2012a) and spaceborne (Baccini *et al.*, 2008; Lefsky, 2010; Simard *et al.*, 2011) LiDAR instruments. To estimate canopy height from LiDAR data, the height of the ground surface has to be first determined (Lovell *et al.*, 2003; Hyypä *et al.*, 2008). As the vertical coordinates of returns are measured to a datum (e.g. the Australian Height Datum), returns need to be normalised for terrain before canopy height can be estimated. This is achieved by partitioning returns into those either backscattered from the ground or not, where ground returns are often identified as the last return from an outgoing pulse. For data captured over dense vegetation the last return may be from within the canopy, a filtering process can be used to improve identification of the ground returns (Evans and Hudak, 2007). Once ground points have been successfully identified, a Triangular Irregular Network (TIN) vector is created, which when rasterised, produces a Digital Elevation Model (DEM). Using the TIN or DEM, the normalised height of returns relative to the ground surface can be estimated by subtracting the ground height from the return height. From the normalised point cloud, canopy height for an areal unit can be inferred as the vertical distance

between the ground to a top-of-canopy metric, for example, maximum or a percentile of height (Næsset, 1997, 2005, 2007).

There are a number of sources of error that need to be considered when estimating canopy height with ALS. The sample of small footprint instruments is often incomplete (i.e. not all visible surfaces are sampled) and therefore the crown apex may not be detected, which leads to an underestimation of canopy height (Lim et al., 2003; Lovell et al., 2003; Hyypä et al., 2008). The number of identified ground returns may be insufficient to derive a reliable DEM and therefore an accurate estimate of canopy height (Takahashi et al., 2008). This again is influenced by sampling frequency as well as pulse penetration through the canopy (Goodwin et al., 2006) and scan angle (Hyypä et al., 2008).

2.1.2. Canopy vertical structure

The quantity, arrangement and connectivity of foliar and woody material, along the vertical axis of a forest canopy, can be defined as canopy vertical structure (Brokaw and Lent, 1999). Canopy vertical structure is traditionally measured *in situ* and at the plot scale. Popular techniques to attribute for canopy vertical structure include measuring the distance to intercepting foliage from underneath the canopy, either with a calibrated telephoto camera (MacArthur and Horn, 1969; Lefsky et al., 1999a; Lovell et al., 2003; Goodwin et al., 2006) or more recently a TLS (Jupp et al., 2008; Lovell et al., 2012). This generates a Canopy Height Profile (CHP), where the probability of an intercept is plotted against height (MacArthur and MacArthur, 1961). Alternative techniques include ocular assessment (Van Den Meersschaut and Vandekerckhove, 1998; Hnatiuk et al., 2009), use of inventory measurements and allometry (Maltamo et al., 1997; Baker and Wilson, 2000; Zenner and Hibbs, 2000) or three dimensional geometric crown models (Van Pelt and North, 1996; Drake et al., 2002; Scanlan et al., 2010). As with canopy height however, measurements based on *in situ* field data are logistically expensive to collect and often only capture a relatively small subsample (Means et al., 1999; Whitehurst et al., 2013).

Canopy vertical structure can be estimated with remote sensing, for example by derivation of vertically resolved gap probability (P_{gap}). Gap probability is the proportion of sky (or ground for downward looking instruments) relative to the proportion of canopy at a certain height within

the canopy, so that at the top of the canopy P_{gap} is equal to 1. Assuming a Poisson distribution of vegetation, a logarithmic transformation can be applied to P_{gap} estimates to account for occlusion of vegetation in the shadow of observed surfaces (MacArthur and Horn, 1969). Canopy cover L at height z along the canopy vertical axis can be estimated with:

$$L(z) = -\log(P_{\text{gap}}(z))$$

EQ. 3

The derivative of L is equivalent to the “apparent” canopy height profile (Lovell et al., 2003). This differs from the “actual” canopy height profile due to the inability to accurately quantify the total volume of vegetation i.e. plant volume area density (PAVD). Derivation of the “actual” canopy height profile can be achieved either analytically (Ni-Meister et al., 2001) or with remote sensing (Ni-Meister et al., 2001; Jupp et al., 2008), where assumptions about the angular distribution of vegetation and the non-Poisson distribution of foliage (i.e. “clumping”) are made. TLS instruments have been able to estimate PAVD as they can operate at the “hinge” angle of 57.5° where foliage density measurements are insensitive to leaf angle distribution (Jupp et al., 2008). As the aim of this investigation is to characterise vegetation structure over large areas using synoptic remote sensing techniques, which tend to have a limited viewing angle, quantification of plant volume area density as a metric of canopy vertical structure is beyond the scope of this study.

Owing to the ability to penetrate the upper canopy surface and measure three-dimensional vegetation structure, a number of studies have utilised LiDAR to estimate canopy vertical structure. As with canopy height, both discrete return (Lovell et al., 2003; Riaño et al., 2003; Jaskierniak et al., 2011) and waveform recording (Lefsky et al., 1999a; Means et al., 1999; Drake et al., 2002) LiDAR have been utilised. Application of P_{gap} theory has been extended to discrete return ALS, where the probability of a gap from the top of the canopy to a given height (z) is calculated as a proportion of the total number of outgoing LiDAR pulses:

$$P_{\text{gap}}(z) = 1 - \frac{\sum z_i (z_i > z)}{N}$$

EQ. 4

where z_i are individual returns and N is the total number of returns. ALS derived CC can then be estimated using EQ. 3.

Vertically resolved ALS derived L curves can be noisy owing to leaf clumping and canopy voids (Coops et al., 2007). To stabilise vertically resolved L a parametric distribution function can be fitted to the curve, for example a Weibull function (Lovell et al., 2003; Coops et al., 2007; Mitchell et al., 2012). This not only stabilises the canopy height profile, but can also summarise canopy vertical structure into a few parameters which are correlated with structure metrics. For example, Coops et al. (2007) noted a good agreement between the Weibull location and scale parameters with mid crown depth as a ratio of total height and crown length respectively.

Jaskierniak et al. (2011) fitted bimodal mixture models to ALS derived point clouds captured over multi-strata forests, where single parametric functions would have been inappropriate. An extension of mixture modelling is to use a hard classifier such as k -means clustering to parameterise different canopy strata (Zhang et al., 2011). Other techniques used to characterise canopy vertical structure include analysing ALS return dispersion through the canopy e.g. coefficient of variation or 'rumple' (Zimble et al., 2003; Kane et al., 2011) or interpreting frequency histograms of return height (Næsset, 2004; Maltamo et al., 2005).

2.1.3. Aims and objectives

Characterising forest canopy structure metrics, such as canopy height and canopy vertical structure, over large and heterogeneous forested landscapes requires choosing the most appropriate metrics and techniques that are robust to different forest types. This chapter compares methodologies, across a range of forest types, which have been previously used to attribute canopy height and canopy vertical structure, with the aim of identifying metrics that are universally applicable.

2.2. Materials and methods

2.2.1. Study area

Three 5 km x 5 km study areas (Table 2), situated in Victoria, Australia, were selected based on the following criteria: representativeness of a vegetation type and structural class at a regional scale (Montreal Process

Implementation Group for Australia, 2013) and minimal historical disturbance (e.g. fire or timber extraction). The Low Open Woodland (*LOW*) study area is characterised by a relatively short and discontinuous *Eucalypt* canopy (e.g. *Eucalyptus tricarpa* and *Eucalyptus macrorhyncha*) with a patchy and shrubby understorey. The Mixed Species Forest (*MSF*) study area is characterised by undulating terrain and high *Eucalypt* species diversity. Within steep gullies, patches of temperate rainforest exist where canopy height can be ≥ 50 m. The Tall Closed canopy Forest (*TCF*) is dominated by *Eucalyptus regnans* and *Eucalyptus delegatensis* species that can reach heights of >80 m, canopy cover is dense and there is a tall (sometimes >20 m) shade tolerant strata of *Northofagus* and *Acacia* species. The *TCF* forest type is one of the most carbon dense ecosystems on earth (Keith et al., 2009). Figure 2 compares the canopy structure and terrain of the three study areas.

The unique physiology of *Eucalypt* species and forest stands, such as an erectophile leaf angle distribution, asymmetrical crown configuration, low foliage density and leaf and crown clumping (Jacobs, 1955), may confound the estimation of structural metrics using ALS. For example, this increases the proportion of backscatter from the ground, mid and understorey when measured with ALS (Lee and Lucas, 2007; Armston et al., 2009; Jenkins and Coops, 2011).

2.2.2. Forest inventory data

A total of twenty seven 0.04 ha circular forest inventory plots were installed across the three study areas. Plots were established in the summer/autumn 2011-12 following Victorian Department of Environment, Land, Water and Planning (DELWP) Victorian Forest Monitoring Programme (VFMP) protocols (Department of Sustainability and Environment, 2012). Measurements for all trees within the forest inventory plot with a Diameter at Breast Height (DBH) >10 cm included DBH, species and live status. For a subset of trees (including the three tallest), height was also recorded where tree height was measured using a range finder (TruPulse 200B, Laser Technology, Colorado, USA). Canopy height was then estimated as (i) the maximum recorded tree height (INV_{max}), (ii) Lorey's height ($INV_{Lorey's}$) (EQ. 2) and (iii) Dominant height



Figure 1. Representative forest structure at the three study areas, derived using ALS point cloud data. For each study area, three panes illustrate: a 1500 m x 30 m transect highlighting topography and structural heterogeneity [top], a downward looking view showing canopy cover (green is canopy and brown is ground) [left] and a side view showing canopy vertical structure [right]. The location of the downward and side ward panes are displayed in the transect plot. Refer to key pane for pane dimensions.

Table 1. Description of study areas.

Site	Abb.	Coordinates	Elevation (m)	Slope (°)	Description	Dominant forest types* (>10% of study site)	Canopy height (m)	Mean annual rainfall (mm)^
Low open woodland	<i>LOW</i>	36° 45' 11" 144° 57' 57"	170 – 300	5	Short, open forest with low canopy density and sparse, shrubby understorey.	Heathy Dry Forest, Box Ironbark Forest	<30	584
Mixed species forest	<i>MSF</i>	37° 28' 26" 148° 20' 19"	200 – 400	16	Medium height, dense forest characterised by diversity of eucalypt species. Dense non-eucalypt understorey.	Riparian Forest, Damp Forest, Shrubby Dry Forest	20 – 50	833
Tall closed forest	<i>TCF</i>	37° 41' 22" 145° 41' 5"	975 – 1220	15	Very tall, dense, high productivity forest. Consisting of 1939 regrowth and localised areas of 'old growth' forest (Lindenmayer et al., 2000).	Montane Wet Forest, Wet Forest, Riparian Forest, Cool Temperate Rainforest, Montane Damp Forest	20 – 90	1122

(INV_{dom}) which is equal to the mean height of the three tallest trees in a 0.04 ha plot (Lovell et al., 2003).

2.2.3. Airborne laser scanning data

Small-footprint ALS was acquired over an area of 5 km x 5km for each study area (Table 3). A Riegl LMS-Q560 laser scanner was used, this digitised the full waveform of backscattered energy in real-time. Post-processing with Riegl RiAnalyze® (version 4.1.2) identified discrete peaks (“returns”) in the backscattered signal using a Gaussian Pulse Estimation technique (Riegl, 2006), resulting in a maximum of 7 returns per outgoing pulse. Analysis of discrete return data was performed (as opposed to the full-waveform) as this is still the operational norm for large area forest characterisation (Wulder et al., 2012b). Flight line overlap was removed using Airborne Research Australia’s RASP software (Lieff, 2009), where for each 2 m x 2 m voxel, returns with the scan angle closest to nadir were retained. Returns were then classified into either ground or non-ground. Using ground returns only, a TIN ground surface was computed and subtracted from all returns to calculate relative return height. Return classification and height calculations were achieved using LAStools v.120913 (Isenburg, 2012).

Twenty-seven ALS “plots” concurrent with the forest inventory plots were extracted from the full-dataset. ALS plots were circular with a radius of 11.8 m to replicate the VFMP large tree plot dimensions (DEPI, 2012). Following the suggestion of previous studies, a threshold of 2 m was applied to discard returns from the shrub layer and coarse woody debris, as well as misclassified ground-returns (Nilsson, 1996; Næsset, 1997; Maltamo et al., 2005; Jaskierniak et al., 2011).

Table 3. Flight and sensor specifications for the ALS acquisition at the three study areas.

Specifications	
<i>Capture specifications</i>	
Date	15/4 - 18/4/2012
Flying height	<600 m agl
Pulse density (overlap removed)	10 pulses m ⁻²
Swath overlap	50%
Absolute vertical accuracy	±20 cm
Absolute horizontal accuracy	±30 cm
Mean footprint diameter	30 cm
<i>Instrument specifications</i>	
Instrument	Riegl LMS-Q560 laser scanner (Horn, Austria)
Operating wavelength	1550 nm
Beam divergence	0.5 mrad
Max off-nadir scan angle	±22.5°
Outgoing pulse rate	240 kHz

2.3. Data processing

2.3.1. Canopy height

Four commonly derived analogues of canopy height were computed from the ALS data. These were; maximum return height (ALS_{max}) (Næsset, 1997), the height of the 95th percentile of return height (ALS_{95}) (Jenkins, 2012), the arithmetic mean of return height $\geq 80^{\text{th}}$ percentile of return height ($ALS_{>80}$) (Tsfamichael et al., 2010) and “predominant height” (ALS_{CHM}) (Lovell et al., 2003). ALS_{CHM} was estimated from a 1 m x 1 m Canopy Height Model (CHM) of maximum return height, the CHM was divided into areas equal to one crown width (~ 7 m) where the mean of local maxima was computed as ALS_{CHM} .

ALS derived canopy heights were compared with the three forest inventory measurements of canopy height by calculating the Root Mean Square Error (RMSE) and computing a linear regression of inventory measured (dependent variable) and ALS estimated (independent variable) canopy height.

2.3.2. Canopy vertical structure

As discussed in Section 2.1.2, characterisation of the canopy vertical structure can be achieved with the application of a parametric continuous Probability Density Function (PDF). However in certain forests, for example where a shade tolerant of sub-dominant strata is present below the upper canopy (e.g. *TCF* study area, Figure 2), the use of a single function may not adequately characterise the vertical distribution of ALS returns. In this instance a mixture model of two or more parametric functions is best suited (Jaskierniak et al., 2011). Here, the work of Jaskierniak et al. (2011) is extended to forests where the number of parametric functions (k) required is not fixed for a particular forest type and is assumed unknown *a priori*.

More formally, the probability of an ALS return coming from a certain height along the vertical profile, from the ground to the canopy top (z_{\max}), can be summarised with a nonparametric PDF of the response variable $f(z)$. For example, the probability of a return X at height z_X is:

$$Pr[0 \leq z_X \leq z_{\max}] = \int_0^{z_{\max}} f(z) dz$$

EQ. 5

Alternatively, the canopy height profile can be parameterised with a single continuous parametric PDF. A parametric PDF has the form $f(\theta)$ where θ consists of two parameters μ and σ for location and scale respectively, some PDFs have up to an additional two shape parameters (ν and τ). For example, the commonly used 3-parameter (location, scale and a shape parameter) Weibull PDF can be used to characterise z (Lovell et al., 2003; Coops et al., 2007; Mitchell et al., 2012);

$$f(z|\mu, \sigma, \nu) = \frac{\nu}{\sigma} \left(\frac{z - \mu}{\sigma}\right)^{\nu-1} e^{-\left(\frac{z - \mu}{\sigma}\right)^\nu}$$

EQ. 6

A canopy with two or more strata could present a situation where a single function may not be satisfactorily representative e.g. a single function mapped to two or more distinct stratum. Complex distributions can therefore be expressed by a combination of PDFs, where a single PDF

characterises the vertical distribution of a single strata. This is known as a mixture modelling:

$$f(z|\psi) = \sum_{k=1}^K \pi_k f_k(z|\theta_k)$$

EQ. 7

where the mixture model $f(z)$ depends on parameter set $\psi = (\theta, \pi)$, where $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, $\pi = (\pi_1, \pi_2, \dots, \pi_k)$ and $K = 1, 2, \dots, k$ model components; $f_k(z|\theta_k)$ is the PDF of z where θ_k are the PDF parameters and $0 < \pi_k < 1$ is the prior probability for component k . Here, k was iterated from 1 – 2 reflecting the possible occurrence of up to two canopy strata (e.g. single dominant strata or dominant and shade tolerant strata), at the *TCF* study area the possibility of a third canopy strata was also tested.

Mixture modelling was achieved in R (R Development Core Team, 2014) using the *Generalized Additive Models for Location, Scale and Shape* (GAMLSS) package (Stasinopoulos et al., 2012). GAMLSS estimates the response variable distribution parameters with a ‘semi-parametric’ regression model, where the regression model may include non-parametric smoothing functions (hence semi-parametric) (Stasinopoulos et al., 2012). Optimal parameters were chosen using the Expectation-Maximisation (EM) method, EM is an iterative method for finding the maximum likelihood of the mixture parameters. A total of 44 continuous PDFs are available in the GAMLSS package, here this was reduced to a set of 27 most appropriate for Australian sclerophyll forest canopies (see Table 4 in Jaskierniak et al. (2011)).

Firstly, on a per plot basis, all 27 distributions were fitted to the ground normalised ALS data as a single function. Recursively applying mixture models for all combinations of functions, where $k = [1, 2, 3]$, across all 27 plots, would be computationally time consuming. Therefore the most suitable mixture model where $k = 2$ were determined by first pairing all 27 functions with the Normal PDF (NO). Subsequently and on a per plot basis, the PDFs used for the four best fitting NO combination mixture models were utilised as the first model component in the next step. Each PDF was combined with the 27 functions to find the most appropriate function for the second model component. For the *TCF* study area, the four best fitting

mixture models were $k = 2$ were again combined with all 27 functions to determine the most appropriate third function. It should be noted that if the return height probability density function is not clearly bi- or multi-modal, the order which PDFs are utilised within GAMLSS mixture models cannot be specified (e.g. which function is associated with a specific strata).

An information-theoretic approach, the Akaike Information Criterion (AIC), was used to test goodness-of-fit (Akaike, 1974). $AIC = -2\ln(L) + 2p$ where L is the maximum likelihood of the mixture model and p is the sum of the number of model parameters. Therefore AIC penalises a model for an increasing number of model of parameters, with the aim of reducing over-fitting. To inter-compare a single model i against the set of all other possible models R , AIC values were weighted ($wAIC_i$) using the best fitting model, per plot, as a benchmark (AIC_{min}) i.e. the model with lowest AIC score regardless of k .

$$\Delta_i = AIC_i - AIC_{min}$$

EQ. 8

$$wAIC_i = \frac{e^{(-0.5\Delta_i)}}{\sum_{r=1}^R e^{(-0.5\Delta_i)}}$$

EQ. 9

Weighting AIC values allows confidence in which model is most appropriate for characterising an individual plot. For example, if all models from a set had similar $wAIC$ values then there would be little confidence as to which is most appropriate. Conversely, if one model had a far larger $wAIC$ value (e.g. $wAIC \rightarrow 1$), then that model could be selected with confidence.

2.4. Results

2.4.1. Canopy height

When estimating INV_{max} , ALS_{max} and ALS_{95} produce the most accurate results (Figure 3). A strong linear association ($r^2 > 0.75$, $p < 0.001$) and an RMSE < 5 m (Figure 4) is achieved estimating INV_{dom} with ALS_{95} , $ALS_{>80}$ and ALS_{CHM} . However, ALS_{95} performs well when estimating INV_{dom} across the

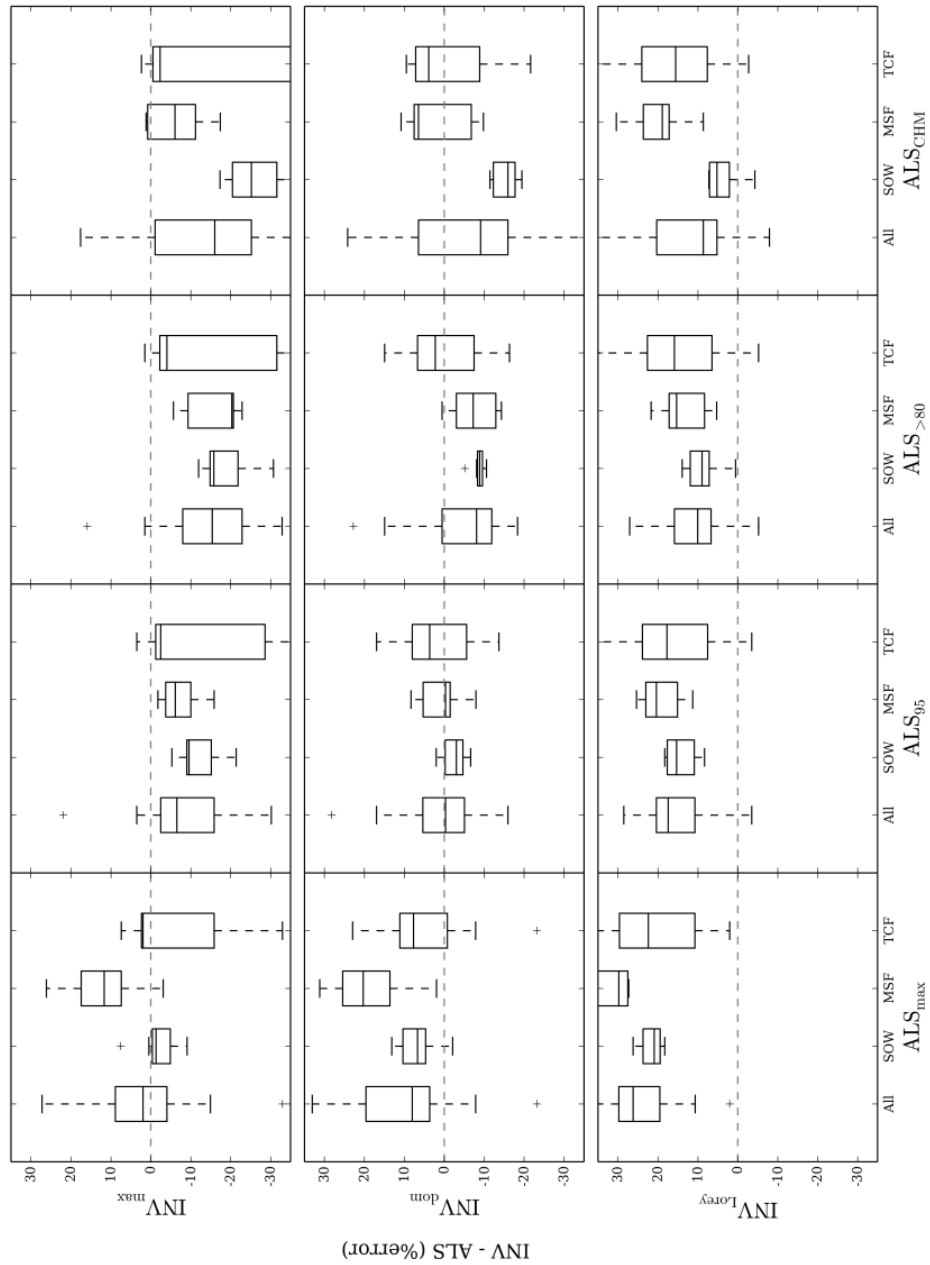


Figure 3. Boxplots of residual error when using ALS metrics to estimate inventory (INV) measured canopy height. Within each pane, the four groups represent all plots and plots from the *Low Open Woodland*, *Mixed Species Forest* and *Tall Closed canopy Forest* (left to right). Within each group, different point cloud components are displayed.

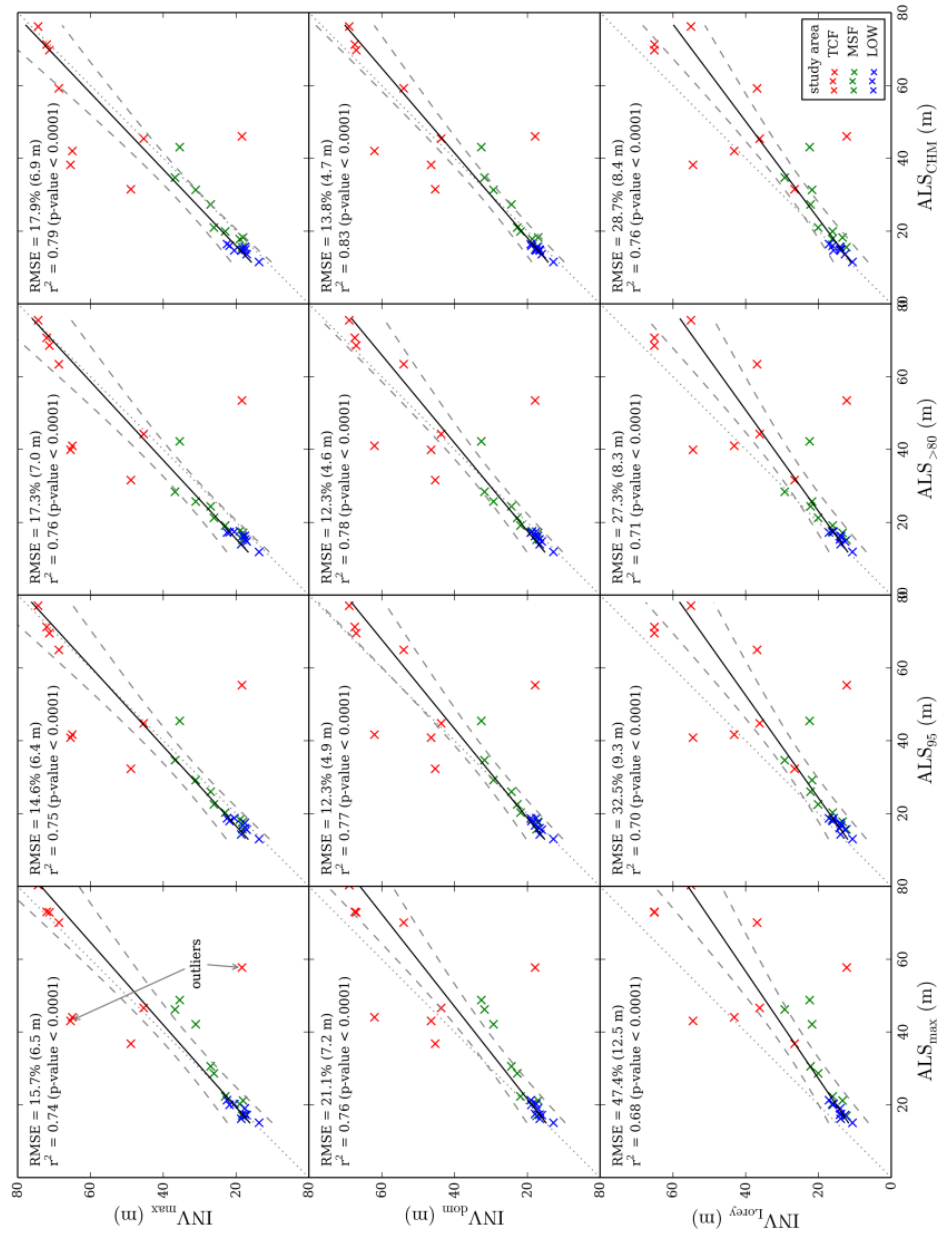


Figure 4. Scatter plots comparing different techniques for inventory measured and ALS derived canopy height at 27 plots. Also displayed are the metrics derived using three different point cloud components and linear regression using 'all' returns (-) including 95th percentile confidence intervals (--). RMSE values are calculated with the outliers (95th percentile of residuals) removed.

three different forest types where mean error is close to zero (Figure 3). Using ALS_{CHM} returns a strongest linear association with regard to INV_{dom} (Figure 4), however ALS_{CHM} underestimates INV_{dom} by $\sim 10\%$ (Figure 3). Furthermore generation of this statistic is more computationally expensive, particularly when generated across a large area. None of the ALS metrics estimate INV_{Lorey} particularly well, ALS_{CHM} returns the most accurate estimate although this still underestimates INV_{Lorey} by $\sim 10\%$.

It is clear that the canopy height of shorter forests can be estimated more precisely with ALS when compared to inventory measurements, and as canopy height increases so does error between measurements (Figure 4). Three outliers were identified at the 95th percentile confidence interval (Figure 4). All three plots were located at the *TCF* site; at one plot ALS maximum height was greater than inventory height due to the germination points of large trees, that comprised the dominant canopy, falling outside the inventory plot. At the other two plots, ALS maximum height was significantly less than inventory height and suggested reasons are steep terrain and dense vegetation causing geo-registration errors.

For wall-to-wall maps of canopy height of the three study areas at a 30 m resolution, refer to Section 0, Figure 11.

2.4.2. *Canopy height profiles*

When considering $k = 1$, of the seventeen continuous probability density functions (PDF) tested, the generalised gamma PDF returned the lowest AIC value for 16 of the 27 ALS plots. The relative complexity of the generalised gamma PDF (3 parameters) allows it to characterise the high negative skew characteristic of ALS derived canopy profiles, when compared to simpler functions from the same family e.g. the 2-parameter Weibull. When considering $k = 2$ to characterise the canopy height profile, the generalised gamma again featured prominently, often in combination with another generalised gamma function (Figure 5). For plots at the *TCF* site, parameterising the canopy height profile with $k = 3$ was also tested i.e. assuming three canopy strata. Mixture models were comprised of a number of different combinations of parametric functions, including models that used the generalised gamma and the Box-Cox- t functions. However, there was no general consensus between plots for a study area specific mixture model.

Metrics of canopy vertical structure suitable for large area forest attribution

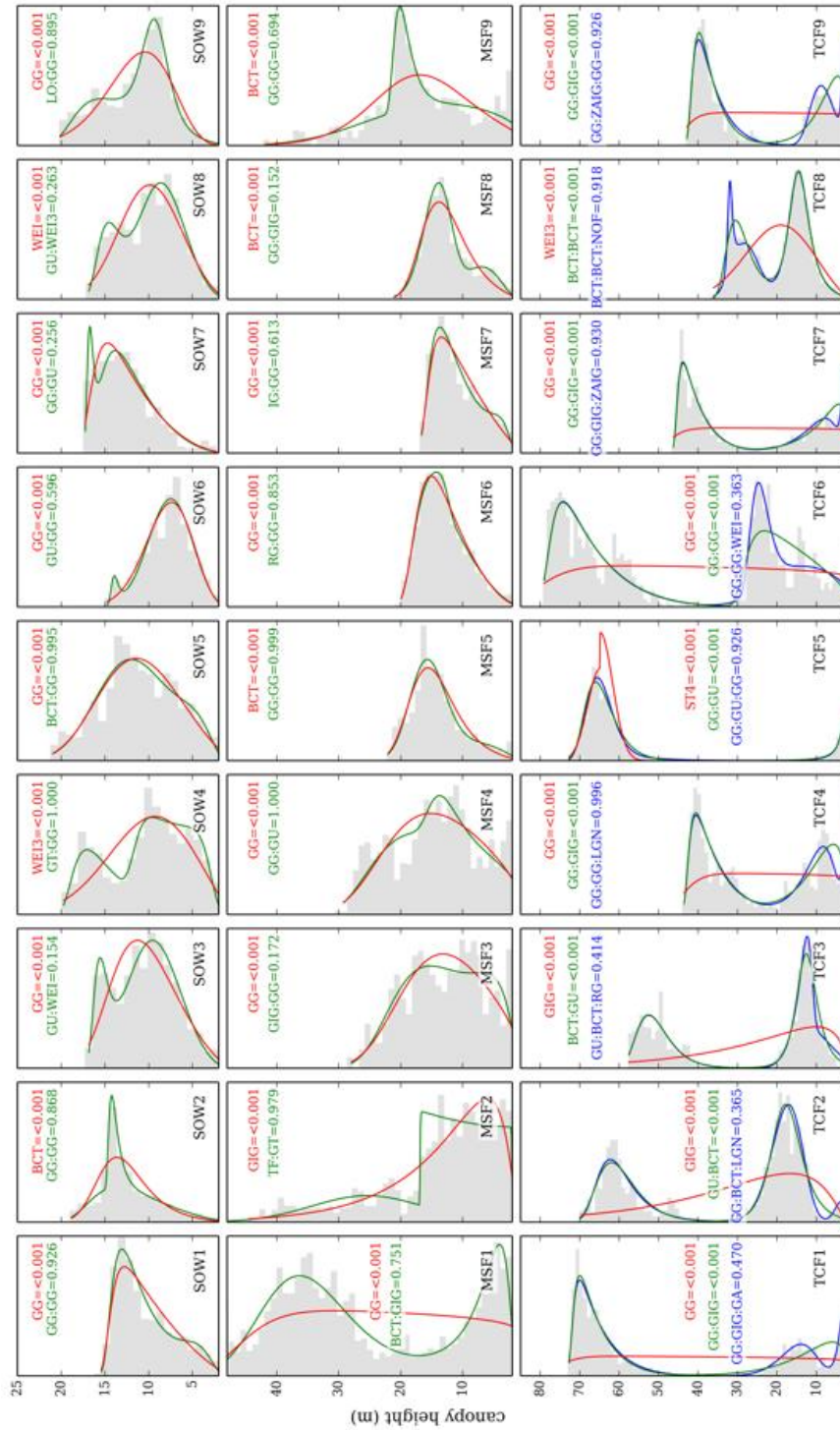


Figure 5. (previous page) Canopy height profiles characterised using mixture models where k was iterated from 1 - 3. Values are $wAIC$ calculated using EQ. 9. Function abbreviations are: *GG* generalised gamma, *BCT* Box-Cox- t , *GU* Gumbel, *WEI* Weibull, *GT* generalised y , *WEI3* Weibull (μ the mean), *LO* logistic, *GIG* generalised inverse Gaussian, *TF* t Family, *RG* reverse Gumbel, *IG* inverse Gaussian, *GA* gamma, *LGN* log normal, *ST4* skew t type 4, *ZAIG* zero adjusted inverse Gaussian and *NOF* normal family. The order of functions listed for each plot is not the order they were necessarily applied.

In all cases, models with the greater number of components returned the best fit i.e. $k = 2$ at *LOW* and *MSF* and $k = 3$ at *TCF*. If considering the weighted AIC values ($wAIC$), one mixture model tended to be a far better fit than other single functions or mixture models tested (e.g. $wAIC > 0.5$). Where $wAIC$ is < 0.5 , for example plot *MSF8* (Figure 5), the certainty of any single mixture model from the set providing the best fit to characterise the plot is reduced. Over fitting, defined here as two or more functions used to characterise a single canopy strata, is not desirable when attributing forest over large areas, as parameters describing specific strata cannot be directly inter-compared between plots. It is therefore suggested that AIC is not suited as a model selection criteria, as it does not penalise sufficiently for increasing model complexity.

Applying parametric functions over large areas is computationally expensive, particularly when attempting to select the most suitable model from a set of potential models (Jaskierniak et al., 2011). In an attempt to reduce computation time when applying over a larger area (in this case along 200 m transects as in Figure 6), k was assumed for each site and the set of potential mixture models was limited to the most appropriate from Figure 5. This again indicates the generalised gamma is the most suitable single PDF for characterising single strata forest (Figure 6 A1 – A8) and features often in mixture models. “Mixed” plots, which are comprised of two different dominant height cohorts, are characterised by a canopy height profile where two functions are fitted near the top of the canopy (e.g. Figure 6 B3 and C5). For some plots, the assumption of k maybe unrealistic. For example in Figure 6 C8 where the assumption was $k = 3$, a third function is used to parameterise an apparent mid-storey stratum which has a low relative amplitude. It is suggested a more appropriate assumption would be $k = 2$. This is also apparent at plots B1 – B4 (Figure 6)

where $k = 2$ is assumed but a more appropriate value for k may have been 1. This suggests that for native forest k cannot be assumed.

2.5. Discussion

To be able to characterise forest structure across large, continuous and heterogeneously forested areas, metrics as well as the remote sensing techniques used to measure them, need to be transferable between forest types. Furthermore, techniques also need to be robust when applied over forest of where there is no *a priori* knowledge of forest type or structure. The aim of this chapter was to apply metrics, which have been previously used to characterise canopy height and canopy vertical structure, across three structurally distinct forest types.

Canopy height can be accurately estimated from Airborne Laser Scanning (ALS) data using a single metric. Maximum canopy height (INV_{max}) was

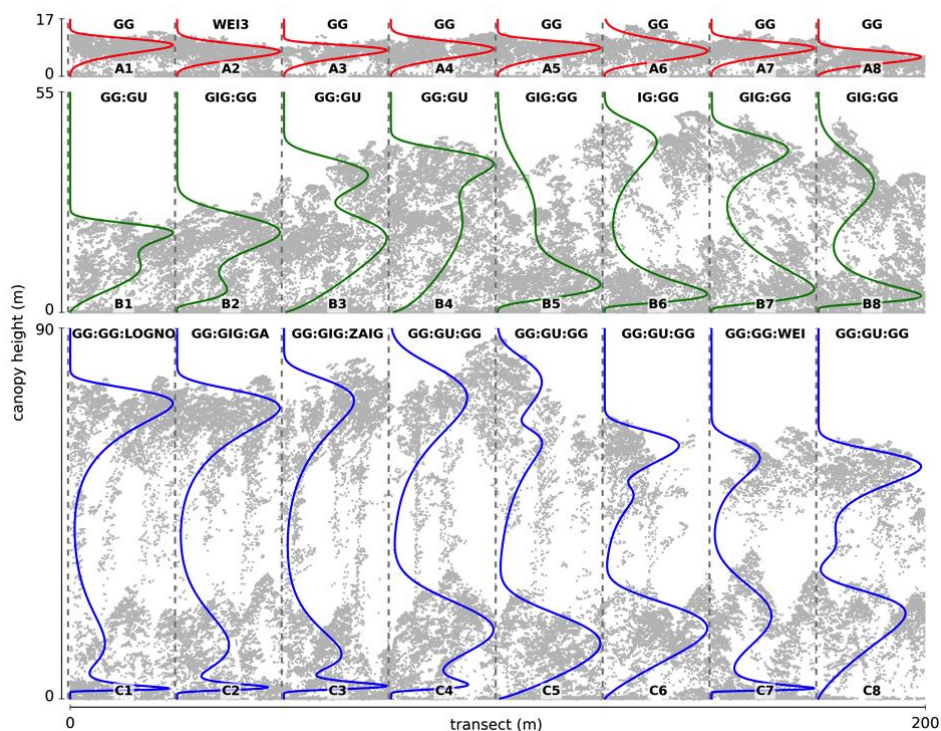


Figure 6. Mixture models fitted to 25 m x 25 m ALS plots along a 200 m x 25 m transect (dashed lines identifies plot boundaries). The three forest types represented are; *LOW* [top], *MSF* [middle] and *TCF* [bottom], where k is assumed to be 1, 2 and 3 respectively. ALS data (grey points) has been ground normalised.

generally well estimated by ALS (ALS_{max}), indicating the sampling frequency of the ALS acquisition (~ 9 pulses m^{-2}) was sufficient for detecting the apex of tallest trees. ALS_{95} returned the most accurate results when predicting INV_{dom} . All ALS metrics overestimated Lorey's height, particularly for taller plots.

A good agreement between *in situ* and ALS is apparent for shorter forests and are similar to errors reported in other studies. For example, Lovell et al. (2003) reported errors of 1 – 2 m for plots where canopy height was ~ 30 m, this similar to errors reported in the comparative *LOW* forest type where the RMSE for INV_{dom} estimated with ALS_{95} is ~ 1 m. To estimate Lorey's height, Hopkinson et al. (2006) used ALS_{MAX} which returned an RMSE of 2.5 m for plots with a canopy height of 4 – 28 m. However in this study ALS_{MAX} significantly overestimated Lorey's height, although overestimation occurred mainly occurred at plots where canopy height was >20 m. Improved estimates of canopy height have been produced using predictive models e.g. stepwise linear regression of ALS derived variables. For example (Næsset, 2007) reported RMSE values of $<6\%$ when predicting dominant canopy height of coniferous forests. However, such predictive models are tailored to a specific forest type and may lack general applicability.

ALS_{max} slightly overestimates INV_{max} for the *MSF* plots, which would suggest that the inventory measurements underestimated canopy height (Andersen et al., 2006). This can be caused by a number of factors including steep terrain and dense vegetation causing difficulties sighting tree tops (Drake et al., 2002; Lee et al., 2004; Goodwin et al., 2006), furthermore the asymmetrical and modular shape of Eucalypt trees may exacerbate accurate *in situ* measurements.

When parameterising the vertical distribution of foliage along the vertical canopy axis, the 2-parameter Weibull function has been widely applied to characterise forest canopies (Lovell et al., 2003; Coops et al., 2007; Mitchell et al., 2012). However, results presented here would indicate that the generalised gamma distribution is more suited to characterising the vertical distribution of ALS returns at forest plots characterised by a single strata; the generalised gamma PDF also features prominently in mixture models of two or more parametric functions. The crown archetype of *Eucalypt* species can be described as an inverted and fluted cone (Sillett et

al., 2010), this results in a concentration of leaves towards the top of the canopy (Jacobs, 1955). The generalised gamma is suited to highly skewed distributions, such as the distribution of foliage within Eucalypt canopies. These results are contrary Jaskierniak et al. (2011), who found that mixtures which included the Gumbel distribution were most commonly selected to characterise Mountain Ash forest plots (similar to the *TCF* study area).

Even for forest plots characterised by a single canopy strata (e.g. *LOW* in Figure 2), mixture models produced a lower AIC value when compared to any single function. If the aim is to characterise forests over large and continuous areas, it would be preferable for a single function to characterise a single strata within the canopy height profile. The over-fitting of 2 or more parametric functions to a single stratum would make estimating the presence and distribution of stratum more difficult. Conversely, using a set number of parametric functions to characterise a canopy height profile (e.g. $k = 2$) may lead to the identification of intra-canopy features (*LOW* plots in Figure 6) as well as making inter-comparison of model parameters, such as mode height, difficult between plots. As AIC penalises for cumulative number of model parameters and not explicitly for number of model components, it is suggested AIC is inadequate for penalising model complexity for this application. Other information-theoretic approaches penalise increasing model complexity more strongly, such as the Schwartz Bayesian Criterion which penalises using the logarithm of number of model parameters (Stasinopoulos and Rigby, 2007).

When considering attribution of a forested landscape, the use of mixture models to determine the canopy height profile is computationally intensive (Jaskierniak et al., 2011). This problem could be addressed in a number of ways, for example, by limiting the number of functions or mixture models to a set of the most appropriate for a particular forest type (Jaskierniak et al., 2011). Model selection could be further enhanced by utilising a look-up table approach, where species composition or structural attributes (e.g. canopy height) are used to preselect the most appropriate models. However, even over a single forest type, particularly old-growth forest and native forest where disturbances can lead to multi-age stands, the number of canopy strata cannot be assumed (Lindenmayer et al., 2000). In this

circumstance a tool that could estimate the number of model components *a priori* would be beneficial. For example Hofton et al. (2000) used inflection points of Gaussian smoothed GLAS waveforms to estimate the number of Gaussian functions required in the deconvolution.

There are a number of limitations to this study worthy of discussion. Firstly, the number of plots considered for each forest type was limited. A larger number of plots may have produced a study area consensus with regard to the set of functions for creating mixture models. Furthermore, plots of only one size were considered. Increasing the plot size would integrate more trees within a sample, this has been shown to change the parameters of the Weibull PDF when applied to canopy height profiles (Wilkes et al., 2013). However the utilised plot size allows comparison with inventory plots, is common amongst remote sensing studies of forest structure and is similar to the resolution of passive remote sensing products which are useful for upscaling (Hudak et al., 2002). Different ALS capture specifications were not considered, for example a comparison of metric retrieval across the a range of scan angles (Lovell et al., 2005). Finally, with regard to use of mixture models to characterise the canopy height profile, the method presented does not account for occlusion of vegetation. In forests with a dense dominant canopy, the proportion of vegetation in the sub-canopy strata may differ significantly to that estimated using ALS. Applying a technique to transform the distribution of the point cloud to account for occlusion (e.g. resampling with a logarithmic transformation) would mitigate for this.

2.6. Conclusion

Techniques to characterise forest structure over large areas are required to be universally applicable across different forest types and where there is no knowledge of forest structure *a priori*. ALS derived metrics of canopy height and canopy vertical structure were tested across three forested landscapes that have distinctly different structural characteristics. When compared to forest inventory measured dominant canopy height, ALS derived metrics performed well, in particular the 95th percentile of non-ground return height estimated dominant canopy height consistently across all three study areas.

To characterise canopy vertical structure, a mixture model approach was taken. However, parameterisation of mixture models with the correct number of model components, that allows consistent characterisation over large areas, or even locally in more structurally complex forest, was not feasible. Over-fitting of mixture models was caused by the AIC model selection criteria not being aggressive enough to increasing model complexity.

3. Chapter 3

Using discrete-return ALS to quantify number of canopy strata across diverse forest types[†]

[†] This chapter is based on:

Wilkes, P., Jones, S.D., Suarez, L., Haywood, A., Soto-Berelov, M., Mellor, A., Woodgate, W., Skidmore, A., *in press*. Using discrete-return ALS to quantify number of canopy strata across diverse forest types. *Methods in Ecology and Evolution*.

3.1. Introduction

Forests are complex ecosystems especially when considering structure and spatial arrangement (Brokaw and Lent, 1999; Franklin and Pelt, 2004; McElhinny et al., 2005). Horizontal structure and spatial arrangement (e.g. extent, fragmentation, fractional cover etc.) has been the subject of much research and there are a number of techniques to describe horizontal arrangement across different scales (Bradshaw and Spies, 1992; Jennings et al., 1999; Plotkin et al., 2002). Techniques to characterise canopy vertical structure, particularly over large and continuous spatial domains, are less well described and are often surmised by an estimation of (variation in) canopy height alone (Zimble et al., 2003; Parker and Russ, 2004). This oversimplification could miss important structural features below the principal canopy such as presence/absence of a mid-storey and/or understorey.

Detail of forest canopy vertical structure provides additional information on within canopy interaction of radiation, temperature, wind speed and humidity (Koike and Syahbuddin, 1993) and consequently provision of ecosystem services including hydrology and habitat (Vierling et al., 2008; Jaskierniak et al., 2011). The characterisation of habitat structure has recently been recognised as an Essential Biodiversity Variable, defined as key measurements for monitoring and reporting of biodiversity change (Pereira et al., 2013). Non-catastrophic disturbances such as low intensity fire or thinning can cause divergence from a simple successional paradigm, leading to the removal of canopy strata or establishment of a multi-age and potentially multi-strata stand (Lindenmayer et al., 2000; Kane et al., 2013). Canopy structure cannot therefore be assumed to be a function of successional stage or other independent drivers (Kane et al., 2010a).

There are a number of definitions of canopy vertical structure applicable to forests (Parker and Brown, 2000). Here we apply the definition where non-uniformity of plant area volume density can be used to identify canopy stratum by locating modal peaks in the density curve (MacArthur and Horn, 1969; Lefsky et al., 1999). By extension a canopy can be stratified into one or more canopy strata (Koike and Syahbuddin, 1993; Whitehurst et al., 2013). As with other attributes of vegetation structure (Lovell et al., 2003), metrics of canopy vertical structure are scale dependent and therefore scale is intrinsic in a metrics definition (Wilkes et al., 2013).

Active ranging sensors, such as Light Detection And Ranging (LiDAR), have long been recognised as having the capability for remote sensing of forest structure. This is due to the ability of active sensors to penetrate the upper canopy surface allowing top-to-bottom measurement of the location of intercepted surfaces (e.g. leaf, branch, stem and ground surface) in three dimensions (Lefsky et al., 1999a; Lim et al., 2003). Over large areas and in operational scenarios, small-footprint ($\varnothing < 3$ m) Airborne Laser Scanners (ALS), configured to capture discrete pulses of backscattered energy (or “returns”), have been used to estimate forest structure attributes (Næsset, 1997; Means et al., 2000; Wulder et al., 2012b).

With regard to the vertical distribution of vegetation, ALS data has been analysed using a variety of statistical techniques, for example, using return height coefficient of variation (Bolton et al., 2013) and the three dimensional equivalent “rumple” (Kane et al., 2010b). More advanced parametric and nonparametric analytical techniques include cluster analysis and other segmentation algorithms (Riaño et al., 2003; Morsdorf et al., 2010; Zhang et al., 2011), application of single mathematical functions e.g. the Weibull (Lovell et al., 2003; Coops et al., 2007) or mixture modelling (Jaskierniak et al., 2011) and classification of points into strata by return height, type or intensity (Miura and Jones, 2010; Morsdorf et al., 2010; d’Oliveira et al., 2012; Zellweger et al., 2013b). Nevertheless, where knowledge of canopy structure is not available *a priori*, it is unlikely a model can be parameterised with the correct number of components (i.e. number of canopy strata) – particularly as structural complexity increases (Jaskierniak et al., 2011; Muss et al., 2011). This is important when attributing heterogeneous forested landscapes where forest structure may be highly variable (Lindenmayer et al., 2000).

Techniques to characterise canopy arrangement where an assumption of structure is not required include polynomial regression, wavelet and Fourier transformations and Gaussian kernel smoothing (Popescu and Zhao, 2008; Hopkinson et al., 2013). Although these techniques can characterise the distribution of foliage in a dynamic way, they do not explicitly attribute for vertical structure. The only existing method to use small footprint ALS, that is automated and makes no prior assumptions of structure is presented by Maltamo et al. (2005), who used a hierarchical histogram method to classify plots into single or multi-strata canopy.

This chapter introduces a new method for characterising forest canopy vertical structure from discrete return ALS that is applicable to large area characterisation. Model output is an ecologically meaningful metric of canopy vertical structure where the Number of Strata (*NoS*) within the canopy vertical profile is estimated. The new metric is applied across structurally diverse forests; from dry sclerophyll woodland to temperate rainforest. Validation is achieved with a comparison of crown volume models derived from field measurements at 27 permanent inventory plots.

3.2. Attributing canopy vertical structure

The new method to estimate Number of Strata (*NoS*) from discrete return ALS is outlined in Figure 7 and is described in more detail in Sections A – C. Processing of ALS data was achieved with the open-source ForestLAS Python package (<https://goo.gl/YkIhGL>).

A. Gap probability derivation

Gap probability (P_{gap}) can be estimated from large-footprint waveform recording LiDAR where the proportion of accumulated energy returned to the sensor from the top of the canopy to a height within the canopy is divided by the total returned backscattered energy (Ni-Meister et al., 2001). The theory has been extended to discrete return ALS where backscattered returns are aggregated into plots (Lovell et al., 2003; Riaño et al., 2003; Hopkinson and Chasmer, 2009) Occluded surfaces beneath the upper canopy can lead to an underestimation of plant area of the lower strata, which can be mitigated by applying a logarithmic transformation to P_{gap} . Projected cover L can therefore be estimated as $L = -\log(P_{\text{gap}})$ (MacArthur and Horn, 1969; Aber, 1979; Lovell et al., 2003; Riaño et al., 2003).

B. Probability Density Function (PDF) from projected cover (L)

If vertically resolved L is equivalent to a cumulative density function of ALS returns (or foliage), the derivative of L can be considered the probability density function (L_{PDF}). This is analogous to the “apparent” canopy height profile (Lovell et al., 2003). With the addition of the probability of a return coming from the ground or below a height threshold (z_t), L_{PDF} can be used

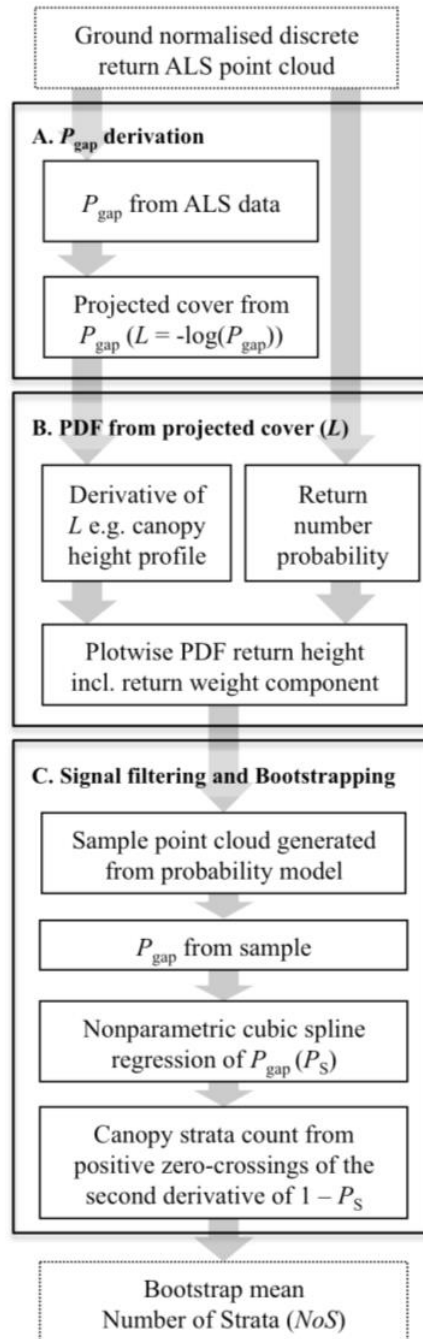


Figure 7. Derivation of canopy vertical structure from ALS derived gap probability (P_{gap})

to estimate the probability of a return X coming from height z within the plot voxel (EQ. 10), where the integral of $L_{PDF} = 1 - L(z_t)$.

$$\Pr[z_{max} > z_X > 0] = \begin{cases} \int_{z_t}^{z_{max}} L_{PDF}(z), z \geq z_t \\ L(z_t), z < z_t \end{cases}$$

EQ. 10

C. Bootstrapping and P_{gap} filtering

ALS is a sampling tool, and therefore may not measure distance to all visible surfaces within the canopy, particularly at lower pulse densities common in operational acquisition (Lim et al., 2003). Therefore, to derive a robust estimate of NoS as well as to quantify variance, NoS is estimated from simulated point clouds that are generated in a bootstrap (Efron, 1979). Points are drawn from EQ. 10 until the simulated point cloud has the same number of returns as the original; in this way each simulated point cloud has a different configuration. In addition, simulated points are also attributed with an estimate of a ‘Number of Returns’ (NoR) metadata value (see Section 3.3.2). This is achieved by calculating the NoR value probability from the original point cloud, in 1 m height intervals, then assigning simulated points an NoR value based on this probability.

From each simulated point cloud P_{gap} is again generated. At this stage, filtering of the P_{gap} curve is applied to remove signal noise, a result of intra-crown voids and foliage clumping, and to generalise canopy structure (Lovell et al., 2003; Coops et al., 2007). Filtering is achieved by application of a nonparametric cubic spline regression (Silverman, 1985). This differs from previous approaches to smooth canopy profiles with cubic splines (Jung and Crawford, 2008; Muss et al., 2011) by applying the transformation directly to the P_{gap} curve. Owing to the relative simplicity of the P_{gap} curve, compared to the canopy height profile, transformation of P_{gap} is more robust to a non-optimal smoothing coefficient. A new curve P_s is automatically generated so that the sum of squares, calculated at the intersections (t or “knots”) of spline section, satisfies the smoothing coefficient α (EQ. 11) (Dierckx, 1993).

$$\alpha = < \sum_{i=1}^n (P_s(t_i) - P_{\text{gap}}(t_i))^2$$

EQ. 11

When α is equal to zero, P_s is equivalent to an interpolating spline fitted through every point in the P_{gap} curve resulting in a close fit to P_{gap} . When $\alpha \rightarrow \infty$, P_s tends to an ordinary least squares solution. For a plot, a correct estimate of the number of strata can be derived from a range of α coefficients (α_s) where $\alpha_{s+1} < \alpha_s < \alpha_{s-1}$, with increasing α the number of strata identified quickly decreases before reaching an asymptote (Figure 8A). When $\alpha = 0$, the derived canopy height profile is dominated by noise, as α increases canopy scale features begin to dominate the shape of the curve until the canopy height profile becomes over-smoothed and canopy features are masked (Figure 8B – E). Determination of an appropriate α coefficient is therefore required.

For the spline regression to correctly interpret the top of the canopy, P_{gap} is extended by a further 10% of maximum canopy height where the “leading edge” is given a P_{gap} value of 1. Large voids in the canopy profile can produce erroneous local maxima with relatively low amplitude compared to the amplitude of modes coincident with canopy strata. These are removed by discounting local maxima in $f(x)'$, where $x = 1 - P_s$, whose amplitude is <5% of the largest maxima.

Positive zero-crossings of $f(x)''$ are indicative of local maxima in foliage density (Hofton et al., 2000; Popescu and Zhao, 2008). Therefore a count of the positive zero-crossings of $f(x)''$ is used to estimate the number of strata and a mean of bootstrapped samples is computed as the metric of canopy vertical structure (Figure 7).

3.3. *Application across a diverse forested landscape*

3.3.1. *ALS acquisition and pre-processing*

Study areas and data acquisition are outlined in Section 2.2. As with the previous chapter, ALS data was ground normalised and points <2 m were removed from analysis.

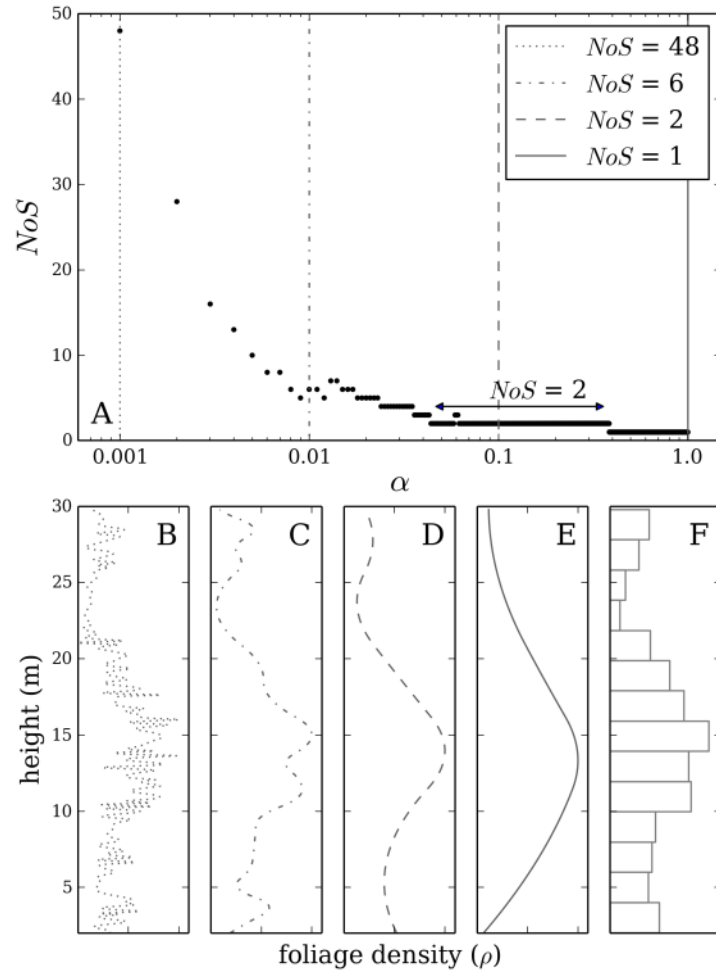


Figure 8. Sensitivity of NoS to increasing α [A] and “apparent” Canopy Height Profiles (CHP) derived by increasing α [B – E]. Also included is the canopy height profile histogram (2 m bins) used in the supervised classification [F] (see Section 3.3.3). Results from a supervised classification suggested this plot is characterised by two canopy strata (equivalent to pane D) and therefore an α_s of 0.04 – 0.38 [A].

Circular ALS plots with a radius of 11.2 m (0.04 ha) were extracted from the full ALS dataset to replicate the DELWP VFMP large tree plot dimensions (DEPI, 2012). ALS plot locations were determined by superimposing a regular grid (100 m intervals) over the three study areas. The grid was then stratified by forest type (Woodgate et al., 1994), from which 30% of sample points (up to a maximum of 30) that intersected each

forest type were randomly selected. To ensure independence of validation data, any selected points that were within 50 m of a permanent forest inventory plot were disregarded. In total, 239 plots were extracted from across the three areas; 59 for the *LOW* area, 76 for the *MSF* area and 104 for the *TCF* area. ALS data coincident with the 27 forest inventory plots was used for validation.

3.3.2. P_{gap} from ALS

Ignoring the assumption of ALS return dependence from a single outgoing pulse when computing P_{gap} can lead to an overestimation of P_{gap} (Lovell et al., 2003; Armston et al., 2013). Owing to factors such as the proprietary derivation of return intensity values, return intensity can be uninformative when estimating the proportion of backscattered energy (Armston et al., 2013). A robust estimate of return intensity was instead derived by weighting individual returns by the Number of Returns (NoR) metadata value (ASPRS, 2008) recorded for each outgoing pulse i.e. $1 / \text{NoR}$. Although this is an oversimplification, for example ignoring the differing surface reflective properties, partial backscatter or transmission losses (Ni-Meister et al., 2001; Hopkinson and Chasmer, 2009), Armston et al. (2013) reported a good agreement when comparing P_{gap} derived with this method and full-waveform data captured over the same plot. Here, P_{gap} is estimated using:

$$P_{\text{gap}}(z) = 1 - \frac{\sum w_i (z_i > z)}{W} \quad \text{EQ. 12}$$

where W is the per plot sum of $1 / \text{NoR}$ (including ground returns) and w_i is $1 / \text{NoR}$ for return i above height z .

3.3.3. Derivation of smoothing coefficient (α)

As mentioned in Section 2C, derivation of a suitable α coefficient used in EQ. 11 is required. To test the universal applicability of α , a forest type, study area and a universal α coefficient were estimated. This was achieved using a supervised classification of 82 plots from across the three study areas in a cross validation with modelled output. Using histograms of weighted return height, binned at 2 m intervals (e.g. Figure 8F), ten

individuals were asked to visually estimate the number of strata (NoS_{sc}) that characterised each of the 82 plots. The mode classifier response was calculated and used to determine α_s for each plot. Next, a suboptimal forest type, study area and universal α ($\tilde{\alpha}$) was estimated using random sub-sampling cross-validation ($N = 50$). For each iteration the most common α for a randomly generated 75% training cohort (Figure 9A) was selected as $\tilde{\alpha}$ (Figure 9B). If two or more $\tilde{\alpha}$ values were equally common, the median $\tilde{\alpha}$ was chosen (Figure 9B). The $\tilde{\alpha}$ coefficient was substituted into EQ. 11 to then compute the number of strata for the withheld samples (Figure 9C). A weighted average (EQ. 13) was then used to compute the most appropriate forest type, study area or universal α .

$$\tilde{\alpha} = \frac{\sum_{i=1}^N (\tilde{\alpha}_i (E_{\max} - E_i))}{\sum_{i=1}^N (E_{\max} - E_i)}$$

EQ. 13

where E_i is the mean of $|NoS_{sc} - NoS|$ for the withheld sample of each iteration i , and E_{\max} is the maximum E_i value.

3.3.4. Bootstrapping simulated point clouds

For each plot, sample point clouds were generated from EQ. 10 in a bootstrap as described in Section 3.2. A sensitivity analysis to determine the appropriate number of bootstrap samples suggested a maximum of 100 samples was sufficient to reduce variance in NoS estimates. This number was significantly reduced for structurally less complex forest plots dominated by a single stratum. In addition to NoS , canopy height, canopy cover and return height coefficient of variation (C_v) was calculated, where for each the mean value from the bootstrap simulations was used. Canopy height was computed as the 95th percentile of return height of all non-ground returns >2 m and canopy cover was estimated as $1 - P_{\text{gap}}(z)$ where $z = 2$ m.

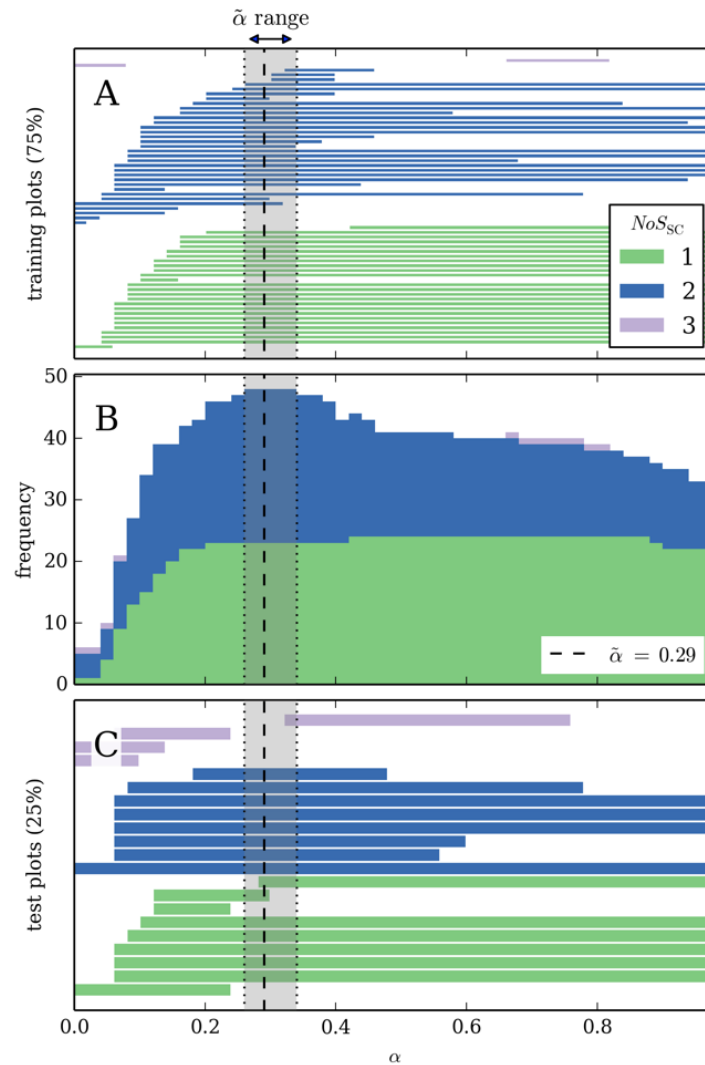


Figure 9. Cross validation to determine a suboptimal smoothing coefficient. For this 75% of plots were subset as training data [A], where the horizontal bars represent α as determined by a supervised classification (No_{SSC}). A suboptimal coefficient $\tilde{\alpha}$ was selected as the median value from the range of most frequently observed α values [B]. $\tilde{\alpha}$ was subsequently substituted into EQ. 11 and applied to the test plots [C].

3.3.5. Validation with field inventory

A Geometric Crown Volume Model (GCVM) approach was used to validate ALS derived canopy vertical structure (Drake et al., 2002; Morsdorf et al., 2010) where the modelling of GCVM follows Scanlan et al. (2010). This approach was chosen as it is not affected by occlusion of the upper canopy, which can be a drawback of *in-situ* ranging techniques (Lovell et al., 2003), it can also be parameterised with existing inventory data. Field measurements and allometry were used to parameterise geometric models of crown volume for individual trees. Crown volume was integrated across each plot producing a histogram of canopy volume. The number of strata was subsequently estimated by visual inspection of the canopy volume histogram and compared to the ALS estimate.

Inventory data was gathered concurrent with the ALS acquisition at 27 plots across the three study areas. At each point a circular 0.04 ha plot was established following the VFMP protocol, measurements included tree species, diameter at breast height and tree height (DEPI, 2012). Additional height to live crown and crown radius measurements were taken at the *LOW* and *TCF* study areas, these were used to derive allometry between maximum crown height and height to live crown base and stem diameter at breast height and crown radius. The exception was for the *TCF* study area where allometrics for height to live crown were taken from Van Pelt et al. (2004).

For crown modelling, an ellipsoidal crown archetype was assumed for all species (Haverd et al., 2012). Crowns were modelled as a solid volume (i.e. assuming a Poisson distribution of foliage) and crown densities were weighted according to position (dominant or sub-dominant) and species. The exception to this was eucalypt trees where the disparity in foliage density between the outer crown shell and the interior was modelled by computing an inner ellipsoid of less dense foliage. Only crowns that were recorded as “live standing” were modelled (~86% of trees) and as the relative *xy* position of trees was not recorded, crown overlap and protrusion of beyond a plot boundary was not considered. Three plots were removed from validation due to a poor agreement between inventory and ALS measured maximum canopy height.

3.4. Results and Discussion

This chapter presents a new technique, which utilises ALS to estimate the number of canopy strata that characterise a forest plot, in an attempt to quantify canopy vertical structure. The new technique was successful in identifying local maxima in vegetation density, coincident with the upper canopy and shade tolerant canopy strata beneath, and therefore generating an estimate of Number of Strata (Figure 10). A metric of canopy vertical structure on its own may not be informative when used in an abstract comparison e.g. comparing a tropical rainforest and dry sclerophyll forest (Parker and Brown, 2000). Nevertheless, the evaluation of a spatially continuous layer of canopy vertical structure across a landscape (Figure 11) can identify patterns useful for understanding disturbance history (Angelo et al., 2010) or the distribution of biomass (Keith et al., 2009). It is also suggested the new method provides a candidate Essential Biodiversity Variable with which to characterise terrestrial habitat structure, as discussed by Pereira et al. (2013). Furthermore, this method offers a means to dynamically parameterise techniques such as mixture modelling or cluster analysis of canopy vertical arrangement (Jaskierniak et al., 2011; Zhang et al., 2011), where a prior estimate of the number of model components is required.

3.4.1. Methodology evaluation

Parameterising the α coefficient with a supervised classification and cross validation was successful ($kappa$ 0.64 – 0.67), particularly for plots where $NoS \leq 2$ (Figure 12). An improved classification ($kappa = 0.72$) can be obtained when plots are classified into a single and multistrata schema (Zimble et al., 2003; Maltamo et al., 2005), although this would ultimately reduce structural information in more complex forests. It is suggested the poor performance where $NoS > 2$ is due to the low number of ALS plots in the training data for of this cohort ($n = 5$).

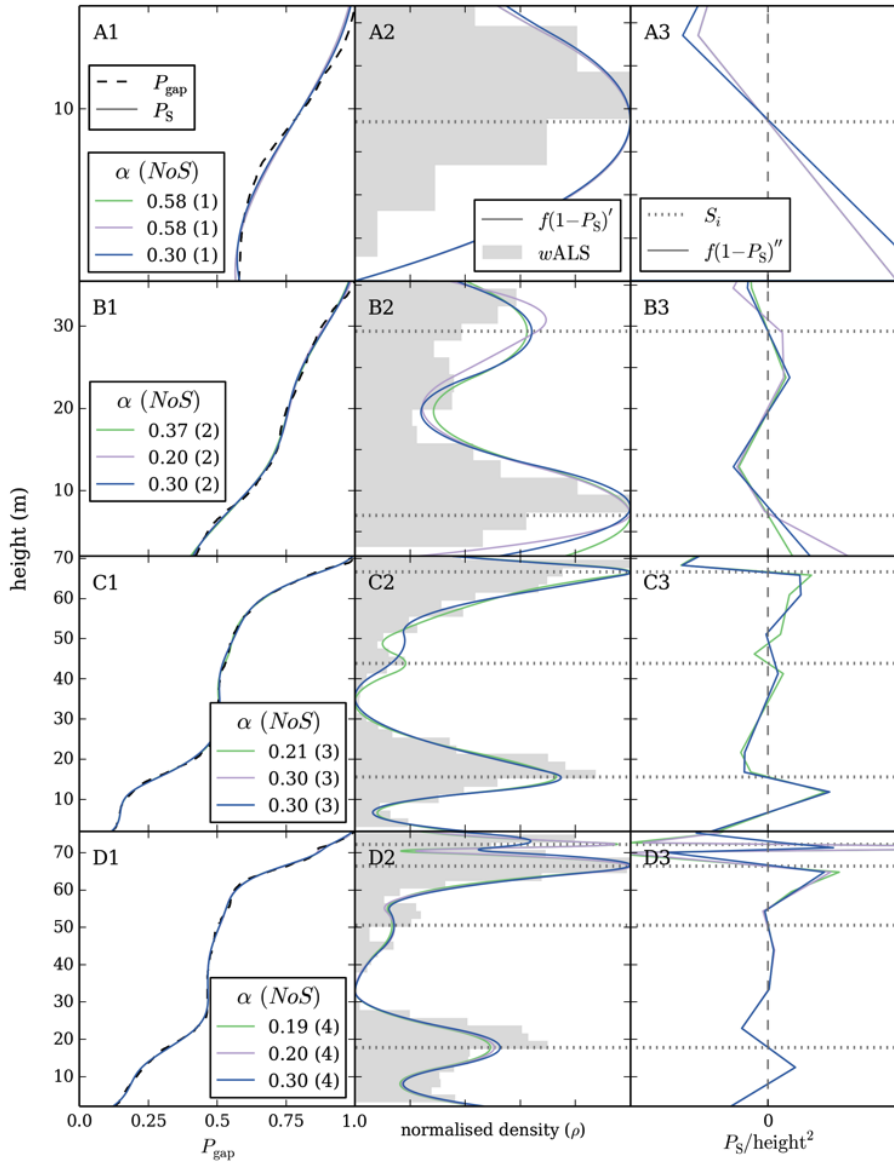


Figure 10. Canopy vertical structure from ALS derived P_{gap} for four structurally diverse forest plots [A - D] using forest type (green), study area (purple) and universal (blue) α . P_{gap} from a single bootstrap sample is transformed with a nonparametric cubic spline regression (P_s) [1]. “Apparent” canopy height profile (Lovell et al., 2003) from $f(x)'$, where $x = 1 - P_s$ [2]. ALS return density maxima identified as strata (S_i) from the positive zero-crossings of $f(x)'$ [3]. A comparison of $f(x)'$ and a histogram of weighted ALS return (wALS) frequency [2] illustrates the generalisation of the canopy height profile possible with this technique.

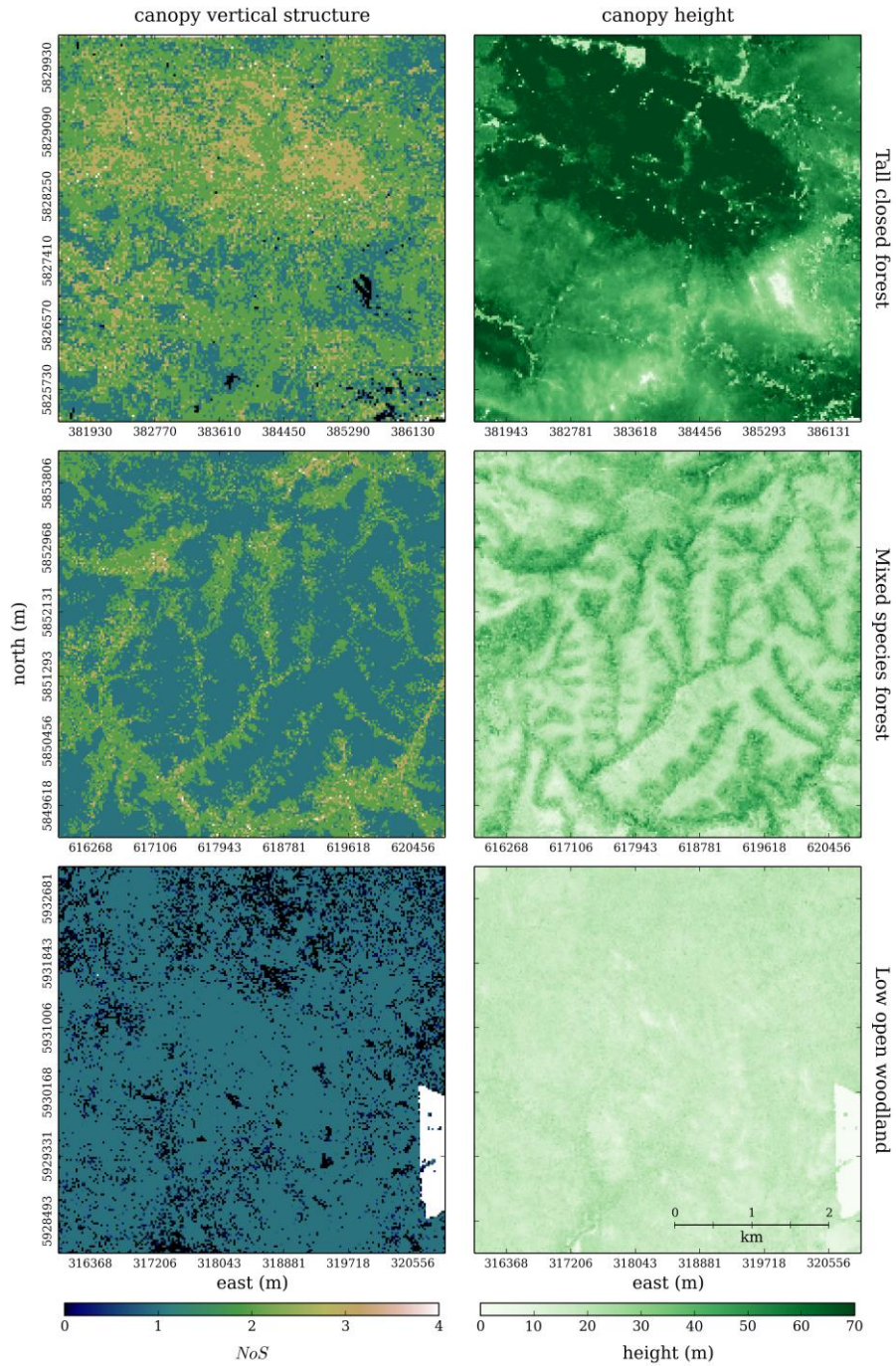


Figure 11. *NoS* and canopy height computed across three diversely forested study areas (**Error! Reference source not found.**) at a 30 m resolution. Map coordinates are Map Grid of Australia Zone 55.

Table 4. Forest type, study area and universal α . *Riparian forest* is found in both the *MSF* and *TCF* study areas. Too few plots were located in *Montane Damp Forest* to compute α .

Forest type		Study area	Universal
<i>Heathy Dry Forest</i>	0.58	<i>LOW</i>	0.58
<i>Box Ironbark Forest</i>	0.54		
<i>Damp Forest</i>	0.37	<i>MSF</i>	0.20
<i>Shrubby Dry Forest</i>	0.13		
<i>Riparian Forest</i>	0.19		0.30
<i>Montane Wet Forest</i>	0.31		
<i>Wet Forest</i>	0.21		
<i>Cool Temperate Rainforest</i>	0.26	<i>TCF</i>	0.30
<i>Montane Damp Forest</i>	-		

Forest type, study area and universal α (Table 4) returned similar P_{gap} curves and canopy profiles when applied to plot data (Figure 10). Furthermore there was no overall improvement in results when a forest type or site specific α was applied (Figure 12). This would indicate a universal α is sufficient to characterise vegetation structure across the three diversely forested study areas, this would suggest no prior assumption of forest structure or type is required (Figure 11). The requirement for forest type assumptions have been previously recognised as a limitation of ALS techniques for deriving forest structure, particularly P_{gap} , over heterogeneously forested areas (Hopkinson and Chasmer, 2009). A suggested reason for the general applicability is the technique generalises canopy structure, therefore for the majority of plots, α_s is relatively large and overlaps for different forest type (Figure 9). Furthermore, the method scales linearly with increasing canopy height (Figure 13) which maintains smoothing characteristics regardless of canopy height.

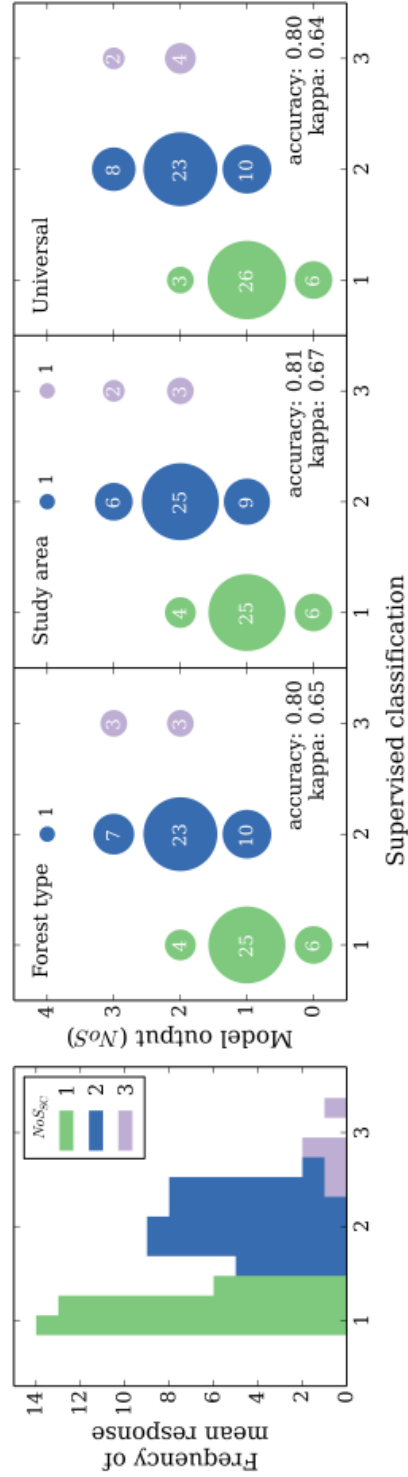


Figure 2. Results from the supervised classification of 82 plots. Frequency histogram of mean classification response (No_{Sc}) classified by mode response (i.e. consensus on number of strata) [A] and confusion matrix comparing supervised classification and model output [B].

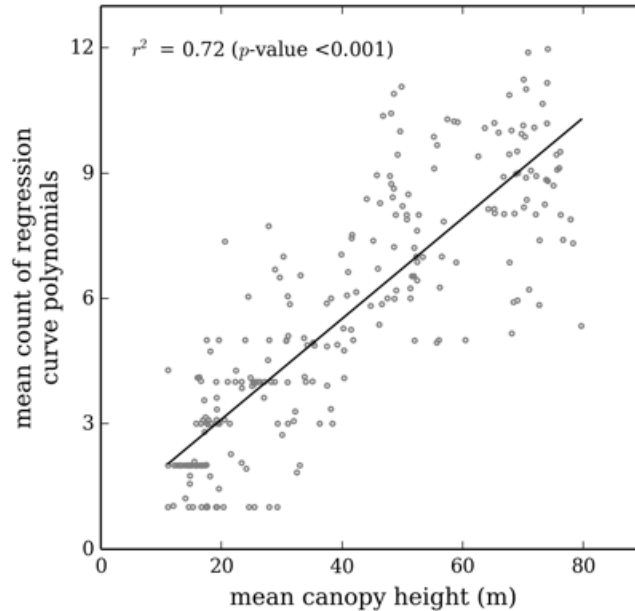


Figure 13. Linear regression of per plot mean height and mean number of spline polynomials. Mean values were derived from the bootstrap simulations.

The standard deviation of number of strata from the bootstrapped output was small for ~80% of plots (<0.4 strata). For the remaining plots, bootstrap iteration produces a sufficiently different canopy profile so that a range of *NoS* estimates were derived; typically ± 1 canopy strata but as much as ± 2 strata in more structurally complex forests. A number of reasons are suggested for a plot returning a range of *NoS* values, these include; (i) under sampling of the sub-canopy in the original acquisition which caused greater variability in simulations, (ii) the influence of partially represented trees on a plot voxel boundary and (iii) the partial convergence of large tree crowns in the dominant canopy or of the dominant and shade tolerant strata. It is suggested that bootstrapping of point cloud configuration to reduce uncertainty in structure estimates has wider application, particularly where acquisition pulse density is low (Figure 14).

Gaussian smoothing is commonly applied to remove noise from canopy height profiles derived from large footprint full-waveform instruments (Hofton et al., 2000). Here the filtering of P_{gap} was opted for instead of smoothing the canopy height profile, as this was more robust to non-

optimal regression coefficients (i.e. α_s was larger). Application of Gaussian smoothing produced a monotonically decreasing P_s , however it resulted in far more false-positives in $f(x)''$. There are a number of techniques to enforce monotonic behaviour when fitting splines (e.g. He and Shi, 1998) which could be explored further if derivation of the canopy height profile is required. Other techniques to identify canopy strata include Gaussian decomposition, however over large areas the iterative nature of this technique can be computationally expensive (Hancock et al., 2015). Here, determination of an appropriate smoothing coefficient was sufficient to identify canopy scale features in $f(x)''$, this negated the requirement for an iterative process and therefore improved algorithm efficiency.

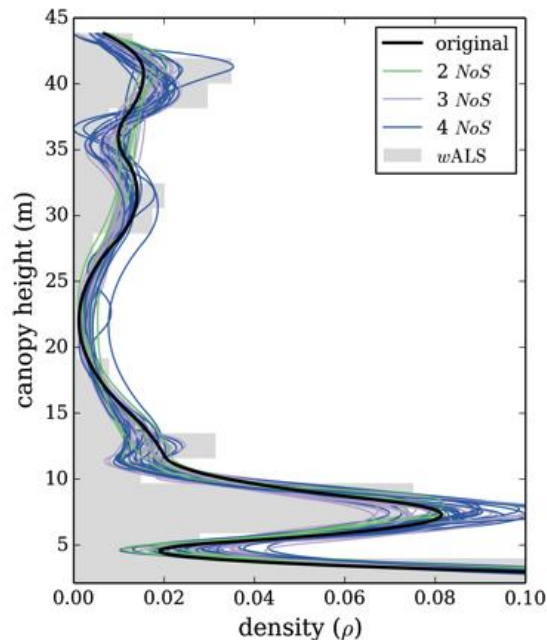


Figure 14. A comparison of canopy height profiles derived using a point cloud with a pulse density of 10 pl m^{-2} (black) and a bootstrap ($N=50$) of a point cloud thinned to 1 pl m^{-2} (see Chapter 4). *NoS* equals 3 for the original point cloud, 4 for the thinned point cloud and 3.12 for the bootstrap mean of the thinned dataset. Also included as a comparison is a histogram of weighted ALS returns (*wALS*).

3.4.2. Validation results

A comparison at 24 plots of ALS and Geometric Crown Volume Modelling (GCVM) derived *NoS* returned a RMSE of 0.69 strata (Table 5). Generally, ALS tended to underestimate *NoS* when compared to GCVM (Table 5). Improvements were made across all study areas when the ALS plot radius was increased to 15 m (RMSE = 0.41). This would suggest that individual trees, particularly large trees both inside and outside the inventory plot, have a significant influence on the canopy height profile at the VFMP plot radius (Figure 15). Increasing the ALS plot radius integrates a larger number of trees into the ALS plot voxel and therefore provides a more robust estimate. Differences in curve shape are evident between the two ALS plot sizes (11.8 m and 15 m radii), where the larger ALS plot produces a more generalised curve (e.g. Figure 15 *TCF1* and *TCF2*). The poorest results were seen at the *TCF* study area (Table 5) and this was attributed to the increasing difficulty (with both techniques) to accurately measure forest structure. For example, GCVM requires accurate measurement of crown dimensions, which in turn requires the ability to accurately sight crown dimensions. This ability can be somewhat diminished in more structurally complex forests such as that found in the *TCF* plots (Drake et al., 2002; Lee et al., 2004).

A qualitative comparison of pseudo-waveforms generated using ALS and GCVM reveals a generally good agreement between techniques, for example the similar location of canopy strata modes and layer boundaries (Figure 15). It is clear where the dominant and shade tolerant strata have converged (e.g. Figure 15 *MSF2*), the ALS generated canopy height profile may only identify a single canopy; this could be mitigated for by decreasing α . In general, ALS tended to underestimate dominant canopy volume and overestimate the density of the shade tolerant strata. A suggested reason for this is the equal weighting applied to ALS returns from the eucalypt and non-eucalypt strata. This may misrepresent the proportion of backscattered energy as a function of plant area volume density from the respective strata, caused by the different leaf projection functions of the eucalypt and non-eucalypt strata (Ross, 1981). This results in eucalypt crowns seeming more permeable when viewed from nadir (Lovell et al., 2003).

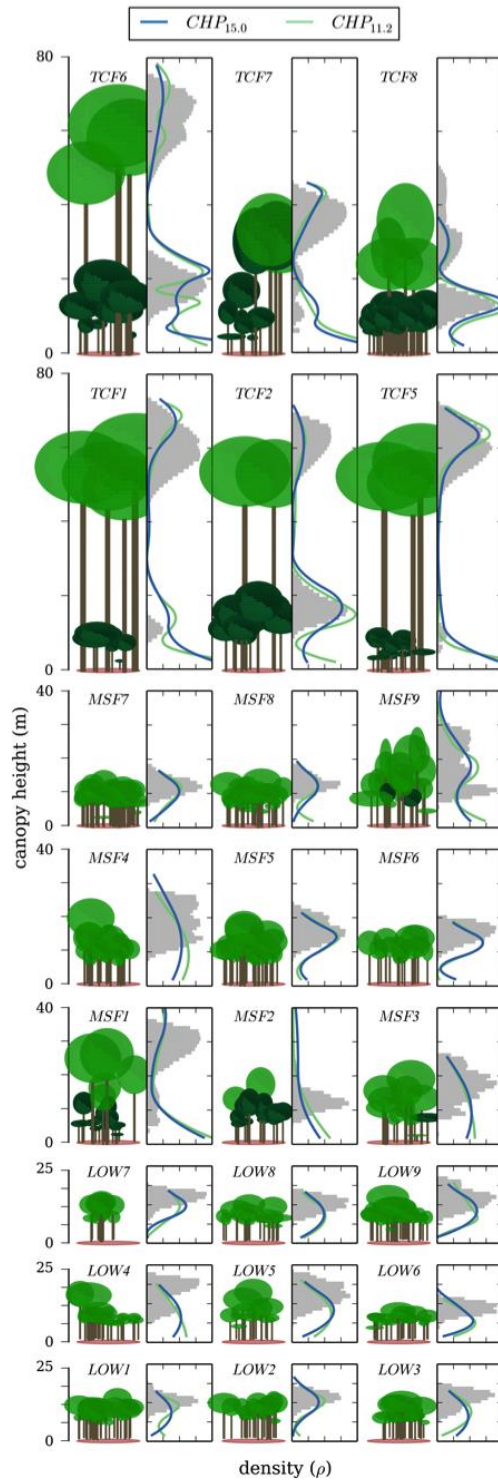


Figure 15. A comparison of 24 canopy height profiles (CHP) generated with GCV and ALS. Light green and dark green crowns represent eucalypt and non-eucalypt species respectively. Trees are randomly placed within the scene as individual stem location was not recorded. Grey histograms were generated using GCV and green and blue CHPs were derived from ALS plots with a radius of 11.8 m and 15 m respectively

Table 5. Comparison of GCMV and ALS derived estimates of *NoS*.

Study site	GCMV – ALS _{radius} (<i>NoS</i>)												
	n	Radius = 11.8				Radius = 15				Radius = 20			
		min	max	μ	RMSE	min	max	μ	RMSE	min	max	μ	RMSE
ALL	24	-2.0	1.3	-0.2	0.69	-1.0	1.0	-0.1	0.41	-1.0	1.0	-0.2	0.47
LOW	9	-2.0	0.0	-0.2	0.66	-1.0	0.1	-0.1	0.33	-1.0	0.0	-0.1	0.33
MSF	9	-1.0	0.0	-0.4	0.60	-1.0	0.0	-0.2	0.36	-1.0	0.0	-0.3	0.52
TCF	6	-1.0	1.3	0.1	0.83	-1.0	1.0	-0.1	0.56	-1.0	1.0	-0.26	0.67

3.5. Canopy vertical structure as an independent metric

Of the 239 plots extracted from across the three study areas, 57% were found to have a single canopy, 31% by two canopy strata and 6% had three or more strata. Eleven plots returned a *NoS* value of <0.5 (i.e. no canopy), these were characterised by a very sparse canopy (canopy cover <0.2) and a single and negatively skewed regression spline; this would indicate the understorey or shrub strata was significantly more dense than the dominant canopy.

NoS was only moderately correlated with other ALS metrics of canopy structure ($r^2 < 0.4$, $p < 0.001$), suggesting the method offers new information on ALS derived vegetation structure. For example, although the probability of an increasing number of strata increases with canopy height (Figure 16A) and indeed only taller forests can accommodate multiple strata (Brokaw and Lent, 1999); yet canopy height was a relatively poor predictor of *NoS* (RMSE ~ 1 strata), particularly when distinguishing plots with between single and two strata plots (Figure 16A). Utilising a proxy metric, such as canopy height, to predict *NoS* may seem appealing over large areas. However this may mask forested areas that have less common and regionally significant characteristics, for example high biomass multi-strata forest caused by partial stand replacement (Keith et al., 2009).

Return height coefficient of variation (C_v) has been used as a metric of canopy vertical structure heterogeneity (e.g. Bolton et al., 2013), however

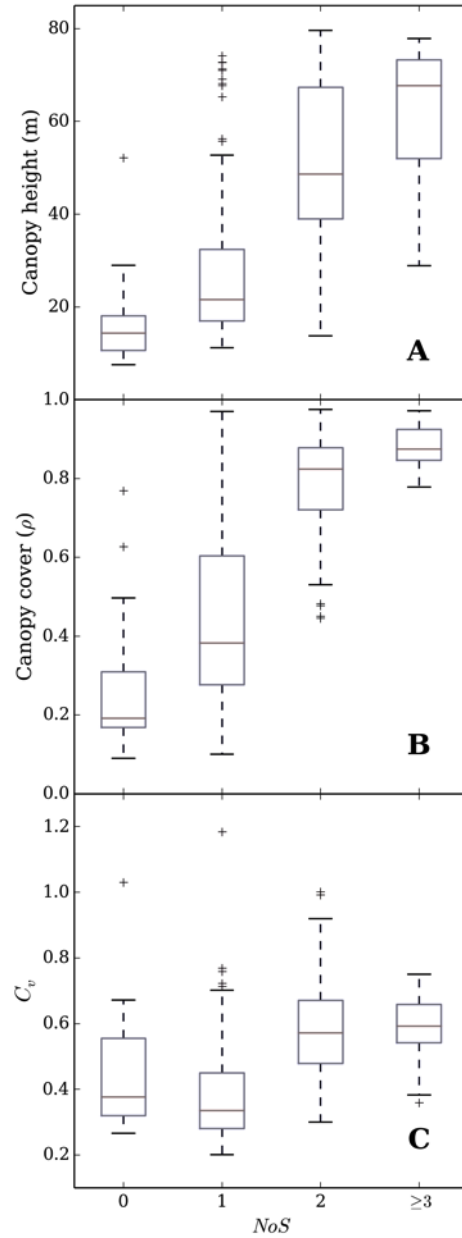


Figure 16. Boxplots summarising the statistical association between *NoS* (rounded to integer values) and canopy height [A], canopy cover [B] and return height coefficient of variation [C]. Boxplot whiskers represent the 95th percentile.

here C_v explains the least variance in NoS ($r^2 = 0.15$, $p < 0.001$) and there is no association between C_v and increasing NoS (Figure 16C). A suggested reason for this is C_v is a measure of relative return dispersion through the canopy and not an explicit metric of canopy vertical structure. Therefore different canopy structure types or capture specifications (e.g. scan angle) may return similar C_v values. For example a plot with a single canopy strata and a canopy height of < 30 m shares a similar C_v value to a plot that has three or more strata and a canopy height of > 50 m (Figure 17).

3.6. Conclusion

This chapter presents a new method for quantifying forest canopy vertical structure, where discrete return ALS is used to derive an estimate of the number of canopy strata. This investigation aimed to develop a technique that could be applied across forested landscapes where no *a priori* assumption of forest structure is required. The proposed technique is well suited for this purpose as the α coefficient used in the cubic spline transformation did not require a specific calibration for forest type or study area. This was demonstrated in forest systems ranging from short woodland with a discontinuous single canopy to tall and structurally complex temperate rainforests. We suggest this method offers additional information for ALS derived vegetation structure that, along with a metric

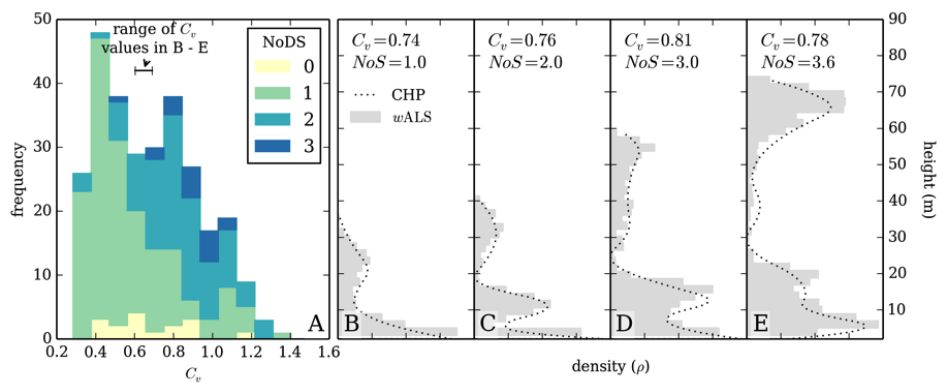


Figure 17. Comparison of coefficient of variation of return height (C_v) and NoS . Cumulative histogram of C_v where NoS are classified into integers for 239 plots [A]. Canopy height profiles for plots with increasing NoS and similar C_v values [B – E] where canopy height profiles are represented by weighted ALS returns ($wALS$) and pseudo-waveform generated from $f(x)'$ where $x = 1 - P_s$.

of canopy height and cover, it could be used as a fundamental descriptor of forest structure (Lefsky et al., 2005b). Furthermore, the technique could also be utilised as an Ecological Biodiversity Variable to characterise habitat structure over large areas (Pereira et al., 2013).

4. Chapter 4

Understanding the effects of ALS pulse density for metric retrieval across diverse forest types[‡]

[‡] This chapter is based on:

Wilkes, P., Jones, S.D., Suarez, L., Haywood, A., Soto-Berelov, M., Mellor, A., Woodgate, W., Skidmore, A., (2015). Understanding the effects of ALS pulse density for metric retrieval across diverse forest types. *Photogrammetric Engineering and Remote Sensing* 81(8), pp. 625 – 635.

4.1. Introduction

The accurate and timely retrieval of vegetation structure metrics is a key component of vegetation management, monitoring and reporting activities, and ecosystem modelling by land management agencies and forest scientists. For example, canopy height is routinely gathered by land management agencies around the world e.g. *Forest Inventory and Analysis* in the US, the Canadian *National Forest Inventory* and the *Victorian Forest Monitoring Programme* in Australia. This information is needed for the assessment of vegetation condition to fulfil statutory and non-legislative reporting obligations such as that agreed by the Santiago Declaration (Miles, 2002). Canopy height is also widely used by forest scientists as a proxy to estimate forest biomass (Drake et al., 2002; Asner et al., 2010) and it is an Essential Climate Variable (Sessa, 2009).

Over the past decade, Light Detecting and Ranging (LiDAR) and in particular small-footprint Airborne Laser Scanner (ALS) systems have progressed from an experimental technique to an operational tool for area based (e.g. plot or stand scale mapping unit) attribution of vegetation structure (Wulder et al., 2012a, 2012b). ALS allows for synoptic capture of large areas of the landscape where remoteness or terrain complexity complicates and increases the cost of establishing inventory plots or where variance in structure is not captured with traditional sampling (McRoberts and Tomppo, 2007; Mora et al., 2013). Furthermore, ALS is now recognised by some authors as a more accurate method of directly measuring vegetation attributes as compared to traditional forest inventory methods (Holmgren and Jonsson, 2004; Maltamo et al., 2006; Magnusson et al., 2007).

Acquisitions over very large areas that transect regional to continental extents are becoming viable as the ability to capture and process large volumes of data improves. For example, the Canadian Forest Service captured data along a ~25,000 km transect (Wulder et al., 2012a) and the Department of Environment, Land, Water and Planning (DELWP) captured LiDAR along ~27,000 km of riparian corridors in Victoria, Australia (Quadros et al., 2011). However, the cost of large-area ALS acquisitions remains significant. There are a number of possibilities for reducing acquisition expenditure, such as using a sample based approach where

LiDAR “plots” are targeted to capture the variance in forest structure (Wulder et al., 2012a) or optimising acquisition parameters, such as sampling frequency, to capture vegetation type and terrain conditions. ALS sampling frequency is referred to as “pulse density” and is defined as the number of emitted laser pulses that intercept a given area of open ground (Evans et al., 2009); pulse density is expressed here as pulses per m² (pl m⁻²). Pulse density is determined by the pulse repetition frequency (PRF) of the instrument and acquisition parameters such as aircraft flying height and aircraft speed. Although a number of factors mean PRF and flying height or speed are not independent, at a constant PRF decreasing pulse density (on the ground) is achieved by increasing flying height and/or aircraft speed (Baltsavias, 1999).

The primary benefit of acquiring ALS at a greater altitude or aircraft speed, and therefore a lower pulse density, is decreased project costs (Jakubowski et al., 2013), although this is potentially offset by a reduction in the accuracy of metric retrieval. For example, decreasing pulse density reduces the probability of intercepting crown apices (Goodwin et al., 2006) or over dense vegetation, decreases the probability of a ground return (Takahashi et al., 2008). Other capture specifications altered by increasing flying height include increasing laser footprint size which integrates the intercepted energy over a larger area that in turn reduces instantaneous laser power at the receiver. However these factors are considered less significant than sampling frequency (Goodwin et al., 2006; van Leeuwen and Nieuwenhuis, 2010). Although there are a number of studies that assess the impact of reducing pulse density on the accuracy of metric retrieval, these are generally limited to a single forest type or small capture area (Goodwin et al., 2006; Magnusson et al., 2007; Gobakken and Næsset, 2008; Takahashi et al., 2008; Næsset, 2009; Jakubowski et al., 2013). Previous studies have concluded that for area based attribution of vegetation structure, successful analysis can be achieved with a pulse density between 0.5 – 1 pl m⁻², even in mixed species multi-strata forests (Hayashi et al., 2014). In this study we extend previous work by assessing and comparing the direct retrieval of vegetation structure metrics from a broad range of forest types and topographies across continental Australia; from open savanna woodland to dense tropical forests (Figure 18). The focus of this investigation is to assess the error and variance of vegetation structure metrics when captured at decreasing pulse densities, as well as to

estimate intra-plot variability and the reproducibility of structural measurements from repeat capture. To facilitate metric estimation accuracy, a new technique for the systematic thinning of point clouds is also introduced.

4.2. Method

4.2.1. Study area and data capture

Six study areas, located across continental Australia (Figure 18), were selected from the Terrestrial Ecosystem Research Network AusCover Facility. Study areas were selected to capture a broad range of forest

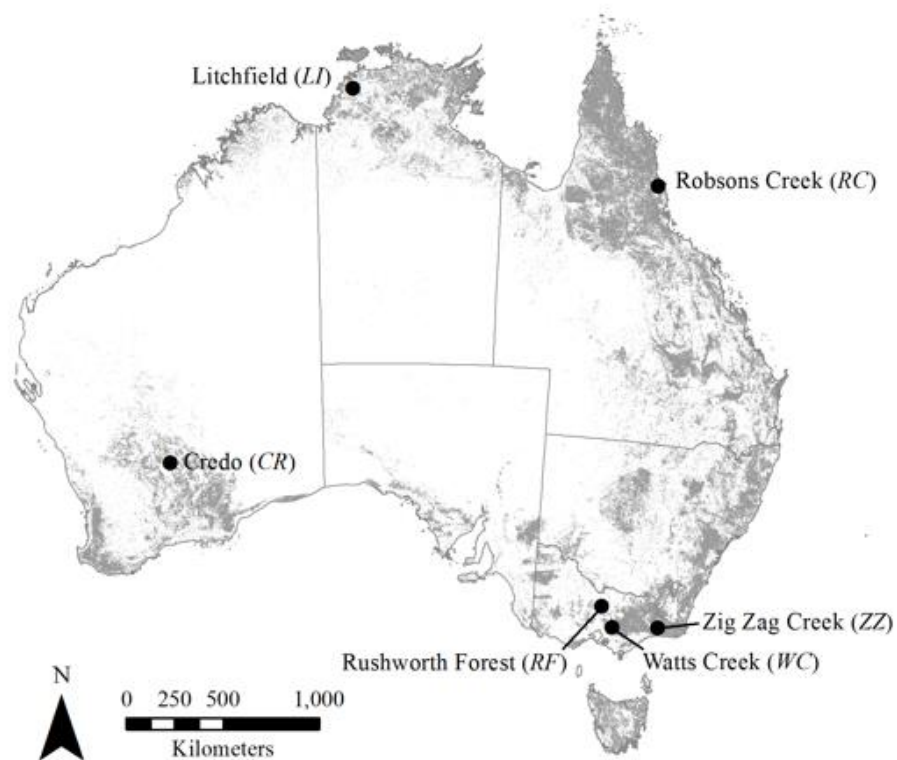


Figure 18. Map displaying the location of the six TERN sites used in this investigation, shaded areas indicate forest extent (Montreal Process Implementation Group for Australia, 2013).

structure and types (**Error! Reference source not found.** and Figure 19). ALS data (TERN/AusCover 2012) was acquired between April 2012 – June 2013 for all areas by a single provider (Airborne Research Australia) utilising the same capture specifications (Table 6) which facilitates comparison between study areas. Flight lines followed a regular north-south pattern spaced 125 m apart. This, in conjunction with a maximum scan angle of $\pm 22.5^\circ$, resulted in an approximate swath overlap of 50%. Owing to steep terrain additional flight lines were required at two study areas (*TCF* and *NVF*). A minimum pulse density of 10 pl m^{-2} was stipulated prior to capture. By maintaining swath overlap in post-processing acquisition pulse density is approximately doubled, though variation in density still occurred. Although capture over a large area at this density with a 50% swath overlap would not be considered operational for a land management agency (Wulder et al., 2012b), oversampling allows a methodical simulation of various pulse densities.

For each study area, 50 plot centres were randomly located across the 5 km x 5 km capture footprint. To ensure spatial independence of extracted plots, plot centres were located at a minimum distance of 250 m apart. From each plot centre, ALS data for a circular plot with a 25 m radius was extracted. To decrease the risk of duplication of returns in a thinned dataset, extracted plots with a pulse density of $<10 \text{ pl m}^{-2}$ were rejected and number of first-returns (i.e. points with a *return number* metadata value of 1) per plot and dividing by plot area (Jakubowski et al., 2013).

Table 6. Description and location of study areas sorted by mean canopy height. See Table 2 for additional study areas.

Site	Coordinates	Elevation	Mean slope	Description	Canopy height	Mean annual rainfall [^]
Notophyll Vine Forest (<i>NVF</i>)	17° 6' 20" 145° 37' 16"	1160 – 700 m	24°	Tropical rainforest with a tall canopy. High species diversity.	25 – 40 m	1890 mm
Savanna (<i>SAV</i>)	13° 10' 39" 130° 47' 23"	210 – 230 m	3°	Sparse open eucalypt forest with isolated patches of monsoon rainforest.	10 – 25 m	1370 mm
Great Western Woodlands (<i>GWV</i>)	30° 11' 22" 120° 39' 17"	440 – 480 m	2°	Open woodland inter-dispersed with open, treeless areas. Small shrub layer prevalent.	5 – 25 m	260 mm

[^] Rainfall data from Hijmans et al. (2005)

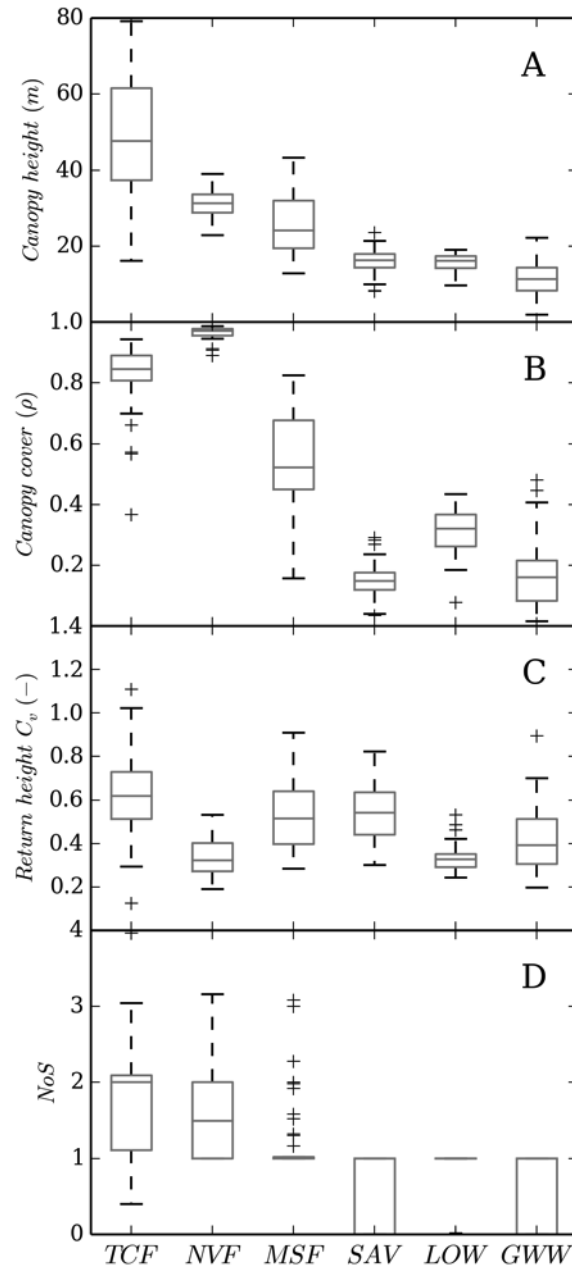


Figure 19. Descriptive statistics for 4 metrics of canopy structure across 6 study areas. 95th percentile of canopy height [A]; proportion of cover calculated as $1 - P_{\text{gap}}(z)$ where z equals 1 m [B]; coefficient of variation (C_v) for non-ground return height[C] and NoS [D].

Table 6. Flight and sensor specifications for the ALS acquisitions.

Specifications	
<i>Capture specifications</i>	
Flying height	300 – 600 m agl
Target pulse density (excl. overlap)	10 pulses m ⁻²
Absolute vertical accuracy	<0.15 m
Absolute horizontal accuracy	<0.15 m
Mean footprint diameter	0.15 m
<i>Instrument specifications</i>	
Instrument	Riegl LMS-Q560 laser scanner (Horn, Austria)
Operating wavelength	1550 nm
Beam divergence	0.5 mrad
Max off-nadir scan angle	±22.5°
Outgoing pulse rate	240 kHz

4.2.2. Data processing

There are a number of existing studies that assess the accuracy of forest metric retrieval from point clouds of different densities. Ideally, analysis of different pulse densities would be undertaken on datasets captured at different flying altitudes or aircraft speeds (Goodwin et al., 2006; Thomas et al., 2006; Magnusson et al., 2007; Morsdorf et al., 2008; Takahashi et al., 2008). This permits analysis of additional variables altered by changing capture specifications such as instantaneous laser pulse power and laser footprint size. However, capture at multiple altitudes and aircraft speeds is limited by cost, particularly over large or discontinuous study areas such as in this study, and alternative modelling methods are required. A number of authors have decreased the number of *points* in a dataset to match a required point density (Maltamo et al., 2006; Tesfamichael et al., 2010; Watt et al., 2013); however this does not necessarily replicate the reduction in *pulse* density caused by change in altitude or aircraft speed if simulating anything other than a *first-return* dataset (Jakubowski et al., 2013). The majority of techniques that simulate a reduction in pulse density do so by superimposing a regular grid over the study area of a specified spatial resolution to attain required pulse densities, returns are then randomly selected from within each voxel (Gobakken and Næsset, 2008; Næsset, 2009; Korhonen et al., 2011; Jakubowski et al., 2013; Hansen

et al., 2015). However, application of this technique may not replicate the regular scan pattern in which data is collected (Baltsavias, 1999), particularly when simulating low pulse densities. Other techniques include the removal of alternate pulses and scan lines (Treitz et al., 2012), stipulating a minimum horizontal distance between returns (Magnusson et al., 2007) or systematically thinning a dataset utilising GPS time (Khosravipour et al., 2014); albeit these techniques are only suitable for generating single or first-return datasets.

Here we introduce a new technique that can (a) systematically sample the original dataset to allow analysis of sampling variance and acquisition reproducibility, and (b) simulate multi-return capture (i.e. as opposed to *first* or *first-and-last* capture) that state-of-the-art small-footprint instruments can produce. Computations were carried out with the ForestLAS Python module except where stipulated. For the 250 extracted plots, nine different plotwise realisations were simulated at 6 different pulse densities (Table 3). This resulted in a total of 15,500 simulations.

As illustrated in Figure 20, thinning was achieved by superimposing a point-grid over the area at a resolution required to achieve a target pulse density (Table 3). First-returns with the shortest Cartesian distance to the grid point were selected and the attributes retained i.e. selected points were not snapped. A search window was utilised when selecting points to filter points that may lie closer to an adjacent grid point (Figure 20), this was optimised at $2/3$ the point-grid resolution to minimise duplication of returns whilst maintaining pulse density (Figure 20A). For each first-return selected, the *number of returns* metadata value (X) was extracted.

Table 3. Pulse density and required pulse spacing.

Pulse density (pulse m ⁻²)	Pulse spacing (m)
0.05	4.47
0.1	3.16
0.5	1.41
1	1.00
2	0.71
4	0.5
10	0.32

An additional X points with a corresponding *return number* metadata value were subsequently selected. For example, if a selected first-return had a *number of returns* value of Y then $[R_2, R_3, \dots R_Y]$ additional returns were selected where the subscript value refers to the *return number*. Additional returns were again selected by shortest Cartesian distance from the grid point. Maximum distance for additional returns was likewise restricted to a search voxel where the extent (SW_e) was determined from an estimate of canopy height (z_{max}) and an assumed maximum scan angle (θ) of 5° (EQ. 14) (Figure 20A). The extracted dataset therefore simulates a near nadir acquisition with a regular scan pattern (Baltasvias, 1999), this standardises simulated capture specifications aiding comparison between plots and study areas.

$$SW_e = 2(z_{max} \cdot \tan\theta)$$

EQ. 14

For the simulated datasets ground points were identified and used to compute a Triangulated Irregular Network (TIN), from which height relative to ground was calculated for all returns. Computing a TIN and relative height for each simulated dataset was necessary so that miscalculation of the ground surface can be accounted for (Magnusson et al., 2007). Identification of ground returns and relative return height calculation was computed using default settings with the *lasground* and *lasheight* tools respectively from the LAStools software package (version 130225) (Isenburg, 2012).

A plot with a radius of 11.8 m was clipped from each thinned and ground normalised dataset to replicate standard forest inventory plot dimensions (Victorian Department of Sustainability and Environment, 2012b). Thinning and TIN creation were computed for the larger dataset to ensure all ground returns within the clipped area had a full neighbourhood from which to generate a TIN. Furthermore, scan pattern towards the edges of the larger dataset became irregular owing to the circular plot shape; clipping removed this effect from the smaller plots.

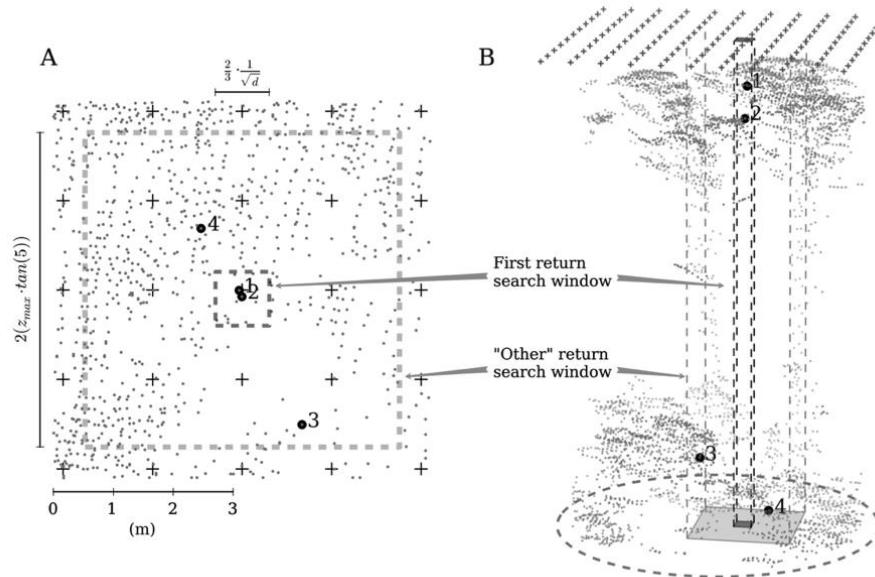


Figure 20. Visualisation describing the point cloud thinning technique using a target pulse density of 0.5 pl m^{-2} as an example. [A] A single first return is selected for each grid point (+) where the ALS return with the lowest Cartesian distance is selected. Selection is restricted to a search-window around each grid point where the search window dimensions are determined by the desired pulse density (d). Selection of X further returns, determined by the *number of returns* metadata field (in this case $X = 4$), are again chosen by their proximity to the grid point. A search-window restricts the maximum distance of “other” returns where the extent is determined by an estimate of the maximum height of the forest ($z_{\max} = 40 \text{ m}$) and an assumed scan angle $\leq 5^\circ$, in this way a nadir acquisition is simulated. NB. Points have been removed from outside the plot boundary to enhance visualisation [B], when point clouds were thinned points from outside the plot boundary could be selected.

Owing to the density of the original dataset and the systematic way in which simulated datasets were constructed, additional realisations could be computed from the original dataset with minimal duplication of returns between realisations. Therefore nine simulated datasets were generated for each plot, where the origin of the sample point grid was offset recursively by $1/3$ of the sampling resolution in both the x and y direction. From the nine realisations, a robust set of descriptive statistics were generated and compared to a value derived from a high pulse density

dataset (see below). Generating different plotwise realisations also allowed the repeatability of ALS capture to be assessed (Bater et al., 2011).

4.2.3. Metrics

Forest structure could be characterised by three categories of primary descriptor: (1) canopy height, (2) canopy cover and (3) canopy vertical structure or canopy vertical structure (Lefsky et al., 2005a; Kane et al., 2010b). Three metrics representing each of these categories were selected: the 95th percentile of non-ground return height as an analogue of dominant canopy height; canopy cover was estimated using $1 - P_{\text{gap}}(z)$ where z equals 1 m; and coefficient of variation (C_v) of return height as a metric of vertical canopy structure (Zimble et al., 2003; Kane et al., 2010b; Bolton et al., 2013). Vertically resolved gap probability $P_{\text{gap}}(z)$ was computed using ALS returns weighted by the return number, as Armston et al. (2013) concluded this produced a more accurate estimate of P_{gap} than a frequency based estimate. As techniques for describing vertical structure are less well described, a metric that characterises vertical canopy arrangement by estimating the Number of canopy Strata (NoS) from the second derivative of transformed P_{gap} was also included. Figure 19 compares the range of values for each of the four metrics at each study area.

To ascertain the benefit of increasing pulse density when characterising forest structure, metrics derived using thinned datasets were compared to a dataset with a common density of 10 pl m⁻² (Næsset, 2009). Mean values were computed from the nine plotwise realisations at six pulse densities from which difference and root mean square difference were calculated. Coefficient of determination values were also calculated where the thinned datasets were the independent variable and the dependent variable was the metric computed from the high pulse density dataset. Furthermore, descriptive statistics for the nine plotwise realisations were calculated to ascertain intra-plot variability.

4.3. Results

Airborne Laser Scanning (ALS) data for three hundred plots were extracted across six vegetation systems that characterise diverse forested landscapes in Australia. For each plot, vegetation metrics representing three primary

descriptors of vegetation structure (Lefsky et al., 2005a; Kane et al., 2010b) were computed at to six different simulated pulse densities (0.05 – 4 pl m⁻²). Results for each descriptor are presented below.

4.3.1. Canopy height

Differences in canopy height estimates are low for high pulse densities and increase with decreasing sampling frequency (Figure 21A). For example, at pulse densities ≥ 0.5 pl m⁻², Root Mean Square Difference (RMSD) is < 0.5 m for the *LOW*, *GWW*, *MSF* and *SAV* study areas, < 1 m at *TCF* and < 1.5 m at *NVF*. At the structurally simple and homogenous *LOW* study area (Figure 19) where pulse density was simulated at 4 pl m⁻², RMSD in height estimation is less than the quoted vertical accuracy of the bare earth ground surface as stated by the data provider. When pulse density < 0.5 pl m⁻², RMSD is generally less than 2 m, however for the *NVF* study area error can be > 10 m (Figure 21A). Furthermore, at low pulse densities differences in height estimation is not necessarily a simple systematic offset, this is highlighted by a decrease in coefficient of determination values with decreasing pulse density (Figure 21A and 22A). As noted in previous studies, canopy height estimates decrease with decreasing pulse density as a result of the incomplete sampling of crown apexes (Goodwin et al., 2006; Morsdorf et al., 2008). However, in this investigation the largest errors occurred as a result of the misidentification of the ground surface, this resulted in a poor TIN model from which to calculate vegetation height. Intra-plot variance in canopy height estimation increases with decreasing pulse density for all study areas (Figure 23A). Standard deviation in canopy height estimates derived from different realisations of the same dataset can exceed 4 m. Calculation of canopy height Coefficient of Variation (C_v) normalises for the large differences in canopy height when comparing plots and study areas (Figure 19A). Variance in C_v of canopy height is greatest for plots in the *SAV* and *CR* study areas, particularly at low pulse densities (Figure 23B). This is attributed to clumped vegetation and low tree density where successive realisations capture a significantly different proportion of vegetation.

Understanding the effects of ALS pulse density for metric retrieval

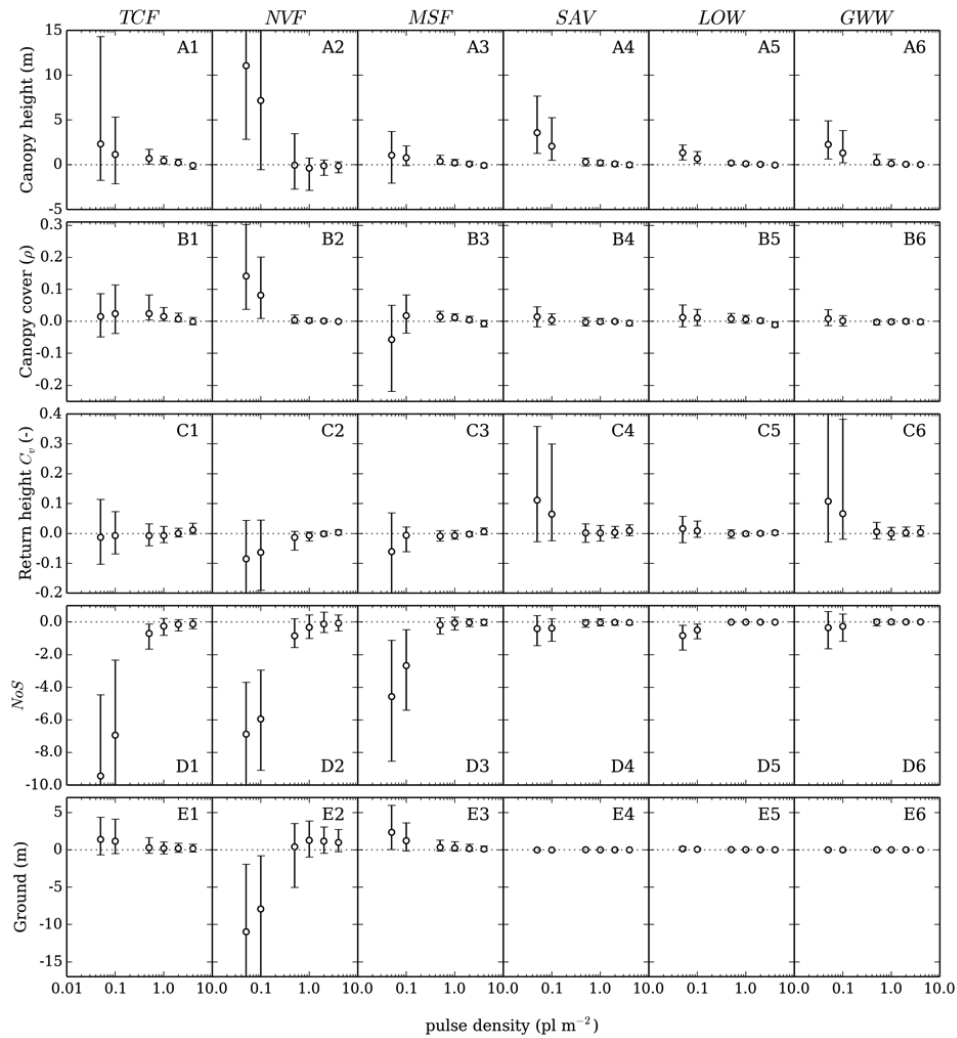


Figure 21. Difference between the plotwise mean value calculated at 6 pulse densities (0.05 - 4 pl m^{-2}) and the value calculated at 10 pl m^{-2} . Four metrics are used to summarise canopy structure; [A] canopy height, [B] canopy cover, [C] coefficient of variation (C_v) of return height and [D] NoS. Additionally difference in ground height is shown [E]. Error bars represent the 95th percentile range.

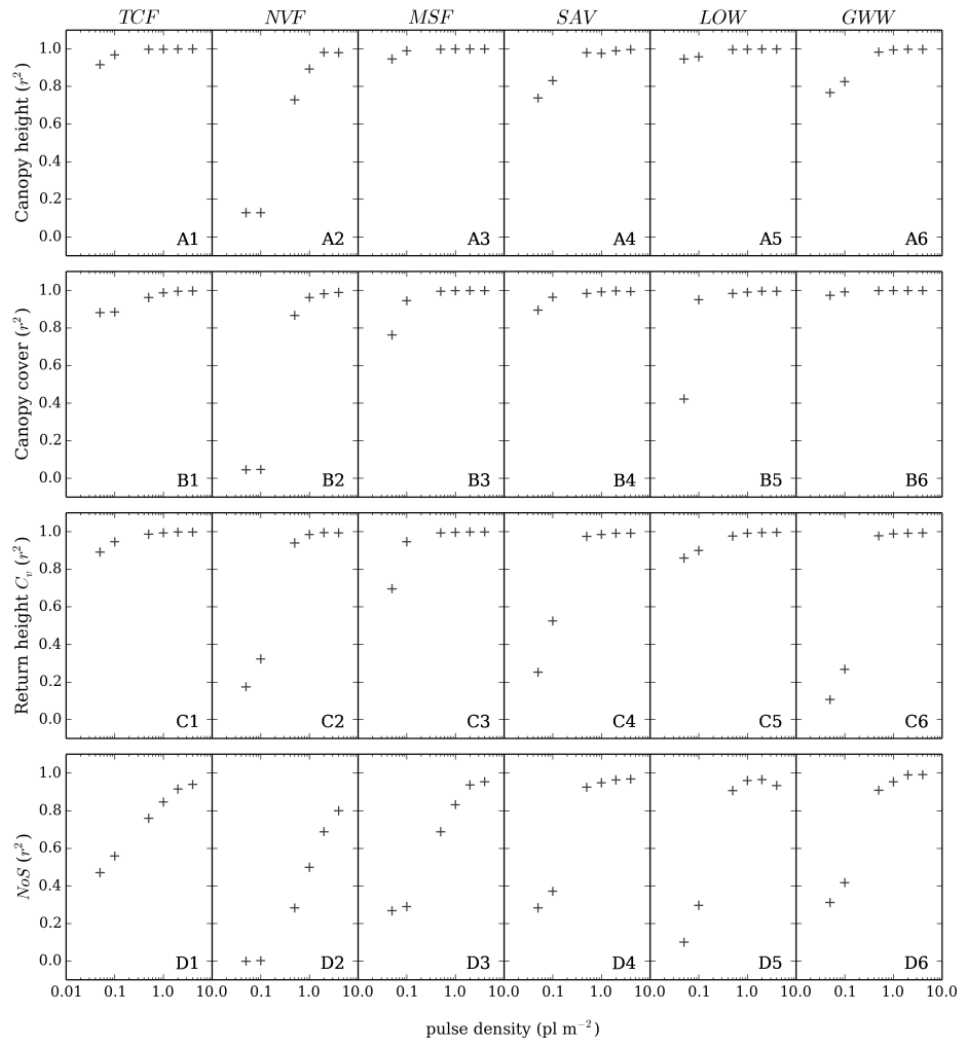


Figure 22. Coefficient of determination where the independent variable was [A] canopy height, [B] canopy cover, [C] return height coefficient of variation (C_v) and [D] NoS at 6 different pulse densities (0.05 – 4 pl m^{-2}). The dependent variable was the metric derived from a pulse density of 10 pl m^{-2} .

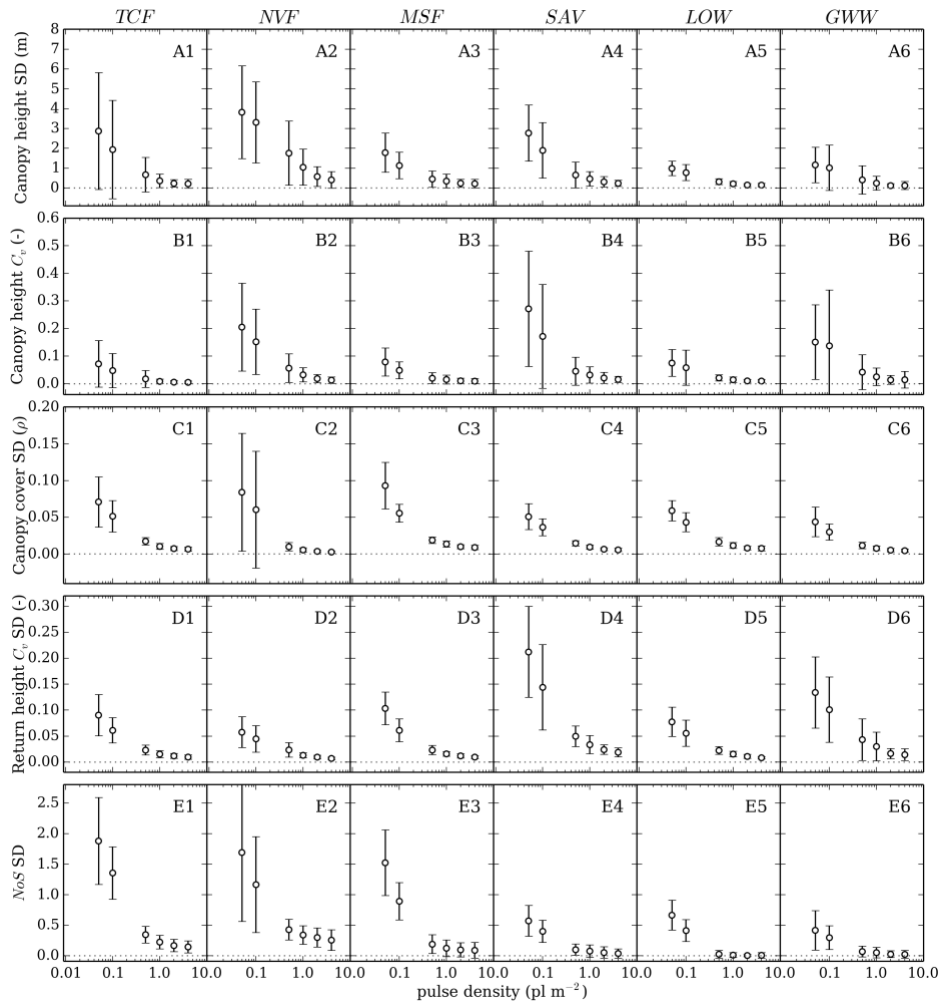


Figure 23. Mean intra-plot variance for vegetation structure metrics where variance is calculated as the standard deviation of 9 realisations drawn systematically from the original dataset. Four metrics were computed; [A] canopy height, [C] canopy cover, [D] coefficient of variation (C_v) of return height and [E] NoS . Additionally, [B] C_v of canopy height was included, which normalises for canopy height when calculating variance. Error bars represent standard deviation of mean variance.

4.3.2. *Canopy cover*

Difference in canopy cover estimates as a function of sample density are smaller when compared with canopy height. For example the difference in cover estimates is close to zero for all point densities across all areas with the exception being *NVF* (Figure 21B). The larger error at *NVF* at pulse densities $<0.5 \text{ pl m}^{-2}$ is again attributed to the poor identification of ground surface. This has shifted the height of the ground datum and therefore altered the proportion of vegetation returns included relative to the nominal height threshold. The trend is further reflected in the low coefficient of determination values at the *NVF* area ($R^2 < 0.2$) (Figure 22B). Replication of cover estimates are also robust to diminishing pulse densities (Figure 23C), where standard deviation of plotwise estimates is $<10\%$ of total cover for 0.05 pl m^{-2} .

4.3.3. *Canopy vertical structure*

Difference in C_v of return height are close to zero for all areas when pulse density $\geq 0.5 \text{ pl m}^{-2}$. When pulse density decreases to $<0.5 \text{ pl m}^{-2}$ the differences become positive for sparsely vegetated areas and negative for other areas (Figure 21C). The largest differences are seen at the *SAV* and *GWW* areas where canopy cover is the lowest (Figure 19B). Here it is suggested that large positive differences are caused by the under representation of vegetation in the return height profile at lower pulse densities. The opposite effect is seen to a lesser degree at the *NVF* study area where returns cluster towards the top of the canopy. Analysis of *NoS* values suggests the canopy height profile is generally well represented when pulse density is $\geq 0.5 \text{ pl m}^{-2}$. This is reflected in an error of <1 canopy strata across all areas (Figure 21D). Pulse density $<0.5 \text{ pl m}^{-2}$ leads to an overestimation in the number of canopy strata as layers appear increasingly fragmented, this trend becomes more apparent with increasing canopy height.

Standard deviation of plotwise C_v of return height increases with decreasing pulse density. Similar to canopy height estimates, variance in C_v of return height is greatest for plots with low canopy cover at pulse densities of $<0.5 \text{ pl m}^{-2}$ (Figure 19). The C_v of return height for plots where canopy cover $>20\%$ have a relatively small standard deviation which suggests the dispersion of returns through the canopy is constant with each

realisation. Standard deviation of plotwise *NoS* is also low for pulse densities >0.5 pl m⁻² where intra-plot variation is <0.5 canopy strata (Figure 23E). In forest types where vertical structure is relatively simple (Figure 19D), standard deviation remains low for lower pulse densities.

4.3.4. Characteristics of thinned point clouds

For the original datasets used in this investigation, mean pulse density was ~ 22 pl m⁻². Generating different realisations of point clouds from an original dataset introduces the potential for shared points between subsets. At pulse density of 4 pl m⁻², $\sim 85\%$ of returns were shared between ≤ 2 of 9 subsets. This is suggested as the upper limit for simulation when applying this technique to a dataset with an original pulse density of 20 pl m⁻². Simulations at a pulse density of 0.01 pl m⁻² were also attempted. However, for forest types where canopy cover was high, poor ground identification limited the number of successful simulations (even within plots) and therefore these results were disregarded. Pulse densities for thinned datasets were an average of 10% less than the prescribed density, which is a result of grid points identifying no returns within the search window. This is caused by irregular pulse spacing which in turn is attributed to transmission losses (Korpela et al., 2012), oblique viewing angles (Lovell et al., 2005) or gaps between flight line overlap. An iterative approach could have been used to dynamically alter search window dimensions until the prescribed pulse density was achieved (Næsset, 2009) although this would have increased the risk of return duplication in subsets.

Examples of thinned point clouds for the six forest types are presented in Figure 24, this illustrates the differences in vegetation cover density and homogeneity apparent between study areas. It is clear that the fidelity of canopy scale features is reduced with decreasing pulse density particularly at the sparsely vegetated sites (*SAV* and *GWV*). For higher pulse densities (10 and 4 pl m⁻²) a reasonable identification of the ground surface was achieved across all areas. As pulse densities decrease, the number of ground returns at the *NVF* and *TCF* areas quickly diminishes. This is in contrast to areas where canopy cover is less (Figure 19) where a regular grid of ground returns is still clearly visible at 0.05 pl m⁻².

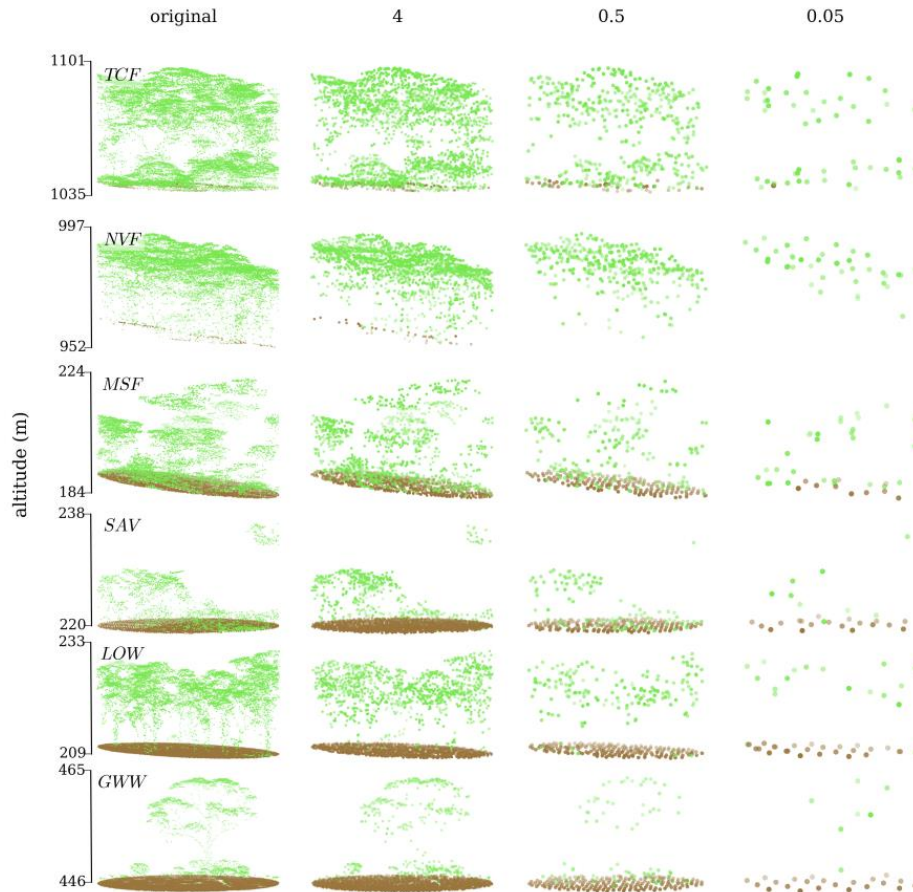


Figure 24. Example point clouds for each forest type displayed at 4 different pulse densities. Points are classified into either ground (black) or non-ground i.e. vegetation (grey). Altitude is relative to the Australian Height Datum. Point size is shown to increase with decreasing density for visualisation purposes only.

4.4. Discussion

The 250 plots across six study areas used in this investigation covered a broad spectrum of vegetation structure where canopy cover ranged from 2 – 98%, canopy height from 2 – 80 m and canopy structure ranged from shrub dominated through to vertically complex systems (Figure 19). Yet for all vegetation types an asymptote in achievable accuracy is reached at a pulse density of 0.5 pl m^{-2} when compared to a high density acquisition

(Figure 21 and Figure 22). This suggests there is only marginal improvement in achievable accuracy by capturing data at higher densities for area based vegetation structure estimates. Furthermore, intra-plot variance is captured at a pulse density of 0.5 pl m^{-2} , although small improvements are evident with increasing pulse density (Figure 23). These results are comparable to results found in similar studies (Goodwin et al., 2006; Treitz et al., 2012; Jakubowski et al., 2013; Watt et al., 2013; Hansen et al., 2015).

Structural measurements derived from pulse densities $<0.5 \text{ pl m}^{-2}$ returned larger differences, particularly at the tropical rainforest (*NVF*) study area which is attributed to a dense canopy cover occluding the ground. Small overestimations of ground height are common in forests where ground vegetation cover is dense (Ni-Meister et al., 2001; Su and Bork, 2006; Magnusson et al., 2007). This was evident at the *TCF* and *MSF* plots where ground height is overestimated by $<2 \text{ m}$ for all simulated pulse densities (Figure 21E). However at the *NVF* area, erroneous ground models are caused by the minimal number of returns that penetrate through the upper canopy (Takahashi et al., 2008). For example at all pulse densities $<2\%$ of returns are classified as ground, which for the pulse density $<0.5 \text{ pl m}^{-2}$ results in an average of ~ 3 identified ground returns in the plot area. In reality a number of simulations had no returns classified as ground which led to the failure of metric calculation (Figure 24). The effect of increasing flying height to attain a lower sampling frequency will only exacerbate poor penetration of pulses to the ground owing to a reduction in pulse intensity (Goodwin et al., 2006; Takahashi et al., 2008). A comparison of the mean height of returns classified as ground for low pulse density datasets reveals that points within the canopy were miss-classified. This increases the height of the ground layer (Figure 21E) and propagates to estimation errors in all metrics. It is acknowledged that a generic approach, using commonly available software, was taken when classifying ground returns and a customised solution may have identified the ground surface more accurately (Evans and Hudak, 2007; Tinkham et al., 2011). Nevertheless, for large area acquisitions (e.g. regional to continental transects) where many different vegetation and terrain types maybe encountered, a bespoke solution for a particular scenario may not be feasible or warranted. The opposite is apparent at study areas where vegetation cover is sparse and can therefore be underrepresented in the

ALS capture at lower pulse densities (Figure 21). This effect is enhanced where vegetation cover is heterogeneous (e.g. clumped) as at the savanna and woodland areas.

Differences in canopy height estimates at diminishing pulse densities are similar to those reported in studies over different forest types. For example, when comparing different flying heights Goodwin et al. (2006) found only small differences in the 99th percentile of canopy height (~1 m) estimated at three pulse densities between ~0.5 and 1 pl m⁻². Jakubowski et al. (2013) reported relatively large errors when using predictive models to estimate field derived canopy cover metrics. Although this investigation does not compare results to field estimates, it is suggested that the weighted $1 - P_{\text{gap}}(z)$ method is robust to diminishing pulse density and could be used to improve predictive models. Previous studies have suggested that if sub-canopy structure is to be assessed satisfactorily then higher pulse densities are required, however these studies have been mostly limited to *first-return* or *first-and-last-return* captures (Thomas et al., 2006; Jakubowski et al., 2013). The inclusion of intermediate returns suggests that the canopy profile can be satisfactorily attributed at 0.5 pl m⁻² when compared to a more dense acquisition (Figure 21C and D). Observed patterns of increasing variance with decreasing pulse density are similar for the four metrics tested and to trends reported in previous studies (Gobakken and Næsset, 2008). Random error in metric estimation increases with decreasing pulse density (Takahashi et al., 2010), this is illustrated by a simultaneous decrease in coefficient of determination values with pulse density (Figure 22).

Systematically applying an offset to the point-grid allowed robust estimates of vegetation metrics independent of sampling location origin. This is particularly useful when analysing low pulse densities where large variation is evident between realisations (Figure 23). Analysis of intra-plot results suggests that variance in metrics is captured at pulse densities ≥ 0.5 pl m⁻². For lower pulse densities, difference between realisations can be more significant. For example at the *TCF* plots, difference in height estimations between realisations had a standard deviation of ~3 m at the lowest pulse density, this drops to <1 m at a pulse density of 0.5 pl m⁻² (Figure 23A). Intra-plot variance was lowest for plots where canopy cover was homogenous and structurally simple e.g. a single canopy (Figure 19

and 24). For example standard deviation of canopy height remained <1 m for the *LOW* study area, even when considering the lowest pulse density simulated (Figure 23A). Conversely at the sparsely vegetated *SAV* study area where vegetation is clumped, the C_v of canopy height and C_v of return height is highly variable (Figure 23B and D respectively). This is due to vegetation cover being heterogeneous and therefore subsequent plotwise realisations capture significantly different proportions of vegetation, resulting in dissimilar interpretations of the canopy profile. When comparing 4 overlapping flight lines acquired on the same day with the same capture specifications, Bater et al. (2011) found a similar difference between height estimates of <1 m at a pulse density of 2 pl m^{-2} . The authors also noted significant differences in metrics calculated from last-returns e.g. describing the ground and lower region of the canopy. This is again less apparent in this study owing to the utilisation of a multi-return recording instrument.

The technique introduced in this paper has allowed a robust comparison of estimating vegetation structure metrics simulating different capture pulse densities across a range of vegetation types. The inclusion of a TIN model calculation is seen as an important step that is not always included in thinning simulations. TIN modelling highlighted that utilising a pulse density $<0.5 \text{ pl m}^{-2}$ in dense forests is unlikely to identify the ground surface adequately to accurately determine vegetation structure. It is recognised the presented technique does not account for factors caused by an increased flying height such as increase in laser footprint size or attenuation of laser power caused by increased atmospheric thickness (Goodwin et al., 2006), as well as acquisition at oblique viewing angles (Lovell et al., 2005) or with a different instrument. Physical models such as that presented by Disney et al. (2010) could potentially be used to more accurately model different acquisition scenarios. However this would require significant effort to recreate landscape scale variance in native forest structure and is therefore beyond the scope of this study.

4.5. Conclusion

Land managers are commissioning small-footprint discrete return airborne laser scanning (ALS) acquisitions over increasingly large areas which may in turn capture a variety of forest types. This study examines the sensitivity

of pulse density (sampling frequency) on primary descriptors of vegetation structure (canopy height, canopy cover, return height coefficient of variation and *NoS*) across a broad range of forest types; from sparsely vegetated savanna woodlands to dense rainforest. ALS was acquired with the same campaign and capture specifications and ALS instrument which facilitated comparisons between forest types. Point clouds were thinned to six different densities (0.05 - 4 pl m⁻²) using a novel technique that systematically selected nine subsets of data from each original plot dataset, this allowed metrics to be computed in a robust way as well as assessing the reproducibility of ALS acquisition. Metrics derived from thinned datasets were compared to a dataset with a pulse density of 10 pl m⁻².

Simulated acquisition with a pulse density of ≥ 0.5 pl m⁻² resulted in minimal differences for all metrics across all forest types when compared to a dataset with a pulse density of 10 pl m⁻². Furthermore, intra-plot variance was significantly lower at higher pulse densities than for less dense simulations. This result suggests there is minimal gain from acquiring ALS data at a pulse density > 0.5 pl m⁻² which could result in potential cost savings for land management agencies. The primary reason for erroneous estimation at lower pulse densities was the poor identification of the ground surface, which propagated to metric estimation, and heterogeneous (e.g. clumped) vegetation being inadequately sampled. The analysis presented here will allow land managers to be confident in specifying lower pulse densities when planning ALS capture for large area vegetation characterisation, even over dense, very sparse, tall or vertically complex forests.

5. Chapter 5

Mapping forest canopy height across large areas by upscaling ALS estimates with freely available satellite data[§]

[§] This chapter is based on:

Wilkes, P., Jones, S.D., Suarez, L., Mellor, A., Soto-Berelov, M., Woodgate, W., Haywood, A. and Skidmore, A., 2015. Generating a continuous map of canopy height over a regional area using a step sampling with LiDAR and multispectral imagery. *Remote Sensing* 7(9) pp. 12563 - 12587.

5.1. Introduction

For large, inaccessible and remote forested areas, the assessment of vegetation structure in an operational framework remains an on-going challenge for land managers (McRoberts et al., 2010; Wulder et al., 2012a, 2012b). Synoptic capture of large forested areas is provided by space borne passive optical remote sensing platforms and has proved useful for the attribution of forest structure (Franklin and Strahler, 1988; Wulder, 1998; Armston et al., 2009; Wulder et al., 2012b; Hansen et al., 2013). However, the inability of passive instruments to sense below the principal canopy limits their applicability for assessing three-dimensional forest structure attributes, such as canopy height (Cohen and Spies, 1992; Lefsky et al., 2002b; Pasher and King, 2011). Over the last two decades Light Detection and Ranging (LiDAR) technologies, and in particular discrete return Airborne Laser Scanners (ALS), have become an operational alternative to traditional forest inventory (Holmgren and Jonsson, 2004; Maltamo et al., 2006; Wulder et al., 2012a). Consequently, there has been recent interest in the fusion of ALS and satellite multispectral imagery for the improved retrieval of vegetation parameters (see review by Torabzadeh et al. (2014)). However, although examples of acquisition of ALS over very large areas exist (Asner et al., 2010; Gregoire et al., 2011; Quadros et al., 2011; Gobakken et al., 2012; Wulder et al., 2012a; Hansen et al., 2014), ALS coverage is often incomplete and follows a transect or linear pattern and is therefore inappropriate for deriving wall-to-wall maps of vegetation structure.

To achieve large area attribution, ALS can be used as a sampling tool in a two-stage approach where ALS is captured over a fraction of the study area. This is achieved by first establishing an empirical statistical model between ALS metrics and spectral reflectance and/or other spatially synoptic datasets. The model is then applied to the reflectance/synoptic data and therefore upscales estimates of canopy structure beyond the confines of the ALS survey extent (Hudak et al., 2002; McNerney et al., 2010; Ørka et al., 2010; Pascual et al., 2010; Cartus et al., 2012; Mora et al., 2013). Examples where assessment has been carried out over large areas of heterogeneous forest include Asner et al. (2010) who estimated forest biomass across ~4 million hectares of Peruvian rainforest and Wulder and Seemann (2003) who estimated canopy height over 700,000 ha of boreal

forest in Canada. Both studies used a linear regression of ALS derived variables with segmented Landsat imagery to predict canopy structure. Continental and global maps of forest structure have also been produced using this method where LiDAR data from the spaceborne the Geoscience Laser Altimeter System (GLAS) sensor was used as a sampling tool in conjunction with coarse resolution satellite imagery (250 - 1000 m) (Baccini et al., 2008; Lefsky, 2010; Simard et al., 2011).

Modelling approaches that have determined a parametric association between response and predictor variables have been successfully applied to the attribution of forest structure (Hudak et al., 2002; Wulder and Seemann, 2003; Pascual et al., 2010). However, in more recent years machine learning techniques have been utilised for remote sensing applications, where the complex statistical associations of multi-source datasets require more advanced approaches to characterise forests over large areas (McInerney et al., 2010; Mellor et al., 2013; Mora et al., 2013). A machine learning technique that has gained in popularity is random forest, an ensemble regression tree technique from the Classification And Regression Tree or CART family (Breiman, 2001). Random forest works by constructing “weak” regression trees (usually in the order of hundreds) from bootstrapped samples of input variables. The “weak” regression trees are then aggregated in an ensemble to produce a robust model that is insensitive to collinear predictor variables and a non-normal distributed response variable (Breiman, 1996). Furthermore, the ease of application (e.g. only two model parameters, see Section 2.3.1) and the ability to run efficiently over large datasets makes random forest an ideal choice for large area attribution (Rodriguez-Galiano et al., 2012b). A number of studies have utilised random forest for mapping forest attributes with remotely sensed data, including biomass (Baccini et al., 2008; Mascaro et al., 2014), species extent (Evans and Cushman, 2009), forest extent (Ørka et al., 2010; Mellor et al., 2012, 2015), canopy cover (Armston et al., 2009; Johansen et al., 2010; Ahmed et al., 2015) and canopy height (Kellndorfer et al., 2010; Stojanova et al., 2010; Simard et al., 2011; Cartus et al., 2012; Peterson and Nelson, 2014; Ahmed et al., 2015).

The majority of studies use a combination of predictor variables that can be roughly split into two cohorts: (a) variables that respond to changes in vegetation and are derived from surface reflectance, and (b) variables that

determine vegetation growth potential (in the absence of disturbance) such as site quality and climate. Particular choices of predictor datasets are dictated by the target variable, scale of analysis and data costs or accessibility; the ready access to free satellite imagery has been recognised as crucial for the long term modelling of environmental systems (Turner et al., 2015). Large area forest assessment should be consistent across the domain of the study, providing an accurate estimate of forest structure regardless of forest type, as well as being locally relevant, for example identifying features in the landscape (Hansen et al., 2013). An earth observing platform that has proved useful in this respect is the Landsat programme, for example, 9 of the 11 random forest studies listed above used Landsat products as a predictor variable in some way.

Random forest is capable of efficiently incorporating a large number of both continuous and categorical variables, as a result of sub-sampling predictor variables at each node when constructing regression trees. From a remote sensing perspective, this enables additional contextual or textural variables (and the large number of variables this can produce) to be easily incorporated into modelling (Rodriguez-Galiano et al., 2012a). The addition of first and second order textural information from satellite imagery has been shown to improve the classification accuracy of forest structure (Franklin et al., 2001; Coburn and Roberts, 2004; Ghimire et al., 2010; Rodriguez-Galiano et al., 2012a). Additionally, the incorporation of variables generated from a remote sensing time series have also improved model performance. For example, when estimating canopy height Ahmed et al. (2015) included a time-since-disturbance variable generated from a Landsat ETM time series, which led to improvements in RMSE of ~20%.

This manuscript extends the work of previous authors by presenting a method for the production of a medium-resolution (30 m) continuous map of canopy height, for a large area (millions of hectares) of highly heterogeneous forest, using freely available datasets as predictive variables and in an open source computing framework. The presented method is intended to be easily adopted (and adapted) by forest scientists and land management agencies for the routine assessment of canopy height. Canopy height was chosen as a candidate metric owing to its importance across many applications including biomass estimation (Lefsky et al., 2001; Lucas et al., 2008b; Asner et al., 2010), habitat assessment (Hyde et al., 2006;

Goetz et al., 2007) and forest inventory (Næsset, 1997, 2007; Wulder et al., 2008a). An estimate of canopy height is also required to fulfil international assessment and reporting obligations such as those outlined in the Montreal process (Miles, 2002).

5.2. *Materials and methods*

5.2.1. *Study area*

The study area is located in the state of Victoria, Australia (Figure 25A) and comprises a total area of 4 million ha, an area similar in size to the country of Switzerland or the US state of Maryland. Land tenure is predominantly public (>70%), the majority of which is located in state forest and national parks; the remainder is privately owned and primarily used for grazing livestock. Within this boundary, forest covers 2.9 million ha (Mellor et al., 2012) where forest is defined as “having the potential to reach >2 m in height and >20% canopy cover” (National Forest Inventory, 1998) (Figure 25D). Canopy height across the study area ranges from 0 – 70 m (Figure 25B).

The forested area extends across seven *Interim Biogeographic Regionalisation for Australia* (IBRA) regions, IBRA regions have distinct ecological, geological and climatological features (Department of Environment, 2012). Vegetation is dominated by dry sclerophyll forest and woodlands which have a relatively sparse canopy and a patchy, scrubby understorey. In the foothills of the Australian Alps there are areas of highly-productive wet forest and rainforest characterised by a tall (>40 m) and closed canopy with high species richness (Gellie et al., 2005). There are also subalpine and alpine areas in the Australian Alps that straddle the middle of the study area, these are characterised by relatively short vegetation.

Two factors confound the estimation of canopy height in the study area when using remote sensing. Firstly, the area is subject to regular disturbance from fuel reduction burns, bush fires and drought. For example, the area experienced the most severe drought in a century in the decade prior to the study, which led to large scale tree mortality (Semple et al., 2010; Van Dijk et al., 2013). Secondly, the physiology of *Eucalypt* trees and stands, such as an erectophile leaf angle distribution, asymmetrical

crown configuration, low foliage density and leaf and crown clumping (Jacobs, 1955) increase the proportion of reflectance coming from the ground, mid and understorey as well as increasing the shadow fraction (Armston et al., 2009; Jenkins and Coops, 2011).

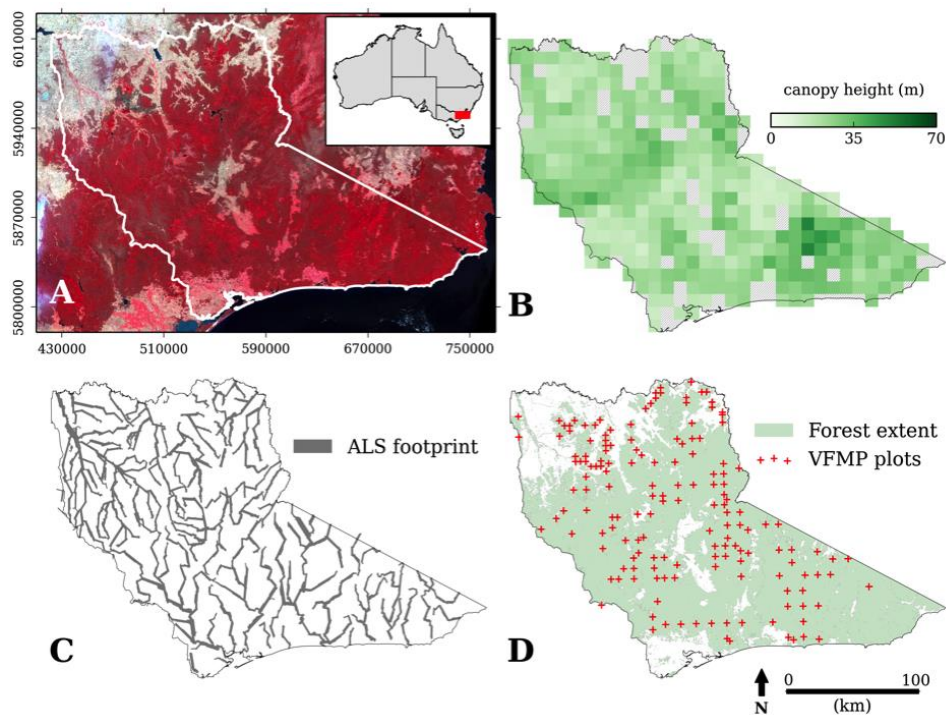


Figure 25. Study area in east Victoria, Australia. A mosaic of 5 Landsat TM false colour composite images covering the study area (outlined in white) and location of the study area within Australia (inset) [A]. Canopy height derived from ALS capture where canopy height values are aggregated into 10 x 10 km cells (grey indicates no data) [B]. The extent of the ALS capture [C]. Forest extent (Mellor et al., 2012) and location of Victorian Forest Monitoring Programme forest inventory plots (VFMP) [D]. Map coordinate system is the projected Map Grid of Australia (MGA) Zone 55.

5.2.2. Data collection

5.2.2.1. Forest inventory

Forest inventory plot data was collected as part of the Department of Environment, Land, Water and Planning (DELWP) Victorian Forest Monitoring Programme (VFMP). A total of 130 forest inventory plots were within the study area including 22 that intersected the ALS acquisition extent. Forest inventory plots were installed between May 2011 – December 2014, where at each sampling location a 0.04 ha plot was established following DELWP protocol (Department of Sustainability and Environment, 2012). Measurements for all trees within the forest inventory plot included diameter at breast height, species and live status. For a subset of trees (including the three tallest) height was also recorded. Dominant canopy height, the mean height of the three tallest live trees in a forest inventory plot, was calculated as the metric summarising canopy height (Lovell et al., 2003).

5.2.2.2. Airborne laser scanning data

Airborne Laser Scanning (ALS) data was acquired as part of the DELWP River Health Programme (Quadros et al., 2011). ALS instrument and survey specifications are presented in Table 1. The ALS data was originally acquired to assess stream bank condition and therefore capture was targeted at the riparian zone, although actual extent ranges from 0.3 – 2.5 km either side of a watercourse. Flight lines followed the course of the rivers and streams and were therefore off cardinal (Figure 25C), this resulted in a substantial and multiple overlap at flight line intersections. A combination of pulse density and the ability of the two ALS instruments utilised to record up to 4 discrete returns per outgoing laser pulse meant the data was suitable for characterising vegetation structure (Wulder et al., 2012b; Wilkes et al., 2015).

The ALS acquisition extent was clipped to the existing forest area (Mellor et al., 2013) and totalled 520,000 ha (Figure 25C). A regular grid with a 250 m spacing (to reduce the spatial autocorrelation of the response variable) was placed over the study area and a total of 12,000 ALS plots were

Table 7. ALS capture and instrument specifications

Specifications	
<i>Capture specifications</i>	
Date	December 2009 – January 2011
Flying height	600 – 1500 m above ground level
Mean pulse density	9.4 pl m ⁻²
Swath overlap	20%
Absolute vertical accuracy	±20 cm
Absolute horizontal accuracy	±30 cm
Mean footprint diameter	~35 cm
<i>Instrument specifications</i>	
Instrument	Leica ALS50-II and ALS60 (Heerbrugg, Switzerland)
Operating wavelength	1064 nm
Max off-nadir scan angle	±15°
Outgoing pulse rate	36.4 Hz

extracted using random stratified sampling. To capture canopy structural variance across the study area, the IBRA bioregion layer was used to stratify the area into distinctive vegetation types. ALS plots that either intersected the edge of the ALS acquisition or had a pulse density <0.5 pulses m⁻² (Wilkes et al., 2015) were removed from analysis. This resulted in ~11,000 ALS plots for model construction and evaluation.

Square plots (50 m x 50 m) were extracted from the ALS dataset, after computation of canopy height plots were clipped to 30 m x 30 m to be consistent with Landsat TM pixel dimensions. Plots were initially extracted at larger plot dimensions to ensure points around the plot edge had a large enough neighbourhood to create a representative ground surface model. Point height data was normalised to ground surface by first classifying points into either ground or non-ground, then using ground classified returns only, creating a triangulated irregular network (TIN) surface. Using the TIN, the ground normalised height for all points was then calculated. Point classification, TIN creation and height normalisation were computed using LAStools (version 130225) (Isenburg, 2012). The 95th percentile of return height for returns classified as non-ground was calculated for each plot as an analogue of dominant canopy height.

5.2.2.3. Satellite imagery and ancillary data

A full list of predictor variables initially processed is presented in Table 2. A total of 10 Landsat Thematic Mapper (TM) images were acquired for two seasons; January – March 2009 (summer) and October – November 2009 (spring). Two seasons were acquired as different vegetation cover and composition characteristics are evident at different times of year. For example, imagery captured in summer maximises the spectral difference between evergreen (overstorey) and cured grass whereas spring imagery captures the green flush (Mellor et al., 2013). Images were geo-rectified and corrected for atmospheric and bi-directional reflectance distribution function effects to obtain surface reflectance (Flood et al., 2013), before being mosaicked. Both image mosaics were captured at a time (pre- and post- summer equinox) when sun angle was relatively high to minimise shadow. Although the summer imagery was captured approximately one year prior to the start of the ALS acquisition, imagery from summer 2010 was significantly cloud affected and therefore unsuitable. A Tasseled Cap (TC) transformation (Crist and Cicone, 1984) was applied to the Landsat TM mosaics, reducing the 6 visible bands to 3 features; brightness, greenness and wetness (Kellndorfer et al., 2010; Rodriguez-Galiano et al., 2012a; Ahmed et al., 2015). From each TC feature, 2 first order texture metrics (mean and variance or contextual and textural metrics respectively) were calculated using a range of kernel sizes (3, 5, 15, 33, 65 and 99 Landsat TM pixels). Maximum kernel size was determined using semivariance analysis of ALS derived canopy height models (30 m x 30 m resolution) captured over three representative forest areas in Victoria. Smaller kernel sizes were used to capture forest structure variance for forests that have a shorter lag and also to characterise forest patches smaller than a continuous canopy e.g. fragmented forests or linear features such as riparian vegetation.

A time series of the Moderate Resolution Imaging Spectroradiometer (MODIS) Normalised Difference Vegetation Index (NDVI) product (MOD13Q1) was utilised to capture changes in vegetation structure in the decade prior to the study period. The MODIS NDVI product was chosen as it is highly correlated with vegetation phenology (Zhang et al., 2003) and also has the highest spatial resolution of MODIS products (250 m). Two scenes (summer and spring) were acquired for each year between 2000 –

Table 8. List of predictor variables with original image resolution in brackets (*kernel sizes: 3, 5, 15, 31, 65, 99 pixels)

Source	
<i>Landsat TM (30 m)</i>	<i>MOD13Q1 (NDVI) time series (2001 – 2010) (250 m)</i>
<i>Tasseled Cap features and NDVI</i>	Summer mean
Summer brightness	Summer standard deviation
Summer greenness	Summer coefficient of variation
Summer wetness	Summer linear regression slope coefficient
Spring brightness	Spring mean
Spring greenness	Spring standard deviation
Spring wetness	Spring coefficient of variation
Summer NDVI	Spring linear regression slope coefficient
Spring NDVI	
<i>Image context/texture*</i>	
Summer mean brightness	<i>SRTM (~30 m)</i>
Summer mean greenness	Elevation
Summer mean wetness	Aspect
Spring mean brightness	Slope
Spring mean greenness	
Spring mean wetness	<i>Climatic (1 km) [66]</i>
Summer brightness variance	Total annual precipitation
Summer greenness variance	Mean annual temperature
Summer wetness variance	
Spring brightness variance	<i>Soils</i>
Spring greenness variance	Soil moisture (1 km) [67]
Spring wetness variance	Major soil type (vector) [68]
	<i>Coordinates (MGA zone 55) (30 m)</i>
	X location
	Y location

2010, where images were captured at the same time each year (first week of February and first week of November respectively). Images were subsequently ordered into a chronological stack (for each season) and four

statistics were computed for each pixel stack: mean, standard deviation, coefficient of variation and the slope coefficient of a linear regression of NDVI with acquisition year.

Site quality and climatological variables can constrain maximum canopy height, for example, a forest plots topographical position or air temperature (Austin, 1987; Jenkins and Coops, 2011). To capture this within the predictive model, additional variables included elevation, slope and aspect derived from the Shuttle Radar Topography Mission (SRTM) 1 arc-second resolution (~30 m) dataset; mean annual temperature and mean total rainfall (Hijmans et al., 2005); soil water balance (Trabucco and Zomer, 2010) and major soil type (National Resource Information Centre, 1991). Additionally Cartesian coordinate layers (X and Y location) were included (Mascaro et al., 2014). All datasets were resampled to a 30 m resolution to match that of Landsat TM and reprojected to MGA zone 55.

5.2.3. Canopy height estimation with random forest

For the estimation of canopy height, random forest was run in 'regression' mode where canopy height was the response (dependent) variable and the satellite and ancillary data were the predictor (independent) variables. In order to preserve the spatial heterogeneity present in native forests managed for conservation (Lindenmayer et al., 2000; McGarigal et al., 2009; Jenkins and Coops, 2011), canopy height output is computed as a continuous surface i.e. not segmented into forest stands. Random forest models were constructed and validated over the entire study area by randomly sampling 5,000 ALS plots from the global dataset in a bootstrap ($N=50$). For each bootstrap iteration, the resulting random forest model was applied to a withheld random sample of 1,000 ALS plots, from which Root Mean Squared Error (RMSE) was estimated. Outliers at the 95th percentile confidence interval were removed from RMSE calculations. To produce a wall-to-wall map of canopy height, individual random forest models from the bootstrapped cross-validation were combined to improve generalisation and then applied to the synoptic datasets. A second validation was achieved by comparing forest inventory measured canopy height from outside ALS acquisition area with model output.

Two further experiments were conducted to assess the suitability and applicability of random forest for estimating canopy height over large

areas: (a) models were constructed and validated using ALS data within smaller geographic extents, and (b) models were constructed from smaller geographic subsets and validated using ALS data from the remaining portion of the study area. For experiment (a) the hypothesis was that improvements can be made in performance by constructing and validating models over smaller geographic areas owing to the reduced range and variability of the response and predictor variables. Experiment (a) was evaluated by dividing the study area into 24 grid squares (50 km x 50 km), then for each grid square a model was constructed using 75% of ALS plots as training data and the remainder for validation.

LiDAR surveys usually cover a much smaller extent than the River Health capture. Experiment (b) therefore tested the possibility of combining disparate acquisitions to estimate canopy height over a larger area. This was achieved by randomly sampling the twenty-four 50 km x 50 km grid squares, where the number of squares included was iteratively increased from 2 to 23 (5 – 95% of the ALS footprint or 1 – 16% of the forested area). ALS plot data from the selected grids was combined and used as a training sample, creating a non-random distribution of training samples. A random sample of 2,000 ALS plots from outside the selected grid squares was used as validation. An additional 18,000 ALS plots were randomly extracted from the River Health dataset (using the method described in Section 2.2.2), and combined with 2,000 from the original sample to create a randomly distributed dataset for the entire study area (i.e. not stratified by IBRA bioregion).

5.2.4. *Random forest implementation*

To facilitate the uptake of this method by land management agencies and forest scientists, computation was achieved using an open source framework (Mellor et al., 2013). ForestLAS (Wilkes, 2015) and LAStools (Isenburg, 2012) were used to extract and process ALS data; GRASS (GRASS Development Team, 2015) and QGIS (QGIS Development Team, 2015) software were used to extract and pre-process predictor variables; data management was achieved with Python (Python Software Foundation, 2015); and random forest was implemented in Python via RPy2 using the R (R Development Core Team, 2014) *randomForest* package (Liaw and Wiener, 2013).

The *randomForest* implementation has two primary user defined parameters; number of candidate variables selected at each node split (*mtry*) and the total number of trees constructed in each forest (*ntree*). The default *mtry* value was used, this is calculated as the total number of predictor variables divided by three (Liaw and Wiener, 2013). Stabilisation of out-of-bag error (the error calculated using the withheld sample from the construction of each regression tree) occurred at ~ 100 trees and was therefore used for *ntree*. Additionally, a sensitivity analysis of the number of training samples was undertaken, indicating an asymptote in achievable accuracy was reached at $\sim 5,000$ plots.

5.2.5. Selecting predictor variables

A more parsimonious model can be obtained by removing highly collinear variables and variables that contribute least to the predictive capability of the model (Svetnik et al., 2004; Murphy et al., 2010). Variable collinearity was tested by constructing a coefficient of determination matrix for all combinations of variables. For each highly correlated pair ($r^2 > 0.9$), correlation coefficients were calculated between the variables and canopy height, where the predictor variable with the highest coefficient of determination was kept. The second step followed the method of Murphy et al. (Murphy et al., 2010) where cohorts of predictor variables were iteratively removed from model construction based on their importance within the model. After each iteration model performance was evaluated.

A total of 19 variables were finally selected (Figure 26) from the original set of 97 (Table 2). Using the cohort of 19 variables resulted in a $< 1\%$ decrease in model accuracy when compared to using the full set. The cohort consisted almost exclusively of reflectance variables, of which fifteen were derived from Landsat TM and 4 were derived from the Tasseled Cap wetness feature alone. Contextual predictor variables featured prominently and were more important than single pixel variables. Landsat TM imagery captured in the summer was more important than spring imagery. Ten year summer and spring mean NDVI was the only MODIS time series variable to feature prominently. Ancillary variables were of less importance, only the X and Y Cartesian coordinates featured significantly. The relatively small range of MSE values from the cross

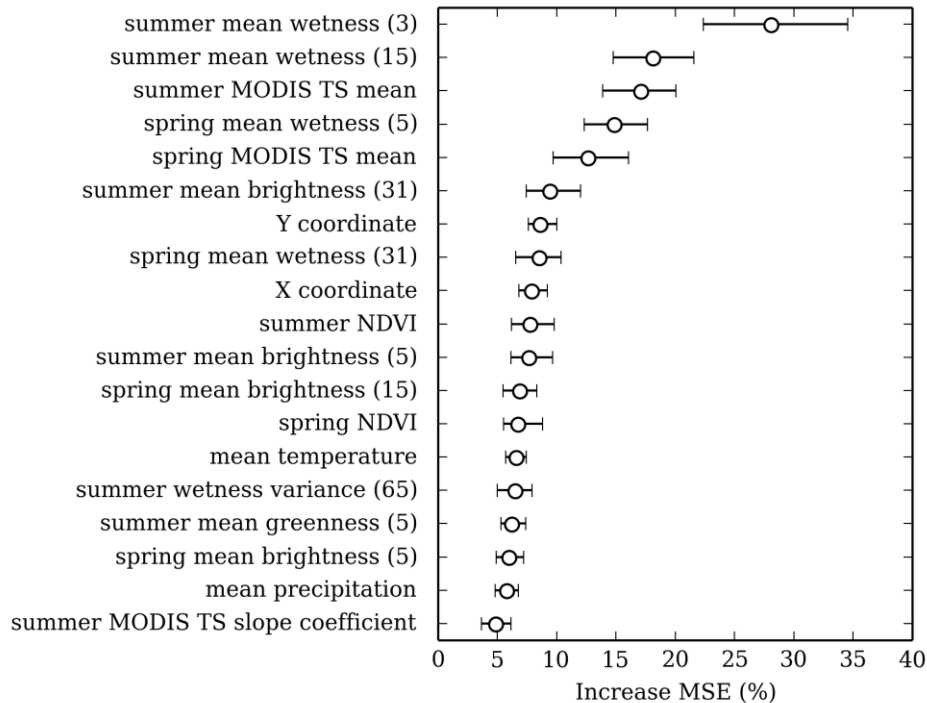


Figure 26. Relative importance of the 18 variables selected for the final random forest model. Confidence intervals (95th percentile) for variable importance were calculated in a bootstrap ($N=50$). Increase mean square error (MSE) is the mean of the squared prediction error when the variable is permuted for a random variable (Cutler et al., 2007). Numbers in brackets indicate the kernel size.

validation for each predictor variable (Figure 26) suggests that the order of the most significant variables was stable.

5.2.6. Systematic error in model output

Exploratory analysis suggested a systematic bias in modelled canopy height output, where the height of shorter and taller plots were over and underestimated respectively. Spatially incorporating an estimate of error (e.g. kriging or cokriging of model residuals) has been utilised previously (Hudak et al., 2002). However, here this was inappropriate owing to the low spatial autocorrelation in model error (Moran's $I = 0.018$, $p < 0.001$). Alternatively, two aspatial methods were tested to mitigate the systematic bias; (a) resampling of the response variable to a uniform distribution and

(b) subtracting from model output a linear model that characterises the systematic error component. When running random forest as a ‘classifier’, inequity in class representation can be addressed by either up-sampling or down-sampling the minority and majority classes respectively (Chen et al., 2004). Here a pseudo-uniform distribution for the response variable was computed by aggregating ALS plots into 10 m height cohorts, then using random sampling (with replacement), increasing or decreasing the number of plots in each class accordingly.

For the second approach, a random forest model and a linear model of systematic error were derived from the training dataset, these were then applied to the withheld dataset or across the whole study area. To ensure independence of training and withheld datasets, the training dataset was divided into two halves. The first half of the training data was used to construct the random forest. Linear coefficients were then determined by regressing model residual error, derived from applying the random forest model to the second training dataset, with ALS estimated canopy height. To then estimate canopy height of the withheld dataset (H_{RF-SE}), the random forest model was applied to the dataset (H_{RF}) and the systematic component was subtracted as follows:

$$H_{RF-SE} = H_{RF} - (\alpha H_{RF} + \beta)$$

EQ. 15

where α and β are the regression coefficients. When applying the technique across the whole study area using the combined random forest model, mean α and β regression coefficients from the cross-validation were used.

5.3. Results

5.3.1. Canopy height estimation

Cross-validation of random forest for estimating canopy height returned a mean RMSE of 31.3% (5.68 m) at the 95th percentile confidence interval, where the model explained 58% of variance in canopy height ($p < 0.0001$). Systematic error was apparent in the over and underestimation of shorter

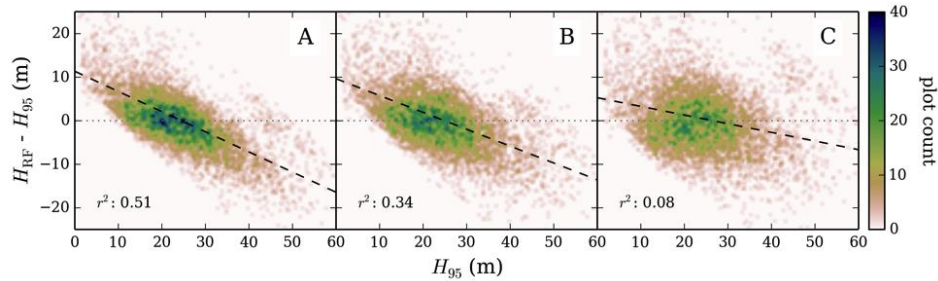


Figure 27. ALS derived canopy height (H_{95}) compared to model residual error for random forest models (H_{RF}); constructed using 5,000 ALS plots of untransformed response data [A], 5,000 ALS plots where the response variable was resampled to a uniform distribution [B], and subtraction of systematic error from modelled canopy height [C]. The coefficient of determination values (r^2) were calculated from a linear regression of measured canopy height and model residuals.

and taller canopy heights respectively (Figure 27A), systematic error accounted for ~ 2 m of total error. This error was caused by forest plots located towards the tails of the response variable distribution having similar predictor variable values to plots closer to the mean (Figure 4). As a result, canopy height values closer to the mean were preferentially modelled to reduce overall prediction error. For example, plots with a canopy height < 10 m have a comparable spectral response to plots where canopy height is 10 – 20 m (Figure 28B). As plots in the 10 – 20 m cohort are more numerous, modelled canopy height for plots where ALS estimated canopy height is < 10 m are allocated to the 10 – 20 m cohort (Figure 28C). The systematic error resulted in additional kurtosis for modelled canopy height when compared to the distribution of the response variable (compare Figure 28A and C). This effectively reduced the range of canopy height from 1.4 – 71.9 m for ALS measured canopy height to 7.9 – 60.7 m for modelled canopy height.

Training random forest with response data that had been resampled to a uniform distribution resulted in a marginal reduction in overall model performance (RMSE = 33% at the 95th percentile confidence interval), however the predicted range of canopy height values increased

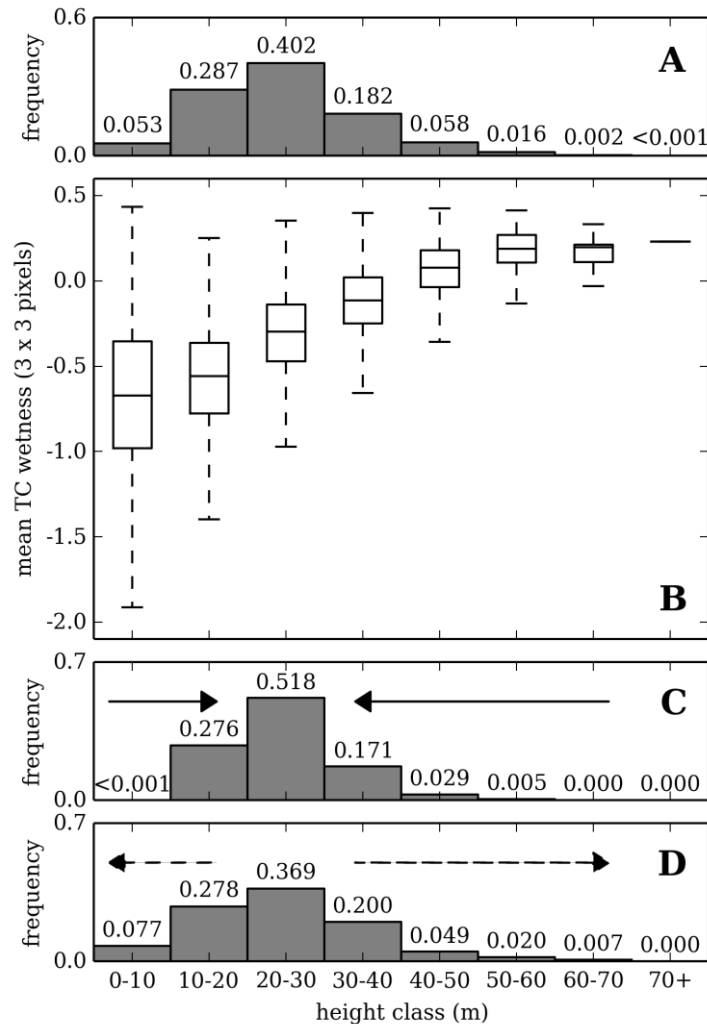


Figure 28. A comparison of the response variable distribution [A], range of Tasseled Cap wetness values (3 x 3 pixels) [B] and the distributions of modelled canopy height (random forest [C] and random forest - systematic error [D]) for different height classes. The solid arrows indicate the direction in which the random forest model output was “squeezed” by inequity in response variable distribution; the dashed arrow indicates the direction canopy height values were rescaled after correcting for systematic error.

to 6.7 – 63.3 m (Figure 27B). When an estimate of systematic error was accounted for the distribution of canopy height closely represents ALS estimate height (Figure 28). Furthermore, the range of canopy height was more representative of the response variable (0.5 – 68.0 m) and mean

error for plots where canopy height was >50 m and <10 m were reduced by 4.2 m and 1.2 m respectively (Figure 28D). Although error became independent of the response variable (Figure 27C), overall model accuracy increased only marginally when compared to the original output (RMSE = 30.4% (6.46 m) at the 95th percentile confidence interval) owing to the rescaling of correctly modelled plots.

A map of canopy height derived from random forest after correcting for the systematic error is presented in Figure 29A. When ALS plots are aggregated into 10 km x 10 km cohorts and compared to the ALS dataset, error was less than $\pm 15\%$ of ALS derived canopy height for 74% of the study area, less than $\pm 10\%$ for 57% and less than $\pm 5\%$ for 33%. Larger errors occur in the taller (south east corner) and shorter (central strip) forests in the study area (Figure 29B and C), which was consistent with the tails of the canopy height distribution being under represented.

Generating a continuous map of canopy height at a 30 m resolution allows for the identification of features in the landscape, for example, caused by land-use and disturbance history. An area of mixed-use forest is presented in Figure 30 where the location, extent and regeneration of (clear-felled) logging coupes are clearly evident. A map generated at a coarser spatial resolution would not identify land-use history with such fidelity. In Figure 30 example A, the poorer model performance for estimating the tails of the canopy height distribution is evident, as it would be expected that canopy height for a recently logged coupe would be closer to 0 m.

5.3.2. *Validation with inventory data*

Further validation of model output was provided by comparing random forest generated canopy height with forest inventory plots from outside the extent of the ALS capture (Figure 31), the random forest model and random forest minus systematic error returns a RMSE of 29% (5.5 m) and 32% (6.3 m) respectively at the 95th percentile confidence interval. Although the model output from random forest produces a more accurate result (Figure 31A), correcting for systematic error improved estimates of taller forest inventory plots in particular (Figure 31B).

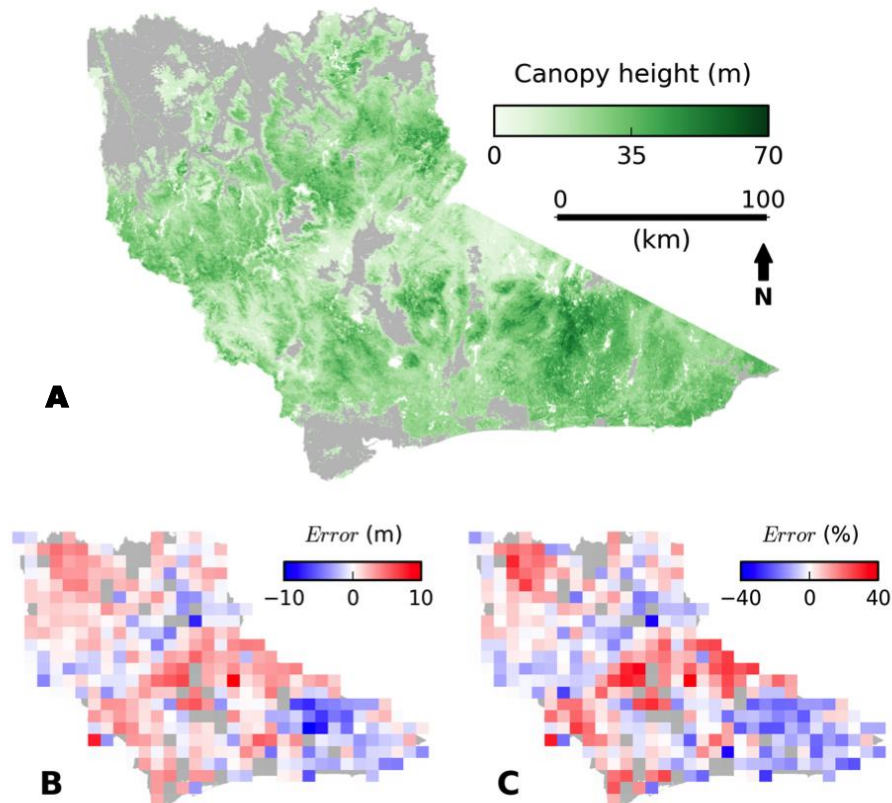


Figure 29. Canopy height at a 30 m resolution (clipped to forest extent) generated using random forest - systematic error [A]. Model output when compared to ALS derived canopy height (10 km x 10 km resolution) where error is represented as height difference [B] and percentage of height [C]. Coordinate system is the projected Map Grid of Australia (MGA) Zone 55

Figure 32 compares ALS and random forest derived canopy height with inventory measurements from within the ALS capture extent. A good statistical association is evident between inventory and ALS measured canopy height where ALS estimates canopy height with a RMSE of 11% (2.9 m). This highlights the suitability of ALS for measuring canopy height over a large area using a single metric, even where forest type is heterogeneous. The association between random forest estimated and inventory measured

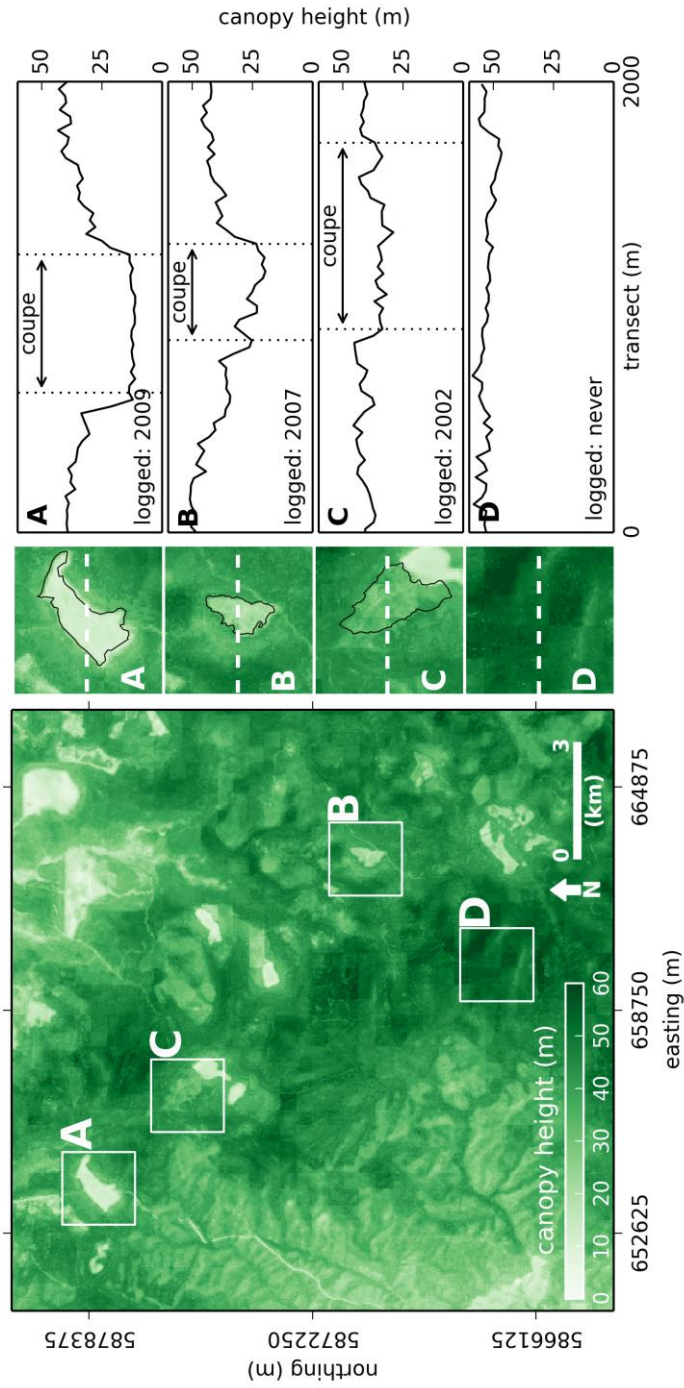


Figure 3. A map of canopy height (30 m resolution) for an area of 16.5 km x 16.5 km, highlighting the land use history of an area of mixed use forest [left]. Three logging coupes that were clear-felled between 2002 and 2009 [A – C] and an area that has never been logged [D] are identified (coupe extents outlined in black) [centre column]. Transects of canopy height across each coupe indicating regrowth since logging, coupe boundaries are also identified [right]. Coordinate system is the projected Map Grid of Australia (MGA) zone 55

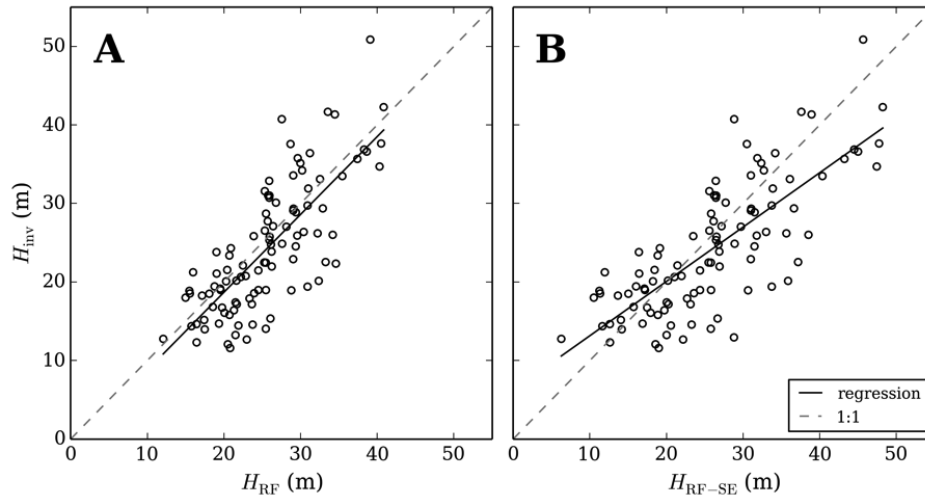


Figure 31. A comparison of inventory measured canopy height (H_{inv}) with random forest (H_{RF}) [A] and random forest corrected for systematic error (H_{RF-SE}) [B] estimated canopy height, at 108 forest inventory plots.

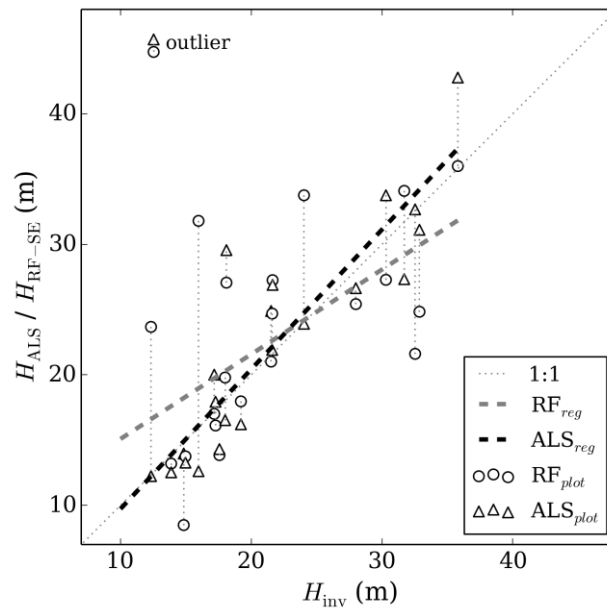


Figure 32. A comparison of inventory (H_{inv}) measured canopy height and ALS (H_{ALS}) and random forest - systematic error (H_{RF-SE}) estimated canopy height, for 22 plots from within the ALS capture area. Vertical dotted lines link the same plot estimated with ALS or random forest.

canopy height is similar to the comparison with forest inventory plots outside the ALS acquisition extent (RMSE = 29% 6.0m)). Interrogation of the outlier in Figure 32 indicates this was the result of a single emergent tree that was significantly taller than the other trees used to estimate dominant canopy height, thereby increasing the ALS estimate of canopy height.

5.3.3. *Training and validation of random forest using smaller geographic areas*

There was generally no improvement in model performance for random forest trained and validated on smaller geographic areas, when compared to the same area trained using the complete dataset (paired t-test; $p=0.4766$). As there is no increase in model performance when training and validating over smaller geographic areas, this would indicate there is no upper limit to modelled area size, as long as training data captures the variance in canopy height and predictor variables. For some locations a large error (>10%) was observed from a random forest model trained and validated on a smaller area when compared to the whole area model, this is attributed to a small training dataset ($N<150$). Improvements in model performance were seen for an area where mean ALS derived canopy height was 37 m. Taller forests were generally less well represented in the study-area wide model, therefore for areas of extreme canopy height values a local model yields improved results.

5.3.4. *Simulating disparate ALS capture for training a random forest*

When training the model with non-randomly distributed sample points (e.g. simulating aggregation of smaller ALS acquisitions) achievable accuracy reaches an asymptote at ~6% of the total forest area (Figure 33). RMSE and variance in estimates is greater than when compared to a stratified random sample approach. It is suggested that the ~3% increase in RMSE is caused by random forest over-fitting to the training data, therefore extrapolation beyond the training areas is impeded (Mascaro et al., 2014). The large estimate variance is due to the training data either not

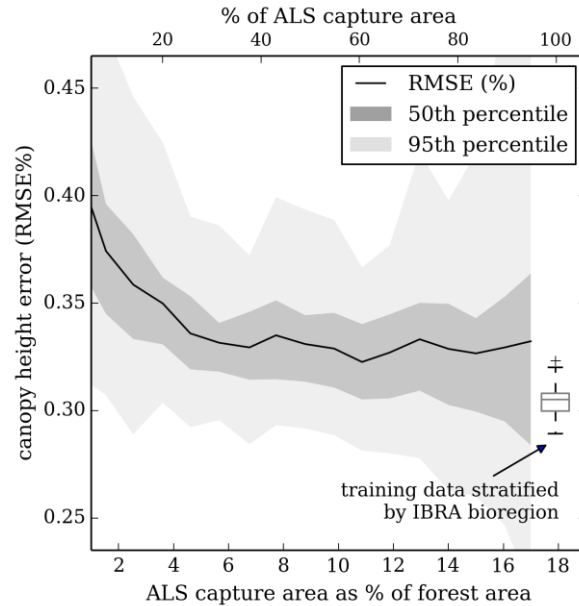


Figure 33. Error in canopy height estimates when constructing random forest models from ALS data selected to represent a combination of a number of disparate (non-random) acquisitions. Model output was validated with ALS plots from outside the training area. For comparison (see boxplot), the results from bootstrapping ($N=50$) random forest trained with a random stratified (by IBRA bioregion) sample from across the whole study area ($\sim 18\%$ of forested area) is included.

capturing the variance within the withheld sample (larger error) or the training and withheld samples having similar canopy height distributions (smaller error). This is exemplified by RMSE of $>40\%$ for a training sample derived from $>90\%$ of the acquisition area (Figure 33), in these instances areas of taller and shorter forest plots were not included in the training cohort.

5.4. Discussion

This manuscript demonstrates a method for assessing canopy height, over a large area, where forest structure is heterogeneous and canopy height ranges from 0 – 70 m, using freely available predictor data and in an open source computing framework. Forest canopy height was modelled from satellite imagery and ancillary data with a two-stage approach, where ALS

data captured over 18% of the study area was used to train the ensemble regression tree model, random forest. Two validation approaches were used, both of which indicate a good agreement between measured and modelled canopy height (RMSE \leq 31% at the 95th percentile confidence interval). Model error is at the upper limit of acceptable error as stipulated by the European Space Agency (ESA) in the upcoming BIOMASS project when estimating forest biomass from canopy height (Seifert et al., 2015). However the ESA target resolution for a canopy height product is much coarser (200 m) than the one presented in this study (30 m).

Previous studies have used regression and machine learning techniques to model canopy height, however these studies have been limited to forests with a maximum canopy height $<$ 30 m or plantations. For example, Mora et al. (2013) used high spatial resolution imagery to estimate canopy height for a 7000 ha area of conifer forest, reporting errors of 21% using a k-Nearest Neighbour method. Using a segmented Landsat image to estimate height over 707,000 ha of coniferous forest, Wulder and Seemann (2003) reported a standard error of 3.3 m. Applying random forest Ahmed et al. (2015) and Cartus et al. (2012) estimated canopy height with an RMSE of \leq 3.5 m and $<$ 1.7 m for managed coniferous and eucalyptus forests respectively. When using regression trees to classify plots in height classes ranging from 0 to $>$ 50 m, Peterson and Nelson (2014) significantly underestimated canopy height for plots $>$ 50 m when compared to inventory data. A comparison of model output (resampled to 1 km) with the Simard et al. Simard et al. (2011) global canopy height product reveals discrepancies between the two approaches of up to \pm 20 m over tall and short forest.

The capture of ALS over large areas is still uncommon and previous studies have shown that, over relatively homogeneously forested landscapes, acceptable results can be obtained from a capture of $<$ 1% of the forest area (Wulder and Seemann, 2003). If a wall-to-wall ALS acquisition is unrealistic, a sample cohort could be created by combining a number of smaller (existing) acquisitions. Results for this study area would suggest that RMSE reaches an asymptote when ALS is acquired over \sim 6% of forested area. However, a non-random sample returned a larger error when compared to a stratified random sample, even when comprised of plots from over a relatively large area. It is suggested that with a targeted

ALS sampling strategy, total area acquired could be reduced. However, this would require detailed *a priori* knowledge of forest structure or segmentation of a synoptic predictor dataset to infer forest structure variability (McRoberts and Tomppo, 2007).

Systematic error was apparent in the tails of modelled canopy height distribution, a similar systematic error was evident in other studies that used random forest to model canopy height and biomass (Baccini et al., 2008; Kellndorfer et al., 2010; Peterson and Nelson, 2014). Two methods were tested to reduce the error; a resampling of the response variable to a uniform distribution and subtracting an estimate of the systematic error component from modelled output. The latter technique proved most successful in recreating the range of canopy heights evident in the training data. However the transformation is non-discriminate when rescaling canopy height values and therefore inevitably introduced noise to the modelled output (e.g. rescaling values correctly modelled by random forest). This is evident from there being a minimal overall improvement in model performance after subtracting the modelled error component.

Overall, reflectance predictor variables were far more important in the model than other data sources. This would suggest that disturbance has a far greater influence on determining canopy height than underlying site condition or climatic processes that constrain maximum canopy height. Mascaro et al. (2014) found the addition of coordinate variables within the model greatly improved accuracy when estimating biomass, however in this instance this was not the case. The preference of reflectance based model drivers may also indicate that the complex set of environmental variables that limit canopy height are not captured within the datasets used, although the low spatial autocorrelation of model error may suggest otherwise. Furthermore, the resolution of the ancillary datasets were generally much coarser and required resampling, therefore these variables would not have adequately captured the within pixel variance. It should be noted that the specific variables and their relative importance will not be universally applicable across all forests outside of the study area, or in previous or subsequent years (Foody et al., 2003). For example, previous and subsequent years would have to be treated as independent and therefore new models created for each (Hudak et al., 2012b).

By far the most important variable was the Tasseled Cap (TC) wetness feature, in particular the mean value calculated for a kernel size of 3 x 3 Landsat TM pixels (90 m x 90 m). Calculating a mean value over a kernel, as opposed to individual pixel values, limits the impact of pixel level noise (Armston et al., 2009). Variable importance calculated for 50 km x 50 km sub areas, across a wide range of forest types and environmental gradients, also consistently ranked TC wetness as the most important variable. The TC wetness feature is driven by contrast between the visible and infrared and short wave infrared wavelengths, highlighting moisture gradients in a scene (Crist and Cicone, 1984). This would indicate that the canopies of taller, denser forests contain more moisture (owing to decreased temperatures and increased evapotranspiration) and shorter forest canopies are more arid, a paradigm that fits the environmental gradients of the study area (Gellie et al., 2005; Jenkins and Coops, 2011). Previous studies have highlighted the strong association between forest structure and the TC wetness feature (Cohen and Spies, 1992; Pascual et al., 2010) and middle/short wave infrared wavelengths (Steininger, 2000; Baccini et al., 2008). However, a linear regression of TC wetness (3 x 3 pixels) and canopy height returns a fairly weak statistical association ($r^2 = 0.35$), highlighting the requirement for a more complex statistical approach.

Although successful results were obtained, there are a number of potential sources of error worthy of discussion. For example, there is up to 2 years between ALS and Landsat TM acquisition and up to 5 years between ALS capture and plot measurements. An assessment of forest inventory plots that have been revisited (a total of 60 state wide) reveals that absolute mean change in canopy height is ~0.5 m per annum. This would therefore suggest that changes in canopy height are minimal at plots that have not been affected by fire or logging in the interim years. Another potential source of error is the limited extent of the ALS capture that was restricted to the riparian zone. Vegetation composition, and therefore structure, is known to differ from non-riparian areas, such as having a lower proportion of Eucalypt species (Lindenmayer, 2000). However, ALS transects were the same width along the entire reach of the river and therefore the proportion of riparian vegetation within the sample decreased in the upper catchments. The strong statistical association of forest inventory data from outside the ALS acquisition extent (and therefore away from riparian vegetation) with modelled canopy height would indicate that the impact of

the limited sample extent was negligible. Forest management practices area poorly represented (e.g. Figure 30 plot A) in the model, where clear-felling practices (1990 - 2010) accounted for <1% of the ALS extent. Under sampling of actively managed areas is partly a result of logging coupes and plantations requiring a 200 m riparian retention strip (VicForests, 2014).

The analysis presented in this manuscript was achieved using an open source computing framework, in addition, model predictor variables are publically available with a global coverage. Therefore, if training data is available (e.g. canopy height measurements from ALS or forest inventory) the presented methodology could be easily adopted by scientists and land management agencies who wish to map canopy height over large areas. Furthermore, with the planned launch of the Global Ecosystem Dynamics Investigation (GEDI) space borne LiDAR mission in 2018, the opportunity to replace costly ALS or inventory acquisitions with a freely available and spatially continuous sampling method is presented (within the temperate and tropical latitudes) (Dubayah and Schaaf, 2015).

5.5. Conclusions

This study presents a method for estimating canopy height at a large-area (i.e. millions of hectares) scale at a 30 m resolution. Application of this method was demonstrated across 2.9 million hectares of heterogeneous forest, comprising a broad range of forest types from open woodland to temperate rainforest, where canopy height ranged from 0 – 70 m. Canopy height was estimated using a two-stage approach, firstly a random forest ensemble regression tree model was trained with ALS derived canopy height, canopy height estimates were then upscaled to the large area using the synoptic datasets. This was achieved utilising existing ALS data (i.e. not a bespoke acquisition) in conjunction with synoptic medium resolution freely available satellite imagery (e.g. Landsat Thematic Mapper (TM) and Moderate Resolution Imaging Spectroradiometer). Root mean square error in estimated canopy height was $\leq 31\%$ (~ 5.6 m) when validated with (a) cross-validation of ALS derived canopy height and (b) a network of forest inventory plots from outside the ALS extent.

Systematic error was evident in model output where taller and shorter forest plots were under and overestimated respectively. This was corrected for by subtracting an estimate of systematic error, derived from a

linear regression of model residuals, from model output. It should be noted that correcting for systematic error did not improve overall model estimates, as plots closer to the mean canopy height were incorrectly rescaled. It is a modellers prerogative to estimate broad trends (e.g. using the random forest output) or more accurately attribute outlier values (e.g. subtracting an estimate of systematic error). The model was predominantly driven by reflectance data, in particular the Landsat TM Tasseled Cap transformed wetness feature. This would suggest that in the study area disturbance is the primary constraint on canopy height.

The method framework is designed to be easily adopted by land management agencies. It was shown that the combination of disparate ALS acquisitions (i.e. a non-random sample) covering ~6% of the forested area could be used to successfully estimate canopy height, although results were slightly worse than for a random stratified sample of the entire study area. In the absence of suitable ALS data, the use of forest inventory plot data to train the random forest model is suggested, if the inventory plot network is sufficiently representative. Alternatively, the planned Global Ecosystem Dynamics Investigation (GEDI) space borne LiDAR will offer a near global sampling of forest structure and furthermore will be freely available. This, coupled with the scalability of random forest, suggests application of the technique at a continental or global scale would be entirely feasible.

6. Chapter 6

Summary and synthesis

6.1. Summary of results

The development of metrics and techniques to assess biospheric attributes, which are universally applicable across large-areas, is becoming increasingly important as the extent and impact of anthropogenic modifications to the biosphere and atmosphere are becoming apparent (Schaaf et al., 2008; Pereira et al., 2013). Remote sensing techniques have been identified as a key tool with which to measure and model metrics, as technological improvements continue to provide more extensive, detailed, accurate and timely information on the biophysical properties of the earth's surface (Hollmann et al., 2013; Skidmore et al., 2015). The requirement for metrics and techniques to assess ecosystems over large and heterogeneous areas is exemplified by the needs of sustainable forest management, and this was the motivation of the thesis.

Large area attribution, particularly in remote or inaccessible areas has proved an on-going challenge for land managers. Heterogeneity in forested landscapes also presents challenges as assessment solutions have to be capable of generalising to varying land cover, possibly without *a priori* knowledge of forest type. Two forest structure metrics were chosen to investigate techniques for upscaling plot scale measurements to large-area attribution; these were canopy height and canopy vertical structure. Chapter 1 provides a broad introduction to the role and importance of forests in a global context, legislative frameworks for their assessment and past and current techniques for forest inventory, with an emphasis on remote sensing and LiDAR. Chapters 2 – 5 focus on answering the research objectives outlined in Section 1.3, key results from which are discussed below. This final chapter provides a summary and synthesis of the major research outcomes, highlighting the benefits of large area analysis of forests and frame the current work in future remote sensing programmes and technologies applicable to a sustainable forestry management context.

Outlined in Table 9 are the airborne laser scanning datasets used in the preparation of this thesis. Determination of appropriate canopy height and canopy height profile metrics was accomplished using the CRC-SI 2.07 dataset. Although the three study areas were located within 300 km of each other, they covered a broad range of forest structure types representative of south eastern Australia; from short open woodland with a discontinuous canopy to very tall and dense cool and temperate

rainforest. To investigate Question 2, the TERN/Auscover datasets were used, this incorporated additional structural variability from sites across Australia, such as savanna and tropical rainforest. The TERN/Auscover sites were not included in the determination of appropriate metrics investigated in Question 1 owing to a lack of detailed forest inventory. To investigate upscaling with passive satellite imagery, the River Health dataset was utilised, as this provided continuous ALS data over a much larger extent. Importantly, vegetation metrics did not require tailoring or recalibration when applied to different ALS instruments, capture specifications or vegetation types. This would indicate the presented metrics and techniques are transferable and robust when generating plot level vegetation structure estimates.

Table 9 Airborne laser scanning datasets utilised

Dataset name	Capture date	Geographic area	Instrument	Mean pulse density (pl m ⁻²)
CRC-SI 2.07	April 2012	Three 25 km ² study areas in Victoria, Australia (Figure 2).	Riegl LMS-Q560	9
TERN/Auscover	2012 - 2013	As above plus an additional three 25 km ² study areas in Queensland, Western Australia and the Northern Territory, Australia (Figure 18).	Riegl LMS-Q560	9
DELWP River Health project	December 2009 – January 2011	5200 km ² capturing the riparian strip in the East Gippsland and Northern Catchment Management Authority areas, Victoria Australia (Figure 25).	Leica ALS50-II and Leica ALS60	4

Question 1. Which metrics of canopy height and canopy vertical structure are suitable for application across forested landscapes?

The empirical statistical association between field measured canopy height and ALS derived analogues were robust across a range of forest types. This would imply a 'broad-brush' metric based approach can be applied to ALS surveys captured over large areas, and a tailored metric or calibrated model (e.g. multiple regression of ALS derived variables) for specific areas or forest types unnecessary. The applicability of using a single metric approach was demonstrated when answering Questions 2 and 3 where canopy height was estimated across a range of forest types as well as over a large area of heterogeneous forest (~3 million ha). For both examples this was achieved using the 95th percentile height of non-ground ALS returns. A comparison of ALS and inventory measurements at 30 points from within the DEWLP River Health study area resulted in a good agreement between the two techniques (RMSE = 3.3 m) with no bias towards forest type.

Using a mixture model to characterise or parameterise canopy vertical structure was less effective when applied across heterogeneous forested landscapes. This was due to mixture models requiring an *a priori* assumption of the number of model components i.e. k or the number of canopy strata, used to characterise foliage distribution. A new technique was introduced that utilised ALS to dynamically attribute canopy vertical structure by estimating the Number of Strata (NoS) present. The technique generated a generalised curve of ALS derived gap probability (P_{gap}) that allowed the identification of canopy scale features. Furthermore, it was shown a universal parameterisation was sufficient to generalise P_{gap} across forest types, negating the requirement for a forest type specific parameterisation. This was demonstrated in Chapter 4, where NoS was calculated for a range of forest types and simulated capture specifications without recalibrating α .

The new technique would allow for existing methods, such as mixture modelling or cluster analysis, to be applied dynamically across a forested landscape. This would improve the fit of the curve to the canopy height profile when compared to the pseudo-waveform generated to calculate NoS and allow parameterisation of the upper canopy and shade tolerant or sub-canopy strata beneath (Figure 34).

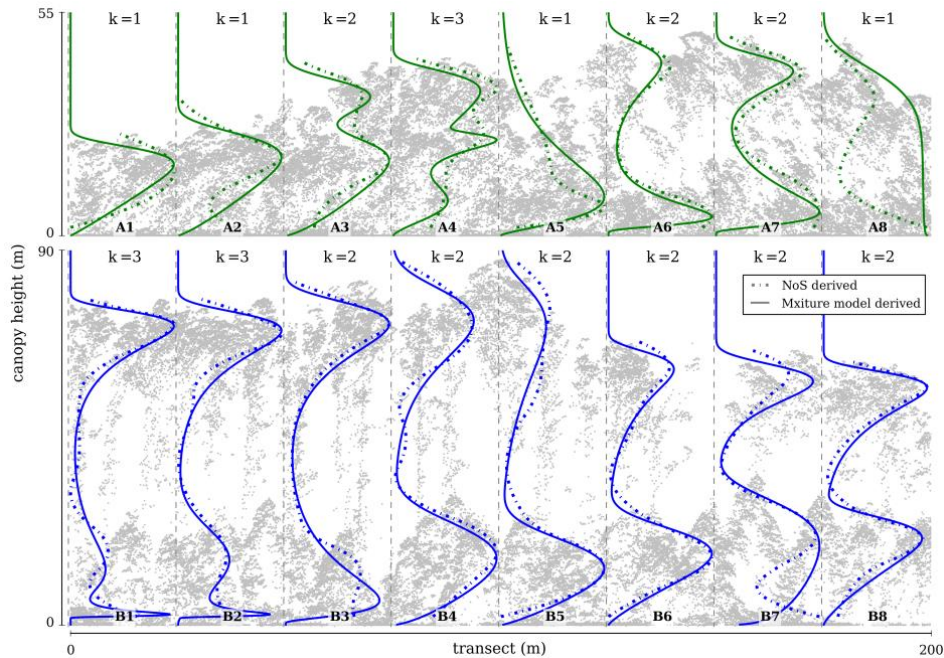


Figure 34. Canopy height profiles, generated with parametric mixture models and pseudo-waveforms from the derivative of $1 - P_s$, fitted to 25 m x 25 m ALS plots along a 200 m x 25 m transect (dashed lines identify plot boundaries). *NoS* values have been used as an *a priori* estimate of the number of model components. Compare with Figure 6.

Validation of *NoS* proved difficult, owing partly to the lack of agreement in techniques to characterise canopy vertical structure or quantify the number of canopy strata (Parker and Brown, 2000) as well as the number and remoteness of inventory plots. The Geometric Crown Volume Model approach, although using an independent dataset, is still an (abstract) model of a forest plot and assumptions, such as an ellipsoidal crown archetype, will have compromised the quality of the comparison. The use of 3D tree reconstruction, either through inventory measurements (Woodgate, 2015) or with terrestrial laser scanning (Burt et al., 2013), could provide a more robust validation dataset. This could be further coupled with a radiative transfer model which would allow parameterisation of both the forest and the ALS acquisition (Disney et al., 2010).

A component of the new technique to assess *NoS* was simulation of point clouds within a bootstrap. As far as it is known, bootstrapping of ALS point cloud configuration, either from a probability density function (as in Chapter 3) or with the original point cloud, is a new approach. This not only allows for the derivation of more robust estimates of vegetation structure, but could also be used to report confidence interval alongside metrics of vegetation structure. This would be particularly useful where pulse density is low, for example where pulse density <0.5 pulses m^{-2} as discussed in Chapter 4.

Question 2. What is the appropriate ALS sampling frequency for attribution of forest structure for different forest types?

Plot scale vegetation metrics can be accurately estimated from ALS acquired at a pulse density of 0.5 pulses m^{-2} (pl m^{-2}), when compared to a more dense acquisition (10 pl m^{-2}). This was demonstrated across a range of vegetation types, from open woodland to dense tropical rainforest. A new technique to systematically thin point cloud data was developed to investigate Question 2. This was required as existing techniques were (a) incapable of recreating the systematic way in which ALS instruments capture data, particularly at low pulse densities, and (b) not suited to multi-return instruments where more than two returns can be recorded for each outgoing pulse.

ALS capture still tends to be a one-off acquisition at a relatively high pulse density (>4 pl m^{-2}), and repeated surveys are unusual, particularly over large-areas and native forests. As the new algorithm selects points in a systematic way, this allows simulation of repeat capture over the same plot. Results indicated that at a pulse density of 0.5 pl m^{-2} canopy structure is well represented and comparable to higher pulse densities. This would suggest that quantification of change in vegetation structure over time by repeat capture is feasible (Hudak et al., 2012b), although this does not take into consideration differences in capture parameters, such as scan angle or atmospheric conditions.

Advances in LiDAR instrument technology have seen an exponential increase in pulse repetition frequencies from 2 KHz to upwards of 800 KHz in little over a decade. The prevalence of commercially available small-footprint full-waveform instruments allows for a greater number of returns

to be identified for a single out-going pulse. Furthermore, computational power and data storage has increased, allowing for efficient processing of large datasets. However, acquisition costs (mobilisation, wages, fuel costs etc.) have remained relatively stable. It is suggested therefore, the trade-off between an acceptable pulse density and the extent of capture may become decoupled as acquisition cost become the dominant factor.

Question 3. How can plot level estimates of canopy structure be scaled to generate continuous regional maps?

Building on the knowledge from the previous two research questions, a map of canopy height was produced for 2.9 million hectares of forest with heterogeneous structure. This was achieved by using ALS data to train a random forest regression model, where synoptic satellite imagery was used as predictor data. As the definition of large-area used in this thesis has no upper limit, it is proposed the technique presented in Chapter 5 could be applied to generate a global scale canopy height product at a resolution not yet achieved. This would be possible owing to the scalability of the random forest technique, which has been demonstrated in a number of previous studies (Clark et al., 2010; Simard et al., 2011). A similar approach for estimating global deforestation was used by Hansen et al. (2013) who utilised the Google Earth Engine cloud computing facility.

A pilot study to predict canopy vertical structure using random forest and satellite imagery was successful when applied to the CRC-SI 2.07 sites (Wilkes et al., 2014). However, attempts to upscale estimates of canopy vertical structure over the River Health area was less successful and a number of reasons are suggested for this. Firstly, the CRC-SI 2.07 data is from three distinct sites where structural attributes and reflectance characteristics are somewhat unique to each site and therefore easier to classify. This is in contrast to the River Health dataset which covered a continuum of forest types and structural characteristics. Secondly, the CRC-SI 2.07 sites were exemplar of the forest type they represented, this meant there were no recent disturbances or anthropogenic land use, therefore variance in vertical forest structure caused by disturbance would be minimised. The River Health extent captured areas previously (and recently) affected by disturbances including planned burns, bushfires, drought and active forest management which adds complexity to the model. Thirdly, canopy vertical structure is more locally variable than

canopy height (Figure 11), making identification of patterns in predictor datasets more challenging for machine learning methods. The inclusion of additional data sources, for example radar tomography (Caicoya et al., 2014), may in the future improve canopy vertical structure attribution over large areas.

A suggested further avenue of research is to investigate a move towards a monitoring framework where maps of canopy height change are produced. It is suggested this could be achieved using training data that is captured with less regularity than predictor variables e.g. the revisit time of permanent inventory plots is ~5 years, whereas cloud free image mosaics could potentially be created seasonally. The challenge moving from an assessment to a monitoring programme would be (potentially large) shifts in predictor variable values caused by dynamic environmental conditions (e.g. drought to wet), meaning each epoch would have to be treated independently (Hudak et al., 2012b).

6.2. *Identifying trends in large-area forest structure*

Tools presented in this thesis allow for analysis of trends in large-area forest structure that may not be apparent at more local scales. Presented in Figure 35 are results from generating canopy height profile attributes from the River Health dataset for 11,000 plots across 2.9 million hectares. This analysis reveals trends in forest structure that an analysis of canopy height alone may not identify. For example it is suggested three broad and distinct forest classes occur within the study area, identified in Figure 35 as A, B and C. Classes were assigned by determining the height of the vegetation density modal peak for the dominant canopy strata. As there is a large overlap between classes when considering canopy height alone, classes would be hard to distinguish without the inclusion of this additional information. When the classification is applied across the study area it identifies classes A, B and C as short open woodland (including alpine vegetation), mixed species forests and temperate rainforest (Figure 36). A suggested reason for the increase in canopy depth with increasing maximum canopy height within each class is the shift from a canopy comprised of a dense population of co-dominant trees to a canopy of fewer dominant individuals. This transition reduces competition for light and therefore changes crown archetype from an inverted cone, where foliage

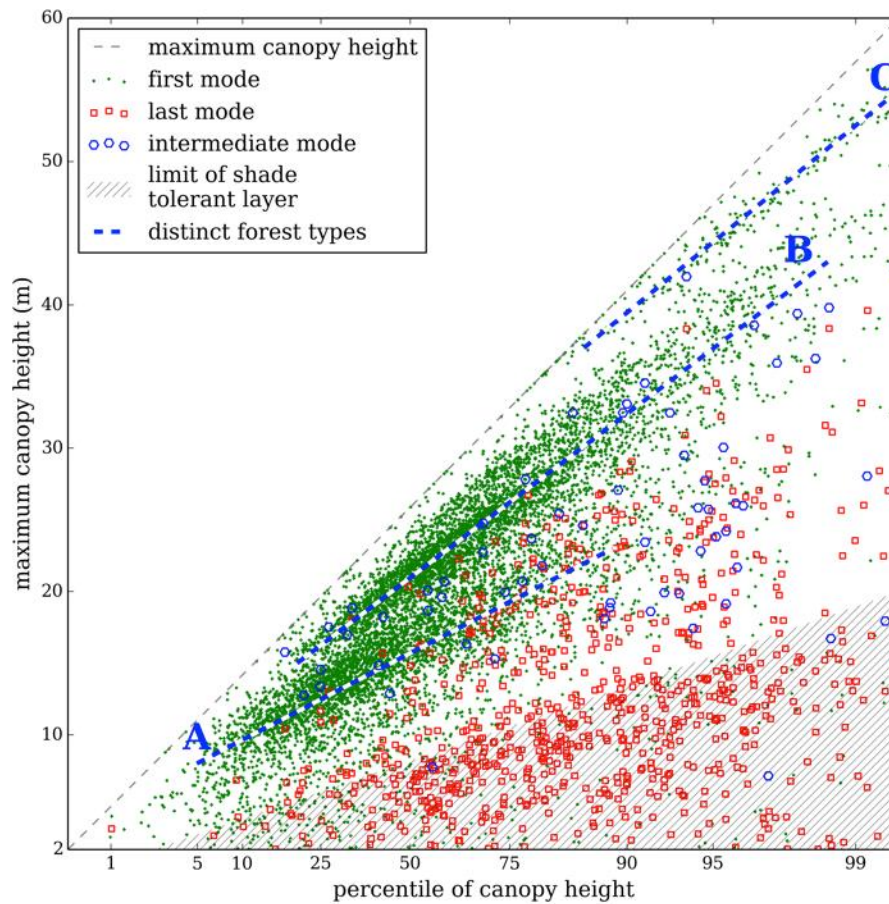


Figure 35. Landscape forest structure captured with ALS transects for a total area of 2.9 million ha. Green, red and blue markers indicate the height of the dominant, shade tolerant and intermediate strata that were calculated using the pseudo-waveform described in Chapter 3. Also illustrated is the suggested limit of the shade tolerant strata and the presence of 3 distinct forest structure types (labelled A - C).

clumps towards the top, to an ellipsoidal shape where the maximum in foliage density is shifted downwards within the crown.

The presence of a second modal peak in the canopy height profile is indicative of a shade tolerant or sub-dominant stratum beneath the dominant canopy. This strata begins to occur in forests where maximum canopy height >10 m, with increasing maximum canopy height the height of the shade tolerant strata increases proportionally (shaded area in Figure

35). This is consistent with the vertical canopy transects presented in Figure 6. It is likely the species composition of these strata changes with increasing maximum canopy height, from larger shrubs in open woodland to a shade tolerant tree strata in taller rainforest (e.g. *Acacia* and *Nothofagus* spp.).

An intermediate strata occurring between the shade tolerant strata and the dominant canopy is uncommon, occurring in <0.5% of all plots and ~3% of plots where canopy height >40 m. The absence of this stratum in tall forest is clearly visible in Figure 2 where a large gap is apparent between the shade tolerant strata and dominant canopy. The presence of an intermediate strata in tall forest is the product of a partial stand replacing disturbance which has led to the initiation of a second generation of dominant canopy species. Locations where an intermediate strata exists are globally important biomass hotspots where carbon density can reach 2,844 tonnes of carbon per hectare (Keith et al., 2009).

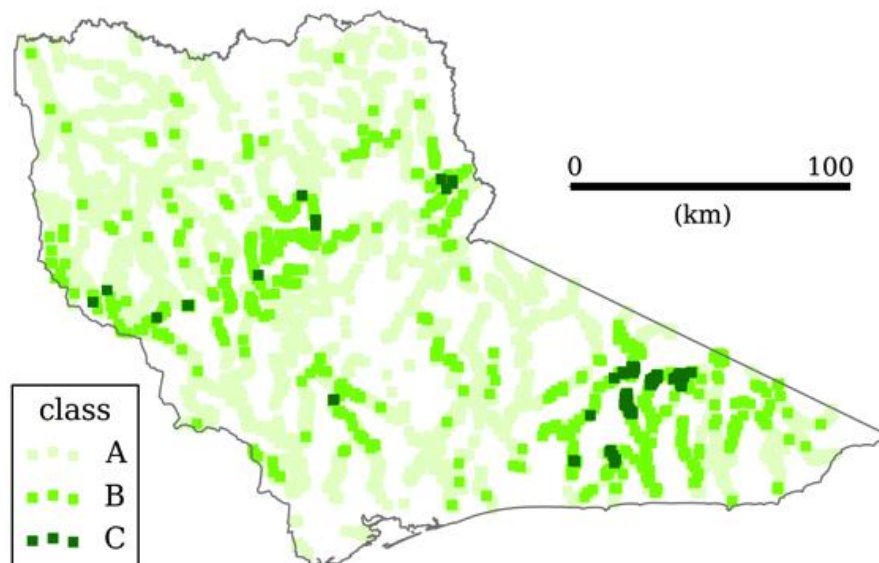


Figure 36. Classification of forest type according to the three types identified in Figure 35.

The application of the canopy vertical structure method presented in Chapter 3 across a large area highlights the additional information that can be extracted from ALS data to characterise forest structure. It also indicates that when parameterising the model so it is suitable for large area analysis, not all plots will be attributed correctly. This is illustrated in Figure 35 where, for certain plots, the last mode is identified high in the canopy profile i.e. above the shade tolerant zone. A suggested improvement to the model would be the capability to correctly identify understorey shrubs e.g. vegetation <2 m or ground (as with the processing of large-footprint full-waveform data) and therefore characterise the complete canopy profile.

6.3. *Remote sensing in sustainable forest management: a future perspective*

This thesis aimed to shift emphasis away from tailored approaches suited to a single study site or vegetation type and towards generic solutions which are applicable across large and heterogeneous forested areas (Figure 37). Although this may have not been fully realised here owing to the “proof of concept” requirement, this thesis presents metrics, acquisition specifications and processing frameworks for generating large area forest attribution. This culminated in a framework for the production of a continuous map of canopy height where, with regard to Figure 37; canopy height was estimated over a large area, using a universally applicable metric, computation was achieved with open-source software, using a machine learning technique which utilised multiple freely available datasets (regarding both cost and accessibility).

A requirement for the paradigm shift is prompted by the need for routine assessment of forests over large areas in a consistent, transferable and transparent manner that has not yet been achieved (Trumbore et al., 2015). The need for assessment is driven by new and existing national and international initiatives and legislative commitments on reporting forest condition as well as monitoring environmental degradation and ensuring the provision of ecosystem services. Examples of these include the UN-REDD (FAO et al., 2008), the Montreal Process (Montréal Process Working Group, 2009), the Essential Climate Variables (Schaaf et al., 2008) and Essential Biodiversity Variables (Pereira et al., 2013).

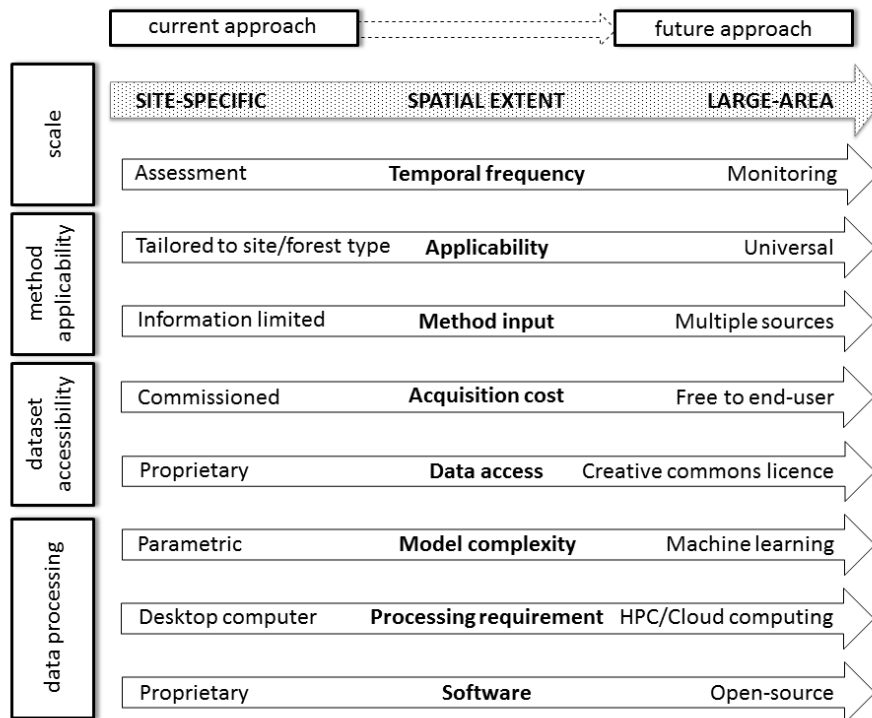


Figure 37. A conceptual model to illustrate the paradigm shift in large-area forest assessment utilising remote sensing.

Setting targets for the accounting of ecosystem services across large areas has been made possible by considerable investment and improving remote sensing technologies (Hollmann et al., 2013; Skidmore et al., 2015). For example, the ESA's Copernicus and BIOMASS projects aim to provide accurate and timely remote sensing data for environmental monitoring. The latter will utilise radar to estimate changes in global forest biomass, this will require a global estimate of canopy height at a resolution of 200 m on a 6 monthly basis (Seifert et al., 2015). Therefore it is suggested the paradigm shift will ultimately extend from an assessment framework to include monitoring in a timely fashion (Lewis et al., 2013).

Recent advances in remote sensing technologies already allow for data to be captured across large areas at applicable resolutions. For example, current global canopy height products highlight the potential for earth observing techniques to characterise vegetation structure and biomass

(Lefsky, 2010; Simard et al., 2011; Los et al., 2012). However, as demonstrated in Chapter 6, these products are yet to be ‘locally relevant’ for Australian forests, a key requirement of large area assessment (Hansen et al., 2013). Furthermore, products such as these are still produced as scientific exemplars, and as yet operational products – useful in a monitoring context – have not been attained. This is partly due to the lack of continuity of earth observation datasets, highlighted as a major issue with monitoring and ultimately tackling ecosystem degradation (Carpenter et al., 2006).

Looking towards the future, the cost of information to the end-user will continue to decrease. Concurrently, the volume and availability of remote sensing data will increase with ever increasing spaceborne remote sensing platforms and terrestrial sensor networks. The recent launch of Landsat 8 has seen the continuation of one of the most beneficial earth observation programmes to forest and landscape science. Planned missions, such as the ESA’s Sentinel constellation and BIOMASS radar programmes, will continue to supply data for forest characterisation into the future. Combining multiple data sources (with different spatial, spectral and angular characteristics) has been recognised as a major challenge of large area assessment and monitoring programmes (Lewis et al., 2013).

For large area vegetation structure assessment, the opportunity presented by spaceborne LiDAR is unrivalled. Studies that have utilised data from the now defunct IceSAT mission have illustrated the utility of spaceborne LiDAR for calibrating models of canopy height and biomass. The planned deployment of the Global Ecosystem Dynamics Investigation LiDAR (GEDI) on the International Space Station in 2018 presents an opportunity to once again measure and map canopy structure on a global scale. A major advantage of GEDI is the near continuous spatial coverage of laser footprints (Dubayah and Schaaf, 2015), this is compared to IceSAT’s repeated ground track coverage which had a relatively limited coverage. The move towards spaceborne ranging sensors and radar tomography may eventually negate the requirement for airborne platforms as a sampling tool for vegetation structure assessment.

With the reduction in instrument costs and the increasing ubiquity of the *Internet of Things*, terrestrial wireless sensor networks will allow for ever increasing profusion of data at the plot scale, where measurements will still

be required for calibration and validation purposes. Examples of sensor networks include VEGNET, an *in situ* LiDAR system that scans the forest canopy on a nightly basis and records changes in plant area index (Culvenor et al., 2014), *in situ* digital repeat photography for measuring canopy greenness (Keenan et al., 2014) and eddy covariance methods that measure fluxes of mass and energy between the canopy and atmosphere (Baldocchi, 2014). With the increased availability of cheap and light-weight laser scanners, ALS as a calibration or validation tool may soon be replaced by terrestrial laser scanning (Calders, 2015) or the use of unmanned airborne vehicles (Wallace et al., 2012).

With the potential of data from numerous satellite and terrestrial sensors, computing capability to process large and complex datasets will also need to respond accordingly. This will represent a shift from desktop machines to high performance computing (HPC) servers housed in centralised data centres where large-scale parallel computing is possible. Utilisation of such technology has already begun, for example Hansen et al. (2013) processed >650,000 Landsat scenes with the Google Earth Engine (GEE). The GEE is a cloud computing service which allows users to process large amounts of archived satellite remote sensing data through a browser based Javascript interface. Facilities such as the GEE therefore open up the satellite archive, making it available to scientists around the world regardless of location and computing infrastructure.

The requirement for data sharing has been recognised as key to ensuring continued progress against ecosystem degradation (Wulder et al., 2008b; Roy et al., 2014; Turner et al., 2015). Open-access data policies are becoming more common as organisations are realising the benefits of data sharing. For example since December 2008, Landsat data from United States Geological Survey has been available under an open data policy which allows free access and redistribution of Landsat data products. This has increased the utilisation of Landsat data significantly, from <30,000 scenes downloaded in 2007 to >5,500,000 scenes accessed in 2014 (Turner et al., 2015). Within Australia, the TERN/Auscover facility captures and disseminates remote sensing and biophysical data under a Creative Commons Licence. Open-access policies are also extending to the publication of scientific computer programming source code in open-access

repositories, where in most cases this is written for open-source software packages.

7. References

- Aber, J. D. (1979). A Method for Estimating Foliage-Height Profiles in Broad-Leaved Forests. *Journal of Ecology*, 67(1), 35–40.
- 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 Journal of Photogrammetry and Remote Sensing*, 101, 89–101. doi:10.1016/j.isprsjprs.2014.11.007
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705
- Aldred, A., & Bonnor, G. (1985). Application of airborne lasers to forest surveys. In *Information Report PI-X-51*. (p. 62). Chalk River, Ontario, Canada: Petawawa National Forestry Institute.
- American Society of Photogrammetry and Remote Sensing. (2008). LAS Specification 1.2, 1–13.
- Andersen, H. E., Reutebuch, S. E., & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing*, 32(5), 355–366. doi:10.5589/m06-030
- Angelo, J. J., Duncan, B. W., & Weishampel, J. F. (2010). Using Lidar-Derived Vegetation Profiles to Predict Time since Fire in an Oak Scrub Landscape in East-Central Florida. *Remote Sensing*, 2(2), 514–525. doi:10.3390/rs2020514
- Armston, J. D., Denham, R. J., Danaher, T. J., Scarth, P., & Moffiet, T. N. (2009). Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. *Journal of Applied Remote Sensing*, 3(1), 033540. doi:10.1117/1.3216031
- Armston, J. D., Disney, M. I., Lewis, P., Scarth, P., Phinn, S., Lucas, R. M., Bunting, P., & Goodwin, N. R. (2013). Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sensing of Environment*, 134, 24–38. doi:10.1016/j.rse.2013.02.021

References

- Asner, G. P., Powell, G. V., Mascaro, J., Knapp, D. E., Clark, J. K., Jacobson, J., Kennedy-Bowdoin, T., Balaji, A., Paez-Acosta, G., Victoria, E., Secada, L., Valqui, M., & Hughes, R. F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(38), 16738–16742. doi:10.1073/pnas.1004875107
- Austin, M. P. (1987). Models for the analysis of species' response to environmental gradients. *Vegetatio*, *69*(1-3), 35–45. doi:10.1007/BF00038685
- Axelsson, C., Jones, S., Haywood, A., Suarez, L., & Wilkes, P. (2012). Key attributes for monitoring and assessment of Australian forests : a land management perspective. In *GSR2*. Melbourne, Australia, 10 - 12th December.
- Baccini, A., Laporte, N. T. T., Goetz, S. J., Sun, M., & Dong, H. (2008). A first map of tropical Africa ' s above-ground biomass derived from satellite imagery. *Environmental Research Letters*, *3*(4), 1–9. doi:10.1088/1748-9326/3/4/045011
- Baker, P. J., & Wilson, J. S. (2000). A quantitative technique for the identification of canopy stratification in tropical and temperate forests. *Forest Ecology and Management*, *127*, 77–86.
- Baldocchi, D. D. (2014). Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method. *Global Change Biology*, 3600–3609. doi:10.1111/gcb.12649
- Baltsavias, E. (1999). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, *54*(2-3), 199–214. doi:10.1016/S0924-2716(99)00015-5
- Bater, C. W., Wulder, M. A., Coops, N. C., Nelson, R. F., Hilker, T., & Næsset, E. (2011). Stability of Sample-Based Scanning-LiDAR-Derived Vegetation Metrics for Forest Monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(6), 2385–2392.
- Bechtold, W. A., & Patterson, P. L. (2005). *The enhanced forest inventory and analysis program - national sampling design and estimation procedures*.

- Asheville, North Carolina: United States Department of Agriculture Forest Service.
- Beets, P. N., Brandon, A., Fraser, B. V., Goulding, C. J., Lane, P. M., & Stephens, P. R. (2010). New Zealand. In E. Tomppo, T. Gschwantner, M. Lawrence, & R. E. McRoberts (Eds.), *National Forest Inventories Pathways for Common Reporting* (pp. 391 – 410). doi:10.1007/978-90-481-3233-1
- Bolton, D. K., Coops, N. C., & Wulder, M. A. (2013). Measuring forest structure along productivity gradients in the Canadian boreal with small-footprint Lidar. *Environmental Monitoring and Assessment*, 185(8), 6617–34. doi:10.1007/s10661-012-3051-9
- Brack, C. (2007). National forest inventories and biodiversity monitoring in Australia. *Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology*, 141(1), 104–112. doi:10.1080/11263500601153875
- Bradshaw, G. A., & Spies, T. A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology*, 80(2), 205–215.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. doi:10.1007/BF00058655
- Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.
- Brenner, A. C., Zwally, H. J., Bentley, C. R., Csathó, B. M., Harding, D. J., Hofton, M. A., Minster, J. B., Roberts, L., Saba, J. L., Thomas, R. H., & Donghui, Y. (2003). Derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights. *Geoscience Laser Altimeter System (GLAS) Algorithm Theoretical Basis Document, version 4.*(September 2003), 93p. doi:10.1117/12.414116
- Brokaw, N., & Lent, R. (1999). Vertical structure. In M. L. Hunter (Ed.), *Maintaining biodiversity in forest ecosystems* (pp. 373–399). Cambridge, UK: Cambridge University Press.
- Burgman, M. A. (1996). Characterisation and delineation of the eucalypt old-growth forest estate in Australia: a review. *Forest Ecology and Management*, 83(3), 149–161. doi:10.1016/0378-1127(96)03725-5

- Burt, A., Disney, M. I., Raunonen, P., & Armston, J. D. (2013). Rapid Characterisation of Forest Structure From TLS And 3D Modelling. In *IGARSS* (pp. 3387–3390). Melbourne, Australia, 21st - 26th July.
- Caicoya, A. T., Bayer, D., Biber, P., Heym, M., Papathanassiou, K., Pardini, M., & Alonso, M. T. (2014). Perspectives and interpretation of forest 3D structure for future spaceborne SAR missions. In *ForestSAT*. Riva Del Garda, Italy, 4th - 7th November.
- Calders, K. (2015). *Terrestrial laser scanning for forest monitoring (PhD Thesis)*. Wageningen University.
- Carpenter, S. R., De Fries, R., Dietz, T., Mooney, H. A., Polasky, S., Reid, W. V., & Scholes, R. J. (2006). Millennium ecosystem assessment: research needs. *Science*, *314*, 257–258.
- Cartus, O., Kellndorfer, J. M., Rombach, M., & Walker, W. (2012). Mapping canopy height and growing stock volume using airborne lidar, alos palsar and landsat ETM+. *Remote Sensing*, *4*(11), 3320–3345. doi:10.3390/rs4113320
- Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data*. University of California, Berkeley.
- Clark, M. L., Aide, T. M., Grau, H. R., & Riner, G. (2010). A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment*, *114*(11), 2816–2832. doi:10.1016/j.rse.2010.07.001
- Coburn, C., & Roberts, A. (2004). A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing*, *25*(20), 4287–4308. doi:10.1080/0143116042000192367
- Cohen, W. B., & Spies, T. A. (1992). Estimating structural attributes of Douglas-fir/western hemlock forest stands from landsat and SPOT imagery. *Remote Sensing of Environment*, *41*(1), 1–17. doi:10.1016/0034-4257(92)90056-P
- Coops, N. C., Hilker, T., Wulder, M. A., St-Onge, B., Newnham, G. J., Siggins, A., & Trofymow, J. A. (2007). Estimating canopy structure of Douglas-fir

- forest stands from discrete-return LiDAR. *Trees*, 21(3), 295–310. doi:10.1007/s00468-006-0119-6
- Costanza, R., Arge, R., Groot, R. De, Farberk, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., Neill, R. V. O., Paruelo, J., Raskin, R. G., & Suttonkk, P. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387(May), 253–260. doi:10.1038/387253a0
- Crist, E. P., & Cicone, R. C. (1984). A Physically-Based Transformation of Thematic Mapper Data---The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(3), 256–263. doi:10.1109/TGRS.1984.350619
- Culvenor, D. S., Newnham, G., Mellor, A., Sims, N., & Haywood, A. (2014). Automated In-Situ Laser Scanner for Monitoring Forest Leaf Area Index. *Sensors*, 14(8), 14994–15008. doi:10.3390/s140814994
- Cutler, D. R., Edwards, T. J., Beard, K., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11), 2783–2792.
- d'Oliveira, M. V. N., Reutebuch, S. E., McGaughey, R. J., & Andersen, H. E. (2012). Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sensing of Environment*, 124, 479–491. doi:10.1016/j.rse.2012.05.014
- Department of Environment. (2012). *Australia's bioregions (IBRA)*.
- Department of Sustainability and Environment. (2012). *Victorian Forest Monitoring Program Guidelines For Ground Plot Measurement Standard Operating Procedure 13: Measuring a Large Tree Plot*. Melbourne, Australia.
- Di Gregorio, A., & Jansen, L. J. M. (1998). *Land Cover Classification System (LCCS): classification concepts and user manual*. Food and Agriculture Organization of the United Nations, Rome.
- Dierckx, P. (1993). *Curve and Surface Fitting with Splines*. Oxford: Oxford University Press.

References

- Disney, M. I., Kalogirou, V., Lewis, P., Prieto-blanco, A., Hancock, S., & Pfeifer, M. (2010). Simulating the impact of discrete-return lidar system and survey characteristics over young conifer and broadleaf forests. *Remote Sensing of Environment*, 114(7), 1546–1560. doi:10.1016/j.rse.2010.02.009
- Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., Blair, B. J., & Rica, C. (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, 81, 378–392.
- Dubayah, R. O., & Schaaf, C. B. (2015). Global Ecosystem Dynamics Investigation. In *GOFC/GOLD Biomass workshop*. 24th - 26th February, Brisbane.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.
- Evans, J. S., & Cushman, S. A. (2009). Gradient modeling of conifer species using random forests. *Landscape Ecology*, 24(5), 673–683. doi:10.1007/s10980-009-9341-0
- Evans, J. S., & Hudak, A. T. (2007). A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 45(4), 1029–1038.
- Evans, J. S., Hudak, A. T., Faux, R., & Smith, A. M. (2009). Discrete Return Lidar in Natural Resources: Recommendations for Project Planning, Data Processing, and Deliverables. *Remote Sensing*, 1(4), 776–794. doi:10.3390/rs1040776
- FAO. (2000). *Forest Resources Assessment WP 33; FRA 2000 ON DEFINITIONS OF FOREST AND FOREST CHANGE*. Rome.
- FAO. (2010). *Global Forest Resources Assessment 2010*. Retrieved April 30, 2015, from <http://www.fao.org/forestry/fra/fra2010/en/>
- FAO. (2015). *Criteria and indicators for sustainable forest management*. Retrieved May 13, 2015, from <http://www.fao.org/forestry/ci/en/>

- FAO, UNDP, & UNEP. (2008). UN collaborative programme on reducing emissions from deforestation and forest degradation in developing countries. *Framework Document FAO, UNDP, UNEP*, (June), 27.
- 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 Sensing*, 5(1), 83–109. doi:10.3390/rs5010083
- Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., ... Snyder, P. K. (2005). Global consequences of land use. *Science (New York, N.Y.)*, 309(5734), 570–574. doi:10.1126/science.1111772
- 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 Sensing of Environment*, 85(4), 463–474. doi:10.1016/S0034-4257(03)00039-7
- Franklin, J. F., Denison, W., McKee, W. A., Swanson, F., Maser, C., Sedell, J., & Juday, G. (1981). *Ecological Characteristics of Old-Growth Douglas-Fir Fore*. United States Department of Agriculture Forest Service.
- Franklin, J. F., & Pelt, R. Van. (2004). Spatial Aspects of Structural Complexity in Old-Growth Forests. *Journal of Forestry*, (April/May), 22–28.
- Franklin, J. F., & Spies, T. A. (1991). Composition, Function, and Structure of Old-Growth Douglas-Fir Forests. In L. F. Ruggiero, K. B. Aubry, A. B. Carey, & M. H. Huff (Eds.), *Wildlife and Vegetation of Unmanaged Douglas-fir Forests. USDA Forest Service General Technical Report PNW-GTR-285* (pp. 71 – 80).
- Franklin, J., & Strahler, A. H. (1988). Invertible canopy reflectance modeling of vegetation structure in semiarid woodland. *IEEE Transactions on Geoscience and Remote Sensing*, 26(6), 809–825. doi:10.1109/36.7712
- Franklin, S. E., Wulder, M. A., & Gerylo, G. R. (2001). Texture analysis of IKONOS panchromatic data for Douglas-fir forest age class

- separability in British Columbia. *International Journal of Remote Sensing*, 22(March 2015), 2627–2632. doi:10.1080/01431160120769
- Gellie, N. J. H., & James, N. (2005). Native Vegetation of the Southern Forests: South-east Highlands, Australian Alps, South-west Slopes, and SE Corner bioregions. In *Cunninghamia* (Vol. 9, pp. 219–254).
- Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, 1(1), 45–54. doi:10.1080/01431160903252327
- Gobakken, T., & Næsset, E. (2008). Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, 38(5), 1095–1109. doi:10.1139/X07-219
- Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O. M., Gregoire, T. G., Ståhl, G., Holm, S., Ørka, H. O., & Astrup, R. (2012). Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment*, 123, 443–456. doi:10.1016/j.rse.2012.01.025
- Goetz, S. J., Steinberg, D., Dubayah, R. O., & Blair, B. J. (2007). Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment*, 108(3), 254–263. doi:10.1016/j.rse.2006.11.016
- Goodwin, N. R., Coops, N. C., & Culvenor, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, 103(2), 140–152. doi:10.1016/j.rse.2006.03.003
- GRASS Development Team. (2015). Geographic Resources Analysis Support System (GRASS) Software.
- Gregoire, T. G., Ståhl, G., Næsset, E., Gobakken, T., Nelson, R. F., & Holm, S. (2011). Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41(1), 83–95. doi:10.1139/X10-195

- Gschwantner, T., Schadauer, K., Vidal, C., Lanz, A., Tomppo, E. O., Di Cosmo, L., Robert, N., Duursma, D. E., Lawrence, M., Cosmo, L., Robert, N., Duursma, D. E., & Lawrence, M. (2009). Common tree definitions for national forest inventories in Europe. *Silva Fennica*, *43*(2), 303–321. doi:10.14214/sf.463
- Hancock, S., Armston, J. D., Li, Z., Gaulton, R., Lewis, P., Disney, M., Danson, F. M., Strahler, A. H., Schaaf, C., Anderson, K., & Gaston, K. J. (2015). Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy. *Remote Sensing of Environment*, *164*, 208–224. doi:10.1016/j.rse.2015.04.013
- Hansen, A. J., Phillips, L. B., Dubayah, R. O., Goetz, S. J., & Hofton, M. A. (2014). Regional-scale application of lidar: Variation in forest canopy structure across the southeastern US. *Forest Ecology and Management*, *329*, 214–226. doi:10.1016/j.foreco.2014.06.009
- Hansen, E., Gobakken, T., & Næsset, E. (2015). Effects of Pulse Density on Digital Terrain Models and Canopy Metrics Using Airborne Laser Scanning in a Tropical Rainforest. *Remote Sensing*, *7*(7), 8453–8468. doi:10.3390/rs70708453
- Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. A, Tyukavina, A., Thau, D., Stehman, S. V, Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(November), 850–853. doi:10.1126/science.1244693
- Haverd, V., Lovell, J. L., Cuntz, M., Jupp, D. L., Newnham, G. J., & Sea, W. (2012). The Canopy Semi-analytic P gap And Radiative Transfer (CanSPART) model: Formulation and application. *Agricultural and Forest Meteorology*, *160*, 14–35. doi:10.1016/j.agrformet.2012.01.018
- Hayashi, R., Weiskittel, A., & Sader, S. A. (2014). Assessing the Feasibility of Low-Density LiDAR for Stand Inventory Attribute Predictions in Complex and Managed Forests of Northern Maine, USA. *Forests*, *5*(2), 363–383. doi:10.3390/f5020363
- Haywood, A., & Stone, C. (2011). Mapping eucalypt forest susceptible to dieback associated with bell miners (*Manorina melanophys*) using

- laser scanning, SPOT 5 and ancillary topographical data. *Ecological Modelling*, 222(5), 1174–1184. doi:10.1016/j.ecolmodel.2010.12.012
- He, X., & Shi, P. (1998). Monotone B-Spline Smoothing. *Journal of the American Statistical Association*, 93(442), 643–650. doi:10.1080/01621459.1998.10473717
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. doi:10.1002/joc.1276
- Hill, R., & Thomson, A. (2005). Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, 26(17), 3763–3779. doi:10.1080/01431160500114706
- Hinsley, S., Hill, R., Fuller, R., Bellamy, P., & Rothery, P. (2009). Bird species distributions across woodland canopy structure gradients. *Community Ecology*, 10(1), 99–110.
- Hnatiuk, R. J., Thackway, R., & Walker, J. (2009). Vegetation. In *Australian Soil and Land Survey Field Handbook* (3rd Editio., pp. 73 – 125). Melbourne: CSIRO Publishing.
- Hofton, M. A., Minster, J. B., & Blair, B. J. (2000). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4), 1989–1996. doi:10.1109/36.851780
- Hollaus, M., Wagner, W., Eberhöfer, C., & Karel, W. (2006). Accuracy of large-scale canopy heights derived from LiDAR data under operational constraints in a complex alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(5), 323–338. doi:10.1016/j.isprsjprs.2006.05.002
- Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, a., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendaal, M., & Wagner, W. (2013). The ESA climate change initiative: Satellite data records for essential climate variables. *Bulletin*

- of the American Meteorological Society*, 94(10), 1541–1552. doi:10.1175/BAMS-D-11-00254.1
- Holmgren, J., & Jonsson, T. (2004). Large scale airborne laser scanning of forest resources in Sweden. *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI(8/W2), 157–160.
- Hopkinson, C., & Chasmer, L. (2009). Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, 113(1), 275–288. doi:10.1016/j.rse.2008.09.012
- Hopkinson, C., Chasmer, L., Lim, K., Treitz, P., & Creed, I. (2006). Towards a universal lidar canopy height indicator. *Canadian Journal of Remote Sensing*, 32(2), 139–152. doi:10.5589/m06-006
- Hopkinson, C., Lovell, J. L., Chasmer, L., Jupp, D. L., Kljun, N., & van Gorsel, E. (2013). Integrating terrestrial and airborne lidar to calibrate a 3D canopy model of effective leaf area index. *Remote Sensing of Environment*, 136, 301–314. doi:10.1016/j.rse.2013.05.012
- Hudak, A. T., Bright, B., Negron, J., McGaughey, R. J., & Hicke, J. A. (2012a). Predicting Live and Dead Tree Basal Area in Bark Beetle-Affected Forests from Discrete-Return LiDAR. In *SilviLaser*. Vancouver, Canada, 16th - 19th September.
- 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 Sensing of Environment*, 82, 397–416.
- Hudak, A. T., Strand, E. K., Vierling, L. A., Byrne, J. C., Eitel, J. U., Martinuzzi, S., & Falkowski, M. J. (2012b). Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sensing of Environment*, 123, 25–40. doi:10.1016/j.rse.2012.02.023
- Hurt, G. C., Dubayah, R. O., Drake, J. B., Moorcroft, P. P. R., Pacala, S. W. S., Blair, B. J., & Fearon, M. G. M. (2004). Beyond Potential Vegetation: Combining LiDAR Data and a Height-Structured Model for Carbon Studies. *Ecological Applications*, 14(3), 873–883.
- Hyde, P., Dubayah, R. O., Walker, W., Blair, B. J., Hofton, M. A., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using

References

- multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, 102(1-2), 63–73. doi:10.1016/j.rse.2006.01.021
- Hyypä, J., Hyypä, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29(5), 1339–1366.
- Isenburg, M. (2012). LAStools - efficient tools for LiDAR processing.
- Jacob, A. L., Jane, S., & Lewis, S. (2014). Forests are more than sticks of carbon. *Nature*, 507, 306. doi:10.1038/507306c
- Jacobs, M. R. (1955). *Growth habits of the Eucalypts*. Canberra: Commonwealth Forestry and Timber Bureau.
- Jakubowski, M. K., Guo, Q., & Kelly, M. (2013). Tradeoffs between lidar pulse density and forest measurement accuracy. *Remote Sensing of Environment*, 130, 245–253. doi:10.1016/j.rse.2012.11.024
- Jaskierniak, D., Lane, P. N. J., Robinson, A., & Lucieer, A. (2011). Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sensing of Environment*, 115(2), 573–585. doi:10.1016/j.rse.2010.10.003
- Jenkins, R. B. (2012). Airborne laser scanning for vegetation structure quantification in a south east Australian scrubby forest-woodland. *Austral Ecology*, 37(1), 44–55. doi:10.1111/j.1442-9993.2011.02248.x
- Jenkins, R. B., & Coops, N. C. (2011). Landscape Controls on Structural Variation in Eucalypt Vegetation Communities: Woronora Plateau, Australia. *Australian Geographer*, 42(1), 1–17. doi:10.1080/00049182.2011.546316
- Jennings, S. B., Brown, N. D., & Sheil, D. (1999). Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Forestry*, 72(1), 59–73. doi:10.1093/forestry/72.1.59
- Johansen, K., Phinn, S., & Witte, C. (2010). Mapping of riparian zone attributes using discrete return LiDAR, QuickBird and SPOT-5

- imagery: Assessing accuracy and costs. *Remote Sensing of Environment*, 114(11), 2679–2691. doi:10.1016/j.rse.2010.06.004
- Jung, J., & Crawford, M. (2008). A two-stage approach for decomposition of ICESat waveforms. *IEEE Geoscience and Remote Sensing Symposium*, III-680–III-683.
- Jupp, D. L., Culvenor, D. S., Lovell, J. L., Newnham, G. J., Strahler, A. H., & Woodcock, C. E. (2008). Estimating forest LAI profiles and structural parameters using a ground-based laser called “Echidna”. *Tree Physiology*, 29(2), 171–81. doi:10.1093/treephys/tpn022
- Kane, V. R., Bakker, J. D., McGaughey, R. J., Lutz, J. A., Gersonde, R. F., & Franklin, J. F. (2010a). Examining conifer canopy structural complexity across forest ages and elevations with LiDAR data. *Canadian Journal of Forest Research*, 40(4), 774–787. doi:10.1139/X10-064
- Kane, V. R., Gersonde, R. F., Lutz, J. A., McGaughey, R. J., Bakker, J. D., & Franklin, J. F. (2011). Patch dynamics and the development of structural and spatial heterogeneity in Pacific Northwest forests. *Canadian Journal of Forest Research*, 2291, 2276–2291. doi:10.1139/X11-128
- Kane, V. R., Lutz, J. A., Roberts, S. L., Smith, D. F., McGaughey, R. J., Povak, N. A., & Brooks, M. L. (2013). Landscape-scale effects of fire severity on mixed-conifer and red fir forest structure in Yosemite National Park. *Forest Ecology and Management*, 287, 17–31. doi:10.1016/j.foreco.2012.08.044
- Kane, V. R., McGaughey, R. J., Bakker, J. D., Gersonde, R. F., Lutz, J. A., & Franklin, J. F. (2010b). Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research*, 40(4), 761–773. doi:10.1139/X10-024
- Keenan, T. F., Darby, B., Felts, E., Sonnentag, O., Friedl, M., Hufkens, K., O’Keefe, J. F., Klosterman, S., Munger, J. W., Toomey, M., & Richardson, A. D. (2014). Tracking forest phenology and seasonal physiology using digital repeat photography: a critical assessment. *Ecological Applications*, 140206175103002. doi:10.1890/13-0652.1

- Keith, H., Mackey, B. G., & Lindenmayer, D. B. (2009). Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(28), 11635–11640.
- Kellndorfer, J. M., Walker, W. S., LaPoint, E., Kirsch, K., Bishop, J., & Fiske, G. (2010). Statistical fusion of lidar, InSAR, and optical remote sensing data for forest stand height characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM+, and ancillary data sets. *Journal of Geophysical Research: Biogeosciences* (2005–2012), *115*.G2. doi:10.1029/2009JG000997
- Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T., & Hussin, Y. A. (2014). Generating Pit-free Canopy Height Models from Airborne Lidar. *Photogrammetric Engineering & Remote Sensing*, *80*(9), 863–872. doi:10.14358/PERS.80.9.863
- Koike, F., & Syahbuddin. (1993). Canopy structure of a tropical rain forest and the nature of an unstratified upper layer. *Functional Ecology*, 230–235.
- Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, *115*(4), 1065–1080. doi:10.1016/j.rse.2010.12.011
- Korpela, I. (2008). Mapping of understory lichens with airborne discrete-return LiDAR data. *Remote Sensing of Environment*, *112*(10), 3891–3897. doi:10.1016/j.rse.2008.06.007
- Korpela, I., Hovi, A., & Morsdorf, F. (2012). Understory trees in airborne LiDAR data — Selective mapping due to transmission losses and echo-triggering mechanisms. *Remote Sensing of Environment*, *119*, 92–104. doi:10.1016/j.rse.2011.12.011
- Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, *83*(4), 395–407. doi:10.1093/forestry/cpq022

- Lee, A. C., & Lucas, R. M. (2007). A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests. *Remote Sensing of Environment*.
- Lee, A. C., Lucas, R. M., & Brack, C. (2004). Quantifying Vertical Forest Stand Structure Using Small Footprint LiDAR to Assess Potential Stand Dynamics. *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, XXXVI(8/W2), 213–217.
- Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, 37(15), 1–5. doi:10.1029/2010GL043622
- Lefsky, M. A., Cohen, W. B., Acker, S. A., & Parker, G. G. (1999a). Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Ecological Applications*, 361(May), 339–361.
- Lefsky, M. A., Cohen, W. B., & Harding, D. J. (2001). Lidar remote sensing of above ground biomass in three biomes. *International Archives of Photogrammetry and Remote Sensing*, XXXIV-3, 155–160.
- Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002a). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography*, 11(5), 393–399.
- Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002b). Lidar Remote Sensing for Ecosystem Studies. *BioScience*, 52(1), 19. doi:10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2
- Lefsky, M. A., Harding, D. J., Cohen, W. B., Parker, G. G., & Shugart, H. (1999b). Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland , USA. *Remote Sensing of Environment*, 67, 83–98.
- Lefsky, M. A., Hudak, A. T., Cohen, W. B., & Acker, S. A. (2005a). Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. *Remote Sensing of Environment*, 95(4), 532–548. doi:10.1016/j.rse.2005.01.010

- Lefsky, M. A., Hudak, A. T., Cohen, W. B., & Acker, S. A. (2005b). Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment*, 95(4), 517–531. doi:10.1016/j.rse.2005.01.004
- Lewis, P., Gomez-Dans, J., & Disney, M. I. (2013). A Vision for a Land Observing System. In *Living Planet Symposium*. Edinburgh, UK, 9th - 13th September: European Space Agency.
- Liaw, A., & Wiener, M. (2013). *Package "randomForest."*
- Lieff, W. (2009). RASP 0.93 - Reference Manual. Airborne REsearch Australia.
- Lim, K., Treitz, P., Wulder, M. A., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27(1), 88–106. doi:10.1191/0309133303pp360ra
- Lim, S., Trindera, J., & Turner, R. (2011). Full Waveform Lidar Remote Sensing for Forest Inventory in New South Wales, Australia. In *Proc. Int. Symp. LiDAR & Radar: Technologies & Applications*. Nanjing, China, 26th - 29th May.
- Lindenmayer, D. B. (2000). Factors at multiple scales affecting distribution patterns and their implications for animal conservation – Leadbeater’s Possum as a case study. *Biodiversity and Conservation*, 9, 15–35.
- Lindenmayer, D. B., Cunningham, R., Donnelly, C., & Franklin, J. F. (2000). Structural features of old-growth Australian montane ash forests. *Forest Ecology and Management*, 134(1-3), 189–204. doi:10.1016/S0378-1127(99)00257-1
- Los, S., Rosette, J., Kljun, N., North, P. R., Chasmer, L., Suárez, J., Hopkinson, C., Hill, R., van Gorsel, E., Mahoney, C., & Berni, J. (2012). Vegetation height and cover fraction between 60° S and 60° N from ICESat GLAS data. *Geoscientific Model Development*, 5, 413–432. doi:10.5194/gmd-5-413-2012
- Lovell, J. L., Haverd, V., Jupp, D. L., & Newnham, G. J. (2012). The Canopy Semi-analytic Pgap And Radiative Transfer (CanSPART) model: Validation using ground based lidar. *Agricultural and Forest Meteorology*, 158-159, 1–12. doi:10.1016/j.agrformet.2012.01.020

- Lovell, J. L., Jupp, D. L., Culvenor, D. S., & Coops, N. C. (2003). Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, 29(5), 607–622. doi:10.5589/m03-026
- Lovell, J. L., Jupp, D. L., Newnham, G. J., Coops, N. C., & Culvenor, D. S. (2005). Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management*, 214(1-3), 398–412. doi:10.1016/j.foreco.2004.07.077
- Lucas, R. M., Lee, A., Armston, J., & Breyer, J. (2008a). Advances in forest characterisation, mapping and monitoring through integration of LiDAR and other remote sensing datasets. In *SilviLaser*. Edinburgh, UK, 9th - 13th September.
- Lucas, R. M., Lee, A. C., & Bunting, P. J. (2008b). Retrieving forest biomass through integration of CASI and LiDAR data. *International Journal of Remote Sensing*, 29(5), 1553–1577. doi:10.1080/01431160701736497
- MacArthur, R. H., & Horn, H. S. (1969). Foliage Profile by Vertical Measurements. *Ecological Society of America*, 50(5), 802–804.
- MacArthur, R. H., & MacArthur, J. (1961). On bird species diversity. *Ecology*, 42(3), 594–598.
- Magnusson, M., Fransson, J. E. S., & Holmgren, J. (2007). Effects on Estimation Accuracy of Forest Variables Using Different Pulse Density of Laser Data. *Forest Science*, 53(6), 619–626.
- Maltamo, M., Eerikäinen, K., Packalén, P., & Hyyppä, J. (2006). Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, 79(2), 217–229. doi:10.1093/forestry/cpl007
- Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., & Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, 216(1-3), 41–50. doi:10.1016/j.foreco.2005.05.034
- Maltamo, M., Uutera, J., & Kuusela, K. (1997). Differences in Forest Stand Structure Between Forest Ownership Groups in Central Finland.

References

- Journal of Environmental Management*, 51(2), 145–167.
doi:10.1006/jema.1997.0140
- 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(1), 12–16.
doi:10.1371/journal.pone.0085993
- McElhinny, C., Gibbons, P., Brack, C., & Bauhus, J. (2005). Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management*, 218(1-3), 1–24.
doi:10.1016/j.foreco.2005.08.034
- McGarigal, K., Tagil, S., & Cushman, S. A. (2009). Surface metrics: An alternative to patch metrics for the quantification of landscape structure. *Landscape Ecology*, 24(3), 433–450. doi:10.1007/s10980-009-9327-y
- McInerney, D. O., Suarez-Minguez, J., Valbuena, R., & Nieuwenhuis, M. (2010). Forest canopy height retrieval using LiDAR data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland. *Forestry*, 83(2), 195–206. doi:10.1093/forestry/cpq001
- McRoberts, R. E., Cohen, W. B., Næsset, E., Stehman, S. V., & Tomppo, E. O. (2010). Using remotely sensed data to construct and assess forest attribute maps and related spatial products. *Scandinavian Journal of Forest Research*, 25(4), 340–367.
doi:10.1080/02827581.2010.497496
- McRoberts, R. E., & Tomppo, E. O. (2007). Remote sensing support for national forest inventories. *Remote Sensing of Environment*, 110(4), 412–419. doi:10.1016/j.rse.2006.09.034
- Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L., & Hendrix, C. J. (2000). Predicting Forest Stand Characteristics with Airborne Scanning Lidar, 66(11), 1367–1371.
- Means, J. E., Acker, S. A., Harding, D. J., Blair, B. J., Lefsky, M. A., Cohen, W. B., Harmon, M. E., & McKee, W. A. (1999). Use of large-footprint scanning

- airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, 67, 298–308.
- 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 Journal of Photogrammetry and Remote Sensing*, 105, 155–168. doi: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 Sensing*, 5(6), 2838–2856. doi:10.3390/rs5062838
- Mellor, A., Jones, S. D., Haywood, A., & Wilkes, P. (2012). Using random forest decision tree classification for large area forest extent mapping with multi-source remote sensing and GIS data. In *16th Australasian Remote Sensing and Photogrammetry Conference, 27-28th August* Australasian Remote Sensing and Photogrammetry Conference. Melbourne, Australia, 21st - 26th July.
- Miles, P. D. (2002). Using biological criteria and indicators to address forest inventory data at the state level. *Forest Ecology and Management*, 155, 171–185. doi:10.1016/S0378-1127(01)00557-6
- Millennium Ecosystem Assessment. (2005). *Living Beyond Our Means: Natural Assets and Human Well-being. Statement from the Board*. Washington, DC: Island Press. doi:10.2111/RANGELANDS-D-13-00013.1
- Mitchell, P. J., Lane, P. N. J., & Benyon, R. G. (2012). Capturing within catchment variation in evapotranspiration from montane forests using LiDAR canopy profiles with measured and modelled fluxes of water. *Ecohydrology*, 5(6), 708–720. doi:10.1002/eco.255
- Miura, N., & Jones, S. D. (2010). Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment*, 114(5), 1069–1076. doi:10.1016/j.rse.2009.12.017

- Miura, N., & Jones, S. D. (2012). Estimation of Coarse Woody Debris using Airborne LiDAR. In *16th Australasian Remote Sensing and Photogrammetry Conference, 27-28th August*. Melbourne, Australia.
- Montreal Process Implementation Group for Australia. (2013). *Australia's State of the Forests Report 2013*. Forestry Commission, Tasmania.
- Montréal Process Working Group. (2009). *Criteria and Indicators for the Conservation and Sustainable Management of Temperate and Boreal Forests*.
- Mora, B., Wulder, M. A., Hobart, G. W., White, J. C., Bater, C. W., Gougeon, F. A., Varhola, A., & Coops, N. C. (2013). Forest inventory stand height estimates from very high spatial resolution satellite imagery calibrated with lidar plots. *International Journal of Remote Sensing*, 34(12), 4406–4424. doi:10.1080/01431161.2013.779041
- Morsdorf, F., Frey, O., Meier, E., Itten, K. I., & Allgöwer, B. (2008). Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, 29(5), 1387–1406. doi:10.1080/01431160701736349
- Morsdorf, F., Marell, A., Koetz, B., Cassagne, N., Pimont, F., Rigolot, E., & Allgöwer, B. (2010). Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. *Remote Sensing of Environment*, 114(7), 1403–1415.
- Murphy, M. A., Evans, J. S., & Strofer, A. (2010). Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. *Ecology*, 91(1), 252–261.
- Muss, J. D., Mladenoff, D. J., & Townsend, P. A. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, 115(3), 824–835. doi:10.1016/j.rse.2010.11.008
- Næsset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(2), 49–56. doi:10.1016/S0924-2716(97)83000-6

- Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19(2), 164–179. doi:10.1080/02827580310019257
- Næsset, E. (2005). Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, 98(2-3), 356–370. doi:10.1016/j.rse.2005.07.012
- Næsset, E. (2007). Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research*, 22(5), 433–442. doi:10.1080/02827580701672147
- Næsset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, 113(1), 148–159. doi:10.1016/j.rse.2008.09.001
- National Forest Inventory. (1998). *Australia's State of the Forests Report 1998*. Canberra.
- National Resource Information Centre. (1991). *Atlas of Australian Soils (digital version)*.
- Nilsson, M. (1996). Estimation of Tree Heights and Stand Volume Using and Airborne Lidar System.pdf. *Remote Sensing of Environment*, 56, 1–7.
- Ni-Meister, W., Jupp, D. L., & Dubayah, R. O. (2001). Modeling Lidar Waveforms in Heterogeneous and Discrete Canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), 1943–1958.
- Ørka, H. O., Wulder, M. A., Gobakken, T., & Næsset, E. (2010). Integrating airborne laser scanner data and ancillary information for delineating the boreal-alpine transition zone in Hedmark County , Norway. In *SilviLaser 2010* (pp. 179–186). Freiburg, Germany.
- Park Williams, A., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C. A., Meko, D. M., Swetnam, T. W., Rauscher, S. A., Seager, R., Grissino-Mayer, H. D., Dean, J. S., Cook, E. R., Gangodagamage, C., Cai, M., & McDowell, N. (2012). Temperature as a potent driver of regional

- forest drought stress and tree mortality. *Nature Climate Change*, 3(3), 292–297. doi:10.1038/nclimate1693
- Parker, G. G., & Brown, M. (2000). Forest Canopy Stratification-Is It Useful? *The American Naturalist*, 155(4), 473–484. doi:10.1086/303340
- Parker, G. G., & Russ, M. E. (2004). The canopy surface and stand development: Assessing forest canopy structure and complexity with near-surface altimetry. *Forest Ecology and Management*, 189(1-3), 307–315. doi:10.1016/j.foreco.2003.09.001
- Pascual, C., García-Abril, A., Cohen, W. B., & Martín-Fernández, S. (2010). Relationship between LiDAR-derived forest canopy height and Landsat images. *International Journal of Remote Sensing*, 31(5), 1261–1280. doi:10.1080/01431160903380656
- Pasher, J., & King, D. J. (2011). Development of a forest structural complexity index based on multispectral airborne remote sensing and topographic data This article is one of a selection of papers from Extending Forest Inventory and Monitoring over Space and Time. *Canadian Journal of Forest Research*, 41(1), 44–58. doi:10.1139/X10-175
- Patenaude, G., Hill, R., Milne, R., Gaveau, D., Briggs, B., & Dawson, T. (2004). Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, 93(3), 368–380. doi:10.1016/j.rse.2004.07.016
- Pereira, H. M., Walters, M., Ferrier, S., Geller, G. N., Jongman, R., Scholes, R. J., Bruford, M., & Reyers, B. (2013). Essential Biodiversity Variables. *Science*, 339(January), 277–278.
- Peterson, B., & Nelson, K. (2014). Mapping Forest Height in Alaska Using GLAS, Landsat Composites, and Airborne LiDAR. *Remote Sensing*, 12409–12426. doi:10.3390/rs61212409
- Plotkin, J. B., Chave, J., & Ashton, P. S. (2002). Cluster analysis of spatial patterns in Malaysian tree species. *The American Naturalist*, 160(5), 629–644. doi:10.1086/342823

- Popescu, S. C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*, 112(3), 767–781. doi:10.1016/j.rse.2007.06.011
- Python Software Foundation. (2015). Python.
- QGIS Development Team. (2015). QGIS Geographic Information System. Open Source Geospatial Foundation Project.
- Quadros, N., Frisina, R., & Wilson, P. (2011). Using Airborne Survey to map Stream Form and Riparian Vegetation Characteristics across Victoria. In *SilviLaser*. Hobart, Australia, October 16th - 19th.
- R Development Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria.
- Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles*, 13(4), 997–1027. doi:10.1029/1999GB900046
- Riaño, D., Meier, E., Allgöwer, B., Chuvieco, E., & Ustin, S. L. (2003). Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment*, 86(2), 177–186. doi:10.1016/S0034-4257(03)00098-1
- Riegl. (2006). Airborne Data-Processing Software or Full Waveform Analysis RiANALYZE 560. Horn, Austria.
- Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M. P. M., & Jeganathan, C. (2012a). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93–107. doi:10.1016/j.rse.2011.12.003
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012b). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104. doi:10.1016/j.isprsjprs.2011.11.002
- Ross, J. (1981). *The Radiation Regime and Architecture of Plant Stands*. Dordrecht, Netherlands: Junk, Dr. W.

- Roy, D., Wulder, M. A., Loveland, T., Woodcock, C., Allen, R., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D., Kennedy, R., Scambos, T., Schaaf, C. B., Schott, J., Sheng, Y., Vermote, E., Belward, A., Bindschadler, R., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, *145*, 154–172. doi:10.1016/j.rse.2014.02.001
- Sasaki, N., & Putz, F. E. (2009). Critical need for new definitions of “forest” and “forest degradation” in global climate change agreements. *Conservation Letters*, *2*(5), 226–232. doi:10.1111/j.1755-263X.2009.00067.x
- Scanlan, I., McElhinny, C., & Turner, P. (2010). A Methodology for Modelling Canopy Structure: An Exploratory Analysis in the Tall Wet Eucalypt Forests of Southern Tasmania. *Forests*, *1*(1), 4–24. doi:10.3390/f1010004
- Schaaf, C. B., Barry, R. G., Brady, M., Brown, J., CEOS, Christiansen, H. H., Cihlar, J., Clow, G., Csiszar, I., Dolman, H., Famiglietti, J., Fritz, G., Gobron, N., Grabs, W., Haeberli, W., Healy, E., Herold, M., ... Zemp, M. (2008). *Terrestrial Essential Climate Variables for Climate Change Assessment, Mitigation and Adaptation*. (R. Sessa & H. Dolman, Eds.) *Terrestrial essential Climate Variables* (GtOS 52.). Food and Agriculture Organization of the United Nations.
- Seifert, F., Scipal, K., Quegan, S., & Le Toan, T. (2015). The Earth Explorer Biomass Mission - status and preparatory activities. In *GOFC/GOLD Biomass workshop*. 24th - 26th February, Brisbane.
- Semple, B., Rankin, M., Koen, T., & Geeves, G. (2010). A Note on Tree Deaths during the Current (2001–?) Drought in South-eastern Australia. *Australian Geographer*, *41*(March 2015), 391–401. doi:10.1080/00049182.2010.498042
- Sessa, R. (Ed.). (2009). Assessment of the status of the development of the standards for the terrestrial essential climate variables: biomass. In *Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables* (Version 10., pp. 1–18). Rome.
- Sillett, S. C., Van Pelt, R., Koch, G. W., Ambrose, A. R., Carroll, A. L., Antoine, M. E., & Mifsud, B. M. (2010). Increasing wood production through old

- age in tall trees. *Forest Ecology and Management*, 259(5), 976–994. doi:10.1016/j.foreco.2009.12.003
- Silverman, B. (1985). Some Apects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting.pdf. *Journal of the Royal Statistical Society, Series B*, 47(1), 1–52.
- Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*, 116(G4), 1–12. doi:10.1029/2011JG001708
- Skidmore, A. K., Pettorelli, N., Coops, N. C., Geller, G. N., Hansen, M. C., Lucas, R. M., Múcher, C. A., O'Connor, B., Paganini, M., Pereira, H. M., Schaepman, M. E., Turner, W., Wang, T., & Wegmann, M. (2015). Agree on biodiversity metrics to track from space. *Nature*, 523, 403–405.
- Smith, W. B. (2002). Forest inventory and analysis: a national inventory and monitoring program. *Environmental Pollution (Barking, Essex: 1987)*, 116 Suppl , S233–S242.
- Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7).
- Stasinopoulos, M., Rigby, B., & Akantziliotou, C. (2012). Instructions on how to use the gamlss package in R: Second Edition.
- Steininger, M. K. (2000). Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia. *International Journal of Remote Sensing*, 21(March 2015), 1139–1157. doi:10.1080/014311600210119
- Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., & Džeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, 5, 256–266. doi:10.1016/j.ecoinf.2010.03.004
- Su, J., & Bork, E. (2006). Influence of Vegetation, Slope, and Lidar Sampling Angle on DEM Accuracy. *Photogrammetric Engineering & Remote Sensing*, 72(11), 1265–1274. doi:10.14358/PERS.72.11.1265

- Sugden, A., Fahrenkamp-Uppenbrink, J., Malakoff, D., & Vignieri, S. (2015). Forest Health in a Changing Environment. *Science, August*, 800 – 801.
- Svetnik, V., Liaw, A., Tong, C., & Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. *Multiple Classifier Systems*, 334–343. doi:10.1007/978-3-540-25966-4_33
- Swatantran, A., Dubayah, R. O., Roberts, D. A., Hofton, M. A., & Blair, B. J. (2011). Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sensing of Environment*, 115(11), 2917–2930. doi:10.1016/j.rse.2010.08.027
- Takahashi, T., Awaya, Y., & Hirata, Y. (2008). Effects of flight altitude on LiDAR-derived tree heights in mountainous forests with poor laser penetration rates. *The Photogrammetric Journal of Finland*, 21(1), 86–96.
- Takahashi, T., Awaya, Y., Hirata, Y., Furuya, N., Sakai, T., & Sakai, A. (2010). Stand volume estimation by combining low laser-sampling density LiDAR data with QuickBird panchromatic imagery in closed-canopy Japanese cedar (*Cryptomeria japonica*) plantations. *International Journal of Remote Sensing*, 31(5), 1281–1301. doi:10.1080/01431160903380623
- Tesfamichael, S. G., van Aardt, J. A., & Ahmed, F. (2010). Estimating plot-level tree height and volume of *Eucalyptus grandis* plantations using small-footprint, discrete return lidar data. *Progress in Physical Geography*, 34(4), 515–540. doi:10.1177/0309133310365596
- Thomas, V., Treitz, P., McCaughey, J. H., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. *Canadian Journal of Forest Research*, 36(1), 34–47. doi:10.1139/x05-230
- Tickle, P. K., Lee, A. C., Lucas, R. M., Austin, J., & Witte, C. (2006). Quantifying Australian forest floristics and structure using small footprint LiDAR and large scale aerial photography. *Forest Ecology and Management*, 223(1-3), 379–394. doi:10.1016/j.foreco.2005.11.021

- Tinkham, W. T., Huang, H., Smith, A. M., Shrestha, R., Falkowski, M. J., Hudak, A. T., Link, T. E., Glenn, N. F., & Marks, D. G. (2011). A Comparison of Two Open Source LiDAR Surface Classification Algorithms. *Remote Sensing*, 3(3), 638–649. doi:10.3390/rs3030638
- Tomppo, E., Gschwantner, T., Lawrence, M., & McRoberts, R. E. (Eds.). (2010). *National Forest Inventories Pathways for Common Reporting Media*. Heidelberg: Springer. doi:10.1007/978-90-481-3233-1
- Tomppo, E. O., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., & Katila, M. (2008). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, 112(5), 1982–1999. doi:10.1016/j.rse.2007.03.032
- Torabzadeh, H., Morsdorf, F., & Schaepman, M. E. (2014). Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems – A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 97, 25–35. doi:10.1016/j.isprsjprs.2014.08.001
- Trabucco, A., & Zomer, R. (2010). *Global Soil Water Balance Geospatial Database*. CGIAR Consortium for Spatial Information. Retrieved February 20, 2015, from <http://www.cgiar-csi.org>
- Treitz, P., Lim, K., Woods, M., Pitt, D., Nesbitt, D., & Etheridge, D. (2012). LiDAR Sampling Density for Forest Resource Inventories in Ontario, Canada. *Remote Sensing*, 4(12), 830–848. doi:10.3390/rs4040830
- Trumbore, S., Brando, P., & Hartmann, H. (2015). Forest health and global change. *Science*, 349(6250), 814 – 818.
- Turner, R. (2007). An overview of Airborne LiDAR applications in New South Wales state forests. In *Australian and New Zealand Institute of Foresters* (pp. 1–22). Coffs Harbour.
- Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, a K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L. P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M., & Woodcock, C. E. (2015). Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation*, 182, 173–176. doi:10.1016/j.biocon.2014.11.048

- Van Den Meersschaut, D., & Vandekerkhove, K. (1998). Development of a stand-scale forest biodiversity index based on the State Forest Inventory. In M. Hansen & T. Burk (Eds.), *Integrated Tools for Natural Resources Inventories in the 21st Century* (pp. 340 – 349). Boise, Idaho: USDA.
- Van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., De Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., & Viney, N. R. (2013). The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, *49*, 1040–1057. doi:10.1002/wrcr.20123
- Van Leeuwen, M., & Nieuwenhuis, M. (2010). Retrieval of forest structural parameters using LiDAR remote sensing. *European Journal of Forest Research*, *129*(4), 749–770. doi:10.1007/s10342-010-0381-4
- Van Pelt, R., & North, M. P. (1996). Analyzing canopy structure in Pacific Northwest old-growth forests with a stand-scale crown model. *Northwest Science*, *70*, 15–30.
- Van Pelt, R., Sillett, S. C., & Nadkarni, N. M. (2004). Quantifying and Visualizing Canopy Structure in Tall Forests: Methods and a Case Study. In M. D. Lowman & H. B. Rinker (Eds.), *Forest Canopies* (second, p. 544). Elsevier Academic Press.
- VicForests. (2014). *Operating Procedures Regulatory Handbook Version 2.1*.
- Victorian Department of Sustainability and Environment. (2012a). FPMRIS Guidelines for Ground Plot Measurement, version 1.2.
- Victorian Department of Sustainability and Environment. (2012b). *Permanent growth plots*. Retrieved September 25, 2012, from <http://www.dse.vic.gov.au/forests/publications/research-reports/growth-and-yield/permanent-growth-plots>
- Vierling, K., Vierling, L. A., Gould, W. A., Martinuzzi, S., & Clawges, R. M. (2008). Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, *6*(2), 90–98. doi:10.1890/070001

- Wallace, L., Lucieer, A., Watson, C., & Turner, D. (2012). Development of a UAV-LiDAR system with application to forest inventory. *Remote Sensing*, 4(6), 1519–1543. doi:10.3390/rs4061519
- Watt, M. S., Adams, T., Aracil, S. G., Marshall, H., Watt, P. J., & Gonzalez Aracil, S. (2013). The influence of LiDAR pulse density and plot size on the accuracy of New Zealand plantation stand volume equations. *New Zealand Journal of Forestry Science*, 43(1), 15. doi:10.1186/1179-5395-43-15
- Whitehurst, A., Swatantran, A., Blair, B. J., Hofton, M. A., & Dubayah, R. O. (2013). Characterization of Canopy Layering in Forested Ecosystems Using Full Waveform Lidar. *Remote Sensing*, 5(4), 2014–2036. doi:10.3390/rs5042014
- Wilkes, P. (2015). ForestLAS Python packakge.
- Wilkes, P., Jones, S. D., Suarez, L., Haywood, A., Mellor, A., Soto-berelov, M., Suárez, L., Mellor, A., Soto-berelov, M., & Woodgate, W. (2013). MAUP and Lidar derived canopy structure. *IEEE Geoscience and Remote Sensing Symposium*, (June 2015).
- Wilkes, P., Jones, S. D., Suarez, L., Haywood, A., Mellor, A., Woodgate, W., Soto-berelov, M., & Skidmore, A. K. (2015). Understanding the effects of ALS pulse density for metric retrieval across diverse forest types. *Photogrammetric Engineering & Remote Sensing*, 81(8), 625 – 635.
- Wilkes, P., Jones, S., Suarez, L., Haywood, A., & Mellor, A. (2014). Estimating vegetation vertical structural complexity at a regional scale. In *ForestSAT*. Riva Del Garda, Italy, 4th - 7th November.
- Woodgate, P. W., Peel, W., Ritman, K., Coram, J., Brady, A., Rule, A., & Banks, J. (1994). *A Study of the Old-Growth Forest of East Gippsland*. Melbourne, Australia.
- Woodgate, W. (2015). *Uncertainty in Remote Sensing of Leaf Area Index in forests (PhD thesis)*. RMIT University.
- Wulder, M. A. (1998). Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography*, 22(4), 449–476. doi:10.1177/030913339802200402

- Wulder, M. A., Bater, C. W., Coops, N. C., Hilker, T., & White, J. C. (2008a). The role of LiDAR in sustainable forest management. *The Forestry Chronicle*, *84*(6), 807–826.
- Wulder, M. A., & Seemann, D. (2003). Forest inventory height update through the integration of lidar data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, *29*(5), 536–543.
- Wulder, M. A., White, J. C., Bater, C. W., Coops, N. C., Hopkinson, C., & Chen, G. (2012a). Lidar plots — a new large-area data collection option: context, concepts, and case study. *Canadian Journal of Remote Sensing*, *38*(05), 600–618. doi:10.5589/m12-049
- Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2008b). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment*, *112*(3), 955–969. doi:10.1016/j.rse.2007.07.004
- Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., Hilker, T., Bater, C. W., & Gobakken, T. (2012b). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, *121*, 196–209. doi:10.1016/j.rse.2012.02.001
- Zellweger, F., Braunisch, V., Baltensweiler, A., & Bollmann, K. (2013a). Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *Forest Ecology and Management*, *307*, 303–312. doi:10.1016/j.foreco.2013.07.023
- Zellweger, F., Morsdorf, F., Purves, R. S., Braunisch, V., & Bollmann, K. (2013b). Improved methods for measuring forest landscape structure: LiDAR complements field-based habitat assessment. *Biodiversity and Conservation*, *23*(2), 289–307. doi:10.1007/s10531-013-0600-7
- Zenner, E. K., & Hibbs, D. E. (2000). A new method for modeling the heterogeneity of forest structure. *Forest Ecology and Management*, *129*, 75–87.
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using

-
- MODIS. *Remote Sensing of Environment*, 84(3), 471–475.
doi:10.1016/S0034-4257(02)00135-9
- Zhang, Z., & Liu, X. (2012). Support vector machines for tree species identification using LiDAR-derived structure and intensity variables. *Geocarto International*, (August 2012), 1–15.
doi:10.1080/10106049.2012.710653
- Zhang, Z., Liu, X., Peterson, J., & Wright, W. (2011). Cool temperate rainforest and adjacent forests classification using airborne LiDAR data. *Area*, 43(4), 438–448. doi:10.1111/j.1475-4762.2011.01035.x
- Zhao, K., Popescu, S. C., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment*, 115(8), 1978–1996.
doi:10.1016/j.rse.2011.04.001
- Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., & Gerard, P. D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, 87(2-3), 171–182. doi:10.1016/S0034-4257(03)00139-1

8. Summary

The attribution of forest structure forms an integral part of international monitoring and reporting obligations with regard to sustainable forest management. Furthermore, detailed information about forest structure allows land managers and forest scientists to determine a forest's ability to provide ecosystem services. Currently, forest attribution is achieved using a network of forest inventory plots that are revisited periodically. This approach comprises a sparse sample, both temporally and spatially, that may not capture variance in forest structure. This is particularly true in dynamic native forests where variability in forest structure can be high. In recent years the capability of remote sensing techniques has been realised for sustainable forest management applications. Advantages of a remote sensing approach include synoptic and high temporal coverage as well as reduced costs to the end-user. Furthermore, recent advancement in active sensors, such as Light Detection and Ranging Instruments (LiDAR) have allowed for detailed three-dimensional forest measurement of structure across large areas.

This thesis presents new metrics, techniques and acquisition specifications for the attribution of forest canopy over large areas (e.g. comprising two or more forest types where forest structure maybe unknown *a priori*) using active and passive remote sensing. In particular, the focus is on attributes that quantify the vertical structure of forests; canopy height and canopy vertical structure. Canopy height is a commonly measured multipurpose attribute that is utilised, for example, to estimate biomass. Attribution of the canopy height profile, although less common, is important for mapping habitat suitability, biomass and fire susceptibility. Current techniques to attribute forests tend to be tailored to a particular forest type or location and therefore application of these models across large areas is unreliable. Here the aim is to develop metrics and techniques that are transferable between different forest types and applicable to forests where there is no prior knowledge of forest structure.

Here a multi-scale remote sensing approach was taken, where plot scale measurements were upscaled to attribute large areas. Initially, existing LiDAR derived metrics applicable at the plot scale were tested at three 5 km x 5 km study areas in Victoria, Australia where forests cover a broad

range of structural types. Results indicate existing metrics of canopy height were applicable across the range of forest types, for example the 95th percentile of LiDAR derived height estimated inventory measured canopy height with a RMSE of 12% (~5 m). An existing mixture modelling technique to attribute the canopy height profile was found unsuitable when applied across heterogeneously forested landscape. This was due to the inability to parameterise the model correctly without *a priori* knowledge of forest structure e.g. presence or absence of shade tolerant layers. For this reason a new technique was developed utilising a nonparametric regression of LiDAR derived gap probability that generalised the canopy profile. Taking the second derivative of the regression curve identified locations within the canopy that correspond with canopy strata, this therefore allowed a dynamic attribution of canopy vertical structure. Model output was validated with a crown volume modelling approach at 24 plots, where crown models were parameterised with inventory data and allometry. Results indicate this technique can estimate the number of canopy strata with a RMSE of 0.41 strata. Furthermore, the new technique met the transferability criteria, as a universal regression coefficient was transferable between forest types with different structural attributes.

As LiDAR acquisition that cover large areas will inevitably encounter a range of forest types, parameters for attributing canopy structure that were transferable between forest types were investigated; in particular sampling frequency. To effectively assess a range of pulse densities would require repeat capture over a study area at a range of flying heights, which would be prohibitively expensive. For this reason a new technique was developed that systematically thinned point clouds. This technique differs from previous approaches by allowing simulation of multi-return instruments as well as repeat capture of the same plot. Six sites from around Australia were utilised which covered a broad range of forest types, from open savanna to tropical rainforest. For a suite of metrics, the ability of progressively less dense point clouds ($4 - 0.05 \text{ pl m}^{-2}$) to estimate canopy structure was estimated by comparison with higher density data (10 pl m^{-2}). Results indicate that canopy structure can be adequately attributed with data captured at 0.5 pl m^{-2} . When pulse densities are $<0.5 \text{ pl m}^{-2}$, the inability to adequately identify ground resulted in poor metric estimation, this was particularly evident in high biomass forest. Conversely

at lower pulse densities in savanna systems, the inability to characterise sparse vegetation resulted in poor attribution of the canopy profile.

Techniques derived at the plot scale were then applied to estimate canopy height across 2.9 million hectares of heterogeneous forest. Canopy height in the study area ranged from 0 – 70 m and comprised forest types from open woodland to tall closed canopy rainforest. LiDAR derived canopy height was used to train ensemble regression trees (random forest), where predictor datasets included synoptic passive optical imagery and other ancillary spatial datasets, such as Landsat TM and MODIS. Results suggest canopy height can be estimated with a RMSE of 30% (5.5 m) when validated with an independent inventory dataset. This is a similar error to that reported in previous studies for less complex forests and is within the European Space Agency target for canopy height estimation. However, model output did show a systematic error, where the height of short and tall forests were over and underestimated respectively. This was corrected by subtracting a modelled estimate of error from the random forest output. Production of a canopy height map over a large area allowed for a consistent product that covered a broad range of forest types, derivation at a 30 m resolution allowed the identification of landscape features such as logging coupes. The presented technique utilised an open source computing framework as well as freely available predictor datasets to facilitate uptake of by land management agencies and forest scientists.

9. Samenvatting

De toekenning van een bosstructuur vormt een integraal onderdeel van de internationale controle en rapportage verplichtingen met betrekking tot duurzaam bosbeheer. Bovendien geeft meer gedetailleerde informatie over bosstructuur landbeheerders en wetenschappers de mogelijkheid meer inzicht te krijgen in ecosystemen. Momenteel wordt het predikaat bos toegekend door inventarisatie van een netwerk van percelen die periodiek worden herzien. Deze werkwijze vertegenwoordigt slechts een klein gebied, zowel in tijd en ruimte, en bestrijkt zo misschien minder de variëteit in de bosstructuur.. Dit is zeker het geval voor dynamische inheemse bossen waar variabiliteit in bosstructuur juist hoog kan zijn. In de afgelopen jaren heeft men zich de mogelijkheden van het gebruik van Remote Sensing technieken gerealiseerd voor duurzame toepassingen in bosbeheer. Voordelen van Remote Sensing zijn de overzichtelijkheid en de hoge temporele dekkingsgraad, alsmede de lagere kosten voor de eindgebruiker. De recente ontwikkelingen in de actieve sensoren, zoals bijvoorbeeld Light Detection and Ranging Instruments (LiDAR), hebben het mogelijk gemaakt gedetailleerde en gestructureerde driedimensionale metingen te verrichten in grote bosgebieden.

Dit proefschrift presenteert nieuwe methodieken en technieken, en wervings specificaties voor het benoemen van bos over grote gebieden (bijvoorbeeld waar de bosstructuur wellicht onbekend is) met behulp van actieve en passieve Remote Sensing. Hierbij wordt de nadruk gelegd op attributen die de verticale structuur van de bossen kwantificeren, te weten de verticale structuur van de kruinhoogte en -structuur. De kruinhoogte wordt gewoonlijk gemeten vanwege zijn multifunctionele eigenschappen, bijvoorbeeld om biomassa te kunnen inschatten. De toewijzing van het kruinhoogteprofiel, hoewel minder vaak voorkomend, is belangrijk voor het in kaart brengen van een geschikt leefgebied, biomassa en brandgevoeligheid. Huidige technieken die gebruikt worden voor bossen worden vaak aangepast aan een bepaalde bos-omgeving en zijn derhalve voor toepassing in grotere bosgebieden onbetrouwbaar. Het doel is daarom om gegevens en technieken te ontwikkelen die overgedragen kunnen worden tussen verschillende types bos en die kunnen worden toegepast zonder voorkennis over de bosstructuur.

In het onderzoek werd remote sensing op verschillende niveaus gebruikt, waarbij metingen van proefvlakken werden opgeschaald voor gebruik in grote gebieden. Aanvankelijk zijn bestaande, van LiDAR afgeleide, metingen van proefgebieden getest op drie percelen van 5 x 5 km in Victoria, Australië, waar bossen een breed scala aan structuren vertegenwoordigen. De resultaten geven aan dat de bestaande statistieken van de kruinhoogte toepasbaar zijn in de verschillende bostypen; bijvoorbeeld het 95e percentiel van LiDAR afgeleide, geschatte hoogte mat de kruinhoogte met een RMSE van 12% (~ 5 m). Een bestaande combinatie van modelleer technieken die werd gebruikt om de kruinhoogte te meten, werd ongeschikt bevonden wanneer toegepast in een heterogeen type bos landschap. Dit was te wijten aan het onvermogen van het model om correcte parameters te geven wanneer er geen kennis vooraf was op het gebied van bosstructuur, zoals bijvoorbeeld de aan- of afwezigheid van schaduw tolerante lagen. Om deze reden werd een nieuwe techniek ontwikkeld die gebruik maakte van een non-parametrische regressie van van LiDAR afgeleide waarschijnlijkheid om een algemeen kruinprofiel te genereren. De tweede afgeleide nemend van met de regressiecurve geïdentificeerde locaties in de kruin die overeenkomen met de bovenlagen, maakte een dynamische toewijzing van de verticale structuur mogelijk. Modeluitkomsten werden gevalideerd via modellering van het kruinvolume in 24 percelen, met behulp van de inventarisatie van de gegevens en allometry. De resultaten geven aan dat deze techniek een schatting kan maken van het aantal vegetatielagen met een RMSE van 0,41. Verder voldeed de nieuwe techniek aan de overdraagbaarheid van de criteria, omdat de universele regressiecoëfficiënt overdraagbaar was tussen de soorten bos met verschillende structurele kenmerken.

LiDAR houdt er rekening mee dat het in het geval van grote gebieden, geconfronteerd wordt met een scala van verschillende soorten bos, en dus werden de parameters voor de toekenning van de kruinstructuur die overdraagbaar zijn tussen de verschillende soorten bos onderzocht; met name de frequenties. Om de reeks van dichtheden effectief te kunnen beoordelen is het nodig om frequente beelden te hebben van een studiegebied vanaf wisselende vlieghoogtes, wat onbetaalbaar zou zijn. Om deze reden werd een nieuwe techniek ontwikkeld, die systematisch wolken punten uitdunde. Deze techniek verschilt van de vorige benaderingen door de simulatie van multi-reeks instrumenten als ook de herhaling van het in

beeld brengen van hetzelfde perceel op frequente tijden. Zes locaties rond Australië werden gebruikt met een breed scala aan soorten bos, van open savanne tot tropisch regenwoud. Voor een reeks, werd het vermogen van steeds minder dichte puntenwolken ($4-0,05 \text{ pl m}^{-2}$) om kruinstructuur in te schatten vergeleken met gegevens over een hogere dichtheid (10 pl m^{-2}). De resultaten geven weer dat de kruinstructuur voldoende kan worden weergegeven wanneer gegevens worden vastgelegd op $0,5 \text{ pl m}^{-2}$. Wanneer de dichtheden $<0,5 \text{ pl m}^{-2}$ bereiken, wordt het onmogelijk de grond te identificeren, wat resulteerde in een slechte schatting, vooral in bossen met een hoge biomassa. Omgekeerd was het bij lagere dichtheden in savanne systemen, onmogelijk om schaarse vegetatie te karakteriseren, wat resulteerde in een slechte herkenning van het profiel.

Technieken die afgeleid zijn van de plot reeks werden vervolgens toegepast om de kruinhoogte van 2,9 miljoen hectare heterogeen bos te schatten. De vegetatiehoogte in het studiegebied varieerde tussen 0-70 m en bestond uit verschillende soorten bos: van zeer open bos tot dicht regenwoud. De LiDAR metingen voor de kruinhoogte werden gebruikt voor ensemble regressie bomen (in een gemengd bos), gebaseerd op synoptische passieve optische beeldvorming en soortgelijke ondersteunende datasets zoals Landsat TM en MODIS. De resultaten suggereren dat de kruinhoogte kan worden geschat met een RMSE van 30% (5,5 m) wanneer de resultaten gevalideerd worden met een onafhankelijke inventarisatie dataset. Deze uitkomst is gelijkwaardig aan die uit eerdere studies van minder complexe bossen, en wordt gebruikt door de Europese ruimtevaartorganisatie ESA om de kruinhoogte te schatten. De modeluitkomsten toonden echter een systematische fout, waarbij de hoogte van de lage en hoge bomen respectievelijk over- en onderschat werden. Dit werd gecorrigeerd door de aftrek van een standaard inschatting van deze fout. Productie van een kaart met de boomhoogte over een groter gebied leverde een goed product dat een breed scala aan boomsoorten behelsde. De 30m resolutie zorgde ervoor dat veel landschapselementen geïdentificeerd konden worden, zoals houtkap percelen. De gepresenteerde techniek maakte gebruik van een open source computing framework alsmede vrij verkrijgbare datasets, om de toepasbaarheid voor landbeheerders en ook wetenschappers te vergroten.

10. ITC Dissertation List

http://www.itc.nl/research/phd/phd_graduates.aspx