

The use of airborne LiDAR to assess tree species and forest biomass in
subtropical forests

by

Lin Cao

B.Sc., Nanjing Normal University, China, 2006

M.Sc., Nanjing Forestry University, China, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2015

© Lin Cao, 2015

Abstract

The subtropical forest biome accounts for approximately a quarter of the area of China and is particularly important for local economies, and for maintaining biodiversity and the carbon balance of forest ecosystems. Despite their importance, there is still considerable uncertainty about the characterization and spatial distribution of tree species, as well as the carbon budgets of these forests, many of which have been altered by anthropogenic activities. Remote sensing has the potential to provide quantitative, spatially explicit information for mapping and monitoring forest ecosystems. It is also a cost-effective tool to provide temporally uniform and “wall-to-wall” observations over time. Light Detection and Ranging (LiDAR) is an active remote sensing laser technology that provides an advantage over most other remote sensing technologies in its ability to provide detailed three-dimensional information of forest canopy structure, which is particularly useful for studying forest biophysical and structural properties. The aim of this dissertation is to investigate novel approaches for using and examining the effectiveness of LiDAR technologies, in order to classify tree species and estimate forest biomass and dynamics, across a study site within the subtropical region of southeast China. Specifically, airborne LiDAR was evaluated for its ability to: (i) discriminate tree species using small-footprint full-waveform LiDAR metrics; (ii) estimate forest biomass components by discrete-return and full-waveform LiDAR metrics; (iii) spatially extrapolate the estimation of forest biomass components, and (iv) predict and map biomass dynamics using multi-temporal LiDAR data. The results of this dissertation confirm that LiDAR-based approaches can make significant contributions to analyze the structure, composition and distribution of tree species across the study site, and provide effective methodologies and techniques for developing high resolution, spatially explicit estimations of forest biomass (and its dynamics). These methods have important applications to sustainable forest management, forest carbon cycling studies and carbon accounting projects.

Preface

In this dissertation I established the research objectives and questions, conducted the data analysis and wrote the manuscripts and chapters. The co-authors provided advice on methodology and made editorial comments as required. In Chapter 4, co-author Txomin Hermosilla processed the full-waveform LiDAR data and calculated the metrics. Discrete return LiDAR data in 2007 were collected by the Jiangsu Surveying and Mapping Engineering Institute, China. Full-waveform LiDAR data in 2013 were collected by the Institute of Forest Resource Information Techniques (IFRIT), Chinese Academy of Forestry (CAF), China. Field plot measurements were collected by members of the Faculty of Forestry, Nanjing Forestry University, China. Forest stand inventory data in 2007 and 2012 were collected by the East China Forestry Investigation Planning and Design Institute, China.

A version of Chapter 3 has been submitted for publication as:

Cao, L., Coops, N.C., Innes, J.L., Dai, J.S., Ruan, H.H., She, G.H., Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data.

A version of Chapter 4 has been published in:

Cao, L., Coops, N.C., Hermosilla, T., Innes, J.L., Dai, J.S., She, G.H., 2014. Using small-footprint discrete and full-waveform airborne LiDAR metrics to estimate total biomass and biomass components in subtropical forests. *Remote Sensing*. 6, 7110–7135.

A version of Chapter 5 has been published in:

Cao, L., Coops, N.C., Innes, J.L., Dai, J.S., She, G.H., 2014. Mapping above- and below-ground biomass components in subtropical forests using small-footprint LiDAR. *Forests*. 5, 1356–1373.

A version of Chapter 6 has been submitted for publication as:

Cao, L., Coops, N.C., Innes, J.L., Sheppard, S.R.J., Ruan, H.H., She, G.H., Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data.

Table of Contents

Abstract	ii
Preface.....	iii
Table of Contents	iv
List of Tables	viii
List of Figures	x
List of Acronyms	xiii
Acknowledgements.....	xv
Dedication	xvii
1 Introduction.....	1
1.1 General background, objectives and chapter overview	1
1.2 An overview of airborne LiDAR.....	7
1.3 Airborne LiDAR for tree species classification	9
1.4 Airborne LiDAR for forest biomass estimation	13
1.4.1 Estimation of biomass components (using discrete-return LiDAR data)	13
1.4.2 Estimation of biomass components (using full-waveform LiDAR data)	17
1.4.3 Modeling and estimating biomass dynamics (using multi-temporal LiDAR data)	21
2 Study Area and Data	24
2.1 Study area	24
2.2 LiDAR data	26
2.3 Plot data	26
2.4 Stand inventory data	31
3 Tree Species Classification using Full-waveform LiDAR	33
3.1 Introduction.....	33

3.2	Materials	34
3.2.1	Study area.....	34
3.2.2	Field data.....	34
3.2.3	LiDAR data	34
3.3	Methods.....	35
3.3.1	Data pre-processing	36
3.3.2	Individual tree detection	37
3.3.3	Full-waveform LiDAR metrics.....	41
3.3.4	Metrics selection and classification	44
3.4	Results.....	46
3.5	Discussion.....	51
3.5.1	Segmentation of trees.....	51
3.5.2	Metrics selection and their explanation	51
3.5.3	Assessments of classification and the effects of voxel sizes	53
4	Estimation of Biomass Components using Discrete and Full-waveform LiDAR Metrics..	55
4.1	Introduction.....	55
4.2	Materials and methods	55
4.2.1	Study site.....	55
4.2.2	Field data.....	55
4.2.3	LiDAR data.....	56
4.2.4	Data pre-processing	56
4.2.5	LiDAR metrics.....	57
4.2.5.1	Discrete data based metrics.....	57
4.2.5.2	Full-waveform based metrics.....	58
4.2.6	Statistical analyses	59

4.3 Results.....	61
4.4 Discussion.....	69
5 Mapping Biomass Components using LiDAR Data.....	74
5.1 Introduction.....	74
5.2 Methods.....	74
5.2.1 Study site and materials.....	74
5.2.2 LiDAR metrics.....	76
5.2.3 Statistical analyses.....	77
5.2.4 Biomass mapping.....	78
5.3 Results.....	79
5.4 Discussion.....	85
6 Estimation of Biomass Dynamics using Multi-temporal LiDAR.....	87
6.1 Introduction.....	87
6.2 Materials and methods.....	87
6.2.1 Study area.....	87
6.2.2 LiDAR data.....	87
6.2.3 Plot data.....	88
6.2.4 Stand inventory data.....	90
6.2.5 LiDAR metrics.....	90
6.2.6 Predictive modeling and spatial extrapolation.....	91
6.3 Results.....	93
6.4 Discussion.....	99
7 Conclusions.....	105
7.1 Unique outcomes.....	108
7.2 Limitations.....	109

7.3 Future directions and recommended implications.....	111
References.....	117

List of Tables

Table 1.1 Summary of previous studies that classified tree species using full-waveform LiDAR data.	11
Table 1.2 Summary of previous studies that estimated forest biomass and biomass components using discrete-return LiDAR data.	16
Table 1.3 Summary of previous studies that estimated forest characteristics using full-waveform LiDAR.	19
Table 1.4 Summary of previous studies that predicted biomass change using multi-temporal airborne LiDAR data.	23
Table 2.1 The summary of scanner properties and flight parameters of the two LiDAR surveys. .	26
Table 2.2 Description of the forest characteristics for the 12 plots (with individual tree locations) for the research in chapter 3.	29
Table 2.3 Summary statistics of predicted ground data in 2007 and field estimated data in 2013 (for all 66 plots).	30
Table 2.4 The age class intervals and age groups for the dominant tree species at the research site.	32
Table 3.1 Description of all detected trees (in 6 species) with the statistics of the characteristics. .	39
Table 3.2 Description of the composite waveform metrics.	43
Table 3.3 Comparison of the tree species classification accuracies (of the selected metrics) and demonstration of the most important metrics (determined by the indices of MDA) for different voxel sizes.	48
Table 3.4 Confusion matrix for the detected trees on all the plots in three classification levels, i.e., using all six tree species, four main species and two generalized classes. The results of using the “selected metrics” and the “most important metrics” are compared. The values of the fields in the confusion matrix are given as percentage of predicted classes of the total number of reference classes, and the correctly classified results are highlighted with a grey background.	50
Table 4.1 The accuracy assessment results for three sets of predictive models among various forest types.	60
Table 5.1 Summary of the mean and range of the derived biomass attributes of the plots; $H_{\text{Lorey's}}$ (m): Lorey’s height; DBH_{avg} (cm): average DBH; W_f (Mg ha ⁻¹): foliage biomass; W_b (Mg ha ⁻¹): live branch biomass; W_s (Mg ha ⁻¹): trunk wood biomass; W_a (Mg ha ⁻¹): total	

above-ground biomass; W_r (Mg ha ⁻¹): root biomass; W_t (Mg ha ⁻¹): total biomass.....	75
Table 5.2 Summary of the calculated attributes of the selected stands; W_f (Mg ha ⁻¹): foliage biomass; W_b (Mg ha ⁻¹): live branch biomass; W_s (Mg ha ⁻¹): trunk wood biomass; W_a (Mg ha ⁻¹): total above-ground biomass; W_r (Mg ha ⁻¹): root biomass; W_t (Mg ha ⁻¹): total biomass.	76
Table 5.3 Summary of LiDAR metrics computed from LiDAR point cloud.....	77
Table 5.4 Summary of the biomass components prediction models and plot-level accuracy assessment results; W_f (Mg ha ⁻¹): foliage biomass; W_b (Mg ha ⁻¹): live branch biomass; W_s (Mg ha ⁻¹): trunk wood biomass; W_a (Mg ha ⁻¹): total above-ground biomass; W_r (Mg ha ⁻¹): root biomass; W_t (Mg ha ⁻¹): total biomass.	80
Table 5.5 Plot-level validation results: summary of the plot-level observed mean, the mean and standard deviation of the differences of the cross-validation results for the separated models.	81
Table 5.6 Stand-level validation results: summary of the observed mean, the mean and standard deviation for the differences in stand predictions using the selected regression equations. ...	82
Table 6.1 The fitted dummy variable models (and accuracy assessments) of DBH growth (2007-2013) for each tree species.....	89
Table 6.2 The fitted dummy variable models by the sample trees (and accuracy assessments) of tree height for each tree species.....	89
Table 6.3 Summary of LiDAR metrics for biomass stock and biomass change estimation.	90
Table 6.4 The best models for height, biomass stock and change estimation, and the statistics of accuracy assessments.....	94
Table 6.5 Summary of the cross-validation results for the R^2 , relative RMSE (rRMSE) and differences in cross-validation models for height, biomass stock and change estimation.	96

List of Figures

Figure 1.1 Map of global ecological zones. The subtropical regions lie between 23.5 and 40 degrees of latitude in both hemispheres (FRA, 2000).....	2
Figure 1.2 Interconnection of dissertation chapters (red), data sources (blue) and applications of the findings for forest resources management (FRM) and ecological studies. Notes: FWF=full-waveform, DR= discrete-return.	6
Figure 1.3 Schematic diagram of an airborne LiDAR system (Zhang and Cui, 2007).....	7
Figure 1.4 Illustration of full-waveform and discrete return LiDAR systems. The entire return pulse is digitized and recorded in a full-waveform LiDAR system. Only the peaks are recorded in a discrete return LiDAR system (Lefsky et al., 2002).....	8
Figure 1.5 Above and below ground forest biomass components (Popescu and Hauglin, 2014)...	13
Figure 2.1 Yushan forest study site and the locations of sample plots of three main forest types. The upper-right sub-figure shows the topography of the site.....	25
Figure 2.2 Photographs and hemispherical images of the six common tree species in the study area.	25
Figure 2.3 Map of the forest stands (in 2007) stratified by forest types (classified by the dominant tree species in the stands) (a) and age groups (b).	31
Figure 3.1 An overview of the workflow for tree species classification. DSM: Digital Surface Model; DTM: Digital Terrain Model; CHM: Canopy Height Model.	35
Figure 3.2 The process of extracting point cloud and creating composite waveforms from the observed waveforms. Two samples (i.e. one coniferous tree and one broadleaved tree) were used to show the processes and results.....	37
Figure 3.3 Height distributions (Mean of the proportion of returns within 0.3m height intervals) of all the detected coniferous trees (a) and broadleaved trees (b).	40
Figure 3.4 Illustration of a sample composite waveform (has been normalized by the DTM), and the metrics derived from the waveform. HOME: height of median energy; WD: waveform distance; HTMR: height/median ratio; NP: number of peaks (NP); ROUGH: roughness of outermost canopy; FS: front slope angle; RWE: return waveform energy.	42
Figure 3.5 Example map of CHM with the location of trees (dominant and co-dominant trees), the lidar-detected tree-tops and the estimated tree crowns.....	46
Figure 3.6 The matrix of scatter plots (lower panel) and the correlation coefficient (r) (upper panel) for the low-correlated metrics (with $r < 0.6$) for all detected trees. The histograms of each	

metric were drawn in the diagonal line. The values of r were colored as red for medium degree of correlations ($0.4 \leq r < 0.6$) and as green for low degree of correlations ($0.2 < r \leq 0.4$). See Table 3.3 for codes of the metrics.....47

Figure 3.7 Overall accuracy and the improvement of accuracy when successively including the selected metrics ranked by the indices of Mean Decrease Accuracy (MDA). The ranks (of importance) were shown in X-axis for each sub-figure. The corresponding voxel size was shown at the top of each sub-figure. Note: overall accuracy (in the legend) = overall classification accuracy of 6 tree species.....49

Figure 3.8 Demonstration of the effects of different voxel resolutions on the available (i.e., neither incomplete (a) nor over-synthesized (d)) rates of composite waveforms in one example tree. CV-Wave=composite waveform.....53

Figure 4.1 Technological route of this research.....57

Figure 4.2 Intensity graph of the square of the Pearson's correlation coefficient (R^2) between each biomass components and LiDAR-derived metrics. The values of the coefficients are transformed into pixels within a blue-red color range.....61

Figure 4.3 Plot-level observed and estimated above-ground and root biomass for the combination of forest-type specific models developed from discrete data based metrics, waveform based metrics and their combinations.....63

Figure 4.4 The selected LiDAR metrics of the discrete data model vs. full-waveform model in various forest types.....64

Figure 4.5 The selected LiDAR metrics of the combo model in various forest types.....65

Figure 4.6 Relationship between plot-level field measured and LiDAR-estimated biomass components of three sets of models (i.e. discrete data based model, full-waveform based model and the combo model) within each type of forest. Each biomass components are calculated as the average value of the plots within specific forest type.....67

Figure 4.7 Relationship between the amount of field-measured and LiDAR-estimated biomass components across all field plots (abscissa: all of the field plots; ordinate: the amount of biomass components): (a) the field-measured biomass components are displayed as area. The biomass components estimated by the combo model are displayed as dots; five plots with the largest bias in trunk biomass are highlighted and analyzed against four topographic and vegetation factors; (b) slope (degree) and canopy cover (%) of all plots with the five outlier plots highlighted.....68

Figure 5.1 Scatterplots and 1:1 line (dotted line) of biomass components between the field-measured and the model-estimated results of the selected stands (i.e., (a) foliage biomass; (b) branch biomass; (c) stem biomass; (d) above-ground biomass; (e) root biomass; (f) total biomass). The models were built from the sample plots of three forest types.....83

Figure 5.2 Mean value of the biomass components (Mg ha^{-1}) of the 53 sample plots from (1) field measurements; (2) estimations of the dummy variable including stepwise regression model; and (3) predictions of the general models of three forest types, <i>i.e.</i> , coniferous, broadleaved and mixed forests.....	84
Figure 5.3 Maps of biomass components (Mg ha^{-1}) across the research site (pixel size = 30m), which calculated from the individual models; (A) map of the estimated foliage biomass; (B) branch biomass; (C) stem biomass; (D) root biomass; (E) above-ground biomass (AGB); (F) total biomass.....	84
Figure 6.1 LiDAR point cloud profiles (5 m wide) in 2017(a) and 2013(b) through an undisturbed forest. Note that the heights (z) of point clouds have been normalized.	93
Figure 6.2 Canopy height distributions (vertical interval in 0.3m) calculated by LiDAR first returns in 2007 (gray bars) and 2013 (black line) for six coniferous and broadleaved plots (dominant tree species are shown in the parentheses). The corresponding hemispherical photos are shown in the lower-right for each plot.	93
Figure 6.3 Relationship between ground and LiDAR estimated Lorey's mean height in 2007(a) and 2013(b), above ground biomass (AGB) in 2007(c) and 2013(d), and the change of biomass (<i>i.e.</i> , ΔAGB) estimated by the indirect approach (e) and direct approach (f).	95
Figure 6.4 The spatial prediction of the changes of AGB across the research site (between 2007 and 2013). The upper-right sub-figure shows the scatterplot between LiDAR estimated AGB in 2007 and in 2013 for all the cells across the research. Some highlighted locations of AGB changes are marked on the map, such as, an increase of AGB in regenerated broadleaved forests (a, b, d, e, g) and a decrease of AGB in Chinese fir/tea agroforestry systems (transformed from the Chinese fir stands after silvicultural treatments) (c, f).....	97
Figure 6.5 Histogram and boxplot of above ground biomass (AGB) distributions across the research site in 2007 (a) and 2013(b), the shift (c) and changes (d) between the two times...97	97
Figure 6.6 The boxplots of LiDAR estimated AGB change for all the stands within each age group. The median values were displayed on top of each boxplot.....	98
Figure 6.7 The bar graph of the mean of LiDAR estimated AGB for all the stands within each forest type in 2007, 2013 and their changes. Note: M.Pine=Masson pine; C.Fir=Chinese fir; S.Pine=Slash pine; S.Oak=Sawtooth oak; S.Gum=Sweet gum; O.B = Other broadleaves. ...	98

List of Acronyms

Acronym	Description
3D	Three Dimensional
ABA	Area Based Approach
AGB	Above Ground Biomass
ANN	Artificial Neural Network
BA	Basal Area
CHM	Canopy Height Model
CNFI	Chinese National Forest Inventory
DBH	Diameter at Breast Height
DR	Discrete Return
DSM	Digital Surface Model
DTM	Digital Terrain Model
ESF	Ecological Service Forest
FP	Foliage Profile
FRM	Forest Resources Management
FS	Front Slope
FWF	Full Waveform
GIS	Geographical Information System
GPS	Global Positioning System
HOME	Height of Median Energy
HTMR	Height to Median Ratio
IMU	Inertial Navigational Measurement Unit
ITC	Individual Tree Crown
JSCORS	Jiangsu Continuously Operating Reference Stations
LDA	Linear Discriminant Analysis
LiDAR	Light Detection and Ranging
LVIS	Laser Vegetation Imaging Sensor
MDA	Mean Decrease Accuracy
NDVI	Normalized Difference Vegetation Index
NP	Number of Peaks
PCA	Principal Component Analysis
RADAR	Radio Detection and Ranging
REDD	Reducing Emissions from Deforestation and forest Degradation in developing countries
RF	Random Forest
RMSE	Root Mean Squared Error
ROUGH	Roughness of Outermost Canopy
RWE	Return Waveform Energy

Acronym	Description
UNFCCC	United Nations Framework on Climate Change
VDR	Vertical Distribution Ratio
WD	Waveform Distance

Acknowledgements

Foremost, I would like to thank Dr. John Innes, my supervisor, who gave me the freedom to pursue my research interests, inspiration and confidence, and provided insightful discussions about this research. I am sincerely grateful and indebted to Dr. Nicholas Coops for his unwavering support of my PhD study and research, his patience, enthusiasm, motivation and immense knowledge. I would also like to acknowledge the support of Dr. Stephen Sheppard who provided scientific advice, and shared his comprehensive knowledge and practical suggestions for my research.

My special thanks to Guangyu Wang (UBC), Tongli Wang (UBC), Qinglin Li (Government of B.C., Victoria, Canada) and Baozhang Chen (Chinese Academy of Sciences (CAS), China) for sharing their precious experiences of their PhD studies and providing their brilliant suggestions on the research. Thanks to Dr. Peter Marshall (UBC) for expert advice on the experimental design of tree core sampling, and to Dr. Valerie Lemay for teaching me forest inventory methods, biometric techniques and statistical skills. Thanks to Dr. Peter Krzystek (Munich University of Applied Sciences (MUAS), Germany) and Dr. Txomin Hermosilla (UBC) for providing me with a deeper understanding of full-waveform LiDAR data calibration and processing.

I am also grateful to members of Nanjing Forestry University (NJFU) and the Chinese Academy of Forestry (CAF), China. Thanks to Dr. Guanghui She (NJFU), Dr. Honghua Ruan (NJFU), Dr. Zengyuan Li (CAF) and Dr. Yong Pang (CAF) for sharing their expertise and guidance throughout the research. Special thanks to my friend Jingsong Dai (NJFU), who provided excellent questions and assisted in leading the groups for field work during the hot summers (38-40°C), and to graduate students from the Department of Landscape Plant, Silviculture (especially Dr. Kun Liu) and Forest Resources Management (especially Ting Xu and Xin Shen) at NJFU for field assistance in harsh forest conditions.

This research was funded by the Asia-Pacific Network for Sustainable Forest Management

and Rehabilitation and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

I would also like to thank all members of the Integrated Remote Sensing Studio (UBC, IRSS), collaborators (especially the foresters in Yushan National Forest Park and members in the Jiangsu Surveying and Mapping Engineering Institute), co-authors and reviewers who provided excellent advices, insights, and shared data throughout my PhD research.

Finally, I would like to thank my wife Xiaoqian Yang for her continuous love, understanding and support throughout my PhD study, and to my parents (especially my father, my role model in life, who is proud of his PhD degree in Forest Ecology undertaken with Dr. Hamish Kimmins at UBC, and eagerly waiting to see his son complete a PhD degree from his alma mater) for supporting and encouraging me. I could not have completed my dissertation without them.

Dedication

To my little ginkgo:

三人行，必有我師焉

學而不思則罔，思而不學則殆

學而時習之，不亦說乎

When I walk along with two others, from at least one I will be able to learn.

Learning without thought is labor lost; thought without learning is perilous.

Is it not pleasant to learn with a constant perseverance and application?

- *The Analects of Confucius*

1 Introduction

1.1 General background, objectives and chapter overview

Forests constitute a major part of terrestrial ecosystems, occupying approximately 30% of the world's land area (Dixon et al., 1994). Two-thirds of all plants and animals living on land use forests as a component of their habitat and as a result forests are the most bio-diverse terrestrial ecosystems globally (Schmitt et al., 2009). Forests absorb carbon through photosynthesis and sequester it as biomass, thus contributing significantly to the mitigation of global climate change, and it is estimated that over 80% of the world's aboveground and 40% of belowground carbon are stored in forest ecosystems (Lieth and Whittaker, 1975; Dixon et al., 1994). Because of this, the United Nations Framework on Climate Change (UNFCCC) has created a collaborative program on Reducing Emissions from Deforestation and forest Degradation in developing countries (REDD), in order to establish efficient systems for monitoring forest carbon stocks (Gibbs et al., 2007).

Tropical (and subtropical) forests are species-rich, carbon-dense and highly productive ecosystems (Kant and Berry, 2005) comprising 56% of the world's forests (Fleming et al., 2011). Subtropical forests occur in the transition from the tropics to temperate zones, usually between 23.5 and 40 ° latitude in both hemispheres (Pancel, 1993) (Figure 1.1). Sub-tropical forests have been defined by FAO (2001) as having mean monthly temperatures of over 10° C for more than eight months a year, and consist of two distinct types of subtropical forest, i.e., subtropical humid and subtropical dry, distinguished by the seasonal distribution in rainfall and humidity (Sands, 2005). Most of the subtropical humid forests are located on eastern seaboard, with south-eastern China containing the largest proportion of humid subtropical forests (0.53×10^8 ha) globally (Fang and Chen, 2001). As this area is located in China's most densely populated regions and these regions have the greatest economic growth, the subtropical forests of China are particularly threatened (Wang et al., 2001).

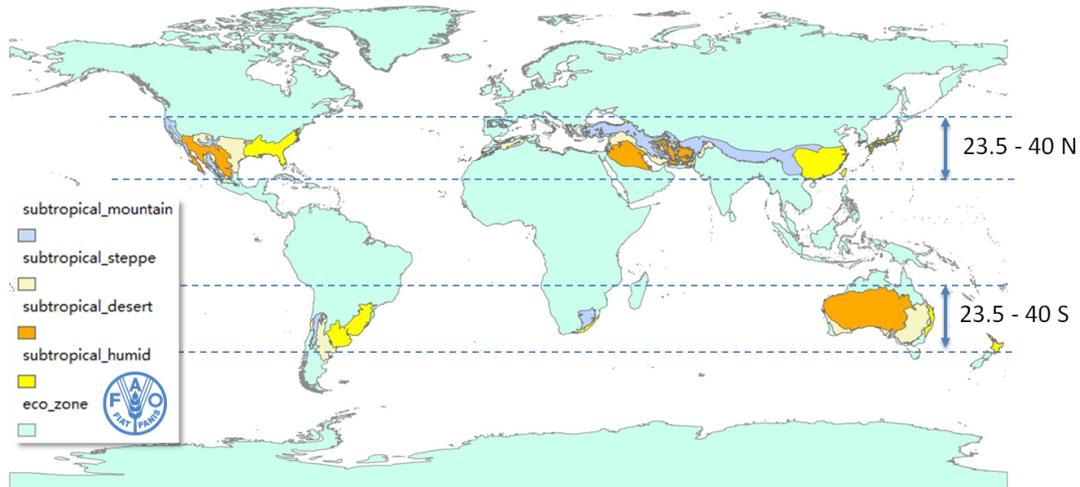


Figure 1.1 Map of global ecological zones. The subtropical regions lie between 23.5 and 40 degrees of latitude in both hemispheres (FRA, 2000).

Tropical (and subtropical) forests contain at least half of all global living species (Fleming et al., 2011), and, as a result, the characterization of tree species and their spatial distribution is critical for sustainable forest management and understanding regional ecological processes. Information on individual tree species is important for forest management activities, such as estimating timber quality attributes, biomass and for assessing species diversity (Jones et al., 2010; Popescu et al., 2003). Species information is also needed when modelling individual tree growth (Falkowski et al., 2010; Yu et al., 2004), assessing tree competition (Biging and Dobbertin, 1992) and elucidating the response of forests to disturbance (Parker et al., 2004), which in turn is important for understanding and management of subtropical forest ecosystems (Li et al., 2012; Plourde et al., 2007). Traditional methods of tree species identification have been based on either field inventory approaches or interpretation of large-scale aerial photographs. Both of these methods are labour intensive and time consuming (Franklin et al., 2000). Remote sensing techniques can provide quantitative spatially explicit information and “wall-to-wall” observations for mapping and monitoring of forest ecosystems (Gibbs et al., 2007). However, although remotely sensed data have been widely explored for forest applications (Franklin, 1994; Key et al., 2001; Martin et al., 1998), the traditional optical remote sensing techniques still fall short of capturing three-dimensional forest structures (Lovell et al., 2003).

In addition to species distribution information in subtropical forest stands, information on their carbon stocks is also important as many countries are considering developing national forest carbon inventories and policies for carbon management that all require accurate methods of biomass estimation (Piao et al., 2005). Despite its importance, there is still considerable uncertainty about carbon budgets within these subtropical forests, despite several national-scale studies using historical forest inventory data (Fang and Chen, 2001). Remote sensing also has the potential to provide quantitative information for estimating biomass (and biomass components) (Hollaus et al., 2014), and has been shown to be a cost-effective tool to provide temporally uniform carbon stock observations over time (Wulder et al., 2010). Remote sensing for forest biomass (i.e., carbon stocks) estimation has traditionally focused on the use of imagery from passive optical and RADAR (Radio Detection and Ranging) sensors in the tropical and subtropical forests (Clark et al., 2011; Du et al., 2014). However, the sensitivity and accuracy of these sensors have been shown to decrease with increasing aboveground biomass and leaf area index, which inhibits reliable carbon stock estimates (Lefsky et al., 2002; Turner et al., 1999; Tsui et al., 2012). Furthermore, optical sensors can only provide limited information on the vertical structure of the forests, and the application of RADAR is limited by terrain, surface moisture and other factors (Wulder, 1998; Rosenqvist et al., 2003).

Light Detection and Ranging (LiDAR) is an active remote sensing laser technology capable of providing detailed, spatially explicit, three-dimensional information on forest canopy structure (Lefsky et al., 2002). LiDAR systems send out pulses of laser light and measure the signal return time to directly estimate the height and vertical structure of forests (Dubayah & Drake, 2000), which is particularly useful for studying biophysical and structural properties over a wide range of forest types (Dubayah and Drake, 2000). Previous studies have demonstrated the potential of LiDAR to classify tree species mostly in boreal forests in Scandinavia (Holmgren and Persson, 2004; Ørka et al., 2009; Yu et al., 2014), temperate forests in central Europe (Heinzel and Koch, 2011; Reitberger et al., 2009; Yao et al., 2012) and North America (Jones et al., 2010; Suratno et al., 2009; Vaughn et al., 2012). The saturation problem in carbon stock estimation in

high-biomass forests can be overcome by LiDAR, as laser pulses can penetrate through even the dense multi-layered canopies and retrieve vertical structure information from return signals, and as a result there tends to be a strong correlation between LiDAR data and many forest biophysical properties (Lu et al., 2012). Previous studies have demonstrated the potential of LiDAR to accurately estimate stand level forest biomass in boreal (Næsset & Gobakken, 2008; Næsset et al., 2013; Kankare et al., 2013), temperate (Lefsky et al., 1999; Lim and Treitz, 2004) and tropical forests (Drake et al., 2002; Asner et al., 2010), as well as biomass at the single tree level (Popescu, 2007; Hauglin et al., 2013).

For the last past decade or so, most conventional airborne LiDAR systems produced small-footprint discrete-return (DR) data (Lindberg et al., 2012). These systems record multi-returns ($n \leq 5$) per transmitted pulse with each return representing the 3D position and intensity of the reflected light (Wulder et al., 2008). However, due to limitations in the electronics, the discrete LiDAR systems can only distinguish surfaces that are sufficiently spaced apart (Parrish et al., 2011). Furthermore, since the returned signal is interpreted internally, the recorded information from these systems is limited to a simple time measurement for each pulse (Wagner et al., 2006). In contrast, small-footprint full-waveform (FWF) LiDAR systems have become available commercially since 2004 (Mallet and Bretar, 2009); these systems provide new opportunities for forestry studies. The FWF systems digitize and record the entire backscattered signal of each emitted pulse, and allow the recording of geometric and physical properties of intercepted objects (Reitberger et al., 2006). The analysis of FWF data has enhanced point extraction (Wu et al., 2011) and provided additional target information (e.g., intensity, pulse width and backscatter cross section etc.) (Höfle et al. 2008; Hollaus et al. 2009). This extended return information has been successfully applied to forestry applications, such as the estimation of forest structure and canopy fuel parameters (Sumnall et al., 2012; Hermosilla et al., 2014; Neuenschwander, 2012), single tree detection (Reitberger et al., 2008; Gupta et al., 2010) and tree species classification (Xu et al., 2012).

The primary objective of this dissertation is therefore to investigate the use and effectiveness

of novel LiDAR technologies to classify tree species and estimate forest biomass and dynamics in forest stands within the subtropical region of southeast China. From this primary objective, four specific research questions have been formulated:

- 1) Can small-footprint full-waveform LiDAR data be used to classify tree species in subtropical forest stands?
- 2) What are the advantages of estimating the forest biomass components by discrete-return and full-waveform LiDAR metrics alone or combined?
- 3) How can airborne LiDAR be used to provide spatially explicit maps of forest biomass components?
- 4) How can multi-temporal airborne LiDAR data be used to estimate and map biomass dynamics over time?

These research questions have been addressed in the context of several case studies in Yushan forest, located in southeast China, within the subtropical region. Following a description of the general study area and data sources in Chapter 2, questions 1, 2, 3 and 4 are then addressed in Chapters 3, 4, 5 and 6 respectively. Specifically, Chapter 3 examines the use of airborne full-waveform LiDAR data to discriminate six subtropical tree species (at three levels of classification) using a voxel-based composite waveform approach and a Random Forest classifier. Chapter 4 presents the use of DR and FWF metrics for estimating forest biomass (i.e., above-ground and total biomass) and its components (i.e., foliage, branch, trunk and root biomass), and the uncertainty analysis of the estimates as well. Chapter 5 spatially extrapolates the estimates of forest biomass components (in 2013) across the site using LiDAR data, and then evaluates the accuracy using a suite of independent stand-level field inventory data. Chapter 6 assesses different approaches to estimate the magnitude and spatial patterns of the above ground biomass (AGB) dynamics (from 2007 to 2013) using multi-temporal airborne LiDAR data. Finally, Chapter 7 concludes this dissertation by summarizing the major findings, potential implications, applications to forest management, and possible future research directions.

The interconnection of dissertation chapters (red), data sources (blue) and applications of the technologies for forest resources management (FRM) and ecological studies are shown in Figure 1.2. The following sections provide an overview of airborne LiDAR technology and a summary review of the key concept and background of chapter 3 to 6.

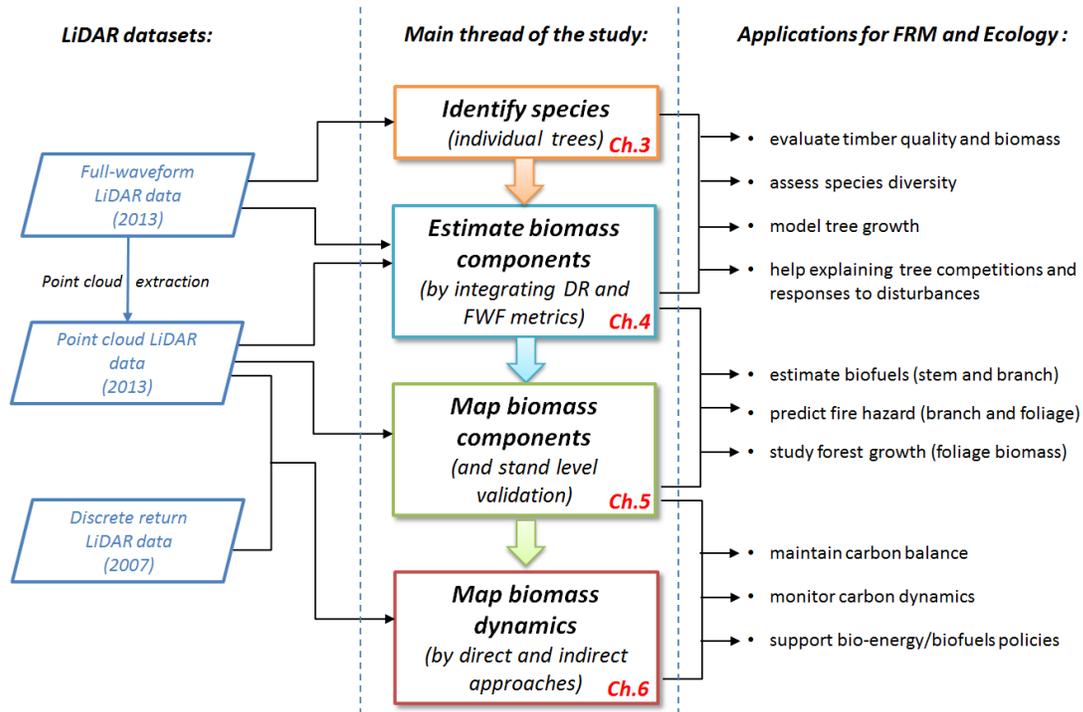


Figure 1.2 Interconnection of dissertation chapters (red), data sources (blue) and applications of the findings for forest resources management (FRM) and ecological studies. Notes: FWF=full-waveform, DR= discrete-return.

1.2 An overview of airborne LiDAR

The LiDAR sensor emits pulses of light to determine the range (or distance) to a target object (Lim et al., 2003). The distance to the target is measured by the elapsed time between the emission of the pulse and the detection of the reflected (or backscattered) signal (Gatziolis and Andersen, 2008):

$$\text{Range (m)} = (\text{Speed of Light} \times \text{Time of Flight}) \div 2 \quad (1.1)$$

The coordinates (x, y, z) of the target object is defined with the known position and orientation of the sensor, and the range measurements.

An airborne LiDAR system comprises a set of instruments: a laser ranging unit, an inertial navigational measurement unit (IMU), a high-precision global positioning system (GPS) receiver and a computer interface (Figure 1.3). The GPS and IMU continuously record the three-dimensional position and attitude (i.e., roll, pitch, and yaw) of the platform, and the computer interface manages communication among devices and data storage (Gatziolis and Andersen, 2008).

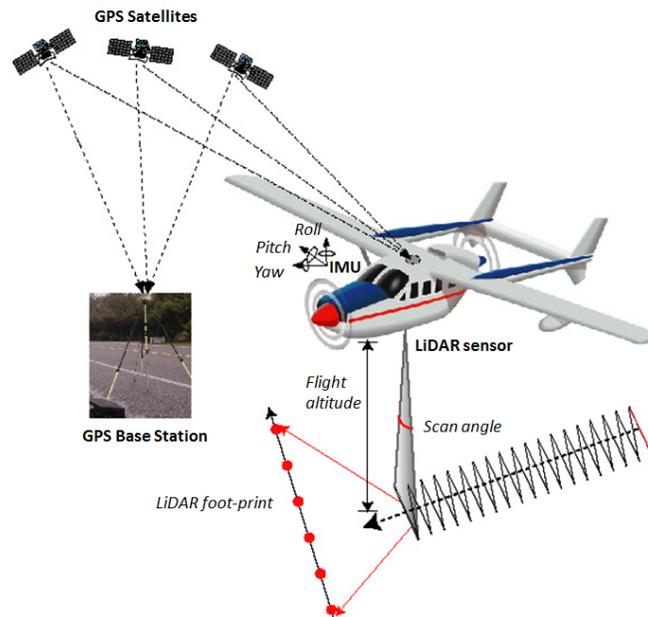


Figure 1.3 Schematic diagram of an airborne LiDAR system (Zhang and Cui, 2007)

Airborne LiDAR systems can be categorized as either discrete return (DR) or full waveform (FWF), and differ with respect to how backscattered laser energy is quantified and recorded by the system's receiver. (Lefsky et al., 2002; Gatzliolis and Andersen, 2008). DR systems quantify reflected energy (at amplitude intervals) and recorded return targets referenced in time and space (White et al., 2013) (Figure 1.4). In a forested area, a laser pulse may be reflected from several surfaces, e.g., the canopy (i.e., branches and leaves) and often the ground. The resulting dataset (of DR LiDAR) is a cloud of 3D points, with the upper points representing the canopy and the lower points representing the ground (Lim et al., 2003). FWF systems record the reflected energy from each laser pulse at very small time intervals. The sensitivity of the FWF sensor is at the level of a few photons per interval (typically 15 cm), and as a result, even small volumes of vegetation can influence the shape of the return waveform (Schuckman and Renslow, 2014) (Figure 1.4). The following sections provide a review of the use of airborne LiDAR for species classification, biomass and biomass change estimations.

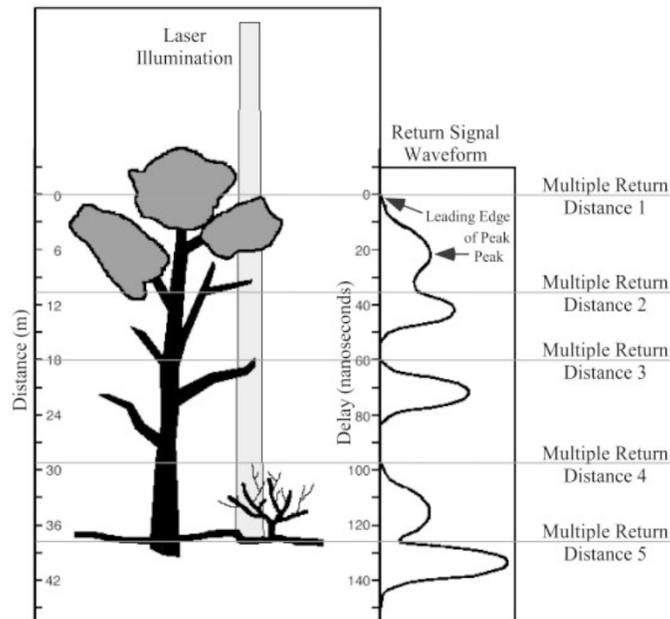


Figure 1.4 Illustration of full-waveform and discrete return LiDAR systems. The entire return pulse is digitized and recorded in a full-waveform LiDAR system. Only the peaks are recorded in a discrete return LiDAR system (Lefsky et al., 2002).

1.3 Airborne LiDAR for tree species classification

Previous studies have demonstrated that attributes extracted from full-waveform (FWF) LiDAR (e.g. amplitude, echo width) and the derived metrics, can be used to enhance the performance of tree species classification. Reitberger et al. (2008) classified coniferous and deciduous trees using FWF metrics (i.e., 3-dimensional coordinates of the reflections, intensity and pulse width) and an unsupervised classifier in the Bavarian Forest National Park, southeast Germany. Using a variety of models, they found an overall accuracy of 85% in a leaf-on situation and 96% in a leaf-off situation. Höfle et al. (2008) found that the mean echo width and backscatter cross section from the segmentations of the canopy surface were able to separate larch (*Larix decidua*), from oaks (*Quercus robur*, *Quercus petraea*) and beech (*Fagus sylvatica*). The benefits of FWF data were also shown in Vaughn et al. (2012), who reported an increment of classification accuracy of approximately 6% for five species (overall accuracy = 85.4%) when incorporating the waveform information. Hollaus et al. (2009) investigated the capability of FWF data for classifying three dominant tree species in a mixed woodland area. The metrics of echo height distributions, mean and standard deviations of echo widths, mean intensities and backscatter cross-sections, were all extracted from the segmented tree crowns, and then used as inputs for the decision-tree based classification. The echo width was demonstrated as being useful for the discrimination of spruce and larch, and the overall accuracy for classifying all the tree species was 75%. An increase in accuracy of 8% was found when discriminating between coniferous vs. deciduous species.

Heinzel and Koch (2011) explored a large number of rasterized FWF metrics, and selected the three most important variables based on the intensity, width and the total number of targets using linear discriminant analysis (LDA). Overall accuracies of 57% for six classes, 78% for four main species and 91% for coniferous and broadleaved trees were achieved. Höfle et al. (2012) explored the possibilities of using FWF data to extract vegetation in urban areas. An overall accuracy of 94% was achieved by using an artificial neural network (ANN) classifier, and a slight enhancement of

2% was found when the classifier was changed to a decision-tree classifier based on metrics including intensity, echo width and the height distribution of echoes within segments. Yu et al. (2014) assessed the possibility of using low density FWF LiDAR data for individual tree detection and species classification in southern Finland. In their research, a Random Forests (RF), which is an ensemble non-parameter classification method, was used for metrics selection and classification. The best result was 73.4% (overall accuracy), and they suggested that including the FWF metrics can improve the classification accuracy of pine, spruce and birch by an increment of 10%. A detailed summary of previous studies that have classified tree species using full-waveform LiDAR data is provided in Table 1.1.

Table 1.1 Summary of previous studies that classified tree species using full-waveform LiDAR data.

Previous study	Study area	Forest type (tree species)	Full-waveform (FWF) sensors	FWF metrics for classification	Classifier	Max number of classes	Highest accuracy of the classification
Reitberger et al. (2008)	Southeast Germany	Temperate forest (alpine spruce forest, mixed mountain and spruce forests)	Riegl LMS-Q560	3-D coordinates of the reflections, intensity and pulse width	K-means /EM classifier	2 (i.e., Coniferous and deciduous trees)	Overall accuracy = 85% (leaf-on) = 96% (leaf-off)
Höfle et al. (2008)	West of Vienna, Austria	Temperate forest (red beech, oak and larch)	Riegl LMS-Q560	Echo width and backscatter cross section	Not reported	3	Not reported
Vaughn et al. (2012)	Seattle, Pacific Northwest, United States	Temperate forest (Douglas-fir, western red-cedar, black cottonwood, big-leaf maple and red alder)	Not reported	Fourier transform characteristics and voxel cluster characteristics	Support Vector Machine (SVM)	5	Overall accuracy = 85.4 % (kappa = 0.817) for all five species
Hollaus et al. (2009)	The federal state of Lower Austria	Temperate forest (red beech, oak and larch)	Riegl LMS-Q560	Echo width, backscatter cross section, the distribution of the echoes	Knowledge-based fuzzy classification	3	Overall accuracy = 85.4 % for all three species
Heinzel and Koch (2011)	Southwestern Germany	Temperate forest (Scots pine, oak, European beech, Norway spruce, Hornbeam and wild cherry)	Riegl LMS-Q560	Intensity, width and the total number of targets	Linear discriminant analysis (LDA)	6	Overall accuracy = 57 % for all six species

Previous study	Study area	Forest type (tree species)	Full-waveform (FWF) sensors	FWF metrics for classification	Classifier	Max number of classes	Highest accuracy of the classification
Höfle et al. (2012)	City center of Vienna, Austria	Temperate urban forest (beech, Norway maple, plane, linden and chestnut)	Riegl LMS-Q560	Intensity, echo width and the height distribution of echoes	artificial neural network (ANN)/ decision-tree classifier	2 (i.e., vegetation or non-vegetation)	Overall accuracy = 96 % for urban vegetation
Yu et al. (2014)	Southern Finland	Boreal forest (Scots pine, Norway spruce and birch)	Leica ALS60	Number of peaks, peak amplitude, distance and distribution of the total received energy etc.	Random forests (RF)	3	Overall accuracy = 73.4 % for all three species

* FWF: full-waveform.

1.4 Airborne LiDAR for forest biomass estimation

1.4.1 Estimation of biomass components (using discrete-return LiDAR data)

LiDAR pulses can penetrate through the forest canopy and sample the vertical distribution of the canopy (e.g., branches and leaves). Previous studies have found ecological and biomechanical links between biomass and vertical structure (Drake et al., 2003; Franco and Kelly, 1998; Givnish, 1986; King and Loucks, 1978), and as a result, a strong correlation between LiDAR data and biomass can be expected (Lefsky et al., 1999; Drake et al., 2002; Lu et al., 2012).

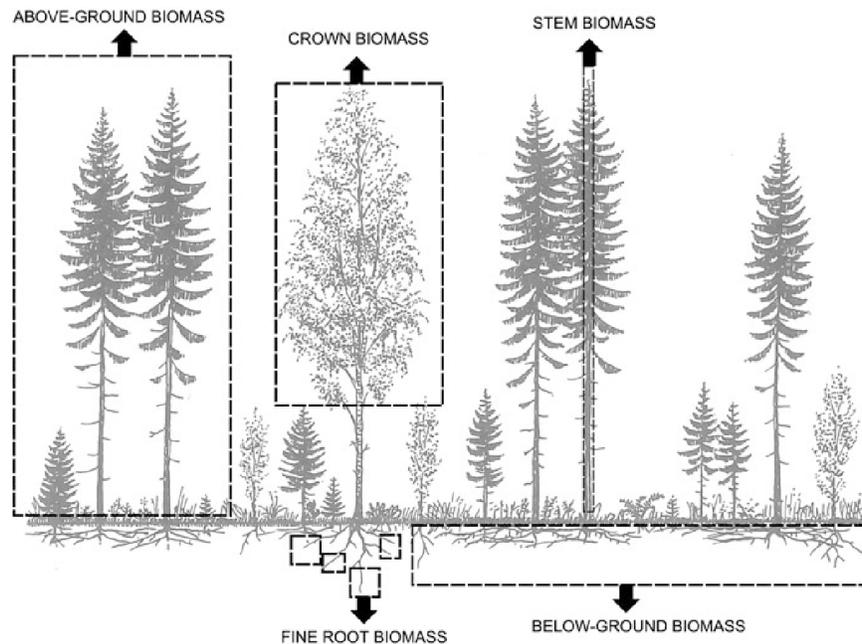


Figure 1.5 Above and below ground forest biomass components (Popescu and Hauglin, 2014)

Using discrete-return (DR) LiDAR data, area-based (ABA) and individual tree crown (ITC) approaches are the most common for estimating biomass and biomass components (see Figure 1.5) in previous studies (Hollaus et al., 2014). At the stand level, Lim and Treitz (2004) evaluated the canopy-based quantile metrics for estimating above ground biomass (AGB) and biomass components in a hardwood forest using the ABA approach, and found that the models (based on each canopy-based quantile) have relative high and similar predictive capabilities for each biomass component, with respect to R^2 (0.82-0.90, the difference between any two models was lower than

9%). Næsset (2004) estimated above-ground and below-ground biomass using LiDAR based canopy height and density metrics in a boreal forest (with young and mature conifers). The results demonstrated that above-ground biomass explained 92% of the variability, whereas the models for below-ground biomass explained 86%. Næsset and Gobakken (2008) extended their research to ten different areas within the boreal forest zone of Norway. The statistical effects of age class were assessed by including dummy variables in the models, and tree species composition was treated as continuous variables. They found that above-ground biomass explained 88% of the variability, whereas the models for below-ground biomass explained 85%. Their results also demonstrated that the proportion of broadleaves had a significant effect on above-ground biomass only, while the effect of age class was significant only in the below-ground biomass model. Stephens et al. (2012) estimated the national carbon stock of New Zealand using a double sampling scheme based on ground plots and LiDAR data. The regression model, fitted by the ground-based measurements and LiDAR metrics (e.g., 30th percentile height and canopy cover), explained 74% of the variation in total carbon. Compared with the ground-based estimates, the LiDAR-based approach improved the accuracy of the national carbon stock estimation in 2008 by 6%. Tsui et al. (2012) estimated above-ground component and total biomass for a coniferous temperate forest using LiDAR-derived metrics and backscatter coefficients from RADAR. The results of the LiDAR-based models indicated that stem biomass had the most accurate relationship with the LiDAR measurements, while crown biomass showed the lowest relationship (relative RMSE = 16%–22%).

At the single tree level, Popescu (2007) used airborne LiDAR data to estimate the aboveground biomass (AGB) and component biomass for loblolly pine (*Pinus taeda* L.) in southeastern United States using the ITC approach. LiDAR-derived single tree-level height and crown diameter were used to estimate tree DBH, AGB and component biomass using regression models. The results showed that 93% of the variability of individual tree biomass, 90% of DBH, and 79–80% of components biomass could be explained by the predictive models. Hauglin et al. (2013) estimated single-tree branch biomass of Norway spruce using LiDAR derived crown

volume, calculated from an interpolated crown surface. Regression models with branch biomass (through destructive sampling) as the dependent variable were fitted to the data. The results demonstrated that the best LiDAR-based models can estimate branch biomass with a higher accuracy than existing allometric biomass equations based on field measurements. Kankare et al. (2013) combined the ABA and ITC approaches to map the AGB in a boreal forest in Finland. The ABA was trained with ITC measurements and resulted in a relatively high accuracy (RMSE=28.5–34.9%) for ABA estimation. They concluded that although it is not feasible to use visually corrected ITC in wall-to-wall forest biomass inventory, it could provide a cost-efficient application for training ABA in biomass mapping. A detailed summary of previous studies that estimated forest biomass and biomass components using discrete-return LiDAR data is provided in Table 1.2.

Table 1.2 Summary of previous studies that estimated forest biomass and biomass components using discrete-return LiDAR data.

Previous study	Study area	Forest type (characteristics)	Modelling approaches	Estimated components of biomass (or carbon)	Average point densities (points m ⁻²)	Type of predictive models	Accuracy of the best predictive model
Lim and Treitz (2004)	Southeast Canada	Temperate forest (tolerant hardwood forest)	ABA	Stem wood, stem bark, live branch and foliage biomass	3-5	Multiple regression models	$R^2 = 0.82-0.90$
Næsset (2004)	Southeast Norway	Boreal forest (young and mature conifers)	ABA	above-ground and below-ground biomass	1.1	Multiple regression models	$R^2 = 0.92$ (aboveground) $R^2 = 0.86$ (belowground)
Stephens et al. (2012)	Across New Zealand	Temperate and subtropical forest (planted and native species)	ABA	Total carbon, above-ground carbon	>3	Double sampling regression	$R^2 = 0.74$ (total) $R^2 = 0.81$ (aboveground)
Tsui et al. (2012)	Vancouver island, western Canada	Coastal temperate forest (actively logged and managed)	ABA	Above-ground, stem and crown biomass	3.74	Linear regression models	Adj $R^2 = 0.82$ (AGB) Adj $R^2 = 0.86$ (stem) Adj $R^2 = 0.72$ (crown)
Popescu (2007)	Southern United States	Loblolly pine stands	ITC	Total, foliage, coarse roots, bark, and stem biomass	2.6	Linear regression models	$R^2 = 0.93$ (total) $R^2 = 0.79-0.80$ (componer)
Hauglin et al. (2013)	Southeast of Norway	Norway spruce	ITC	Canopy base height (CBH) and branch biomass	7-10	Linear regression models (for estimating CBH)	$R^2 = 0.83$ (for CBH)
Kankare et al. (2013)	Southern Finland	Boreal forest (heterogeneously managed forests)	ITC and ABA	Total, stem, canopy and live branch biomass	10	Linear regression models	$R^2 = 0.96$ (total) $R^2 = 0.99$ (stem) $R^2 = 0.85$ (canopy) $R^2 = 0.83$ (live branch)

ABA: area-based approach; ITC: individual tree crown based approach; Adj R^2 = Adjusted R^2 ; AGB: Above-ground biomass; CBH: Canopy base height.

1.4.2 Estimation of biomass components (using full-waveform LiDAR data)

The additional information from small-footprint full-waveform (FWF) data has also been successfully applied to tree species classification (Heinzel and Koch, 2011), and the estimation of forest structure characteristics (e.g., height, DBH, volume and biomass) using the area-based (ABA) and individual tree crown (ITC) approaches (Hollaus et al., 2014). At the stand level, Chauve et al. (2009) used an enhanced peak detection algorithm coupled with echo modelling approaches to assess the capability of FWF data for improving the quality of DTMs and CHM-based heights in black pine (*Pinus nigra*) stands in southern France. The results showed that although the accuracies (of DTMs and CHM-based heights) were slightly improved, 40 to 60 % additional points were detected (mainly in the lower part of the canopy and in the low vegetation), providing a much more detailed structure of forest stands. Lindberg et al. (2012) evaluated the performance of a normalized FWF-derived profile for estimating vegetation volume in a hemi-boreal, spruce-dominated forest in southwest Sweden. The results showed that the normalization process is useful in compensating for the shielding effect of higher vegetation layers, and that the normalized FWF data had the highest accuracy (relative RMSE=27.6%) compared with the other three approaches, i.e., using FWF data (without normalization) (relative RMSE=31.9%), point data derived from FWF (relative RMSE=29.1%), point data from discrete return LiDAR system (relative RMSE=36.5%). Hermosilla et al. (2014) extracted geometric and radiometric information from FWF data based on a spatiotemporal analysis for estimating canopy fuel and forest structure parameters in a Douglas-fir dominated mixed forest in northwest Oregon (United States). A voxel-based approach was used to generate composite waveforms on the normalized full-waveform data before extracting the FWF metrics. The results showed promising accuracy (most of the models explained over 80% of the variability of the forest parameters) of using FWF metrics to characterize the structure and canopy fuel loads of forests. Pirotti et al. (2014) evaluated the performances of three levels of FWF metrics, i.e., peak (from each return echo), pulse (from the whole return signal) and plot level (simulating the large footprint LiDAR data) for estimating basal area (BA) and aboveground biomass (AGB), in a tropical moist forest

in south-western Ghana. The results showed that the basal area (adjusted $R^2=0.62-0.79$) is better predicted than aboveground biomass (adjusted $R^2=0.58-0.65$), and trees with a diameter at breast height equal to or above 20 cm are better predicted than small trees.

At the single tree level, Allouis et al. (2013) assessed the capability of FWF LiDAR to improve the estimation of individual tree-based volume and biomass in a black pine forest in southern France. They found that although the use of additional FWF metrics did not improve the volume, the accuracy of biomass estimation (adjusted $R^2=0.90-0.91$) was improved compared with models developed using CHM-only (adjusted $R^2=0.87$) or CHM+DR (adjusted $R^2=0.87-0.88$) metrics. They also found that when including the FWF metric (i.e., the integral of the cumulative signal, which related to the vertical structure of a tree as a whole), the accuracy of biomass estimation was enhanced. Wu et al. (2009) extracted structural metrics (i.e., crown thickness and canopy height) from LiDAR waveforms for estimating woody and foliar biomass in a wooded savanna in South Africa. The results showed that FWF metrics have strong potential for estimating woody ($R^2=0.71$) and foliar biomass ($R^2=0.73$) at the tree-level in a savanna environment. Sumnall et al. (2012) conducted a comprehensive investigation comparing the estimation of forest structural variables from discrete (system points) and full waveform LiDAR data, using ABA and ITC approaches, in a semi-natural forest in southern England. ITC metrics were derived after individual tree crown delineation, height and intensity-related metrics were derived from DR data (system points), metrics of heights (derived from waveform points), amplitude and pulse widths of the waveforms were all derived from FWF data. Results indicated that FWF metrics appeared to have stronger potential for predicting DBH ($R^2=0.91$), canopy base height ($R^2=0.93$) and number of trees ($R^2=0.80$) than DR metrics ($R^2=0.66, 0.90, 0.70$, respectively). A detailed summary of previous studies that have estimated forest characteristics (e.g., height, volume and biomass etc.) using full-waveform LiDAR data is presented in Table 1.3.

Table 1.3 Summary of previous studies that estimated forest characteristics using full-waveform LiDAR.

Previous study	Study area	Forest type (characteristics)	Modelling approaches	Estimates	Additional information extracted from FWF data	Radiometric calibration	Correction of scan-angle effects	Accuracy of the best predictive model (FWF information being used)
Chauve et al. (2009)	Southern French	Black pine forest (protected afforestation of mature trees)	ABA	Canopy height	Denser point clouds extracted from FWF data			RMSE = 0.21m
Lindberg et al. (2012)	Southwest Sweden	Hemi-boreal forest (spruce dominated)	ABA	Vegetation volume	FWF-derived vegetation volume profile	√		relative RMSE = 27.6%
Hermosilla et al. (2014)	Northwest Oregon, United States	Temperate forest (Douglas-fir dominated mixed forest)	ABA	Canopy fuel and forest structure parameters	Geometric and radiometric metrics form FWF data		√	$R^2 = 0.84$ (AGB) $R^2 = 0.66-0.82$ (Others)
Pirotti et al. (2014)	Southwest Ghana, Africa	Tropical moist forest (disturbed and degraded in many areas)	ABA	Basal area (BA) and AGB	peak (form each return echo), pulse (form the whole return signal) and plot level (simulating the large footprint LiDAR data)			Adj $R^2 = 0.62-0.79$ (BA) Adj $R^2 = 0.58-0.65$ (AGB)
Allouis et al., (2013)	Southern France	Black pine (over-mature stands)	ITC	Volume and AGB	Metrics from FWF signals and vegetation profiles			Adj $R^2 = 0.93-0.95$ (volume) Adj $R^2 = 0.90-0.91$ (AGB)

Previous study	Study area	Forest type (characteristics)	Modelling approaches	Estimates	Additional information extracted from FWF data	Radiometric calibration	Correction of scan-angle effects	Accuracy of the best predictive model (FWF information being used)
Wu et al. (2009)	South Africa	Wooded savanna (near-continuous grass layer, various tree and brush)	ITC	Woody and foliar biomass	Crown thickness and canopy height			$R^2 = 0.71$ (woody) $R^2 = 0.73$ (foliar)
Sumnall et al. (2012)	Southern England	Semi-natural and planted forests (with coniferous and deciduous trees)	ITC and ABA	CBH, DBH and Volume etc.	Amplitude and pulse widths			$R^2 = 0.83$ (Volume) $R^2 = 0.90$ (CBH) $R^2 = 0.91$ (DBH) etc.

ABA: area-based approach; ITC: individual tree crown based approach; FWF: Full-waveform; Adj R^2 = Adjusted R^2 ; BA: basal area; AGB: Above-ground biomass; CBH: Canopy base height.

1.4.3 Modeling and estimating biomass dynamics (using multi-temporal LiDAR data)

Generally, approaches for modeling and predicting change using multi-temporal LiDAR data can be classified into two categories, i.e., direct and indirect approaches (McRoberts et al., 2014). Direct approaches predict biomass change directly using the differences in LiDAR metrics, whereas indirect approaches model biomass at each point in time and predict change as their differences. Bollandsås et al. (2013) evaluated different approaches for estimating biomass change using bi-temporal airborne LiDAR data over a mature boreal forest and found that models based on the direct approach were superior to the indirect approach. Næsset et al. (2013a) used bi-temporal LiDAR data to estimate different levels of AGB change (stratified by management activities) in actively managed boreal forest (comprising both young and mature stands), and also found that the direct approach resulted in more accurate estimates of biomass change than other modelling approaches. In contrast, Økseter et al. (2015) found that the indirect approach yielded the best results in predicting biomass change in a young boreal forest.

Hudak et al. (2012) used repeated airborne LiDAR acquisitions to quantify carbon pools and fluxes in a temperate forest. They found that LiDAR can provide accurate, spatially explicit biomass distribution estimates that can be used to characterize carbon fluxes. Using spatial extrapolation, they found a net gain in AGB in non-harvested forests, but a net loss in harvested forests across the study site. Dubayah et al. (2010) used two full-waveform LiDAR datasets from the NASA Laser Vegetation Imaging Sensor (LVIS) instrument to determine change in height and AGB in tropical forests and found that old-growth forest stands had a net loss of height (-0.33 m) and only a small gain of AGB (2 Mg ha⁻¹). However the secondary forests had a net gain in both height (2.08 m) and AGB (25 Mg ha⁻¹) during a 7-year period. Meyer et al. (2013) used airborne LVIS and discrete-return LiDAR data to analyze the magnitude and spatial patterns of AGB change in an old-growth tropical forest. They found that the accuracies of both the mean biomass stock and biomass change improved with increasing spatial scales (from 0.04 ha to 10 ha). The spatial extrapolation results showed that the old-growth forests had a net loss of AGB (-0.7 Mg ha⁻¹ yr⁻¹), but secondary forests had a net gain (1.8 Mg ha⁻¹ yr⁻¹) across the study site. A detailed summary of previous studies that predicted biomass change using multi-temporal airborne LiDAR data is presented in Table 1.4.

However, to date, most of the research has been conducted in boreal and temperate forests, with very few published studies in tropical forests, except a number of studies using sensors designed for experimental observations (e.g., LVIS), with these data only publicly available for

some intensively studied locations. To the best of my knowledge, no prior published studies assessed the utility of using multi-temporal LiDAR data for estimating and mapping biomass changes in subtropical forest stands. In addition, although Bollandsås et al. (2013) compared the performances of both direct and indirect approaches for estimating changes, their study was only focused on boreal forests making it difficult to transfer these results to different forest biomes and assess which approach is the best. Additional studies assessing these two approaches are needed to be undertaken in a number of different and highly varying forest types before more general conclusions about the correct approaches can be drawn (Økseter et al., 2015).

Table 1.4 Summary of previous studies that predicted biomass change using multi-temporal airborne LiDAR data.

Previous study	Study area	Forest type (characteristics)	Tested modelling approaches*	LiDAR data types (t_1/t_2)	Average point densities (t_1/t_2 points m^{-2})	Data acquisition time (interval, years)	Accuracy of the best bioma change predictive model
Dubayah et al. (2010)	Costa Rica	Tropical forest (old-growth and secondary)	Direct	LVIS/LVIS	- / -	1998-2005 (7)	$\Delta Adj-R^2=0.65$, $RSE=10.54$ $Mg\ ha^{-1}$
Hudak et al. (2012)	Northern Idaho, USA	Temperate forest (heavily managed, with diverse stand ages)	Indirect	DR/DR	0.4 / 12	2003-2009 (6)	$t_1: RMSD = 92.75\ Mg\ ha^{-1}$ $t_2: RMSD = 101.87\ Mg\ ha^{-1}$
Meyer et al. (2013)	Central Panama	Tropical forest (old-growth)	Direct and <u>Indirect</u>	LVIS/DR	- / 5.6	1998-2009 (11)	$t_1: R^2 = 0.70$ $t_2: R^2 = 0.75$
Englhart et al. (2013)	Borneo, Indonesia	Tropical peat swamp forest (intensively disturbed)	Indirect	DR/DR	1.5 / 10.7	2007-2011 (4)	$t_1: R^2 = 0.77$ $t_2: R^2 = 0.81$
Bollandsås et al. (2013)	Southeast of Norway	Boreal forest (mature stands, in mountain area)	<u>Direct</u> and Indirect	DR/DR	3.4 / 4.7	2005-2008 (3)	$\Delta RMSE=4.0-16.5\ Mg\ ha^{-1}$
Næsset et al. (2013a)	Southeast of Norway	Boreal forest (actively managed, young and mature stands)	<u>Direct</u> and Indirect	DR/DR	1.2 / 7.3	1999-2010 (11)	$\Delta R^2=0.40-0.98$
Økseter et al. (2015)	Southeast of Norway	Boreal forest (young stands, in flat area)	Direct and <u>Indirect</u>	DR/DR	1.2 / 6.9	1999-2010 (11)	$\Delta CV-RMSE=7.3\ Mg\ ha^{-1}$

*the modelling approach yields higher prediction accuracy is underlined; t_1/t_2 : the first/second time of LiDAR data acquisition; LVIS: NASA Laser Vegetation Imaging Sensor; DR: discrete return LiDAR data; $Adj-R^2$: Adjusted R-square; RSE: Standard Error; RMSD: Root Mean Square Difference; RMSE: Root Mean Square Error; CV-RMSE: Cross-Validated Root Mean Square Error.

2 Study Area and Data

This chapter describes the general study area and the key datasets that are used throughout this dissertation.

2.1 Study area

All of the research described in this thesis was undertaken in Yushan Forest, a state-operated forest and national forest park located near the city of Suzhou in Jiangsu province, within the subtropical region of southeast China (120°42'9.4"E, 31°40'4.1"N). It covers approximately 1103 ha, with an elevation range of 20–261 m above sea level (Figure 2.1). The main conifers include Masson pine (*Pinus massoniana* Lamb.) and Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.), with a small percentage of introduced Slash pine (*Pinus elliottii* Engelm.) and Japanese blackbark pine (*Pinus thunbergii* Parl.). The primary broadleaved species are Sawtooth oak (*Quercus acutissima* Carruth.), Chinese sweet gum (*Liquidambar formosana* Hance), mixed with Chinese holly (*Ilex chinensis* Sims.), and associated with some evergreen broadleaved tree species belonging to Fagaceae, Lauraceae and Theaceae (Li et al., 2006) (Figure 2.2).

The natural forest in Yushan was harvested before the 1960s, and then the conifers (e.g., Masson pine, Chinese fir and Slash pine etc.) were planted in the 1970s. There has been natural regeneration of native species (e.g., Sawtooth oak, Chinese sweet gum and Chinese holly etc.). The majority of stands were naturally regenerated and not subject to any major natural disturbance or harvest for the past few decades. However, some of the Chinese-fir stands have been transformed to Chinese-fir/tea agroforestry systems following silvicultural treatments. In addition, to reduce the risk of pine caterpillar (*Dendrolimus punctatus* Walker) damage, some minor silvicultural treatments (e.g., selective thinning and replacement of defoliated pine with young indigenous tree species) have been conducted in some of the pine-dominated stands in the past few years (SFMP, 2013). The forest is in close proximity to urban areas, and has experienced anthropogenic activity such as recreation and local tourism. Yushan forest has a plan for

sustainable forest management (SFMP, 2013). The goal of its Development Framework for Sustainable Forests (2013-2020) is to ensure the long-term health of forest ecosystems, and to manage the forest for a broad range of uses and to conserve economic, social, environmental and cultural values (SFMP, 2013).

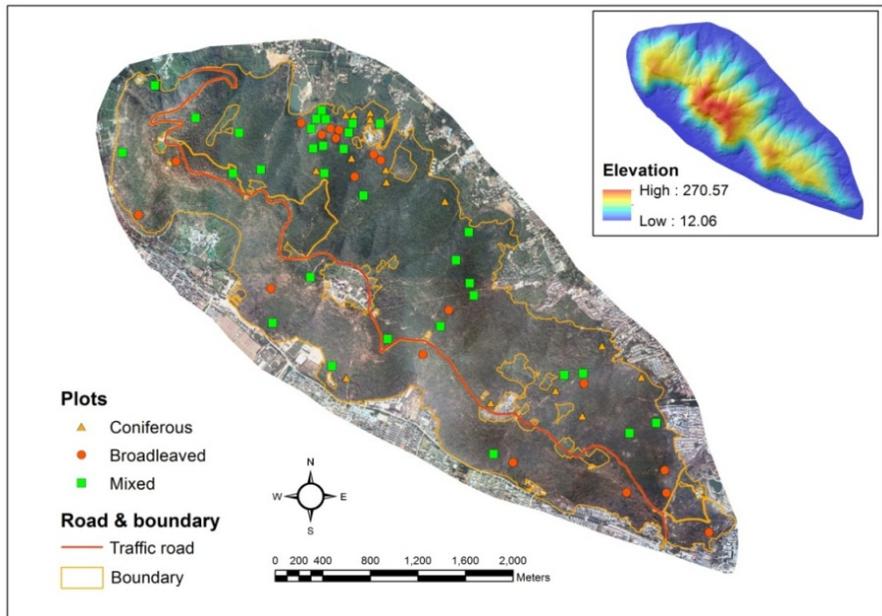


Figure 2.1 Yushan forest study site and the locations of sample plots of three main forest types. The upper-right sub-figure shows the topography of the site.

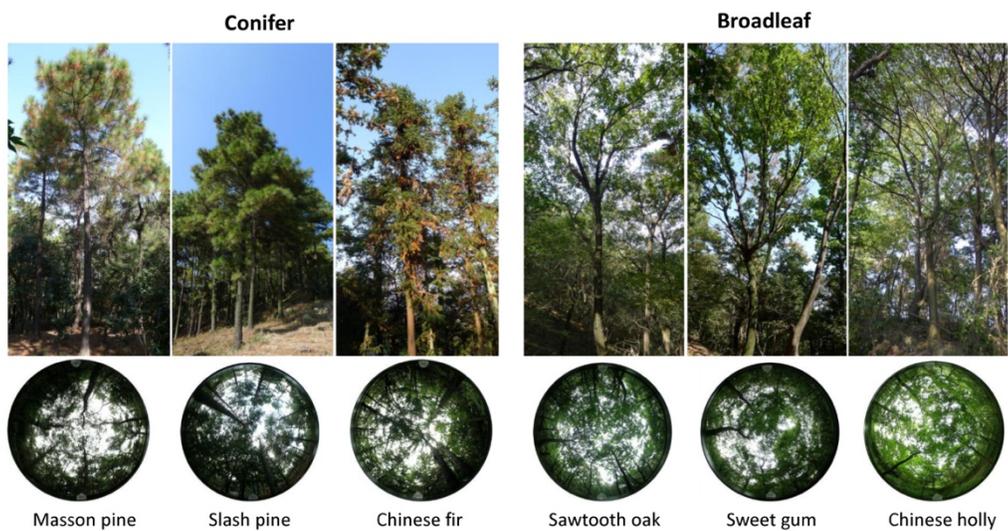


Figure 2.2 Photographs and hemispherical images of the six common tree species in the study area.

2.2 LiDAR data

LiDAR data were acquired in April, 2007 (collected by the Jiangsu Surveying and Mapping Engineering Institute, China) and August, 2013 (collected by the Institute of Forest Resource Information Techniques (IFRIT), Chinese Academy of Forestry (CAF), China) over the research site. An Optech ALTM-3100 discrete-return laser scanner was used in the 2007, operated at 800 m above ground level with a 50 kHz pulse repetition frequency (wavelength = 1064nm). In 2013, a Riegl LMS-Q680i full-waveform scanner was used and data acquired at an altitude of 900 m with a 360 kHz pulse repetition frequency (wavelength = 1550nm). The returned waveforms were recorded with a temporal sample spacing of 1 ns (approximately 15 cm), and point clouds were extracted from the full-waveform data using decomposition algorithms (Wagner, 2010). These LiDAR data were stored in LAS format (American Society for Photogrammetry and Remote Sensing, Bethesda, MD, USA) by the data providers. Other scanner properties and flight parameters of the two LiDAR surveys are provided in Table 2.1.

Table 2.1 The summary of scanner properties and flight parameters of the two LiDAR surveys.

Parameters	Optech ALTM-3100 (2007)	Riegl LMS-Q680i (2013)
Flight altitude (m above ground)	800	900
Beam divergence (mrad)	0.3	0.5
Footprint (cm)	24	45
Wavelength (nm)	1064	1550
Pulse repetition frequency (kHz)	50	360
Scan frequency (Hz)	40	112
Average pulse distance (m)	0.92	0.49
Average point density (m ⁻²)	1.93	8.37

2.3 Plot data

Plot selection for the purpose of this research was guided by a mean height and cover layer (at 30m spatial resolution, derived from historical LiDAR data in 2007) and GIS-based stand inventory data (Yushan Forest Inventory, unpublished data, 2007). A total of 66 square sample plots (30 × 30 m) were established across the study site, covering the dominant species

compositions, age classes, and site indices (Figure 2.1). No plots had evidence of significant mortality or silvicultural operations (e.g., thinning) being undertaken in recent years (i.e., 2007-2013). Based on the stratification the plots were divided into three dominant species compositions: (i) coniferous forest (n = 15, dominated by Masson pine and Chinese fir); (ii) broad-leaved forest (n = 19, dominated by Sawtooth oak and Chinese sweet gum); and (iii) mixed-species forest (n = 32, a mixture of coniferous and broad-leaved species).

Plot data were collected between June-August 2012 and August 2013 under leaf-on conditions. Plot corners were located using Trimble GeoXH6000 Handheld GPS units, and high precision real-time differential signals received from the Jiangsu Continuously Operating Reference Stations (JSCORS), resulting in submeter accuracy (Song et al., 2009). At each plot, for all live trees with a DBH > 5 cm, species, diameter (using diameter tape), tree top height, crown base height (using a Vertex hypsometer) and crown widths (in both cardinal directions) were measured. Crown class, i.e., dominant, co-dominant, intermediate and overtopped, were also recorded (Popescu, 2007). For each dominant tree species in the plot, the DBH (of trees) were allocated to one of three classes and a proportion of dominant trees within each class were cored (two cores were sampled per tree) using increment borers at breast height to quantify tree growth. Individual tree locations were recorded within 12 plots (of the total 66 plots) (for the research in Chapter 3) established in August 2013, the positions of the trees within each plot were determined using the direction and distance of trees relative to the nearest sub-plot (each plot was divided into four equal-size square sub-plots) center. The descriptive statistics of the forest characteristics of the 12 plots (with individual tree location) are summarized in Table 2.2.

For the growth study in Chapter 6, as no field data exist for 2007, tree height and biomass estimations for 2007 were predicted from the field data collected in 2013 using the following approach. In order to compute the change (in AGB) between 2007 and 2013, for all cored trees, species-specific models were developed between DBH increment (measured from the extracted tree cores, as the width of tree ring growth from 2007 to 2013) as a dependent variable, and the 2013 DBH and tree height as independent variables (Du, 1999; Wyoff, 1990). Once developed,

the DBH increment model was then applied to all trees to estimate 2007 DBH. Lastly, once 2007 DBH was computed, species-specific diameter-height relationships developed from the 2013 field collected data were used to estimate height of all trees in 2007.

Once a full set of inventory data were available for 2007 and 2013, several plot-level forest variables were calculated based on the individual tree data, including average DBH, Lorey's mean height (i.e., the basal area weighted height) and above ground biomass (AGB). AGB was calculated for individual trees within each plot based on the field-measured (2013) or predicted (2007) DBH and height, and then summed to obtain plot-level AGB. Species-specific allometric equations referenced from Feng et al. (1999), Qian (2000), Wang and Shi (1990) and Xu et al. (2011) were used (allometric equations developed from tree inventory data from local or nearby provinces were selected for this research) for calculating the AGB based on the field-measured DBH and height. A summary of the predicted ground data in 2007 and field estimated data in 2013 (of all 66 plots) is provided in Table 2.3.

Table 2.2 Description of the forest characteristics for the 12 plots (with individual tree locations) for the research in chapter 3.

Plot No.	Forest type		Height (m)		Crown diameter(m)		Number of trees within each crown class			
	Dominated	Coniferous (%)	Mean	Std.Dev.	Mean	Std.Dev.	Dominant	Co-dominant	Intermediate	Overtopped
1	Coniferous	76	10.20	1.14	4.29	0.72	17	31	13	6
2	Mixed	41	9.84	1.13	3.76	0.68	22	26	21	11
3	Broadleaved	22	11.90	1.66	4.43	1.01	20	25	18	12
4	Mixed	55	7.79	0.82	4.12	0.99	21	24	18	11
5	Coniferous	79	9.43	1.10	3.68	0.65	25	41	30	7
6	Coniferous	76	12.44	1.04	4.35	0.76	21	34	15	5
7	Mixed	52	9.45	1.12	3.31	0.59	25	20	23	13
8	Mixed	38	10.55	1.20	5.01	0.92	16	22	15	9
9	Coniferous	82	10.11	1.16	4.15	0.76	20	31	21	6
10	Broadleaved	17	11.79	1.35	4.74	0.86	20	22	19	9
11	Mixed	49	11.34	1.23	4.31	0.82	17	24	16	11
12	Broadleaved	23	8.42	0.88	4.33	0.95	15	18	23	13

No. = number of trees. Forest type - Coniferous (%) = the proportion of coniferous trees of all trees in the plot.

Table 2.3 Summary statistics of predicted ground data in 2007 and field estimated data in 2013 (for all 66 plots).

Variables	Coniferous forests (<i>n</i> =15)			Broadleaved forests (<i>n</i> =19)			Mixed forests (<i>n</i> =32)		
	Mean	Std.Dev.	Range	Mean	Std.Dev.	Range	Mean	Std.Dev.	Range
Predicted ground data in 2007									
<i>DBH</i> (cm)	12.95	1.19	10.40-17.59	14.27	2.52	10.17-21.28	12.86	1.26	10.27-16.27
<i>h_L</i> (m)	9.78	1.17	8.16-13.28	10.98	2.48	6.95-17.88	10.09	1.24	7.03-12.95
<i>W_{AGB}</i> (Mg ha ⁻¹)	67.14	20.51	38.71-111.02	88.23	43.48	17.47-209.31	72.56	21.77	40.48-122.83
Field estimated data in 2013									
<i>DBH</i> (cm)	13.52	2.83	11.35-20.90	16.23	3.78	11.62-24.69	13.50	2.30	11.42-19.69
<i>h_L</i> (m)	11.18	1.61	8.57-14.76	12.42	2.90	7.70-18.52	11.55	1.54	8.79-14.18
<i>W_{AGB}</i> (Mg ha ⁻¹)	77.87	22.48	40.85-127.39	101.40	42.81	32.03-219.68	86.13	24.07	44.20-141.73

DBH = mean diameter at breast height; *G* = basal area; *h_L* = Lorey's mean height; *W_{AGB}* = mean above ground biomass; Std. Dev. = standard deviation.

2.4 Stand inventory data

Additional sets of forest stand inventory data (2007 and 2012) were also available (collected by the East China Forestry Investigation Planning and Design Institute, China). Within the study site, stands dominated by non-timber species (e.g., fruits, bamboo and tea plantations), and close to anthropogenic activities within the forest (such as stands with large roads intersecting the forest), were removed (note that none of the sample plots in section 2.3 were located within these removed stands). Figure 2.3 shows the forest stands stratified by dominant tree species and age group. The inventory data measured and compiled in summer 2012 served as independent validation data for the biomass component estimation (in Chapter 5), and the inventory data measured in summer 2007 was used to examine the LiDAR-estimated biomass changes in relation to forest type and age (in Chapter 6).

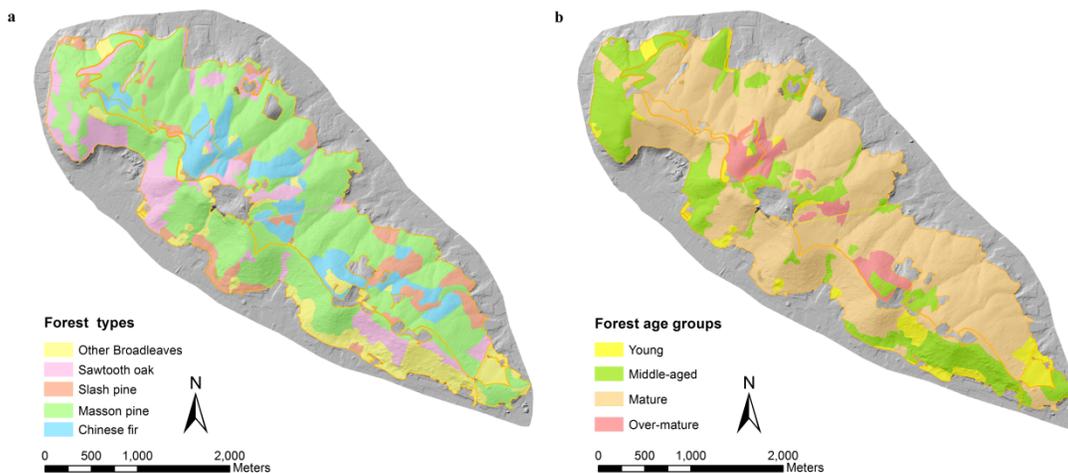


Figure 2.3 Map of the forest stands (in 2007) stratified by forest types (classified by the dominant tree types in the stands) (a) and age groups (b).

Table 2.4 shows the classification of the age groups for the dominant tree species according to Chinese local forest inventory guidelines. The age group is determined by the average age of dominant tree species in the stand, representing broad stages in the life cycle of the sub-tropical forest stand (Xie et al., 2011). The forest age classification system in China varies from other countries due to different growth characteristics of tree species, site quality and harvest intervals

(for the optimal commercial timber values). For example, in the Chinese forest age system, mature stands are those that have reached harvesting age, where as the young and middle aged stands are younger than harvesting age, and the over-mature stands are older than harvesting age (FAO, 1997).

The age class interval is 20 years for the hardwood (e.g., Sawtooth oak) dominated stands and 10 years for conifer (e.g., Masson pine and Slash pine) dominated stands. The fast-growing tree species such as Chinese fir have 5-year age class intervals. The management and silvicultural historical treatments in Yushan forest have subsequently resulted in a strong association between age classes and species compositions. For example, most of the young stands (age < 40 years) comprise regenerated broadleaved trees and many of the middle-aged stands (age = 41 - 60 years) are mostly dominated by Sawtooth oak, whereas mature stands (age = 31 - 60 years) are mostly dominated by Masson pine and over-mature stands (age > 35 years) are dominated by Chinese fir.

Table 2.4 The age class intervals and age groups for the dominant tree species at the research site.

Dominant tree species (in the stand)	Age class interval (years)	Age group (years)			
		Young	Middle-aged	Mature	Over-mature
Masson pine	10	≤ 20	21-30	31-60	>60
Slash pine	10	≤ 10	11-20	21-50	>50
Chinese fir	5	≤ 10	11-20	21-35	>35
Sawtooth oak	20	≤ 40	41-60	61-120	>120
Other broadleaves	20	≤ 40	41-60	61-120	>120

Note: this table is cited from “Guidelines for Forest Resources Inventory in Jiangsu Province, China, unpublished operating manual, 2007”; The age group to which most of each forest stand (determined by dominant tree species) belongs is highlighted as gray.

3 Tree Species Classification using Full-waveform LiDAR

3.1 Introduction

Accurate identification and classification of tree species provides key information for the precise mapping of timber quality attributes, biomass and species diversity (Jones et al., 2010; Popescu et al., 2003) and modeling of individual tree growth (Falkowski et al., 2010; Yu et al., 2004), which is critical for the understanding and managing forest ecosystems (Li et al., 2012; Plourde et al., 2007). The full-waveform (FWF) LiDAR systems digitize and record the entire backscattered signal of each emitted pulse and allow the recording of detailed geometric and physical properties of intercepted objects (Reitberger et al., 2006). The analysis of FWF data has improved height estimations (Duong, 2010), enhanced point extraction (Wu et al., 2011) and provided additional target information (Chauve et al., 2009). This extended return information has been successfully applied to forestry applications, such as the estimation of forest structural characteristics (Hermosilla et al., 2014; Neuenschwander, 2012) single tree detection (Gupta et al., 2010) and tree species classification (Heinzel and Koch, 2011; Xu et al., 2012).

However, to date, much of the research related to FWF LiDAR-based tree species classification has been conducted in temperate and boreal forests (e.g. Hollaus et al., 2009; Heinzel and Koch, 2011; Yu et al., 2014 etc.); published studies from tropical (and subtropical) forests are few. Most of the previous studies performed an analysis on the directly observed waveform, which was commonly assumed to be truly vertical. However, this assumption may lead to uncertain results (for FWF analysis), due to the off-nadir scanning angles and non-vertical waveforms from different flight trajectories (Hermosilla et al., 2014; Pang et al., 2011). So, in this research, a voxel-based composite waveform approach was applied, which aimed to overcome the impacts (of off-nadir scanning angle effects) on the spatio-temporal analysis of waveforms on forest canopies. The overall goal of this study is to explore an automatic process by using small footprint full-waveform data for individual tree detection (ITD) and tree species classification (of

species typically found in the subtropical region of southeastern China). The specific objectives of this research are: (1) to evaluate a locally adapted local maxima algorithm for individual tree detection in subtropical forest stands; (2) to assess the capability to discriminate 6 tree species at 3 levels of classification using a voxel-based composite waveform approach; and (3) to investigate the most important metrics for tree species classification and the optimal voxel size for modelling the composite waveforms.

3.2 Materials

3.2.1 Study area

See details about the study area, main forest types and dominant tree species in Section 2.1.

3.2.2 Field data

A total of 12 square plots (with the recorded individual tree position and other measurements) established in August 2013 were used in this study (see details about these data in Section 2.3).

3.2.3 LiDAR data

Small-footprint full-waveform LiDAR data acquired on 17 August 2013 were used in this study (see details about these data in Section 2.2).

3.3 Methods

An overview of the workflow for tree species classification is shown in Figure 3.1, which provides details of the general methodologies applied in this study. First, A digital terrain model (DTM) was created based on the filtered (above-ground) point clouds (the point clouds were extracted from the pre-processed waveforms using Gaussian decomposition); second, a locally-adapted local maxima algorithm was applied for tree detection and crown delineation based on the canopy height model (CHM); third, the individual tree based composite waveforms were extracted by the delineated crowns – these were then used as a data model for calculating spatiotemporal metrics. Finally, these metrics were utilized as inputs for the Random Forests classifier to discriminate six subtropical tree species at three levels of discrimination.

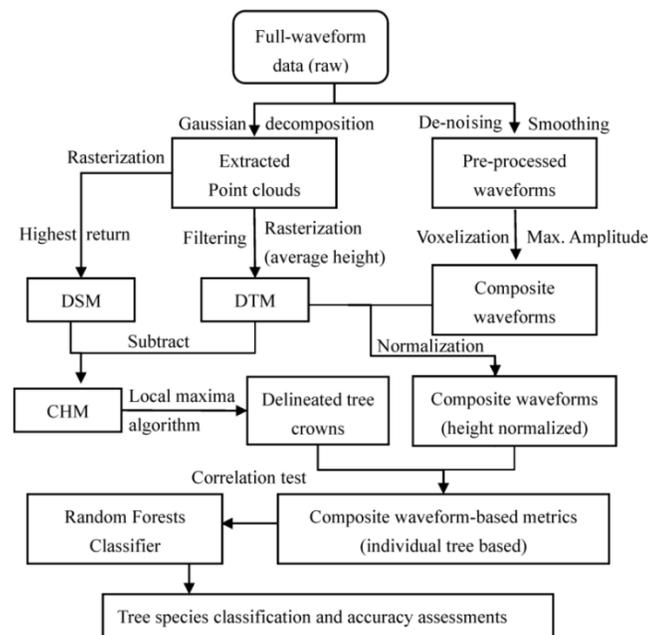


Figure 3.1 An overview of the workflow for tree species classification. DSM: Digital Surface Model; DTM: Digital Terrain Model; CHM: Canopy Height Model.

3.3.1 Data pre-processing

A 1-m digital terrain model (DTM) was created in two steps from the point cloud data. First, the data were filtered to remove the above-ground returns (Kraus and Pfeifer, 1998) and secondly, the DTM was created by calculating the average elevation from the remaining (ground) LiDAR returns within a cell (cells that contain no points were filled by interpolation using neighboring cells). The DTM was derived using FUSION software (McGaughey, 2014). The point cloud was then normalized against the ground surface height and extracted for each plot using the coordinates.

Previous studies have demonstrated that the waveform is stretched as the off-nadir angle increases (Neuenschwander, 2008; Pang et al., 2011; Yang et al., 2011). The waveforms for specific locations may be recorded from several flight trajectories because the acquisition of airborne LiDAR data is usually a composition of many overlapping strips (Bretar et al., 2004; Hermosilla et al., 2014). To avoid the effects of off-nadir pointing on the waveform and to integrate non-vertical waveforms from different flight trajectories, a voxel-based composite waveform approach was applied to the full-waveform data (see Hermosilla et al., 2014 for more information). This approach synthesized multiple raw waveforms into composite waveforms through the vertical space partitioning of forest canopies by voxels, and used the maximum amplitude value to construct new composite waveforms (i.e., each bin in a waveform was geo-located and then the maximum amplitude value was assigned for each voxel). The background noise of each waveform was first suppressed by a de-noising process and then smoothed by a Gaussian filter. Each pre-processed waveform was then spatially located in the three-dimensional space and normalized by subtracting the derived DTM height from the height of each bin in the corresponding positions.

The vertical space was partitioned into voxels according to the footprint size and temporal sample spacing (of 1 ns (approximately 15 cm)), and 8 different voxel sizes were tested to examine their impacts on tree species classification. The maximum amplitude value was calculated and assigned to each intersected voxel. The maximum amplitude was used to preserve

the significance of the major returns of the observed waveforms (Hermosilla et al., 2014). The composite waveforms were finally formed by extracting the preserved amplitude values along each vertical column of voxels, from tree top (waveform beginning) to bottom (ground) within the delineated tree crown projections (Figure 3.2). The composite waveforms within each tree crown were then used as a data model for later calculating the spatiotemporal metrics.

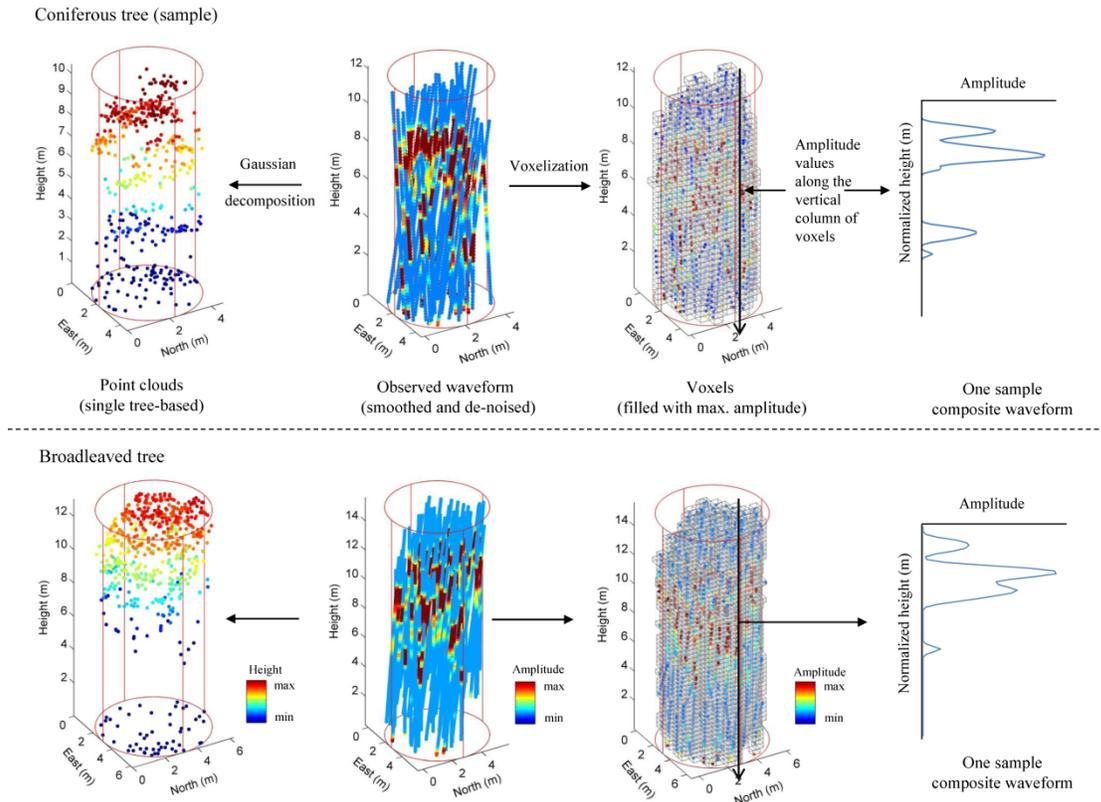


Figure 3.2 The process of extracting point cloud and creating composite waveforms from the observed waveforms. Two samples (i.e. one coniferous tree and one broadleaved tree) have been used to show the processes and results.

3.3.2 Individual tree detection

An approach similar to that of Popescu et al. (2002) and Popescu and Wynne (2004) was applied to detect individual trees. First, a 0.5m canopy height model (CHM) was built from the point cloud data (the CHM was created by subtracting a Digital Terrain Model (DTM) from the Digital Surface Model (DSM), and the DSM was calculated by assigning the highest returns

within the grid cells), and then smoothed by a 3×3 median filter to remove the noise (in the complex canopy upper surface). Second, individual trees were detected from the smoothed CHM using a local maxima algorithm, which identifies high values (likely to be the tree tops) with a variable-size circular window. The appropriate window size is determined by a quadratic model fitted by the height of the trees and their crown widths (Popescu et al., 2003):

$$\text{Crown width}(m) = \beta_0 + \beta_1 \times h^2 \quad (3.1)$$

where h is the tree height (m), and β_0 and β_1 are the model parameters. Locally measured tree heights and crown widths (a total of 673 trees) were used to derive the relationship and to determine appropriate window sizes. Finally, the crown extent was determined by analyzing 16 radial crown profiles of the CHM centered on the tree top, with a sampling point spacing of half a pixel. The crown radius along each profile was calculated by the horizontal distance to a local minimum or to the second point (if a series of three points had heights less than 66% of the tree height) (McGaughey, 2014), and the crown widths were the average of the horizontal distances of 16 radii. The detection of individual trees was processed using FUSION software (McGaughey, 2014).

For each detected tree, the tree position, height and crown diameter were estimated and linked to the field-measured trees. If a detected tree was located within 2m of a field-measured tree, it was designated as correctly detected. Descriptive statistics for the characteristics of all detected trees (classified by the taxonomic species) are shown in Table 3.1.

Table 3.1 Description of all detected trees (in 6 species) with the statistics of the characteristics.

Taxonomic group	Name	Detected trees		Tree height(m)		Crown base height(m)		DBH(cm)		Crown diameter(m)	
		No.	Proportion (%)	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Conifer	Masson pine	85	80	9.73	1.08	5.60	0.46	15.10	2.06	4.20	0.76
	Slash pine	70	84	12.72	1.21	7.16	0.32	21.43	1.28	4.45	0.75
	Chinese fir	81	82	9.24	0.98	5.39	0.41	12.83	1.68	3.77	0.71
Broadleaf	Sawtooth oak	72	77	12.37	0.87	7.62	0.27	18.38	1.08	4.44	0.97
	Sweet gum	73	73	12.26	1.29	7.08	0.54	18.20	2.14	4.72	0.78
	Chinese holly	50	67	8.47	0.87	3.25	0.40	10.86	1.64	4.75	0.88
All species		430	77	10.70	1.03	6.02	0.40	16.13	1.65	4.39	0.81

No. = number of trees. Detected trees - Proportion (%) = Detection rate of the dominant and co-dominant trees in the plots.

In this study, tree crown height distribution was generated by the LiDAR point cloud (extracted from the waveforms) in 2 steps: firstly, all of the non-ground LiDAR returns within each tree crown were collected and binned (with a 0.3m bin vertical spacing) according to normalized height; secondly, each bin was assigned a value as the total number of returns in the bin. A summary of the average crown height distributions of all the detected trees for the six tree species is shown in Figure 3.3.

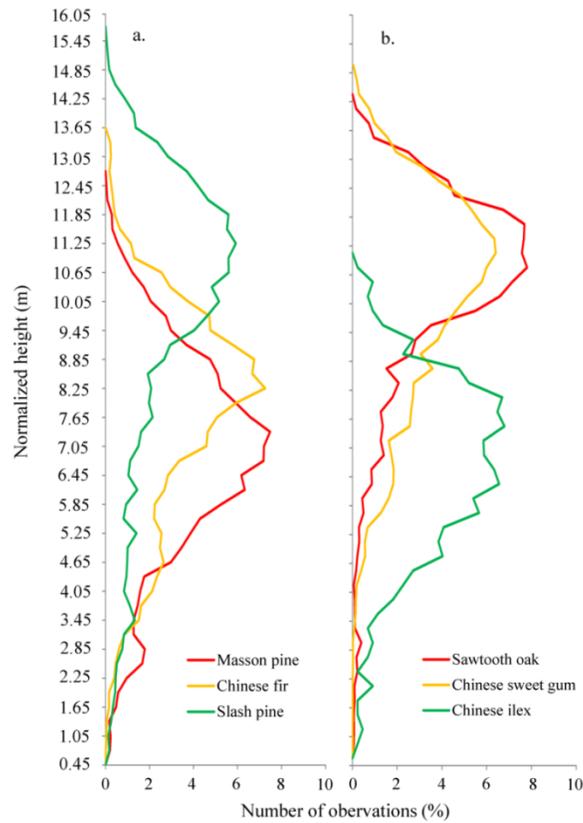


Figure 3.3 Average crown height distributions (mean of the proportion of returns within 0.3m height intervals) of all the detected coniferous trees (a) and broadleaved trees (b).

3.3.3 Full-waveform LiDAR metrics

Full-waveform data based metrics describe the radiometric and geometric attributes of the return waveforms based on the spatio-temporal analysis (Hermosilla et al., 2014). In this research, 14 individual tree-level waveform metrics were incorporated into the composite waveform analysis due to their success in previous studies (Drake et al., 2002; Duong, 2010; Harding, 2005; Neuenschwander, 2012; Ranson et al., 2004; Sun et al., 2008a). These metrics included height of median energy (HOME), waveform distance (WD), height/median ratio (HTMR), number of peaks (NP), roughness of outermost canopy (ROUGH), front slope angle (FS) and return waveform energy (RWE), and were summarized as the mean and standard deviation of all the waveform metrics (of individual composite waveform) within each delineated tree crown. A summary of the metrics with corresponding descriptions is provided in Table 3.2, and Figure 3.4 illustrates the formation of a typical composite waveform and the calculation of the waveform metrics based on it.

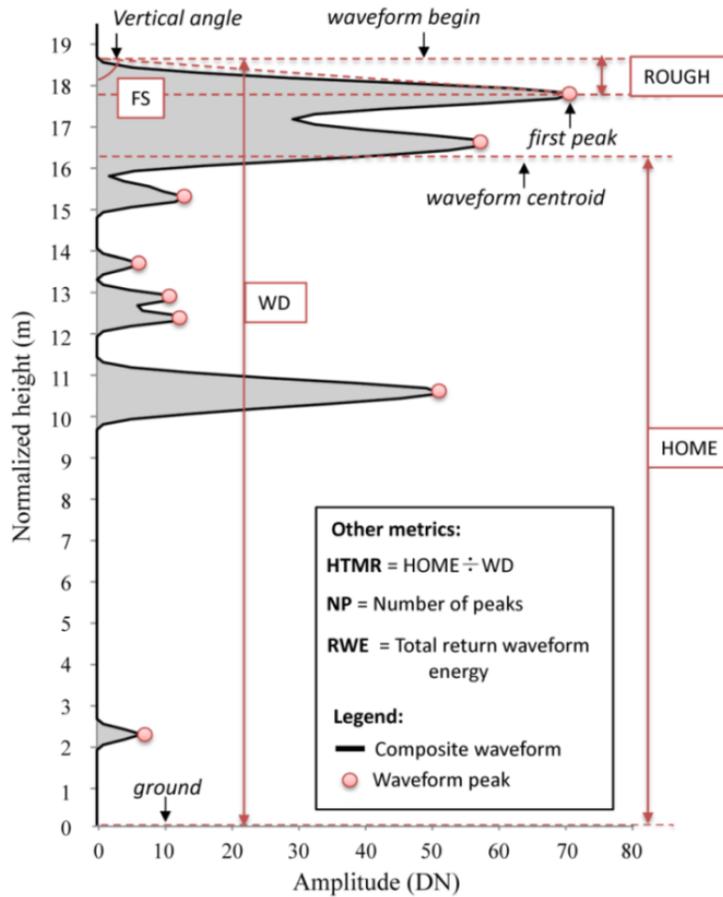


Figure 3.4 Illustration of a sample composite waveform (normalized by the DTM), and the metrics derived from the waveform. HOME: height of median energy; WD: waveform distance; HTMR: height/median ratio; NP: number of peaks (NP); ROUGH: roughness of outermost canopy; FS: front slope angle; RWE: return waveform energy.

Table 3.2 Description of the composite waveform metrics.

Waveform-based metrics	Calculation	Physical explanation
HOME (height of median energy)	The distance from waveform centroid to the ground.	Sensitive to canopy openness and the vertical arrangement of canopy elements (Drake et al., 2002).
WD (waveform distance)	The distance from waveform beginning to the ground.	Relates to the LiDAR canopy height (Sun et al., 2008).
HTMR (height/median ratio)	HOME divided by WD.	Depicts the change of HOME relative to the canopy height (Drake et al., 2002).
NP (number of peaks)	The number of detected peaks within each normalized composite waveform.	Describes the number of height levels corresponding to layers in the forest canopy (Duong, 2010).
ROUGH (roughness of outermost canopy)	The distance from the waveform beginning to the first peak.	Describes the roughness of the upper-most canopy and the spatial arrangement of plant surfaces (Duong, 2010; Harding, 2005).
FS (front slope angle)	The vertical angle from waveform beginning to the peak of canopy return energy.	Provides the variability of the upper canopy (Ranson et al., 2004).
RWE (return waveform energy)	The total received energy, i.e., the area below the waveform between beginning and end.	Describes the canopy surface characteristics, e.g., reflectivity or moisture conditions (Duong, 2010).

In all cases, both the mean (μ) and standard deviation (σ) of these metrics were calculated.

3.3.4 Metrics selection and classification

Metrics selection can reduce the dimensionality of input variables, and is important for creating highly efficient, stable and transferable classification models. All of the variables were analyzed for collinearity and were excluded from the classifier if their correlation coefficients with any other metrics were higher than 0.60.

Random Forest (RF) randomly and iteratively samples the variables to generate a large group (or forest) representing the statistical mode of numerous decision trees, thereby creating a more robust model than a single classification tree (Breiman, 2001). A majority vote is used in RF to combine the decisions over all constructed trees. In RF, a bagging strategy is applied to create new training data by resampling from the original data (Breiman, 1996; Chan and Paelinckx, 2008). Each decision tree is constructed using bootstrap samples selected n times randomly with replacements from a training dataset, and the subsets of predictor variables can prevent problems associated with correlated variables and overfitting. Approximately one-third of the samples that are left out of the training dataset are called “out-of-bag” (OOB) samples, and can be used without bias to estimate the classification error and derive variable importance in many tests (Breiman, 2001). RF internally estimates the prediction error using OOB data; hence there is no need for cross-validation or separate data to evaluate the performances.

The importance of a variable X_m can be evaluated for prediction Y by adding up the weighted impurity decreases $p(t)\Delta i(s_t, t)$ for all nodes t where X_m is used, averaged over all N_T trees in the forest:

$$IMP(X_m) = \frac{1}{N_T} \sum_T \sum_{t \in T: v(s_t) = X_m} p(t) \Delta i(s_t, t) \quad (3.2)$$

where $p(t)$ is the proportion N_t/N of samples reaching t and $v(s_t)$ is the variable used in split s_t . The indices of Mean Decrease Accuracy (MDA), which quantify the degree to which inclusion of a

variable in the model decreases the mean squared error, can be used to assess the variable importance for classification (Breiman, 2001; Liaw and Wiener, 2002).

The number of decision trees (in RF) was set to 100 with 4 variables at each split. The analysis was started with the low-correlated metrics (i.e., the correlation coefficients of these metrics with any other metrics was lower than 0.60) and the one with least importance in each run was discarded until 6 metrics remained. Six metrics were chosen because the inclusion of more metrics did not significantly improve (<3%) the overall classification accuracy (histograms of the improvement of overall classification accuracy were used to determine the number of selected metrics for all the voxel sizes, see Figure 3.7). For each test of classification with different voxel sizes, the classification accuracies were assessed by the “selected metrics”, and the 3 top-ranked metrics were then selected as the final “most important metrics” according to the indices of MDA. In addition, confusion matrices and classification accuracies were also tested in three classification levels, i.e., using all six tree species (highest classification depth), four main species (middle classification depth) and two generalized classes of conifer and broadleaf (lowest classification depth). The results of using the “selected metrics” and the “most important metrics” for classifications are compared under the best voxel size (i.e. $0.5 \times 0.5 \times 0.15\text{m}$, generates the highest classification accuracies, see Table 3.3).

3.4 Results

In total, 430 (76%) of the dominant and co-dominant trees were detected by LiDAR in all of the 12 plots. In general, conifer-dominated plots have a higher detection rate (80-87%) than broadleaved (75-78%) and mixed (68-73%) plots. The analysis also indicated that the crown diameters (RMSE = 0.55m, rRMSE=12.5% observed mean crown diameter) are less accurately estimated than the tree heights (RMSE=0.63m, rRMSE=5.9% observed mean height). Part of the lower accuracy of the estimated crown diameters may be due to the overlapping of tree crowns, as the local-maxima algorithm used in this study was not aimed at measuring overlapping crowns. The GPS positioning and crown measurement errors may also introduce uncertainties to the results (Popescu et al., 2003). One sample plot of the canopy height model (CHM) with the location of LiDAR-detected tree-tops, crown boundaries and their related ground trees can be seen in Figure 3.5.

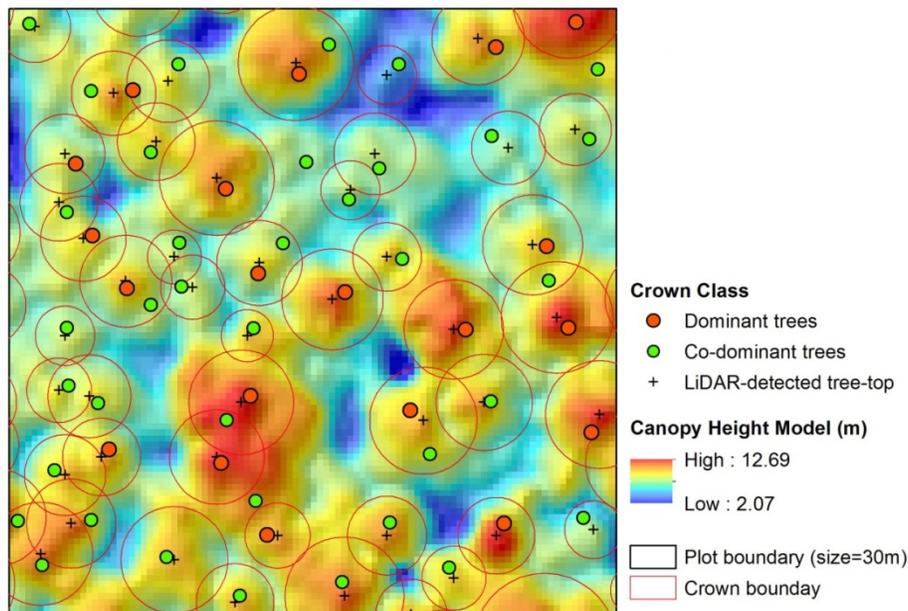


Figure 3.5 Example map of CHM with the location of trees (dominant and co-dominant trees), the LiDAR-detected tree-tops and the estimated tree crowns

The correlation of the selected metrics is shown in Figure 3.6. Nine of the total fourteen metrics were preserved, i.e., average of HOME (height of median energy), ROUGH (roughness of outermost canopy) and RWE (return waveform energy), and the standard deviation of WD (waveform distance), HTMR (height/median ratio), NP (number of peaks), ROUGH, FS (front slope angle) and RWE. These metrics were used for further analysis in this chapter.

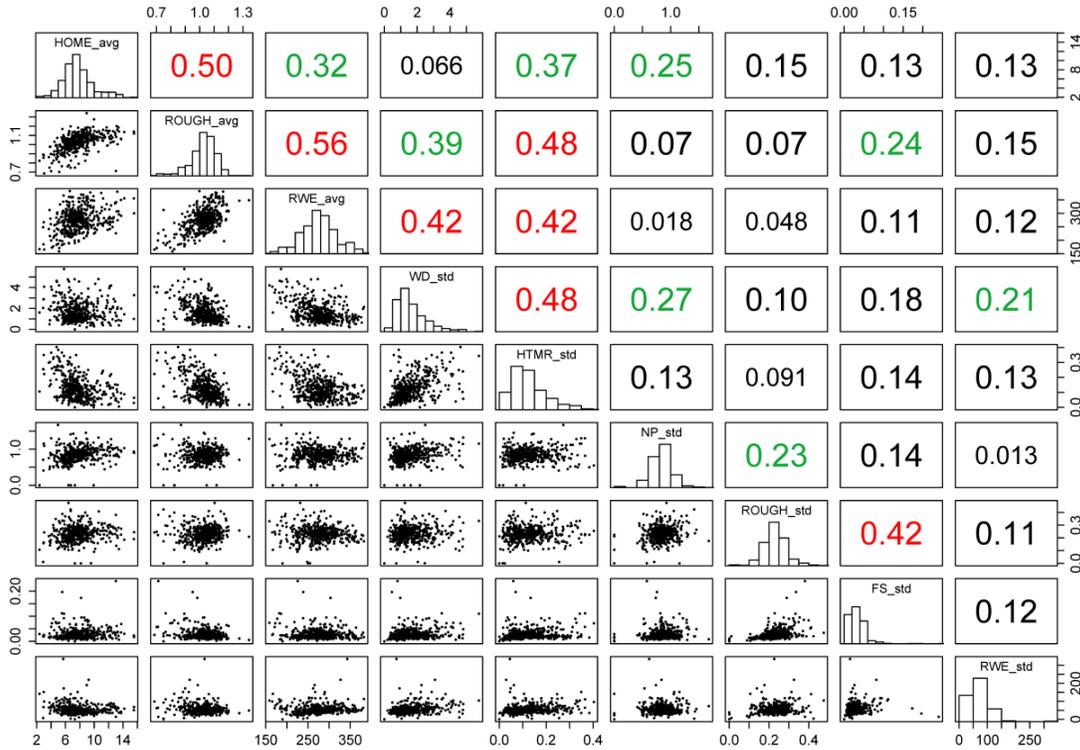


Figure 3.6 The matrix of scatter plots (lower panel) and the correlation coefficients (r) (upper panel) for the low-correlated metrics (with $r < 0.6$) for all detected trees. The histograms of each metric are drawn in the diagonal line. The values of r are colored red for medium degree correlations ($0.4 \leq r < 0.6$) and green for low degree correlations ($0.2 < r \leq 0.4$). See Table 3.3 for codes of the metrics.

The combination of 2 sizes (0.15 and 0.3 m, equal to one and two times the waveform bin spacing) in the vertical dimension and 4 sizes (0.3, 0.5, 1 and 2 m, according to the footprint diameters and tree crown widths) in the horizontal dimension generated 8 different voxel sizes. The classification accuracies of the “selected metrics” were assessed by the kappa coefficients

and overall accuracies, and 3 top-ranked metrics in each trial were selected according to the indices of MDA calculated from the RF analysis. In general, the results demonstrated that the voxels in “High” resolution (i.e., 0.5×0.5×0.15m and 0.5×0.5×0.3m) have the highest classification accuracies (Kappa coefficients = 0.59-0.61, overall accuracies = 66-68%), followed by “Very high” and “Medium” resolutions (Kappa coefficients = 0.55-0.60, overall accuracies = 62-67%). The voxels in the “Low” resolution case had the lowest classification accuracies (Kappa coefficients = 0.52-0.56, overall accuracies = 59-62%). The height-related metrics i.e. the average of the height of median energy (HOME) and standard deviation of waveform distance (WD) were mostly selected as the 2 top-ranked metrics in the group of “Most important metrics” for all 8 trials, followed by the standard deviation of number of peaks (NP) (3 out of 8 were selected), the standard deviation of front slope angle (FS) (3 out of 8) and the average of return waveform energy (RWE) (2 out of 8) (Table 3.3).

Table 3.3 Comparison of the tree species classification accuracies (of the selected metrics) and demonstration of the most important metrics (determined by the indices of MDA) for different voxel sizes.

No.	Voxel Size (m)	Resolution	Kappa coefficient	Overall accuracy	Most important metrics (Ranked by MDA, n=3)
1	0.3×0.3×0.15	Very high	0.58	65%	HOME_avg, WD_std, RWE_avg
2	0.5×0.5×0.15	High	0.61	68%	HOME_avg, WD_std, NP_std
3	1×1×0.15	Medium	0.60	67%	WD_std, HOME_avg, NP_std
4	2×2×0.15	Low	0.56	62%	WD_std, HOME_std, FS_std
5	0.3×0.3×0.3	Very high	0.55	62%	HOME_avg, WD_std, RWE_avg
6	0.5×0.5×0.3	High	0.59	66%	HOME_avg, WD_std, FS_std
7	1×1×0.3	Medium	0.58	64%	WD_std, HOME_avg, FS_std
8	2×2×0.3	Low	0.52	59%	WD_std, HOME_avg, NP_std

In all cases, Number of trees= 430 (6 species). Voxel size = length (X axes) × width (Y axes) × height (Z axes). See Table 3.3 for codes of the metrics; _avg= average value; _std= standard deviation.

The 6 “selected metrics”, ranked by the indices of MDA, were successively included for the accuracy assessment and calculated for all 8 different voxel sizes (Figure 3.7). For all classifications, the first three (especially the first two) metrics significantly improved the overall accuracies, whereas others resulted in improvements of less than 5%.

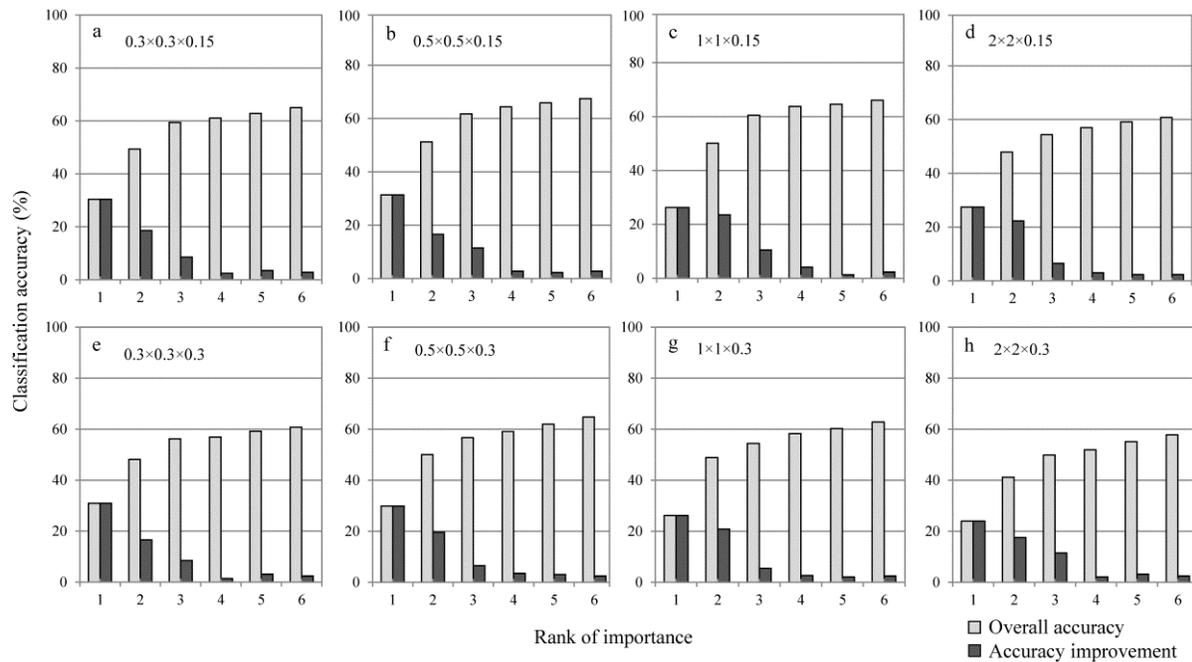


Figure 3.7 Overall accuracy and the improvement of accuracy when successively including the selected metrics ranked by the indices of Mean Decrease Accuracy (MDA). The ranks (of importance) are shown in the X-axis for each sub-figure. The corresponding voxel size is shown at the top of each sub-figure. Note: overall accuracy (in the legend) = overall classification accuracy of 6 tree species.

Table 3.4 shows the confusion matrix and classification accuracies in three classification levels, i.e., using all six tree species, four main species and two generalized classes. Below each sub-table the corresponding overall accuracy and Kappa coefficient are shown. The results using all the “selected metrics” (calculated during the metrics selection process in RF analysis) (shown in the upper sub-tables) and the “most important metrics” (selected according to the indices of MDA in RF) (in the lower sub-tables) for classifications are also compared under the most suitable voxel size (i.e. 0.5×0.5×0.15m, see Table 3.3). Grouping species into coniferous and broadleaved trees resulted in high classification accuracy (overall accuracy = 83.7-85.4%; Kappa coefficient = 0.67-0.71). Accuracy decreased when the number of classes increased.

Table 3.4 Confusion matrix for the detected trees on all the plots in three classification levels, i.e., using all six tree species, four main species and two generalized classes. The results of using the “selected metrics” and the “most important metrics” are compared. The values of the fields in the confusion matrix are given as percentage of predicted classes of the total number of reference classes, and the correctly classified results are highlighted with a grey background.

		M.Pine	C.Fir	S.Pine	S.Oak	C.SG	C.Holy			M.Pine	C.Fir	S.Oak	C.SG			Conifer	Broadleaf
		The selected metrics															
		Predicted classes						Predicted classes								Predicted classes	
Reference classes	M.Pine	68.5	6.2	7.3	2.3	1.7	14	M.Pine	76.7	8.2	8.5	6.6	Conifer	88.3	11.7		
	C.Fir	3.7	74.2	2.4	4.9	7.7	7.1	C.Fir	6.9	79.8	6.1	7.2	Broadleaf	17.5	82.5		
	S.Pine	7.8	3.8	68.1	4.7	13.4	2.2	S.Oak	1.6	5.8	72.3	20.3					
	S.Oak	4.4	5.1	4.2	69.2	15.1	2.0	C.SG	6.2	7.7	14.6	71.5					
	C.SG	1.9	4.3	11.6	13.1	64.3	4.8										
	C.Holy	15.0	7.1	4.3	6.1	5.1	62.4										
		Overall accuracy: 67.8 %				Kappa coefficient: 0.61		75.1 %		0.67		85.4 %		0.71			
		The most important metrics															
		Predicted classes						Predicted classes								Predicted classes	
Reference classes	M.Pine	65.3	6.1	4.9	4.3	4.8	14.6	M.Pine	73.6	9.6	9.4	7.4	Conifer	86.9	13.1		
	C.Fir	5.6	72.2	6.2	6.5	3.6	5.9	C.Fir	8.8	77.8	4.5	8.9	Broadleaf	19.5	80.5		
	S.Pine	10.6	8.5	64.7	5.5	8.4	2.3	S.Oak	1.5	4.2	69.8	24.5					
	S.Oak	4.7	4.4	6.5	66.8	15.1	2.5	C.SG	4.5	8.2	18.4	68.9					
	C.SG	4.8	8.4	9.2	8.1	63.9	5.6										
	C.Holy	19.1	6.5	6.3	5.4	5.2	57.5										
		Overall accuracy: 65.1 %				Kappa coefficient: 0.58		72.5 %		0.63		83.7 %		0.67			

M.Pine: Masson Pine; C.Fir: Chinese fir; S.Pine: Slash pine; S.Oak: Sawtooth oak; C.SG: Chinese sweet gum; C.Holy: Chinese holly.

The selected metrics: HOME_avg, WD_std, NP_std, RWE_avg, FS_std and HTMR_std.

The most important (M-I) metrics: HOME_avg, WD_std and NP_std (see Table 3.3). See Table 3.2 for codes of the metrics.

3.5 Discussion

3.5.1 Segmentation of trees

The detection and delineation of trees can be directly on the point cloud, but more typically on a raster image interpolated from discrete LiDAR data (Koch et al., 2014). In this research, a method similar to Popescu et al. (2002) and Popescu and Wynne (2004) was used to detect individual trees by the CHM, and to delineate tree crowns by fitting radius profiles and analyzing the critical points. The detection rate of all trees in the sample plots was 76%, which is higher than that recorded by Ferster et al., (2009) who found a 41%, for trees taller than 28m in a coastal temperate Douglas-fir, and 71% in the Scandinavian boreal forest (Holmgren and Persson, 2004). This is likely because: 1) the understory and suppressed trees, which are less likely to be detected by the local maxima algorithm used in this research, were not considered; 2) the local maxima algorithm was applied on a CHM created from a higher density point cloud (approximately $15 \text{ pts} \cdot \text{m}^{-2}$ in the study site) (Cao et al., 2014) extracted from full-waveform data by decomposition algorithms; and 3) the size of the search window was determined by a model fitted by the locally collected field measurements, which may enhance the performances of tree top identification.

In this research, conifer-dominated plots (80-87%) had a higher detection rate than broadleaved (75-78%) plots. That was mainly because the crown shape of the coniferous trees tends to be conical and relatively isolated. Broadleaved trees have more rounded crowns that are likely to overlap. Another reason is that some suppressed evergreen trees (e.g. Chinese holly) under the dominant canopies may not be successfully detected by the local maxima algorithm applied in this research.

3.5.2 Metrics selection and their explanation

This research provided an effective procedure for full-waveform metrics selection. The analysis started with all the low-correlated metrics (with $r < 0.6$) and discarded the ones with least importance until the classification accuracy became stable. Then, the 3 top-ranked metrics were selected

(according to the indices of MDA) as the final “most important metrics” from the “selected metrics” under different voxel sizes. For the most suitable voxel size (which had the highest classification accuracy), the most important metrics were (ranked by MDA): average of HOME (HOME_avg), standard deviation of WD (WD_std) and standard deviation of NP (NP_std), and the combination of these three metrics had a relatively strong classification capability that provided an overall accuracy of 65.1% for all the six species (and 83.7% for the general classes of conifer and broadleaf).

The metric of HOME_avg (i.e., the average of height of median energy) appears to have the strongest capability for tree species classification in this research area. Drake et al. (2002) demonstrated that HOME is sensitive to canopy openness and the vertical arrangement of canopy elements in the LVIS-footprint (0.05ha) and plot (0.25-0.5ha) level. In this research, similar principles were found at the individual-tree level by using small-footprint (0.45m) full-waveform LiDAR data. An explanation for this is that for tree crowns with densely packed woody materials and foliage, few LiDAR pulses can penetrate and reach the ground, thereby increasing HOME. Conversely, for more transparent crowns, more LiDAR pulses can pass through, thus, reducing HOME. HOME is sensitive to these differences and as a result it is potentially a good indicator for discriminating tree species. The second ranked metric WD_std (i.e. the standard deviation of waveform distance) relates to the height variation within each tree crown. This relationship could be explained by the different biophysical properties of the crown surface. At the research site, coniferous trees usually consist of a relatively dense and homogenous crown surface, resulting in a lower variation of waveform distance within each tree crown, while the crown surface of broadleaved trees are more heterogeneous and thus have a higher height variation. The third ranked metric, NP_std (i.e., the standard deviation of the number of detected peaks), describes the variation of the total number of targets within a composite waveform, which may be interpreted as the variation of the vertical foliage density within each tree crown. Some previous studies have also indicated that tree species could be discriminated due to their internal structure (Brandtberg, 2007; Reitberger et al., 2006). Different internal structures within tree crowns may lead to a different number of reflections within one pulse, and thus could be used to classify tree species.

3.5.3 Assessments of classification and the effects of voxel sizes

According to the confusion matrix and the final classification accuracies (Table 3.4), the classification of all six species had the lowest accuracy due to the largest number of classes. For the highest level of classification (i.e., with all the six species), the accuracy was relatively high for the coniferous species (i.e., Masson pine, Chinese fir and Slash pine) and the tall deciduous species (i.e., Sawtooth oak and Chinese sweet gum), followed by secondary broadleaved species (i.e., Chinese holly). When classifying the four main species, the pattern is similar to the previous case, and the exclusion of secondary tree species (i.e., Chinese holly) resulted in an improvement in the classification accuracies for Sawtooth oak and Chinese sweet gum. In addition, compared with the classification results using the “selected metrics”, the use of the “most important metrics” did not result in a significant difference for all the classification depths, which agrees with the results in Figure 3.7 that show that the first three metrics (i.e., the most important metrics) significantly improved the overall accuracies, whereas the others only resulted in improvements of less than 5%.

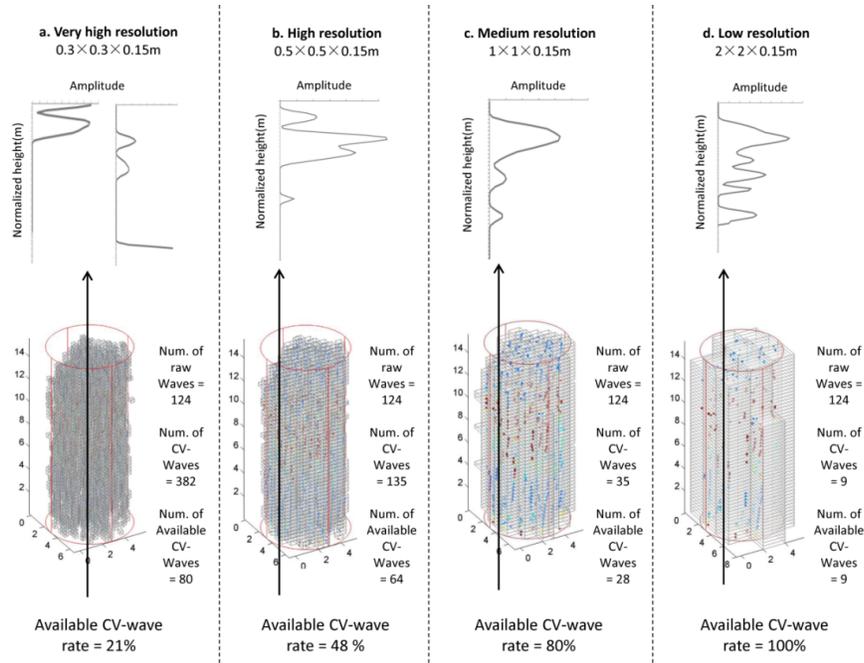


Figure 3.8 Demonstration of the effects of different voxel resolutions on the available (i.e., neither incomplete (a) nor over-synthesized (d)) rates of composite waveforms in one example tree. CV-Wave=composite waveform.

The assessment of different voxel size showed that voxels in the “High” resolution (i.e., $0.5 \times 0.5 \times 0.15$ m and $0.5 \times 0.5 \times 0.3$ m) category had the highest classification accuracies, followed by “Very high” and “Medium” resolutions. The voxels in the “Low” resolution category had the lowest classification accuracies. Under the “Very high” voxel-size condition ($0.3 \times 0.3 \times 0.15$ m), the horizontal resolution of the voxels was even finer than the LiDAR footprint diameter (0.45 m), thereby producing many incomplete composite waveforms. As shown in Figure 3.8.a (i.e., samples of composite waveforms with either incomplete beginning or ending), the available composite (CV) waveform rate (calculated as the number of complete CV-waveforms divided by the total number of CV-waveforms) was only 21% under the “Very high” resolution, although the number of newly formed CV-waveforms (382) was much higher than the directly observed raw waveforms (124). On the other hand, under the “Low” voxel resolution condition ($2 \times 2 \times 0.15$ m), the horizontal resolution of the voxels was around half of the average crown diameter (4.39 m). As demonstrated in Figure 3.8.d, although the available composite (CV) waveform rate was as high as 100%, the number of complete CV-waveforms was very low (9). The final composite waveforms tended to be over-synthesized and as a result many of the detailed geometric and radiometric information were lost. Thus, determination of voxel size should consider the footprint size, crown diameter and the available composite waveform rate.

In this research, the locally adapted local maxima algorithm demonstrated good performance, and when combined with the full-waveform metrics (calculated based on a voxel-based method) and a Random Forest classifier, the approach showed good accuracy discriminating 6 typical subtropical tree species (in 3 classification depths). Some combinations of the full-waveform metrics (based on height of median energy (HOME), waveform distance (WD) and number of waveform peaks (NP)) with high classification importance were shown to be stable among various voxel sizes, and are excellent candidate methods for species classification with high accuracy. In addition, the “High” resolution (i.e., $0.5 \times 0.5 \times 0.15$ m and $0.5 \times 0.5 \times 0.3$ m) voxel size represents the highest classification accuracies. The following chapter examines the capability of discrete and full-waveform data to estimate total biomass and biomass components.

4 Estimation of Biomass Components using Discrete and Full-waveform LiDAR Metrics

4.1 Introduction

The aim of this part of the research is to explore the advantage of discrete and full-waveform data in the estimation of total biomass and specific biomass components. The relationship between forest biomass (i.e. above-ground and total biomass) and its components (i.e. foliage, branch, trunk and root biomass) with small-footprint discrete (DR) and full-waveform (FWF) airborne LiDAR derived metrics were examined in three types of subtropical forest stands (i.e. coniferous, broadleaved and mixed forests). In addition, the estimation capability of DR and FWF metrics based models was evaluated for biomass estimation individually and in combination.

4.2 Materials and methods

4.2.1 Study site

See details about the study site in Section 2.1.

4.2.2 Field data

Data from all of the 66 plots collected between June-August 2012 and August 2013 were used in this study (see details about these data in Section 2.3).

Species-specific allometric equations referenced from Feng et al. (1999), Qian (2000), Wang and Shi (1990) and Xu et al. (2011) were used. These allometric equations were developed from tree inventory data from local or nearby provinces and therefore were used in this research with biomass (i.e., total and above-ground biomass) and biomass components (i.e., stem, root, branch and foliage) calculated for individual trees within each plot based on the field-measured DBH and height, and then summed to obtain total plot biomass.

4.2.3 LiDAR data

The small-footprint full-waveform LiDAR data acquired on 17 August 2013 were used in this study (see details about these data in Section 2.2).

4.2.4 Data pre-processing

A 1-m digital terrain model (DTM) was created in two steps from the DR data. First, the data was filtered to remove the above-ground returns (using an algorithm adapted from Kraus and Pfeifer, 1998), and secondly the DTM was created by calculating the average elevation from the remaining (ground) LiDAR returns within a cell (cells that contain no points were filled by interpolation using neighboring cells). The DR point clouds were then normalized against the ground surface height and extracted for each plot using the coordinates of the lower left and upper right corners. All point cloud data processing was performed using the FUSION software package (McGaughey, 2014).

To avoid the effects of off-nadir pointing on the FWF (e.g., the waveform is stretched as the off-nadir angle increases) and to integrate non-vertical waveforms from different flight trajectories, a voxel-based approach developed by Hermosilla et al. (2014) was applied to the full-waveform data. This approach synthesized multiple raw waveforms into composite waveforms through the vertical space partitioning of forest canopies by voxels, and used the maximum amplitude value to construct new pseudo-vertical waveforms. The background noise of each waveform was first suppressed by a de-noising process and then smoothed by a Gaussian filter. Each pre-processed waveform was then spatially located in three-dimensional space and normalized by subtracting the derived DTM height from height of each bin in the corresponding positions. The vertical space was partitioned into voxels (0.25×0.25×0.15 m) according to the footprint size and temporal sample spacing. A maximum amplitude value was calculated and assigned to each intersected voxel. The composite waveforms were finally extracted as each vertical column of voxels (Hermosilla et al., 2014) (See the technical route of this research in Figure 4.1).

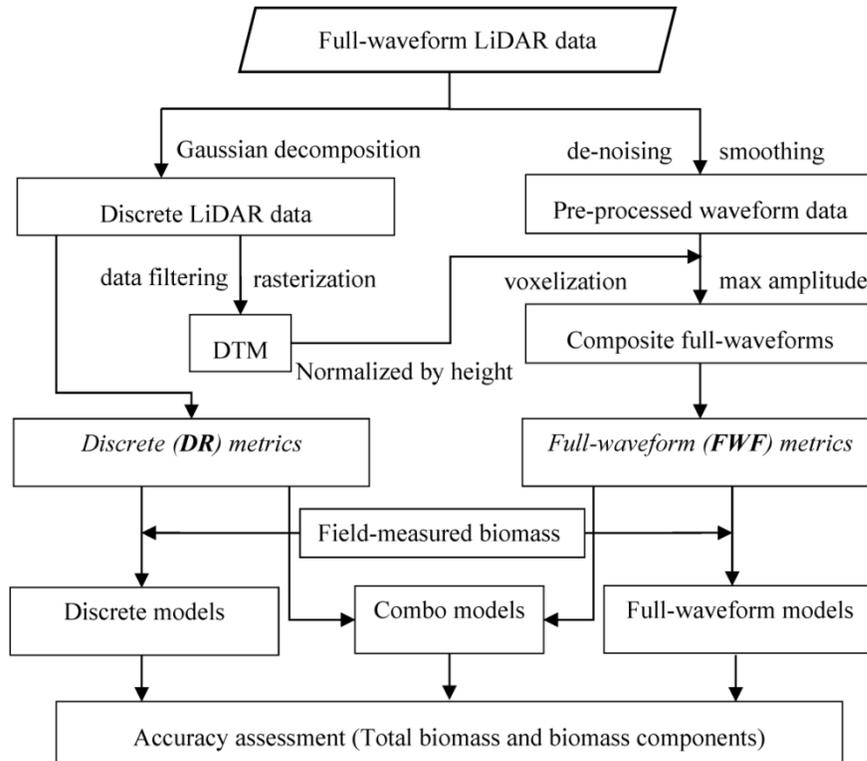


Figure 4.1 Technological route of this research

4.2.5 LiDAR metrics

4.2.5.1 Discrete data based metrics

Discrete data-based metrics are descriptive structure statistics calculated from the height normalized LiDAR point cloud measurements in three-dimensional space (Lefsky et al., 2005). In this study, 15 metrics for each plot within the 30×30 m area of the plots were calculated including: i) selected height measures, i.e., percentile height (h_{10} , h_{25} , h_{50} , h_{75} , h_{90} and h_{95}), maximum height (h_{\max}) and mean height (h_{mean}); ii) variability of height measures, i.e., coefficient of variation of heights (h_{cv}); iii) selected canopy return density measures, i.e., canopy return density (d_2 , d_4 , d_6 and d_8) and iv) canopy cover measures, i.e., canopy cover above 2 m ($CC_{2\text{m}}$) and canopy cover above mean height (CC_{mean}).

All the discrete data-based metrics were generated from first returns, which have been found to be more stable for estimating forest biophysical attributes than all returns (Bater et al., 2011). Metrics of

heights and canopy return densities were computed with a height threshold of 2-m above ground-level to ensure that the non-canopy and below-canopy returns were excluded from metric calculation (Andersen et al., 2005; Hyypä et al., 2012).

4.2.5.2 Full-waveform based metrics

Full-waveform data-based metrics describe the radiometric and geometric attributes of the return waveforms based on the spatio-temporal analysis. In this research, 15 waveform metrics adapted from previous studies of large-footprint full-waveform data (i.e. SLICER, LVIS and GLAS) were incorporated into the pseudo-vertical waveform analysis due to their success in previous studies (Heinzel and Koch, 2011; Xu et al., 2012; Neuenschwander et al., 2009; Neuenschwander, 2012). These metrics included height of median energy (HOME), waveform distance (WD), height/median ratio (HTMR), number of peaks (NP), roughness of outermost canopy (ROUGH), front slope angle (FS), return waveform energy (RWE) and vertical distribution ratio (VDR).

HOME was calculated as the distance from waveform centroid to the ground (i.e. corresponding DTM height), and has been found to be sensitive to canopy openness and the vertical arrangement of canopy elements (Drake et al., 2002). WD, often related to the LiDAR canopy height, was computed as the distance from waveform begin to the ground (Sun et al., 2008). HTMR, which is HOME divided by WD, was used to depict the change of HOME relative to the canopy height (Drake et al., 2002). ROUGH, which is the distance from the waveform beginning to the first peak, was applied to describe the roughness of the upper-most canopy and the spatial arrangement of plant surfaces (Duong, 2010). FS, which is the vertical angle from waveform beginning to the peak of canopy return energy, provided the variability of the upper canopy (Ranson et al., 2004). NP is the number of detected peaks within each composite waveform. RWE represents the total received energy. VDR was calculated as the differences between WD and HOME, divided by WD (Neuenschwander, 2012). The plot level waveform metrics (except VDR) were summarized as the mean and standard deviation of all the waveform metrics (of individual composite waveform) within each plot. A summary of the FWF metrics with corresponding descriptions is shown in Table 3.2.

4.2.6 Statistical analyses

The strength of relationships between biomass (and each of the biomass components) and the 30 LiDAR metrics were tested by the square of the Pearson correlation coefficient (R^2). Multiple linear regression models, which included the two sets of LIDAR metrics (i.e. discrete and full-waveform based metrics) as predictor variables, were developed independently and in combination for estimating the biomass and biomass components among different forest types (i.e. all forests, coniferous forests, broadleaved forests and mixed forests). To ensure the independent variables were not highly correlated, collinearity was first evaluated using Principal Component Analysis (PCA) based on the correlation matrix. Models with a condition number (κ) lower than 30 were accepted to ensure that there was no serious collinearity in the selected models (Naesset, 2002). Standard backward stepwise regression was performed to select variables for the final models with predictor variables left in the models at the 5% significance level. The best fitting models were then selected based on the lowest Akaike information criterion value (Akaike, 1974). Once the best models were chosen, leave-one-out cross validation was used to validate them.

Three sets of predictive models, namely discrete data metrics based models (DR-model), full-waveform metrics based models (FWF-model) and Combo models (includes the combination of both discrete and full-waveform metrics) were fitted. These prediction models were assessed using adjusted coefficients of determination ($Adj-R^2$), cross-validation coefficients of determination ($CV-R^2$), Root-Mean-Square Errors (RMSE), and relative RMSE (rRMSE), defined as the percentage of the ratio of RMSE and the observed mean values (Table 4.1).

Table 4.1 The accuracy assessment results for three sets of predictive models among various forest types.

	Discrete data model				Full-waveform model				Combo model			
	Aj-R ²	C-R ²	RMSE	rRMSE	Aj-R ²	C-R ²	RMSE	rRMSE	Aj-R ²	C-R ²	RMSE	rRMSE
<i>All forests</i>												
W _{AGB}	0.74	0.69	16.12	12.17	0.52	0.45	25.36	28.61	0.75	0.70	15.91	11.46
W _{trunk}	0.59	0.55	18.94	28.27	0.45	0.41	21.88	32.66	0.6	0.56	18.71	27.93
W _{branch}	0.77	0.71	3.97	27.46	0.65	0.62	4.87	33.68	0.79	0.74	3.76	26.01
W _{foliage}	0.29	0.23	4.02	56	0.07	0.04	4.59	63.94	0.32	0.28	3.91	54.47
W _{root}	0.59	0.55	5.68	21.71	0.42	0.38	6.79	25.95	0.56	0.51	5.88	22.47
W _{total}	0.7	0.64	21.11	22.03	0.5	0.43	27.12	28.3	0.7	0.64	20.88	21.79
<i>Coniferous forests</i>												
W _{AGB}	0.84	0.79	11.55	15.84	0.81	0.76	12.45	17.07	0.92	0.87	8.08	11.08
W _{trunk}	0.5	0.46	13.78	25.98	0.65	0.61	11.49	21.66	0.84	0.81	7.78	14.67
W _{branch}	0.88	0.81	2.45	23.25	0.77	0.73	3.36	31.89	0.93	0.88	1.84	17.46
W _{foliage}	0.79	0.73	3.38	36.18	0.73	0.69	3.84	41.1	0.82	0.78	3.12	33.39
W _{root}	0.86	0.82	3.04	13.8	0.76	0.71	3.97	18.02	0.88	0.85	2.82	12.8
W _{total}	0.88	0.83	11.83	14.38	0.79	0.73	15.9	19.33	0.93	0.89	9.29	11.29
<i>Broadleaved forests</i>												
W _{AGB}	0.87	0.84	16.23	15.41	0.58	0.52	28.83	27.38	0.86	0.81	16.31	15.49
W _{trunk}	0.87	0.83	12.55	15.86	0.57	0.53	22.74	28.75	0.89	0.85	11.34	14.33
W _{branch}	0.88	0.82	3.34	16.27	0.61	0.56	6.06	29.53	0.89	0.84	3.23	15.74
W _{foliage}	0.18	0.13	2.05	36.36	0.27	0.22	1.93	34.23	0.33	0.29	1.85	32.81
W _{root}	0.63	0.57	4.08	13.61	0.43	0.38	5.06	16.88	0.65	0.61	4.0	13.34
W _{total}	0.84	0.81	17.84	16.08	0.57	0.51	29.3	26.41	0.84	0.79	18.14	16.35
<i>Mixed forests</i>												
W _{AGB}	0.75	0.72	16.08	18.33	0.59	0.53	20.85	23.77	0.82	0.78	13.94	15.89
W _{trunk}	0.67	0.61	16.1	23.39	0.53	0.49	19.27	27.99	0.77	0.72	13.42	19.49
W _{branch}	0.72	0.65	3.02	24.53	0.58	0.54	3.72	30.21	0.71	0.66	3.08	25.01
W _{foliage}	0.44	0.41	2.92	44.56	0.41	0.37	3.01	45.93	0.48	0.43	2.77	42.27
W _{root}	0.7	0.64	5.3	20.01	0.44	0.40	7.19	27.14	0.73	0.68	4.99	18.84
W _{total}	0.75	0.71	16.89	17.92	0.63	0.58	20.85	22.12	0.81	0.76	14.94	15.85

Note: A_j-R^2 : adjusted R-square; $C-R^2$: The cross-validation coefficient of determination; RMSE: Root mean square error; rRMSE (%): relative RMSE, i.e., percentage of the ratio of RMSE and the observed mean value. W_{AGB} : above ground biomass ($Mg\ ha^{-1}$); W_{trunk} : trunk biomass ($Mg\ ha^{-1}$); W_{branch} : branch biomass ($Mg\ ha^{-1}$); $W_{foliage}$: foliage biomass ($Mg\ ha^{-1}$); W_{root} : root biomass ($Mg\ ha^{-1}$); W_{total} : total biomass ($Mg\ ha^{-1}$).

4.3 Results

Most of the biomass and its components (hereafter written as “biomass components”) had a stronger correlation with the discrete data based metrics than the waveform model based metrics (Figure 4.2). Within the set of discrete data based metrics, except for the “all forests” and “broadleaved forests” foliage biomass components, most of the height-related metrics (especially the percentile heights and the coefficient of variation of heights) and biomass components were strongly correlated. In comparison, measures of canopy return density and canopy cover had a relatively poorer relationship with the biomass components. The strength of the relationships between biomass components and the waveform metrics were also generally weaker. The waveform metrics related to canopy height (e.g. WD), vertical arrangement (e.g. HOME) and roughness of upper-most canopy (e.g. ROUGH) had a more significant relationship with biomass components than others.

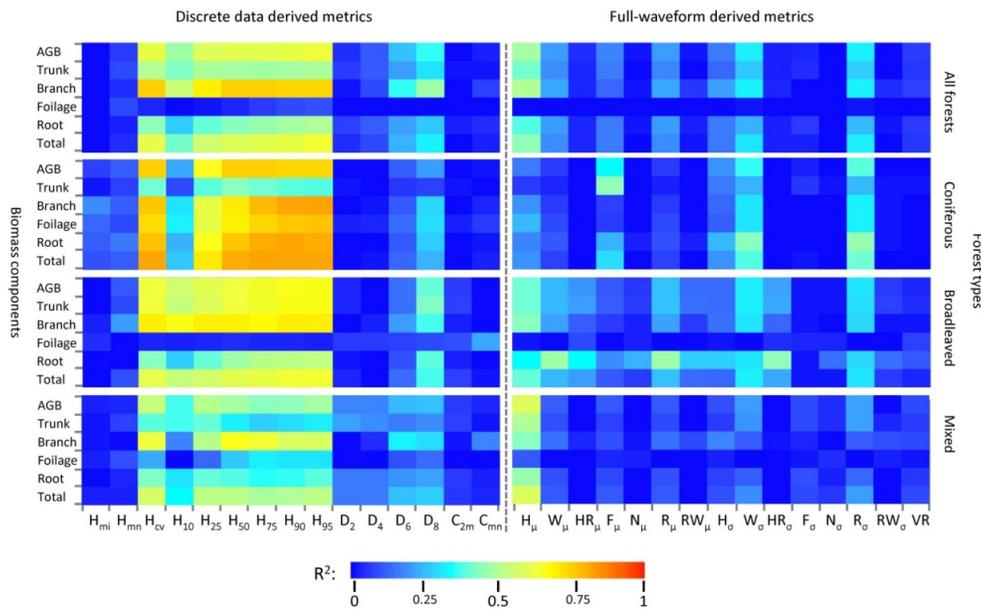


Figure 4.2 Intensity graph of the square of the Pearson’s correlation coefficient (R^2) between each biomass components and LiDAR-derived metrics. The values of the coefficients are transformed into pixels within a blue-red color range.

Note: See Table 4.1 for the codes and descriptions of metrics ($H_{mi}=h_{min}$; $H_{mn}=h_{max}$; $C_{2m}=CC_{2m}$; $C_{mn}=CC_{mean}$; $H_{\mu}=HOME_{\mu}$; $W_{\mu}=WD_{\mu}$; $HR_{\mu}=HTMR_{\mu}$; $F_{\mu}=FS_{\mu}$; $N_{\mu}=NP_{\mu}$; $R_{\mu}=ROUGH_{\mu}$; $RW_{\mu}=RWE_{\mu}$; $H_{\sigma}=HOME_{\sigma}$; $W_{\sigma}=WD_{\sigma}$; $HR_{\sigma}=HTMR_{\sigma}$; $F_{\sigma}=FS_{\sigma}$; $N_{\sigma}=NP_{\sigma}$; $R_{\sigma}=ROUGH_{\sigma}$; $RW_{\sigma}=RWE_{\sigma}$; $VR=VDR$).

Table 4.1 summarizes the accuracies of the three sets of predictive models (i.e., general and forest-type specific models developed from discrete data based metrics, waveform based metrics and their combinations). Overall, most of the biomass components were generally well-predicted ($Adj-R^2=0.42-0.93$, $rRMSE=11.08-33.68\%$) except for foliage biomass ($Adj-R^2=0.07-0.82$, $rRMSE=32.81- 62.94\%$). The combo models performed the best, explaining 60-79% (for the general model) and 65-93% (for each forest type-specific model) of the variability of most biomass components (excluding foliage biomass). Compared with full-waveform models ($Adj-R^2=0.42-0.79$), the discrete data models ($Adj-R^2=0.50-0.88$) had a relatively higher performance, except for the trunk biomass (in coniferous forests) and foliage biomass (in broadleaved forests) (see Table 4.1). The fit for the general model (developed from all plots) was relatively low ($Adj-R^2= 0.07-0.79$). In comparison, the relationships were generally improved for forest type-specific models ($Adj-R^2= 0.18-0.93$). Overall, most of the fitted models had higher correlations in coniferous forests ($Adj-R^2=0.50-0.93$) and broadleaved forests ($Adj-R^2=0.43-0.89$) than in mixed forest types ($Adj-R^2=0.41-0.82$). The relative RMSE was generally in accordance with adjusted R^2 , with lower values in coniferous forests ($rRMSE=11.08-41.1\%$) and broadleaved forests ($rRMSE=13.34-36.36\%$) than in mixed forest types ($rRMSE=15.85-45.93\%$).

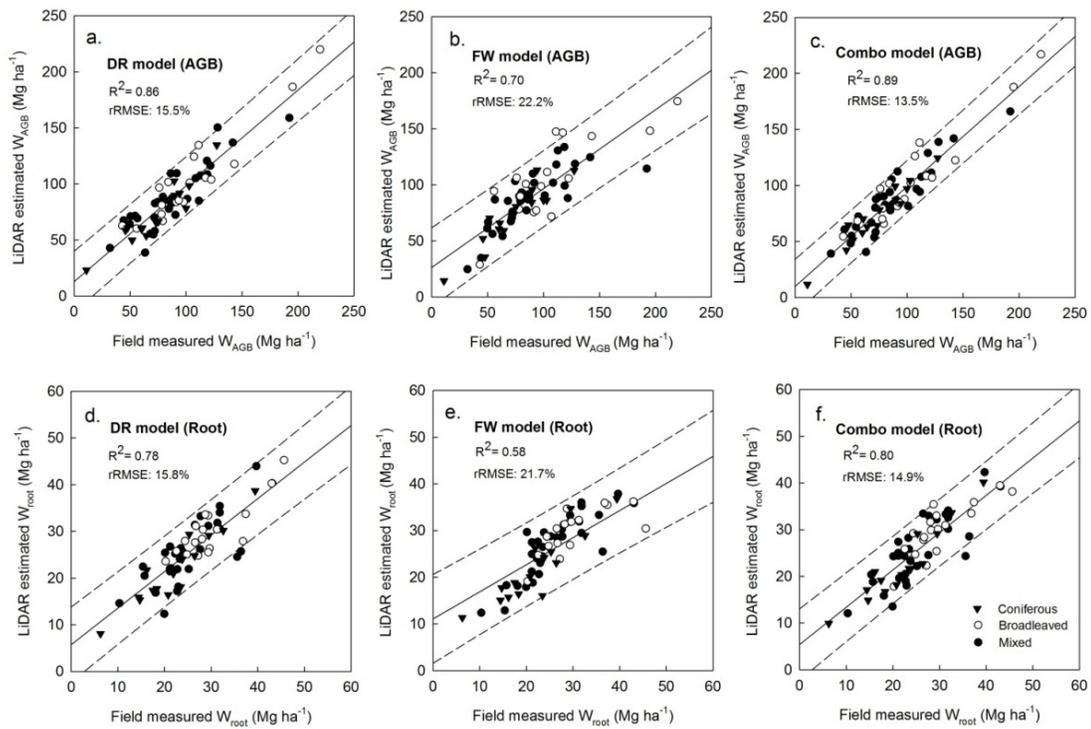


Figure 4.3 Plot-level observed and estimated above-ground and root biomass for the combination of forest-type specific models developed from discrete data based metrics, waveform based metrics and their combinations.

Note: DR model: Discrete data based model; FWF model: Full-waveform data based model; Combo model: Model developed from both of DR and FWF metrics as a pool of candidate variables; AGB: above-ground biomass.

Scatter plots of the plot-level field-measured and LiDAR-estimated above-ground and root biomass are shown in Figure 4.3 with linear fits and prediction confidence intervals (95%). The linear models demonstrate the increased correlations in the Combo models ($R^2=0.89$ for above-ground biomass; $R^2=0.80$ for root biomass) in comparison to the DR model ($R^2=0.88$; $R^2=0.78$) and FWF model ($R^2=0.70$; $R^2=0.58$). The widths of the confidence intervals in the DR model (AGB) and Combo model (AGB) were relatively narrow, indicating a better quality of model fitting. Some outliers, which may be the result of prediction bias, should be noted in the FWF model (AGB) and DR model (Root).



Figure 4.4 The selected LiDAR metrics of the discrete data model vs. full-waveform model in various forest types.
Note: $W_a = W_{AGB}$; $W_s = W_{trunk}$; $W_b = W_{branch}$; $W_f = W_{foliage}$; $W_r = W_{root}$; $W_t = W_{total}$. (See Table 4.1 for biomass components' codes)

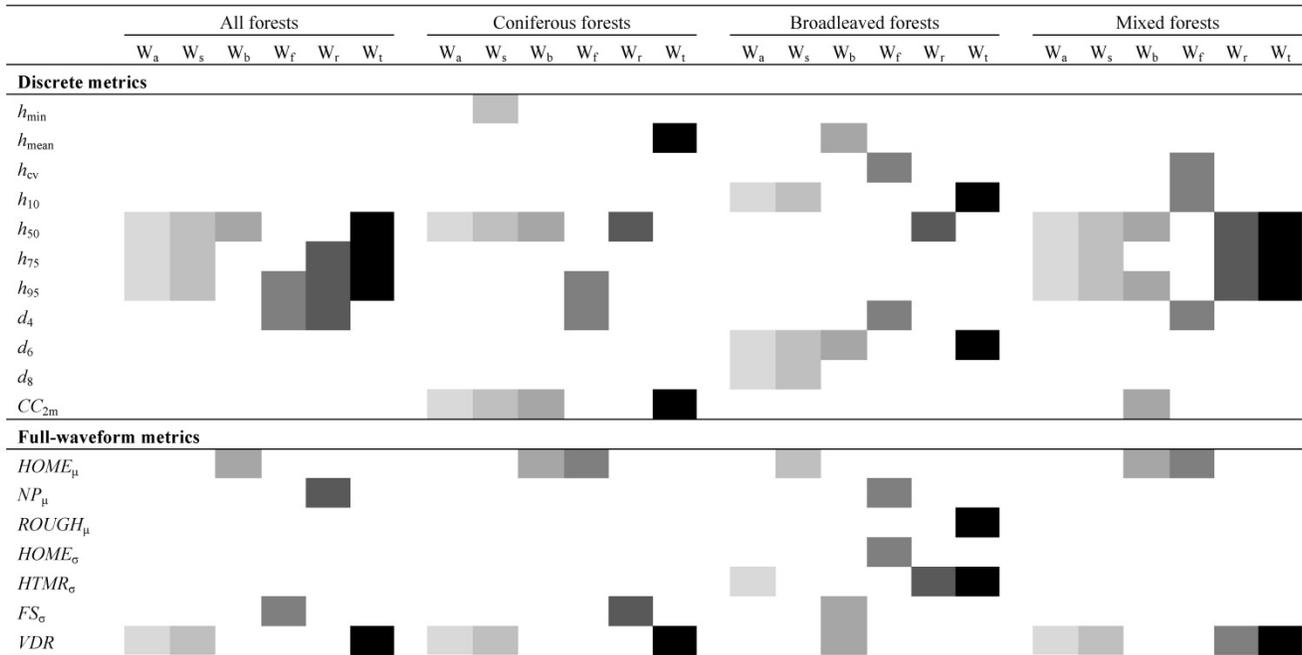


Figure 4.5 The selected LiDAR metrics of the combo model in various forest types.

Note: Each specific biomass component is represented by a cell of different darkness. $W_a = W_{AGB}$; $W_s = W_{trunk}$; $W_b = W_{branch}$; $W_f = W_{foliage}$; $W_r = W_{root}$; $W_t = W_{total}$. (See Table 4.1 for biomass components' codes)

The selected metrics of the two sets of biomass estimation models (i.e., the discrete data models and full-waveform models) are shown in Figure 4.4. For the discrete data models, height-related metrics (h_{mean} , h_{50} , h_{75} and h_{95}) calculated from the point cloud were the most frequently selected by the regression models, followed by canopy return density and canopy cover metrics. The h_{mean} statistic was sensitive to all the biomass components in coniferous forests (selected by all of the 6 conifer models). In comparison, the statistics of h_{50} and h_{95} were more sensitive to the biomass components of “all forests” and mixed forests (selected by 8-9 of 12 models of the “all forests” and mixes), whereas d_6 and d_8 were more sensitive to the biomass components of broadleaved forests (selected by 5 of 6 models of the broadleaves). The h_{75} statistic appeared to be sensitive to above-ground (W_{AGB}), trunk (W_{trunk}) and total biomass (W_{total}), and was often selected by the predictive models (in 3 of the 4 forest types). For the full-waveform models, the statistic of $HOME_{\mu}$ calculated from composite waveforms was the most frequently selected variable. The $HTMR_{\mu}$ and NP_{μ} statistics appeared to be sensitive to foliage biomass since they were selected to predict this component in 3-4 of the 4 forest types. FS_{σ}

values were significantly sensitive to the branch biomass since they were selected in 4 of all the forest types.

Figure 4.5 represents the selected metrics of the combo models, i.e., using both discrete and full-waveform metrics. Each predictive model included at least one variable from the DR and FWF metric pools, whereas more discrete metrics (11 out of the total 15) were used in the combo models than full-waveform metrics (7 out of 15). Height-related metrics of h_{50} and h_{95} remained sensitive to the biomass components of “all forests” and mixed forests (selected by 8-9 of 12 models of the “all forests” and mixes), which agrees with the previous analysis of the discrete data only models. *VDR* was the most frequently used full-waveform metrics in the combo models (it was selected at least once for each forest type-specific models).

Within specific forest types, the biomass components were calculated using the individual models (i.e. forest type-specific models) and then averaged (Figure 4.6). Compared with the field measurements (the 1st left column in each forest type), most of the biomass components were similar to the field-measured values except for an overestimation of branch biomass by all the models within all plots; an underestimation of trunk biomass by full-waveform and combo models within coniferous forests; and an overestimation of trunk biomass by full-waveform and combo models within the mixed forests.

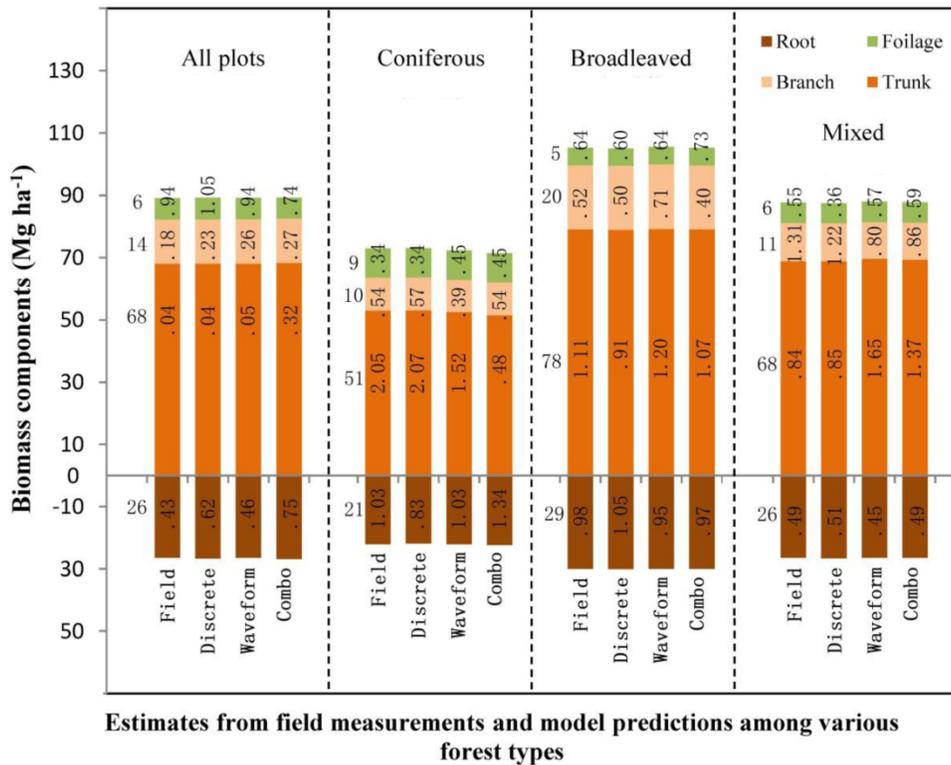


Figure 4.6 Relationship between plot-level field measured and LiDAR-estimated biomass components of three sets of models (i.e. discrete data based model, full-waveform based model and the combo model) within each type of forest. Each biomass component was calculated as the average value of the plots within a specific forest type.

Notes: The values in the first column of each section are baseline data. The numbers inside the cells of the stacked bar are the anomalies. The amount of biomass component = baseline data + anomalies. **Codes:** Field: Field measurements; Discrete: discrete data model estimated biomass; Waveform: full-waveform model estimated biomass; Combo: The combo model estimated biomass.

Across all field plots, the relationship between field-measured and LiDAR-estimated biomass components is shown in Figure 4.7. Overall, most of the estimated biomass components of foliage, branch and root were well-predicted by the combo models. The five plots with the greatest bias in trunk biomass are highlighted and analyzed against four topographic and vegetation factors which may influence the accuracy of biomass estimation (Clark et al., 2004; Hopkinson et al., 2005). Both slope (calculated by DTM) and canopy cover (percentage of first returns above the first return 2-m heights) were derived from discrete LiDAR data, while the mean height and stem number of trees were calculated based on the field-measured data. All five plot outliers had high levels of canopy

cover, and most were located on steep slopes (Figure 4.7.b).

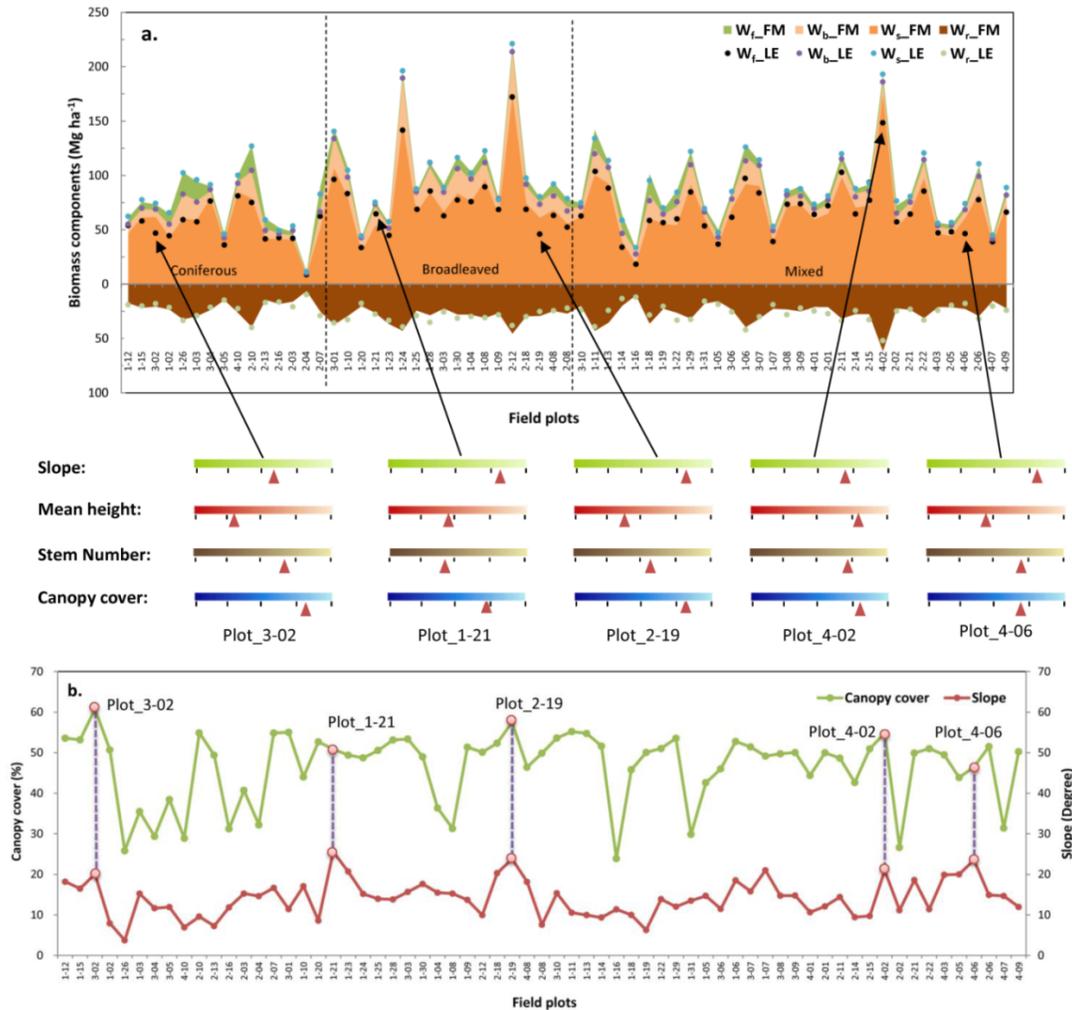


Figure 4.7 Relationship between the amount of field-measured and LiDAR-estimated biomass components across all field plots (abscissa: all of the field plots; ordinate: the amount of biomass components): (a) the field-measured biomass components are displayed as area. The biomass components estimated by the combo model are displayed as dots; five plots with the largest bias in trunk biomass are highlighted and analyzed against four topographic and vegetation factors; (b) slope (degree) and canopy cover (%) of all plots with the five outlier plots highlighted.

Notes: W_f_FM: field-measured foliage biomass; W_b_FM: field-measured branch biomass; W_s_FM: field-measured trunk biomass; W_r_FM: field-measured root biomass; W_f_LE: LiDAR-estimated foliage biomass; W_b_LE: LiDAR-estimated branch biomass; W_s_LE: LiDAR-estimated trunk biomass; W_r_LE: LiDAR-estimated root biomass. The five plots with the greatest bias are singled out, and levels of the topographic (i.e. slope) and vegetation condition (i.e. mean height, stem number and canopy density) factors are marked (by red triangle) within each color ramp bar, showing where the values are within the scaled range of all plots.

4.4 Discussion

Fu et al. (2011) estimated above-ground biomass components using low hit density (point distance=1.08 m) small foot-print discrete LiDAR data in 78 smaller circle plots (radius=7.5 m or 15 m) in mountainous subtropical forests in southwest China. In their study, less variability (compared with results in this research) was explained for above-ground biomass within the broad-leaved (43%) and coniferous (68%) forests, and there was no significant correlation found in the mixed forests. Both the larger plot size (30×30m) (see Zhao et al., 2009 for details of the effects of plot size) and the higher hit density may explain the more significant correlations found in this study. Studies predicting root biomass from small footprint LiDAR, especially in subtropical forests, are rare. Næsset (2004) reported a $R^2 = 0.86$ for below-ground biomass in boreal forests. Pang & Li (2012) fitted regression models for biomass components in temperate forests in Northeast China and explained 85% of the variability in root biomass in conifers, which is slightly higher than reported in this study. The reason for the slightly lower accuracy in this study may be due to the relatively higher species diversity in this study site, which has a far greater variability in stem height and density compared to the above-mentioned boreal and temperate forest stands.

Neuenschwander (2012) extracted waveform metrics (e.g. amplitude, pulse width, integrated canopy energy) from FWF data and found that the FWF-metrics can help with improving the characterization and classification of the vegetation. In this study, most of the combo models (including both DR and FWF metrics) showed a higher variation explanation capability for predicting the biomass components than the DR-metrics (only) models, especially for trunk biomass (in coniferous forests; with a adjusted- R^2 increase of 0.34) and foliage biomass (in broadleaved forests; with a adjusted- R^2 increase of 0.15), which is in agreement with their findings. Sumnall et al. (2012) investigated the estimation of forest structural variables from both discrete (system points) and full waveform LiDAR data. Height and intensity-related metrics were derived from DR data (system points), metrics of heights (derived from waveform points), amplitude and pulse widths of the waveforms were all derived from FWF data. Their results indicated that most of the FWF-models (using both waveform points and full-waveform based metrics as independent variables) had a higher

accuracy than DR models (using system points based metrics as independent variables) for estimating biophysical properties of a semi-natural forest in southern England, which agrees with this research that shows that most of the combo models had a higher performance than DR-models, and the full-waveform metrics contributed to the improvements by characterizing the detailed vertical structure of the forest stands.

For the discrete and full-waveform metrics analyzed in this study, metrics related to vegetation heights, i.e., mean height, upper height percentiles (i.e. h_{50} , h_{75} , h_{95}), HOME (i.e. waveform centroid to the ground) and VDR (calculated from HOME and waveform distance), were the strongest predictors of biomass and its components. Larson (1963) found that the bole form and stem increment are strongly connected to the crown geometry and crown position. The LiDAR sensors measure 3-D structural characteristics of the forest canopy, which provides a good foundation for strong relationships between the LiDAR-derived metrics and forest biomass. Since most LiDAR returns are from dominant trees, the LiDAR mean height likely represents the height of overstory trees. The inclusion of waveform-derived HOME (which depicts the vertical arrangement of canopy elements) and VDR (which describes the change of HOME relative to the canopy height) (Duong, 2010), helped to account for intermediate tree crowns in the midstory and suppressed trees in the understory, therefore potentially improving the characterization of the vertical structure. The predictive capabilities of these metrics were strong (in both the single and combo models) for estimating most of the biomass components, and they therefore provide a precise quantitative description of stand structure characteristics. This finding is consistent with the small-footprint discrete data (Popescu, 2007), small-footprint full-waveform data (Hermosilla et al., 2014) and large-footprint SLICER data (Lefsky et al., 2005; Drake et al., 2002) studies, in which mean height, height percentiles and HOME (and HOME-related waveform metrics) were found to explain most of variability in forest structure characteristics.

Compared to the individual DR and FWF models, the combo models performed the best, explaining a large amount of variability for most biomass components, and indicating significant synergies in small-footprint DR and FWF LiDAR data across the three types of subtropical forests. Fu

et al. (2011) explored several regression models relating both percentile heights and canopy return density variables derived from airborne DR LiDAR for the estimation of above-ground biomass components in a similar forest type. Compared with this study, their results showed relatively low relationships for above ground biomass ($R^2=0.37-0.68$), and no significant correlation between foliage biomass and the DR metrics in the subtropical mixed forests. In this study, FWF models had a higher variation explanation capability for predicting foliage biomass than DR models. The waveform statistic of $HTMR_{\mu}$ and NP_{μ} also seems to be sensitive to foliage biomass, and FS_o was significantly sensitive to branch biomass. The results demonstrate promising results for improving the estimation of crown (i.e. branch and foliage) components, which would be useful for modeling forest growth and fire behavior.

In this study, the forest type-specific models performed better than the general model (developed from all plots). This is consistent with previous studies that have demonstrated that forests with highly variable species composition should be stratified to ensure accurate biomass predictions (Fu et al., 2011; Pang and Li, 2012). Næsset (2004) reported that forest type did not have any significant impact on the estimated biomass models in boreal forest (in southeast Norway, dominated by Norway spruce and Scots pine). However, in this study, the forest conditions in subtropical forest stands contained greater species diversity, making the effects of tree-species composition (classified as forest types) significant. Overall, the models were more accurate for coniferous and broadleaved forests than in mixed forests across all models. The dominant coniferous and broadleaved species at this research site have a relatively even-aged tree structure and higher homogeneous composition, i.e., less variation in tree height and density at the plot level. In contrast, the mixed forest had more variability and less accurate LiDAR-derived heights within plots. Meanwhile, individual tree biomass in coniferous and broadleaved plots was calculated by relatively accurate species-specific allometric equations, while the mixed forest biomass was estimated from a less accurate, generalized allometric equation. Thus, the homogeneous forest conditions coupled with the accurate field-measured biomass was expected to strengthen the linear fit and reduce bias.

Bater & Coops (2009) explored the relationship between satellite-derived NDVI and LiDAR

ground return density in a coastal western hemlock dominated forest, and found that ground return density decreased as NDVI (an estimate of vegetation cover) increased. Clark et al., (2004) also demonstrated that in an evergreen tropical rain forest, RMS error of DEM on steep slopes was 0.67 m greater than on flat slopes. Meanwhile, the LiDAR ground retrieval was complicated by the dense, multi-layered evergreen canopy in old-growth forests, causing an overestimation of 1.95 m. In this study, five plots with the largest bias of trunk biomass were analyzed by topographic and vegetation conditions. All five plot outliers had high levels of canopy cover, and most were located on steeper slopes. Thus, the lower predictive accuracy may be due to less canopy penetration by LiDAR and consequently less accurate definition of the digital terrain model and canopy height model. Bater & Coops (2009) suggested that in areas with high slope and low point density, the prediction uncertainty maps may provide insights into optimal DEM selection and improve the vegetation height estimation.

This study suggests that although the discrete and full-waveform LiDAR data extract forest canopy information in different ways, i.e., discrete data characterize the 3-D canopy structure, while full-waveform data recorded the detailed geometric and radiometric information of the canopy return, both discrete and full-waveform metrics can provide valuable information for accurately estimating specific biomass components in certain forest types. However, the choice of LiDAR metrics impacts the performance of models used for the biomass estimation, and different biomass pools are correlated with different LIDAR metrics. As a result, the combination of DR and FWF metrics and the exploration of new sets of full-waveform metrics contributed to the development of more robust predictive models.

This study implemented two different approaches (i.e., discrete data and full-waveform based) to extract plot-level LiDAR metrics, and evaluated the capacity of these metrics to predict biomass and its components in three types of subtropical stands. The results indicated that the three sets of predictive models performed well across the different subtropical forest types ($\text{Adj-R}^2 = 0.42\text{--}0.93$, excluding foliage biomass). Forest type-specific models ($\text{Adj-R}^2 = 0.18\text{--}0.93$) were generally more accurate than the general model ($\text{Adj-R}^2 = 0.07\text{--}0.79$) with the most accurate results obtained for coniferous stands ($\text{Adj-R}^2 = 0.50\text{--}0.93$) in relatively homogeneous forest conditions. In addition, the

analysis of bias in the estimation of trunk biomass demonstrated that the lower accuracy may be due to reduced canopy penetration by the LiDAR returns, resulting in a less accurate digital terrain model and canopy height model. The following chapter demonstrates the use of small-footprint discrete-return LiDAR data to estimate and map above- and below- ground biomass components in the study site.

5 Mapping Biomass Components using LiDAR Data

5.1 Introduction

The research presented in this chapter reports the use of a small-footprint LiDAR (point cloud) to spatially extrapolate the estimates of above- and below-ground biomass components (in 2013) in subtropical forest stands. The objectives were (1) to build models for above- and below-ground biomass components for forest plots (30 × 30 m) in three forest types (*i.e.*, mixed, broadleaved and coniferous forests) using LiDAR data, and to assess the impact of forest type on the accuracy of the regression models; and (2) to map the above- and below-ground forest biomass components across the site using LiDAR data, and evaluate the accuracy using a suite of independent stand-level field inventory data.

5.2 Methods

5.2.1 Study site and materials

5.2.1.1. Study site

See details about the study site in Section 2.1.

5.2.1.2. Plot data

Plots measured in June-August 2012 (n=53) were used in this study (see details about these data in Section 2.3). A summary of the ground estimated plot-level biomass is shown in Table 5.1.

Table 5.1 Summary of the mean and range of the derived biomass attributes of the plots; $H_{\text{Lorey's}}$ (m): Lorey's height; DBH_{avg} (cm): average DBH; W_f (Mg ha⁻¹): foliage biomass; W_b (Mg ha⁻¹): live branch biomass; W_s (Mg ha⁻¹): trunk wood biomass; W_a (Mg ha⁻¹): total above-ground biomass; W_r (Mg ha⁻¹): root biomass; W_t (Mg ha⁻¹): total biomass.

Variables	<i>Coniferous forest</i> (<i>n</i> = 12)		<i>Broadleaved forest</i> (<i>n</i> = 18)		<i>Mixed forest</i> (<i>n</i> = 23)	
	Range	Mean	Range	Mean	Range	Mean
<i>Biomass-related attributes</i>						
$H_{\text{Lorey's}}$	4.50–14.18	10.54	7.70–18.52	11.96	7.79–14.83	10.63
DBH_{avg}	8.15–20.90	14.19	12.49–22.43	15.19	10.95–20.62	14.22
W_f (foliage)	1.04–23.57	11.69	2.13–8.67	5.12	3.18–19.93	7.89
W_b (branch)	1.62–25.12	12.36	9.72–44.46	19.44	6.42–25.33	13.90
W_s (trunk)	8.35–78.70	48.48	18.65–173.05	72.17	36.61–100.26	67.28
W_a (above-ground)	11.02–127.39	72.52	32.03–219.67	96.76	49.65–141.73	89.07
W_r (root)	6.25–39.42	22.60	10.31–45.62	29.06	15.70–43.05	27.26
W_t (total)	17.27–166.81	95.12	42.34–265.29	125.82	65.35–184.78	116.33
<i>Species composition</i>						
Chinese fir (%)	0–89	29	0	0	0–39	7
Pines (%)	0–90	53	0–29	13	19–52	40
Broadleaved (%)	2–29	18	71–100	87	27–67	53

5.2.1.3. Stand inventory data

An additional set of stand inventory data was available to serve as independent validation data, measured in summer 2012 (by the East China Forestry Investigation Planning and Design Institute, China) following the Chinese National Forest Inventory (CNFI) guidelines. From the approximately 150 stands in the dataset, stands on steep slopes (>30°) dominated by non-timber species (e.g., fruits, bamboo and tea plantations), and close to anthropogenic activities within the forest (such as stands with large roads intersecting the forest), were removed. In addition, if one of the original 53 plots occurred within a stand, that stand was also excluded. The remaining 45 stands were then stratified by age class, site index, and tree species, and again classified as coniferous forest (11 stands), broadleaved forest (15 stands) and mixed forest (19 stands). A summary of selected stand data is shown in Table 5.2.

Table 5.2 Summary of the calculated attributes of the selected stands; W_f (Mg ha⁻¹): foliage biomass; W_b (Mg ha⁻¹): live branch biomass; W_s (Mg ha⁻¹): trunk wood biomass; W_a (Mg ha⁻¹): total above-ground biomass; W_r (Mg ha⁻¹): root biomass; W_t (Mg ha⁻¹): total biomass.

Variables	<i>Coniferous forest</i> (<i>n</i> = 11)		<i>Broadleaved forest</i> (<i>n</i> = 15)		<i>Mixed forest</i> (<i>n</i> = 19)	
	Range	Mean	Range	Mean	Range	Mean
<i>Biomass-related attributes</i>						
W_f (foliage)	2.69–18.06	10.16	3.18–12.52	8.64	4.83–20.76	9.61
W_b (branch)	3.58–18.86	12.08	8.12–29.57	19.48	6.08–22.89	12.96
W_s (trunk)	10.34–67.30	47.32	32.19–135.85	78.75	34.33–95.01	57.40
W_a (above-ground)	16.56–95.26	73.64	54.51–183.45	108.13	50.02–131.73	85.29
W_r (root)	8.31–33.75	21.26	23.04–39.84	29.25	13.84–34.97	25.16
W_t (total)	23.81–146.15	94.18	93.24–216.20	140.96	75.44–160.28	116.37
<i>Species composition</i>						
Chinese fir (%)	0–90	21	0–10	2	0–20	4
Pines (%)	0–100	61	0–30	6	30–50	52
Broadleaved (%)	0–30	19	70–100	94	40–60	47

5.2.1.4. LiDAR data

The small-footprint full-waveform LiDAR data acquired on 17 August 2013 were used in this study (see details about these data in Section 2.2).

5.2.2 LiDAR metrics

LIDAR metrics are descriptive structural statistics calculated from the raw LiDAR point cloud. The metrics for each plot within the 30 × 30 m area of the 53 plots were calculated using the procedures of Lim et al. (2003), Naesset (2002), Naesset (2004) and Ferster et al. (2009), and a summary of the LiDAR metrics with corresponding descriptions is shown in Table 5.3.

LiDAR metrics were generated from first returns (*i.e.*, the first thing hit by a laser pulse, corresponding to the forest canopy) based on research from a number of studies that have found that first returns may have more predictive ability for forest biophysical attributes than all returns (Kim et al., 2009). As a result, percentile heights and canopy return density metrics were computed from first returns with a height of 2 m above-ground-level, reducing potential biases due to low-lying vegetation (Naesset, 2002).

Table 5.3 Summary of LiDAR metrics computed from LiDAR point cloud.

Metrics	Description
Percentile height ($h_5, h_{10}, h_{20}, \dots, h_{95}$)	The percentiles of the canopy height distributions (5th, 10th, 20th . . . 95th) of first returns.
Canopy return density ($d_0, d_1, d_2, \dots, d_9$)	The canopy return density over a range of relative heights, <i>i.e.</i> , percentage (0%–100%) of first returns above the quantiles (0, 10, 20 . . . 90) to total number of first returns.
Mean height (h_{mean})	Mean height above ground of all first returns.
Maximum height (h_{max})	Maximum height above ground of all first returns.
Coefficient of variation of heights (h_{cv})	Coefficient of variation of heights of all first returns.
Canopy cover above 2 meters ($CC_{2\text{m}}$)	Percentages of first returns above 2 m.
Canopy cover above mean (CC_{mean})	Percentages of first returns above the first return mean heights.

5.2.3 Statistical analyses

All of the dependent variables (ground-truth data of the biomass components) and independent variables (LiDAR-derived plot-level metrics) were transformed using the natural logarithm and corrected for bias (Ferster et al., 2009). To ensure that the independent variables were not highly correlated, collinearity was evaluated using Principal Component Analysis (PCA) based on the correlation matrix. Models with condition number (κ) lower than 30 were accepted to ensure that there was no serious collinearity in the selected models (Weisberg, 1985). Individual regression models were developed for all of the biomass components. Stepwise selection was performed to select variables for the final models. Predictor variables were left in the model using an *F*-test with a $p < 0.05$ significance level. Prediction models were assessed using coefficient of determination (R^2), Root-Mean-Square Error (RMSE), which was transformed back to the original scale, and relative RMSE (rRMSE), defined as the percentage of the ratio of RMSE and the observed mean values.

Dummy variables (or class variables) were added to the selected models to assess whether these models differed between forest types (Kutner et al., 2004). Three types of combinations of dependent variables, *i.e.*, Common, Intercept Only and Full, were chosen to evaluate the effects of the intercepts and slopes of the final model fitting. “Common model” implies a reduced model with one common equation regardless of forest type. “Intercept Only” models implies reduced models with common slopes for all forest types, but different intercepts. “Full model” implies both intercepts and slopes

varying by forest type. The partial- F test of Full *versus* Common (FVC) and Full *versus* Intercepts Only (FVIO) was applied to test whether only the intercepts or both intercepts and slopes were different among separated models of different forest types. Cross-validation was used to assess model accuracy. This was an iterative process where one plot was removed from the dataset at a time, and the selected models re-fitted to the $n-1$ remaining plots (Naesset, 2002). Once models were developed, differences between the field-measured stand-level biomass components and the predicted values were evaluated for the 45 selected stands. A 1:1 line was applied to assess accuracy.

5.2.4 Biomass mapping

The second objective of this research was to map forest biomass components using the regression models estimated from the sample plots to improve the understanding of the spatial distribution of each component for the whole site. To do so, a regular grid covering the entire study site was generated, with a cell size of 30×30 m corresponding to the same size of the plots. The selected forest type-specific regression models were used to predict foliage biomass (W_f), live branch biomass (W_b), trunk wood biomass (W_s), above-ground biomass (W_a), root biomass (W_r) and total biomass (W_t) of each grid.

5.3 Results

All the biomass components were fitted against the predictor variables (LiDAR-derived metrics) derived from the first returns. Separate models were built for each of the forest types, *i.e.*, mixed, broadleaved and coniferous forests (Table 5.4). PCA analysis indicated that no serious collinearity problems existed in the models. All of the selected models included less than five predictor variables.

The fit for the common, generic model, for all stands, irrespective of dominant species was relatively low ($R^2 = 0.26\text{--}0.63$). In comparison, the relationship were generally improved for individual models based on dominant species ($R^2 = 0.21\text{--}0.84$). In general, the fitted models had higher correlations in coniferous forest ($R^2 = 0.80\text{--}0.84$) than in other forest types. The results for the biomass components indicated that root biomass (W_r) in coniferous forest ($R^2 = 0.83$) was well predicted, whereas models for root biomass (W_r) in broadleaved forests ($R^2 = 0.53$) and mixed ($R^2 = 0.64$) had a slightly lower level of significance. For the other individual components, namely foliage, branch, and stem biomass, the R^2 values ranged from 0.21–0.83. The poorest correlations were for foliage biomass ($R^2 = 0.21$) in coniferous forests and the highest rRMSE (43.22%) values were also reported for the foliage biomass but in mixed forests. The coefficients of determination for the above-ground and total biomass ranged from 0.58 to 0.84.

Table 5.4 Summary of the biomass components prediction models and plot-level accuracy assessment results; W_f (Mg ha⁻¹): foliage biomass; W_b (Mg ha⁻¹): live branch biomass; W_s (Mg ha⁻¹): trunk wood biomass; W_a (Mg ha⁻¹): total above-ground biomass; W_r (Mg ha⁻¹): root biomass; W_t (Mg ha⁻¹): total biomass.

Dependent	Final models	R ²	RMSE	rRMSE (%)
<i>Common models</i>				
W_f	$\ln W_f = 3.590 + 2.334 \ln h_{cv} + 0.867 \ln h_{25} - 3.021 \ln CC_{mean} + 2.707 \ln d_4$	0.26	4.87	62.44
W_b	$\ln W_b = 1.198 - 0.907 \ln h_{10} + 2.635 \ln h_{25} - 0.633 \ln CC_{mean}$	0.56	4.55	29.47
W_s	$\ln W_s = 2.347 + 1.297 \ln h_{75} - 2.646 \ln CC_{2m} + 2.375 \ln d_2$	0.59	16.96	26.22
W_a	$\ln W_a = 2.464 - 0.634 \ln h_{10} + 1.997 \ln h_{25} - 0.279 \ln CC_{mean}$	0.60	20.74	23.58
W_r	$\ln W_r = 1.713 + 0.432 \ln h_{cv} + 1.036 \ln h_{25}$	0.59	4.85	18.10
W_t	$\ln W_t = 2.803 - 0.625 \ln h_{10} + 1.901 \ln h_{25} - 0.247 \ln CC_{mean}$	0.63	24.40	21.26
<i>Coniferous forest</i>				
W_f	$\ln W_f = -1.892 + 1.976 \ln h_{cv} + 2.902 \ln h_{25} - 3.059 \ln CC_{mean} + 3.055 \ln d_4$	0.80	4.10	35.07
W_b	$\ln W_b = -3.060 - 0.065 \ln h_{10} + 3.048 \ln h_{25} - 0.136 \ln CC_{mean}$	0.81	4.51	36.46
W_s	$\ln W_s = -1.035 + 1.840 \ln h_{75} - 5.718 \ln CC_{2m} + 5.961 \ln d_2$	0.80	10.14	20.93
W_a	$\ln W_a = -0.735 + 0.228 \ln h_{10} + 2.166 \ln h_{25} + 0.076 \ln CC_{mean}$	0.83	18.52	25.54
W_r	$\ln W_r = 0.086 + 0.421 \ln h_{cv} + 1.799 \ln h_{25}$	0.83	4.46	19.74
W_t	$\ln W_t = 0.054 + 0.071 \ln h_{10} + 2.136 \ln h_{25} + 0.026 \ln CC_{mean}$	0.84	22.28	23.42
<i>Broadleaved forest</i>				
W_f	$\ln W_f = 4.980 + 1.227 \ln h_{cv} + 0.327 \ln h_{25} - 4.652 \ln CC_{mean} + 3.685 \ln d_4$	0.21	1.90	37.16
W_b	$\ln W_b = 3.104 + 0.882 \ln h_{10} + 0.173 \ln h_{25} - 0.605 \ln CC_{mean}$	0.71	4.79	24.67
W_s	$\ln W_s = 6.937 + 1.275 \ln h_{75} - 7.328 \ln CC_{2m} + 6.069 \ln d_2$	0.77	18.04	24.99
W_a	$\ln W_a = 3.407 - 0.336 \ln h_{10} + 1.622 \ln h_{25} - 0.471 \ln CC_{mean}$	0.62	24.69	25.52
W_r	$\ln W_r = 2.271 + 0.284 \ln h_{cv} + 0.689 \ln h_{25}$	0.53	4.76	16.37
W_t	$\ln W_t = 3.560 - 0.515 \ln h_{10} + 1.685 \ln h_{25} - 0.382 \ln CC_{mean}$	0.64	27.70	22.02
<i>Mixed forest</i>				
W_f	$\ln W_f = -1.925 + 2.595 \ln h_{cv} + 0.232 \ln h_{25} - 0.643 \ln CC_{mean} + 1.037 \ln d_4$	0.54	3.41	43.22
W_b	$\ln W_b = -1.802 - 1.228 \ln h_{10} + 2.391 \ln h_{25} + 0.398 \ln CC_{mean}$	0.62	4.25	30.60
W_s	$\ln W_s = 1.336 + 1.491 \ln h_{75} - 2.765 \ln CC_{2m} + 2.643 \ln d_2$	0.70	11.28	16.77
W_a	$\ln W_a = 2.295 - 0.451 \ln h_{10} + 1.588 \ln h_{25} - 0.084 \ln CC_{mean}$	0.58	17.10	19.20
W_r	$\ln W_r = 1.444 + 0.295 \ln h_{cv} + 1.091 \ln h_{25}$	0.64	4.27	15.66
W_t	$\ln W_t = 2.481 - 0.428 \ln h_{10} + 1.504 \ln h_{25} - 0.028 \ln CC_{mean}$	0.62	20.73	17.82

Cross-validation of the models indicated that mean differences between the estimated and field-measured values at the 53 sample plots were insignificant (Table 5.5). The standard deviations of the differences ranged from 11.12 to 29.15 Mg ha⁻¹ for above-ground biomass (W_a) and from 16.60 to 24.83 Mg ha⁻¹ for total biomass (W_t). For foliage (W_f), branch (W_b), trunk (W_s), and root biomass (W_r), the standard deviations ranged from 1.44 to 4.78 Mg ha⁻¹ (18.3%–45.7%), 3.25 to 5.87 Mg ha⁻¹ (23.4%–39.4%), 10.16 to 22.43 Mg ha⁻¹ (15.1%–31.1%), and 3.25 to 6.50 Mg ha⁻¹ (14.4%–22.4%), respectively.

Table 5.5 Plot-level validation results: summary of the plot-level observed mean, the mean and standard deviation of the differences of the cross-validation results for the separated models.

Variables (Mg ha ⁻¹)	Coniferous forest (n = 12)			Broadleaved forest (n = 18)			Mixed forest (n = 23)		
	#OM	#MD	#SD	#OM	#MD	#SD	#OM	#MD	#SD
W_f	11.69	-1.09 NS	4.78 (40.9%)	5.12	-0.87 NS	2.34 (45.7%)	7.89	0.48 NS	1.44 (18.3%)
W_b	12.36	-0.51 NS	4.87 (39.4%)	19.44	-1.02 NS	5.87 (30.2%)	13.90	-0.21 NS	3.25 (23.4%)
W_s	48.48	-1.96 NS	11.96 (24.7%)	72.17	-0.42 NS	22.43 (31.1%)	67.28	-1.02 NS	10.16 (15.1%)
W_a	75.52	-0.13 NS	13.91 (18.4%)	96.76	-2.33 NS	29.15 (30.1%)	89.07	-1.76 NS	11.12 (12.5%)
W_r	22.60	-0.94 NS	3.25 (14.4%)	29.06	-1.12 NS	6.50 (22.4%)	27.26	-0.04 NS	3.32 (12.2%)
W_t	95.12	0.02 NS	16.60 (17.5%)	125.82	-4.98 NS	20.76 (16.5%)	116.33	-1.07 NS	24.83 (21.3%)

Level of significance: NS = not significant (>0.05); * $p < 0.05$, ** $p < 0.01$; #OM: The arithmetic mean of the observed value; #MD: Mean of the differences in cross-validation of the selected regression equations in Table 5.4; #SD: Standard deviation of the differences in cross-validation of the selected regression equations in Table 5.4.

The forest type-specific equations were applied to the 45 reference stands to predict the stand-level forest biomass components. The comparisons indicated that 9 of the 12 developed models in coniferous and broadleaved forests did not have significant mean differences between predicted and stand inventory measures ($p < 0.05$), whereas only one model for mixed forests was not significantly different (Table 5.6). For foliage biomass (W_f), a mean difference of 0.24 Mg ha⁻¹ indicated an underestimation by LiDAR in broadleaved stands. In mixed and coniferous stands, the LiDAR-estimated foliage biomass (W_f) overestimated the field-measured values by 1.49–1.52 Mg ha⁻¹. The predicted branch

biomass (W_b) overestimated the ground-truth by 3.01–3.05 Mg ha⁻¹ in mixed and coniferous stands, but underestimated by 0.37 Mg ha⁻¹ in broadleaved stands, although these differences in broadleaved and coniferous stands were insignificant. Stand-measured trunk biomass (W_s) values were significantly overestimated by 3.02–12.50 Mg ha⁻¹, which corresponded to a bias of 5.8%–18.6%, whereas an underestimation of between 9.1% and 17.4% was found for foliage biomass (W_f) in coniferous and mixed stands, and an overestimation of 6.5% for broadleaved stands. Above-ground biomass (W_a) and total biomass (W_t), were the only two models where differences were not significantly different for coniferous and broadleaved forests.

Table 5.6 Stand-level validation results: summary of the observed mean, the mean and standard deviation for the differences in stand predictions using the selected regression equations.

Variables (Mg ha ⁻¹)	Coniferous forest (n = 11)			Broadleaved forest (n = 15)			Mixed forest (n = 19)		
	#OM	#MD	#SD	#OM	#MD	#SD	#OM	#MD	#SD
W_f	7.08	1.52 NS	2.63 (37.1%)	9.25	-0.24 NS	2.15 (22.2%)	6.37	1.49 *	1.88 (29.5%)
W_b	7.83	3.01 NS	3.10 (39.6%)	19.48	-0.37 NS	6.18 (31.7%)	9.57	3.05 *	3.92 (41.0%)
W_s	38.93	10.54 *	12.93 (33.2%)	67.21	12.50 **	32.85 (48.9%)	52.32	3.02 NS	10.50 (20.1%)
W_a	59.48	9.90 NS	10.43 (17.5%)	94.12	15.81 *	31.55 (33.5%)	66.42	8.90 **	12.32 (18.5%)
W_r	24.94	-2.27 NS	2.28 (9.14%)	27.47	1.78 NS	7.43 (27.0%)	29.49	-4.33 **	3.35 (11.4%)
W_t	96.63	10.14 NS	15.70 (16.2%)	141.72	15.20 NS	19.03 (13.4%)	115.47	14.05 *	25.62 (22.2%)

Level of significance: NS = not significant (>0.05); * $p < 0.05$, ** $p < 0.01$; #OM: The arithmetic mean of the observed value; #MD: Mean of the differences between the model-predictions and the field inventory data; #SD: Standard deviation of the differences between the model-predictions and the field inventory data.

Figure 5.1 shows LIDAR-based biomass predictions *versus* the stand-measured biomass estimates in the 45 reference stands. As indicated, overall model fit was good for most of the biomass components with relationships close to the 1:1 line. The mean values of the individual components for each sample plot, illustrated in Figure 5.2, were extrapolated across the entire site using the regression models estimated from the sample plots (Figure 5.3). As expected, relatively low biomass components were found on steep slopes, especially the mountain ridges of both southern and northern slopes. Higher values of biomass components were found in most valleys, especially on the northern slopes where the dominant species prefer cooler summer temperatures and sufficient water supply. In some of the north-central areas, very high value stands of both above-ground and root biomass were associated with

broadleaved forests (e.g., Sawtooth oak). Coniferous-dominated forests contributed high biomass in areas with good site quality (e.g., forest dominated by Masson pine in some valleys with thick soil layers) and relatively flat areas (e.g., stands dominated by Slash pine at the foot of the southern part of the study area). There was a relatively large area of lower biomass, especially root biomass, in the northwest–central part of the study area, reflecting the presence of low-quality Chinese fir plantations on ridge tops with poor site quality and some agroforestry plantations (Chinese fir accompanied by tea) with large canopy gaps and low stem density.

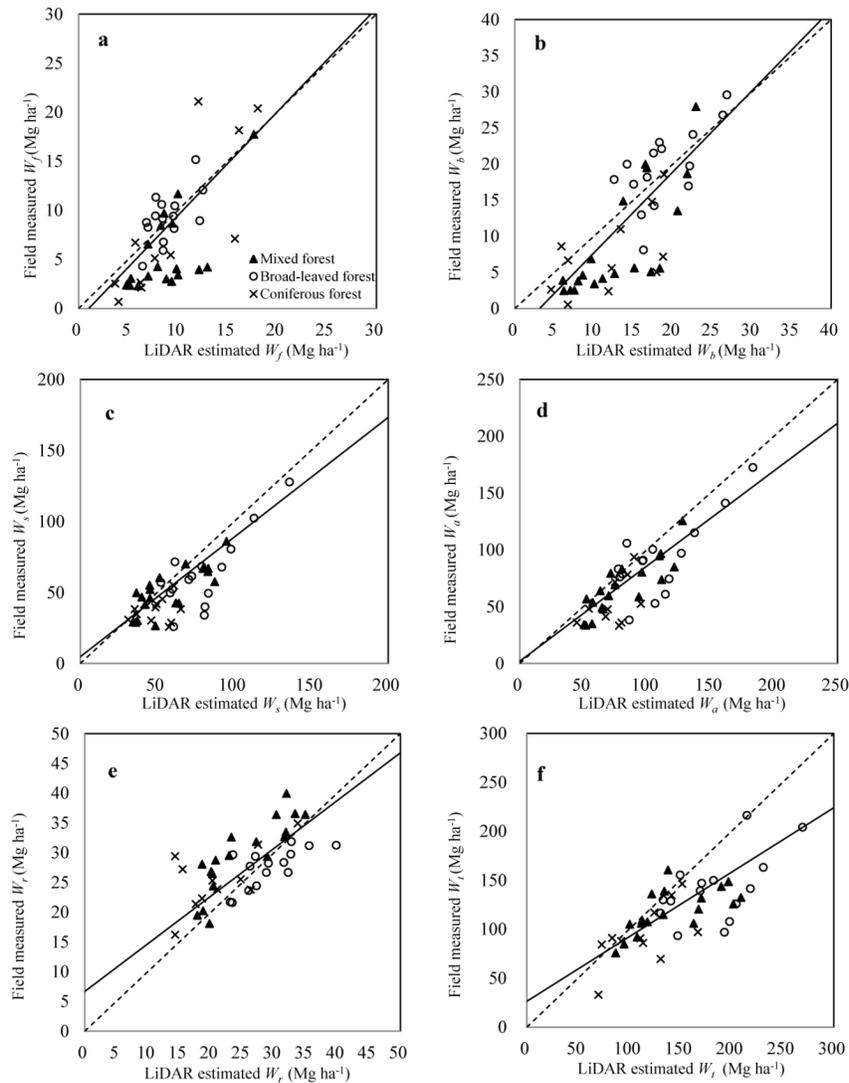


Figure 5.1 Scatterplots and 1:1 line (dotted line) of biomass components between the field-measured and the model-estimated results of the selected stands (i.e., (a) foliage biomass; (b) branch biomass; (c) stem biomass; (d) above-ground biomass; (e) root biomass; (f) total biomass). The models were built from the sample plots of three forest types.

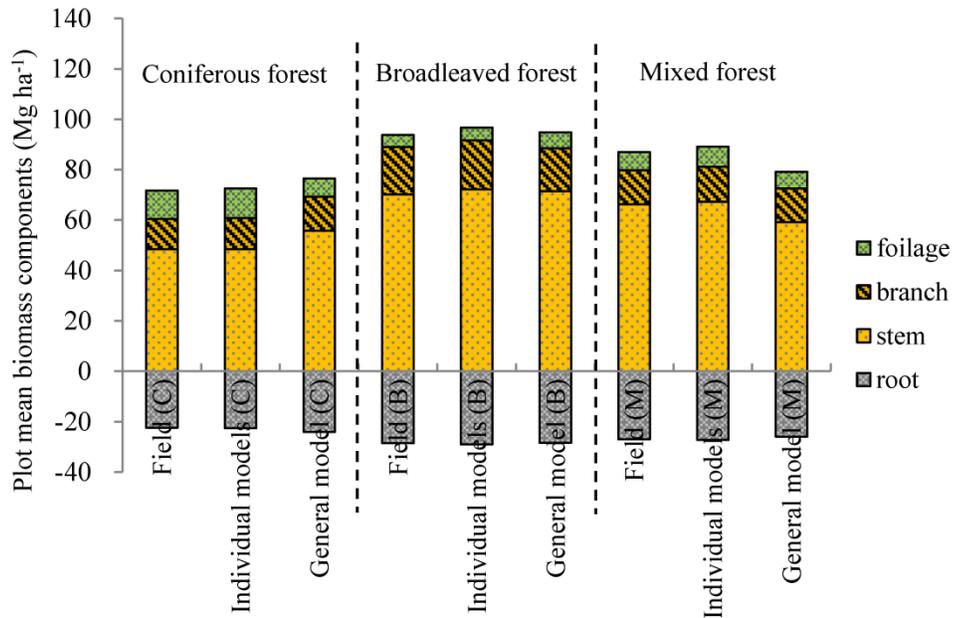


Figure 5.2 Mean value of the biomass components (Mg ha^{-1}) of the 53 sample plots from (1) field measurements; (2) estimations of the dummy variable including stepwise regression model; and (3) predictions of the general models of three forest types, *i.e.*, coniferous, broadleaved and mixed forests.

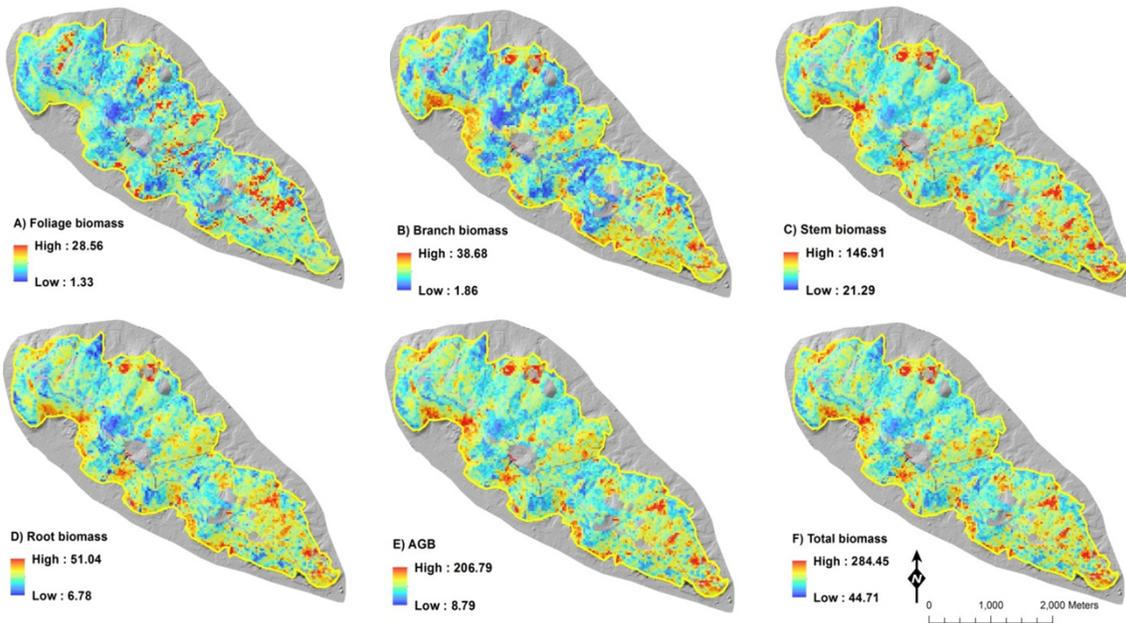


Figure 5.3 Maps of biomass components (Mg ha^{-1}) across the research site (pixel size = 30m), which calculated from the individual models; (A) map of the estimated foliage biomass; (B) branch biomass; (C) stem biomass; (D) root biomass; (E) above-ground biomass (AGB); (F) total biomass.

5.4 Discussion

Dummy variables are numerical variables used in regression analysis, to distinguish different treatment groups and to represent subgroups of the sample (Kutner et al., 2004). The dummy variable technique provides a method for including categorical data into regression analysis, which does not require a relatively large number of observations (Heathington and Isibor, 1972). Næsset and Gobakken (2008) estimated above- and below- ground biomass within the boreal forest zone of Norway. The statistical effects of age class were assessed by including dummy variables in the models. They found that the effect of age class was significant only in the below-ground biomass model. In this study, dummy variables were added to the models to assess whether these models differ between forest types, and the results showed that the effect of forest type was significant in all the biomass components estimation models and the dummy variable-included models ($R^2=0.21-0.84$) generally performed better than the general model ($R^2=0.26-0.63$).

Previous studies have demonstrated that in forests with highly variable species composition, habitat-based models are required to ensure accurate biomass predictions (Naeset, 2002; Fu et al., 2011; Pang and Li, 2012). Likewise in this study, the inclusion of class variables based on dominant species addressed this issue successfully and improved estimations significantly. For foliage, branch, and stem biomass, most of the R^2 values of the individual species models were above 0.54. The coefficients of determination for above-ground and total biomass ranged from 0.58 to 0.84, and the R^2 for root biomass in coniferous forest was as high as 0.83. Lim *et al.* (2003) estimated biomass equations in hardwood forests in Canada by small-footprint LiDAR. The highest proportion of explained variability in the estimated log-transformed linear models was 85%, which is similar to the results in this study. Results in this study indicate that biomass predictions for coniferous stands were more accurate than broadleaved and mixed forest types. In coniferous forests, 80%–84% of the variability could be explained, whereas, 21%–77% and 54%–70% could be explained in broadleaved and mixed forests respectively. Næsset (2004) reported that forest type did not have any significant impact on the estimated biomass models. However, the relatively mixed species composition in this study site led to significant effects of tree-species composition (classified as forest types) on the biomass models.

The dominated coniferous species in this research site are Masson pine and Chinese fir (Table 5.1), which, in this forest, have a relatively even-aged tree structure and higher homogeneous composition, *i.e.*, less variation in tree height and density at the plot level. In contrast, the broadleaved forest dominated by Sawtooth oak and Chinese sweet gum, have more variability in tree height and density

and as a result have more variable LiDAR-derived metrics within plots. Most of the selected models comprised three independent variables (except the foliage biomass models, which include four). All of the models contained at least one variable related to canopy height (h_{cv} , h_{10} , h_{25} , h_{75}) derived from the first returns, indicating sensitivity and high potential for predicting biomass components. LiDAR data allows estimation of the 3-D structural characteristics of the forest canopy, which provides a good foundation for strong relationships between the LiDAR-derived metrics and forest biomass. Since most LiDAR returns are from dominant trees, the LiDAR upper percentile heights (e.g., h_{75}) likely represent the height of overstory trees. The inclusion of variables which describe height variations (e.g., h_{cv}) and lower percentile heights (e.g., h_{10} and h_{25}), accounted for intermediate tree crowns in the midstory and suppressed trees in the understory, therefore potentially improving the characterization of vertical structure. Fu *et al.* (2011) estimated above-ground biomass components using the lower hit density (distance=1.08 m) small-footprint LiDAR data in 78 smaller circle plots (radius = 7.5m or 15m) in southern China. In similar forest types, 43% and 68% of the variability in above-ground biomass was explained within broadleaved and coniferous forests, respectively, but there was no significant correlation between any of the biomass components and the LiDAR-derived metrics in mixed forests. Both the increased plot size and the higher hit density may be reasonable for the more significant correlations found in this study. Studies predicting root (below-ground) biomass from small-footprint LiDAR are rare, especially in subtropical forests stands. Næsset (2004) reported a $R^2 = 0.86$ for below-ground biomass prediction in boreal forests. Pang and Li (2012) fitted regression models for biomass components in temperate forests in northeastern China and explained 85% of the variability in root (below-ground) biomass in conifers, which is higher than reported in this study (53%–83%). The reason for the lower accuracy in this study may be explained by the relatively complicated forest conditions in the study site with root biomass of some less common tree species calculated from generalized allometric equations.

Overall, small-footprint LiDAR has shown strong capability to estimate the amount of biomass components in the study site within the subtropical region of southeast China, and the dummy variable technique is likely to compensate the issue of limited number of field plots and improve the estimations of biomass components. Forest type-specific models performed better than the general model. Results varied between forest types; in this study, the most accurate results were obtained for coniferous stands in relatively homogeneous forest conditions. In addition, some of canopy height metrics appeared to have stronger capabilities for estimating biomass components. The following chapter demonstrates the use of multi-temporal LiDAR to estimate biomass dynamics in the study site.

6 Estimation of Biomass Dynamics using Multi-temporal LiDAR

6.1 Introduction

In this part of the research, the capacity of using multi-temporal airborne LiDAR data, field surveys and statistical models to estimate forest biomass dynamics was assessed over the study site, and the magnitude and spatial patterns of the above ground biomass (AGB) changes were analyzed over a 6-year period. The objectives of this study are: (1) to evaluate the accuracies of direct and indirect approaches for AGB change estimation in a subtropical forest; (2) to extrapolate the change of AGB across the entire study site; and (3) to examine the relationships between the change of AGB and the forest type (classified by dominated tree species) and age (classified by tree age groups) within the forest stands. This research provides new insights as it examines both direct and indirect approaches to predicting change in above-ground biomass. In addition, understanding carbon (and biomass) dynamics across varied forest landscapes (with different forest types and age) is useful for predictive modeling and carbon management, since it connects forest ecosystem processes with landscape-scale carbon pools and fluxes (Hudak et al., 2012). Lastly this study focuses on the prediction and spatial extrapolation of biomass changes in certain forest types within the subtropical forest zone, a relatively poorly studied ecosystem, for forest biomass change estimation.

6.2 Materials and methods

6.2.1 Study area

See details about the study site in Section 2.1.

6.2.2 LiDAR data

LiDAR data were acquired in April, 2007 and August, 2013 over the research site (see details about this data in Section 2.2).

A normalization approach was applied to eliminate any systematic shifts in the heights (z) between the two LiDAR acquisitions (Næsset & Gobakken, 2005) with the 2007 data co-registered to the 2013 data (which has higher point density). A road that transects the entire research area was used for the calibration (see Figure 2.1). A total of 163 circle plots (2m in radius) were sampled along the road center-line at 50 m intervals. All last return ground points (from 2007) were matched to their nearest 2013 ground return using a 0.5 m search window. The mean difference in height between 2007 and 2013 was 10.3 cm (i.e. the 2007 datasets was shifted upwards as compared to the 2013 datasets). As a result the 2007 LiDAR point cloud was shifted by this mean difference, similar to the procedure used by Næsset and Gobakken (2005).

A 1 m digital terrain model (DTM) was created by the last return points from 2013. The data was first filtered to remove the above-ground returns (Kraus and Pfeifer, 1998), and then the DTM was created by calculating the average elevation from the remaining (ground) LiDAR returns within a cell (cells that contain no points were filled by interpolation using neighboring cells). The LiDAR point clouds from both 2007 and 2013 were then normalized against the ground surface height (by DTM) and extracted for each plot using their coordinates.

6.2.3 Plot data

All of the plots sampled between June-August 2012 and August 2013 were used in this study (see details about this data in Section 2.3).

As no field data exists for 2007, tree height and biomass estimations for 2007 were estimated from the field data collected in 2013 using the following approach. In order to compute the change (in AGB) between 2007 and 2013, for all cored trees, species-specific models were developed between DBH increment (measured from the extracted tree cores, as the width of tree ring growth from 2007 to 2013) as a dependent variable, and the 2013 DBH and tree height as independent variables (Du, 1999 and Wyoff, 1990). Dummy variables (or class variables) were added to the models (as the dependent variables) to assess whether models differed among tree species (Kutner et al., 2004). Once developed (Table 6.1), the DBH increment model was then applied to all trees to estimate 2007 DBH. Lastly,

once 2007 DBH was computed, species-specific diameter-height relationships (Table 6.2) developed from the 2013 field collected data were used to estimate height of all trees in 2007.

Table 6.1 The fitted dummy variable models (and accuracy assessments) of DBH growth (2007-2013) for each tree species.

Species group	Names	Predictive models	R^2	RMSE(cm)
Conifers	Masson pine	$\Delta d = 0.967 - 0.047H + 0.034D$ ($n = 142$)	0.658	0.098
	Chinese fir	$\Delta d = 0.642 + 0.003H - 0.008D$ ($n = 89$)		
	Slash pine	$\Delta d = 1.795 - 0.014H + 0.043D$ ($n = 90$)		
Broadleaves	Sawtooth oak	$\Delta d = -0.114 + 0.012H + 0.078\ln(D)$ ($n = 104$)	0.647	0.116
	Sweet gum	$\Delta d = 1.357 - 0.003H + 0.023\ln(D)$ ($n = 79$)		
	Other broadleaves	$\Delta d = 0.418 + 0.008H + 0.051\ln(D)$ ($n = 91$)		

Δd = DBH increment (cm); D = diameter at breast height (cm); H = tree top height (m).

Table 6.2 The fitted dummy variable models by the sample trees (and accuracy assessments) of tree height for each tree species.

Species group	Names	Predictive models	R^2	RMSE
Conifers	Masson pine	$h_t = 4.585 + 0.339D$ ($n = 703$)	0.761	1.283
	Chinese fir	$h_t = 3.345 + 0.438D$ ($n = 508$)		
	Slash pine	$h_t = 7.834 + 0.263D$ ($n = 335$)		
Broadleaves	Sawtooth oak	$h_t = -2.948 + 4.978\ln(D)$ ($n = 452$)	0.734	1.431
	Sweet gum	$h_t = -3.830 + 5.292\ln(D)$ ($n = 417$)		
	Other broadleaves	$h_t = -2.177 + 4.394\ln(D)$ ($n = 405$)		

h_t = tree top height; D = diameter at breast height.

Once a full set of ground data were available for 2007 and 2013, several plot-level forest variables were calculated based on the individual tree data, including average DBH, Lorey's mean height (i.e., the basal area weighted height) and above ground biomass (AGB). AGB was calculated for individual trees within each plot based on the field-measured (2013) or predicted (2007) DBH and height, and then summed to obtain plot-level AGB. Species-specific allometric equations, developed from tree inventory data from local or nearby provinces were selected for this research, for calculating the AGB (Feng et al., 1999; Qian, 2000; Wang and Shi, 1990; Xu et al., 2011).

6.2.4 Stand inventory data

An additional set of forest stand inventory data (2007) was used to examine the LiDAR-estimated biomass changes in relation to forest type and age (see details about these data in Section 2.4)

6.2.5 LiDAR metrics

LiDAR point clouds for 2007 and 2013 were extracted and metrics calculated for each plot using the procedures of Næsset (2002), Hudak et al. (2012) and Bollandsås et al. (2013). These metrics were generated from first returns as they have been found to be more sensitive to changes in the forest canopy (Næsset & Gobakken, 2005). The LiDAR returns below 2 m were omitted to reduce potential biases due to low-lying vegetation (Nilsson, 1996; Næsset & Gobakken, 2005). A summary of the LiDAR metrics with corresponding descriptions is provided in Table 6.3. Differences between the metrics were calculated from the two LiDAR data acquisitions (2007 and 2013) and labeled as “ Δ -metrics” (Bollandsås et al., 2013). As a result, three datasets (each with 22 metrics) were created, i.e., metrics for LiDAR acquisitions in 2007 and 2013, and the Δ -metrics (for the 6 year period).

Table 6.3 Summary of LiDAR metrics for biomass stock and biomass change estimation.

LiDAR metrics	Description
Percentile heights ($h_{10}, h_{20}, h_{30}, \dots, h_{90}$)	The percentiles of the canopy height distributions (10th, 20th, 30th, . . . , 90th)
Canopy return density ($d_0, d_1, d_2, \dots, d_9$)	The canopy return density over a range of relative heights, i.e. percentage (0–100%) of LiDAR returns above the quantiles (0, 10, 20, . . . , 90) to total number of returns
Maximum height (h_{\max})	Height of the highest LiDAR return
Mean height (h_{mean})	Average height of non-ground LiDAR returns
Coefficient of variation of heights (h_{cv})	Coefficient of variation of heights of non-ground LiDAR returns

6.2.6 Predictive modeling and spatial extrapolation

Both direct and indirect approaches were applied for modeling and estimating changes in AGB. For the direct approach, the differences between plot-level metrics calculated from the two LiDAR datasets were directly related to changes in AGB. The linear model of biomass change can be written in the following generalized form (Bollandsås et al., 2013):

$$\Delta W_i = \alpha_0 + \sum_{k=1}^n \beta_k \Delta x_{ki} + \tau_i \quad (6.1)$$

where ΔW_i is the ground estimated change of AGB in the i th plot, α_0 is a constant term, n is the total number of variables in the model, β_k is a coefficient estimated for independent variables, Δx_{ki} is the differences between metrics (or “delta values”) in the i th plot, and τ_i is the error term (Bollandsås & Næsset, 2010).

For the indirect approach, AGB for 2007 and 2013 were predicted using separate models for each period (i.e., 2007 and 2013), and the difference calculated between the two estimates to quantify AGB changes. The general form of the AGB estimation model for single time point measurement is written as:

$$W_i^t = \alpha_0^t + \sum_{k=1}^n \beta_k^t x_{ki}^t + \varepsilon_i^t \quad (6.2)$$

where W_i^t is the ground estimated AGB at one time point ($t =$ the year of 2007 or 2013) in the i th plot and x_{ki}^t is the derived plot-level metrics at one time point (Bollandsås et al., 2013). Finally, the ΔW_i is calculated as the difference between W_i^{2013} and W_i^{2007} .

Linear regression models were fitted separately to predict Lorey’s mean height and AGB at each time point (2007 and 2013) and the changes of AGB. To ensure that the independent variables were not highly correlated, collinearity was evaluated using Principal Component Analysis (PCA) based on the correlation matrix. Models with condition number (κ) lower than 30 were accepted to ensure that there was no marked collinearity in the variables (Weisberg, 1985). Stepwise selection was performed to select variables for the final models. Predictor variables were left in the model using an F -test with

a $p < 0.05$ significance level. Prediction models were assessed using the coefficient of determination (R^2), Root-Mean-Square Error (RMSE) and relative RMSE (rRMSE) defined as the percentage of the ratio of RMSE and the observed mean values. The best fitting models were then selected based on the lowest Akaike information criterion value (Akaike, 1974). Once the best models were chosen, leave-one-out cross-validation was used to validate prediction accuracies of the models.

The spatial prediction of AGB change was calculated as the differences between AGB distributions in 2007 and 2013. For estimating the AGB distribution in 2007 and 2013, regular grids covering the entire study site were generated, with a cell size of 30×30 m corresponding to the size of the plots. In addition, the spatial distribution of AGB change was compared with forest stand inventory data, stratified by the forest type (classified by the dominated tree species within each stand) and age group (Yushan Forest Inventory, unpublished data, 2007).

6.3 Results

LiDAR point cloud (height has been normalized) profiles (5m wide) in 2007 (1.93 points m⁻²) and 2013 (8.37 points m⁻²) within undisturbed forest profiles are shown in Figure 6.1. Six samples of canopy height distributions and hemispherical photos for coniferous and broadleaved dominated plots are shown in Figure 6.2.

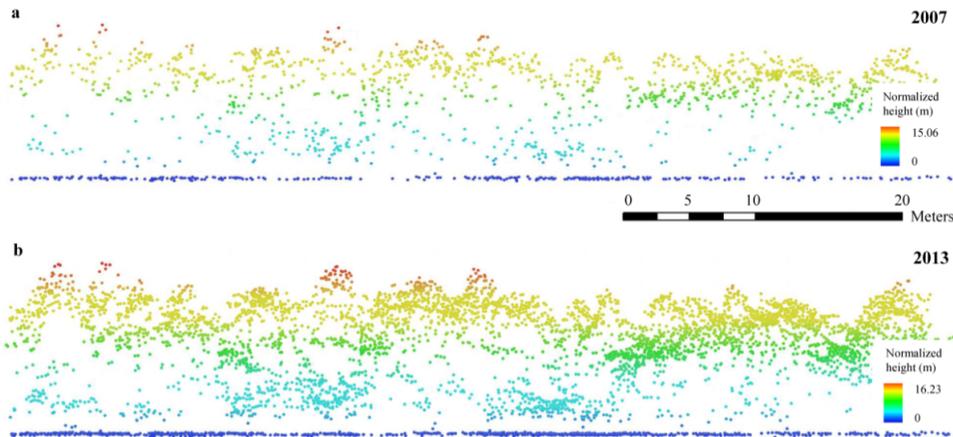


Figure 6.1 LiDAR point cloud profiles (5 m wide) in 2017(a) and 2013(b) through an undisturbed forest. Note that the heights (z) of point clouds have been normalized.

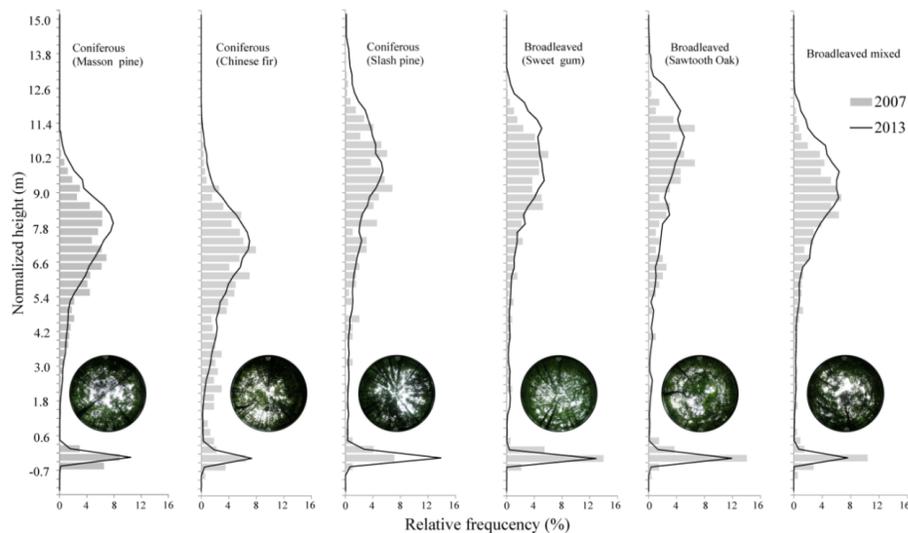


Figure 6.2 Canopy height distributions (vertical interval in 0.3m) calculated by LiDAR first returns in 2007 (gray bars) and 2013 (black line) for six coniferous and broadleaved plots (dominant tree species are shown in the parentheses). The corresponding hemispherical photos are shown in the lower-right for each plot.

The Lorey's mean height (hereafter "height") and AGB estimation models for 2007 and 2013 and for AGB change (i.e. Δ AGB between 2007 and 2013) are shown in Table 6.4. All of the height models have two independent variables and the AGB and Δ AGB models have three. In general, the height models ($R^2 = 0.81$ - 0.84) explained relatively higher levels of variation (R^2) than the AGB and Δ AGB models ($R^2 = 0.67$ - 0.79). Within the biomass estimation models, the Δ AGB model ($R^2 = 0.67$) explained less variability in the response variable than the AGB models in 2007 and 2013 ($R^2 = 0.74$ - 0.79). The canopy height distribution metrics of mean height (h_{mean}), median height (i.e., h_{50}) and delta value of mean height (Δh_{mean}) were most often selected by the models. A small number of metrics of canopy return density from the upper canopy (i.e., d_7 and d_9) were selected for estimating AGB in 2007 and 2013.

Table 6.4 The best models for height, biomass stock and change estimation, and the statistics of accuracy assessments.

Models	Parameters and coefficients	R^2	RMSE	rRMSE (%)	κ
(1) h_L^{2007}	$2.904 + 0.536 \times h_{50} + 0.299 \times h_{95}$	0.81	1.09	10.31	12.78
(2) h_L^{2013}	$0.847 + 0.903 \times h_{50} + 5.176 \times h_{cv}$	0.84	0.97	8.28	5.94
(3) W_{2007}	$4.162 + 11.506 \times h_{\text{mean}} - 71.873 \times d_7 + 183.052 \times d_9$	0.74	12.85	16.94	12.28
(4) W_{2013}	$-18.987 + 12.041 \times h_{50} - 52.057 \times d_7 + 347.596 \times d_9$	0.79	13.48	15.21	11.91
(5) ΔW_{07-13}	$-0.982 + 11.302 \times \Delta h_{\text{mean}} - 2.387 \times \Delta h_{75} + 2.512 \times \Delta h_{95}$	0.67	2.91	22.65	14.04

See Table 6.3 for codes of the LiDAR metrics. h_L^{2007} : Lorey's mean height in 2007; h_L^{2013} : Lorey's mean height in 2013; W_{2007} : above ground biomass in 2007; W_{2013} : above ground biomass in 2013; ΔW_{07-13} : change of above ground biomass (2007-2013); rRMSE: the percentage of RMSE divided by the ground estimated mean; κ : condition number.

Scatter plots of the plot-level ground estimated and LiDAR predicted height, AGB (in 2007 and 2013) and Δ AGB (using both direct and indirect approaches) are shown in Figure 6.4. For both height and AGB estimations in 2007 and 2013, the relationships are close to the 1:1 line. The height ($R^2=0.84$) and AGB ($R^2=0.79$) estimations in 2013 have higher correlations than for the same variables in 2007. Compared with the fitted model for Δ AGB using the indirect approach ($R^2=0.62$, rRMSE=27.55%), the fitted model for Δ AGB using the direct approach ($R^2=0.67$, rRMSE=22.65%) had a relatively higher performance.

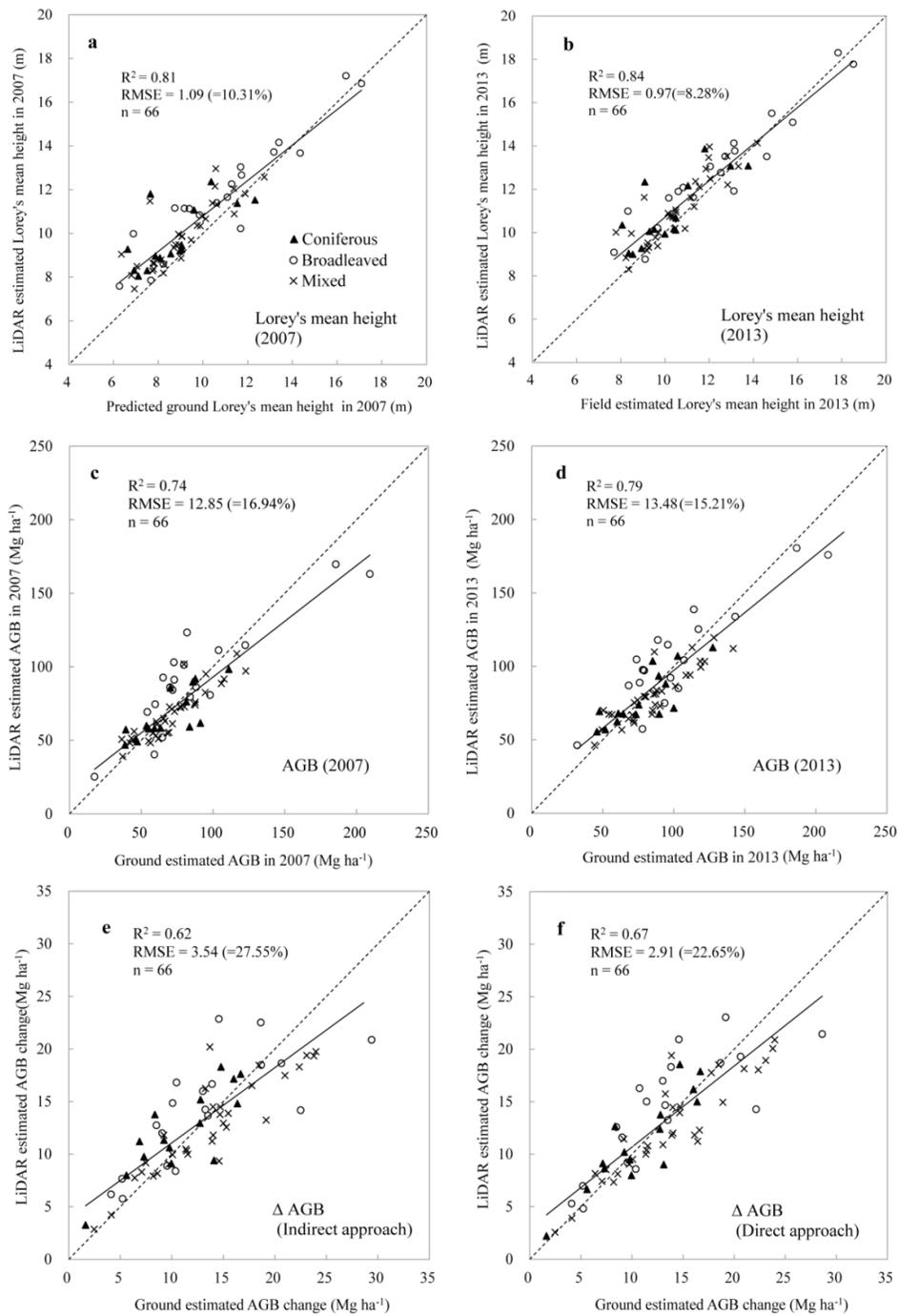


Figure 6.3 Relationship between ground and LiDAR estimated Lorey's mean height in 2007(a) and 2013(b), above ground biomass (AGB) in 2007(c) and 2013(d), and the change of biomass (i.e., Δ AGB) estimated by the indirect approach (e) and direct approach (f).

Cross-validation of the predictive models showed that the mean differences between LiDAR-predicted and ground-estimated height, AGB (in 2007 and 2013) and Δ AGB were not statistically significant in the cross-validated models (Table 6.5). In general, the height models ($CV-R^2 = 0.75-0.80$, $CV-rRMSE = 12.34-14.12\%$) had higher accuracy than the AGB and Δ AGB models ($CV-R^2 = 0.59-0.74$, $CV-rRMSE = 18.12-28.35\%$). Within the biomass estimation models, the AGB models in 2007 and 2013 ($CV-R^2 = 0.68-0.74$, $CV-rRMSE = 18.12-20.48\%$) had higher accuracy than the Δ AGB model ($CV-R^2 = 0.59-0.63$, $CV-rRMSE = 25.64-28.35\%$). Compared with the results of the indirect approach for Δ AGB ($CV-R^2=0.59$, $CV-rRMSE = 28.35\%$) (Table 6.5. Eq. (5)), the direct approach ($CV-R^2 = 0.63$, $CV-rRMSE = 25.64\%$) had a higher accuracy (Table 6.5. Eq. (6)).

Table 6.5 Summary of the cross-validation results for the R^2 , relative RMSE (rRMSE) and differences in cross-validation models for height, biomass stock and change estimation.

Models ^a	Variables	CV- R^2	CV-rRMSE (%)	Differences in cross-validation		
				Mean ^b	Std. Dev.	Range
(1)	h_L^{2007}	0.75	14.12	0.68 NS	0.98	-2.76 – 3.35
(2)	h_L^{2013}	0.80	12.34	0.35 NS	0.87	-1.84 – 2.63
(3)	W_{2007}	0.68	20.48	-1.23 NS	13.95	-46.31 – 30.18
(4)	W_{2013}	0.74	18.12	-1.39 NS	16.14	-48.60 – 34.85
(5)	ΔW_{07-13}^{dir}	0.63	25.64	-0.43 NS	2.75	-7.17 – 6.96
(6)	ΔW_{07-13}^{indir}	0.59	28.35	-0.78 NS	4.32	-9.03 – 8.89

See Table 6.3 for codes of the variables; CV- R^2 : R-square of cross-validation; CV-rRMSE: Relative Root Mean Square Error (RMSE) of cross-validation; Std. Dev.: standard deviation; ΔW_{07-13}^{indir} : the cross validated results of Δ AGB using direct approach; ΔW_{07-13}^{indir} : the cross validated results of Δ AGB using indirect approach. ^a Models: refer to the models (1-5) in Table 6.4; ^b Level of significance: NS = not significant ($p > 0.05$).

Figure 6.4 shows the spatial estimates of Δ AGB distribution (grid size: 30×30m) across the research site, calculated as the difference between the AGBs of 2007 and 2013. Figure 6.5 shows the histograms of AGB for the study site in 2007 and 2013, and the shift and change of AGB between the two time periods. The average Δ AGB was $11.41 \text{ Mg ha}^{-1} \pm 5.39$ (SD = standard deviation) from 2007 ($71.98 \text{ Mg ha}^{-1} \pm 24.24$ (SD)) to 2013 ($83.39 \text{ Mg ha}^{-1} \pm 26.45$ (SD)) (see Figure 6.5.a and b). Figure 6.5.c shows the distribution of AGB shifted from 2007 with a range of 28.74 - 172.92 Mg ha^{-1} to 31.08

- 181.30 Mg ha⁻¹ in 2013, and resulting in a Δ AGB with a range of -14.16 - 41.72 Mg ha⁻¹ (Figure 6.5.d).

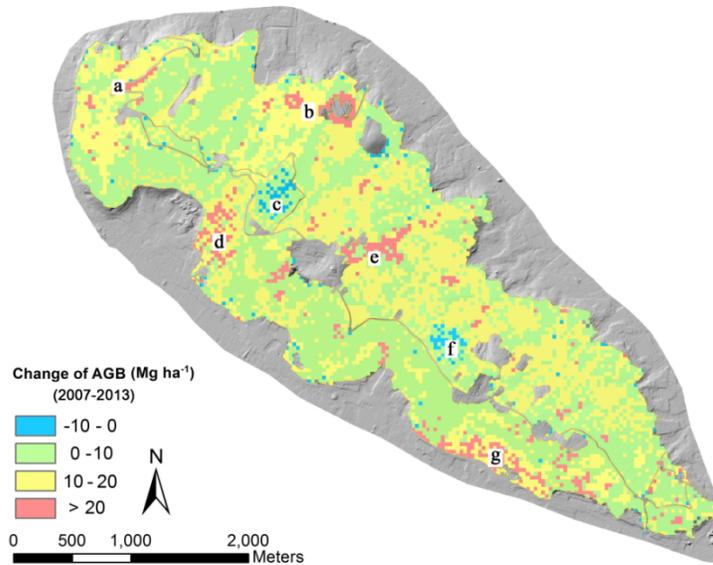


Figure 6.4 The spatial prediction of the changes of AGB across the research site (between 2007 and 2013). The upper-right sub-figure shows the scatterplot between LiDAR estimated AGB in 2007 and in 2013 for all the cells across the research. Some highlighted locations of AGB changes are marked on the map, such as, an increase of AGB in regenerated broadleaved forests (a, b, d, e, g) and a decrease of AGB in Chinese fir/tea agroforestry systems (transformed from the Chinese fir stands after silvicultural treatments) (c, f).

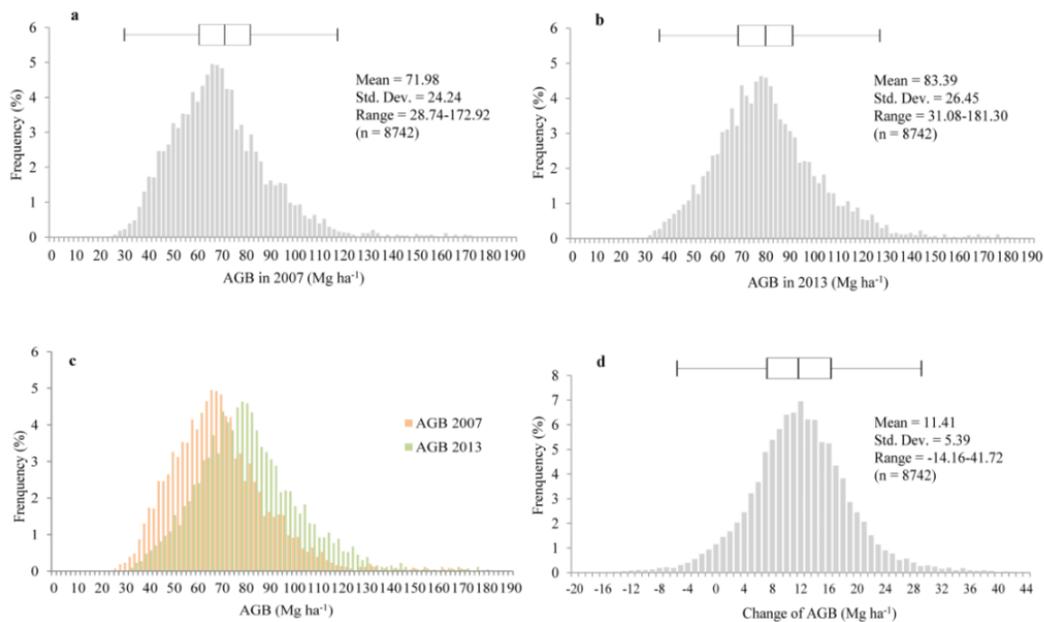


Figure 6.5 Histogram and boxplot of above ground biomass (AGB) distributions across the research site in 2007 (a) and 2013(b), the shift (c) and changes (d) between the two times.

Figure 6.6 shows the LiDAR estimated AGB change for all the stands grouped by age. Figure 6.7 shows LiDAR estimated AGB for all the stands within each forest type in 2007, 2013 and the changes. For the change of biomass (Δ AGB) within age groups, middle-aged stands had the highest gain of AGB (13.76 Mg ha^{-1}), followed by mature (10.31 Mg ha^{-1}) and young (8.86 Mg ha^{-1}) stands. Over-mature stands had the lowest gains of AGB (4.58 Mg ha^{-1}) from 2007 to 2013 (Figure 6.6). Stands dominated by Sawtooth oak had the highest gain of AGB (13.85 Mg ha^{-1}), followed by Slash pine (11.57 Mg ha^{-1}), Masson pine (10.16 Mg ha^{-1}) and other stands dominated by broadleaved trees (10.38 Mg ha^{-1}). Stands dominated by Chinese fir had the lowest gain of AGB (5.73 Mg ha^{-1}) during the 6-year period (Figure 6.7).

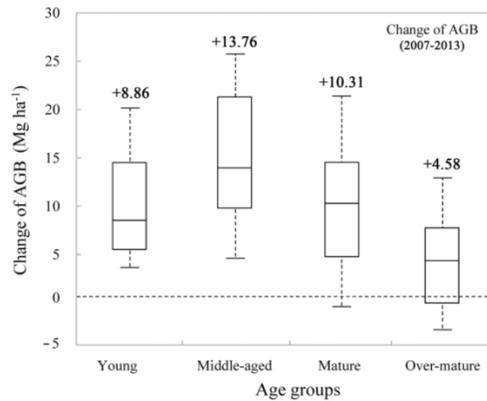


Figure 6.6 The boxplots of LiDAR estimated AGB change for all the stands within each age group. The median values were displayed on top of each boxplot.

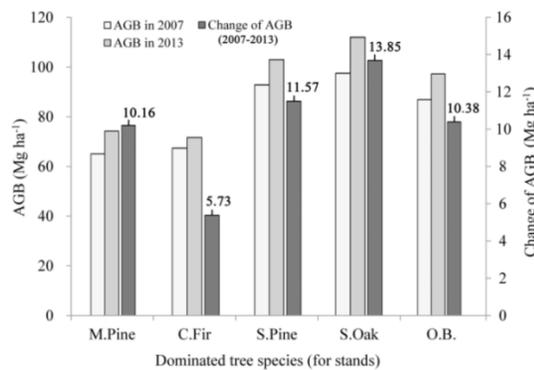


Figure 6.7 The bar graph of the mean of LiDAR estimated AGB for all the stands within each forest type in 2007, 2013 and their changes. Note: M.Pine=Masson pine; C.Fir=Chinese fir; S.Pine=Slash pine; S.Oak=Sawtooth oak; S.Gum=Sweet gum; O.B = Other broadleaves.

6.4 Discussion

For forest biomass change detection studies, it would be desirable to have field-measured data that correspond to LiDAR survey at each point in time. However, in this study, the ground-truth data for AGB in 2007 was estimated by the field-estimated AGB (calculated from the field-measured DBH and height) in 2013. It was anticipated that the species-specific diameter growth models (fitted on the field-measured samples of DBH, DBH increment, tree height and a site factor) and the diameter-height relationships (fitted on the field-measured samples of DBH and height) would give better predictions of the DBH and height in 2007, compared to using general forest growth models from the published literature. The results demonstrating the relatively good performance of Lorey's mean height and AGB predictions in 2007 as shown in Table 6.4 and 6.5 indicated this is likely the case.

In this study, since different sensors were used for repeated LiDAR surveys, the selection of a standard ground reference level is critical when calculating height values (Næsset & Gobakken, 2005). St-Onge and Vepakomma (2004) argued that the same terrain model should be used in multi-temporal LiDAR studies to avoid errors in canopy height changes caused by modelled terrain differences. Yu et al. (2004) also reported that, in certain cases, bias from using different terrain models (on two occasions) was even larger than the real canopy height growth. As a result, in this study, after the height calibration of LiDAR data for 2007, only one terrain model (interpolated by the 2013 LiDAR data, which had higher density) was used for the two LiDAR datasets. With the rapid development of LiDAR sensors, multi-temporal LiDAR datasets are likely to be different in acquisition parameters, such as the differences of point cloud densities between two occasions. In this study, the point density of LiDAR surveys in 2013 (8.37 points m⁻²) was approximately three times greater than in 2007 (1.93 points m⁻²). However, according to the results of the cross-validation (Table 6.5), point density did not significantly affect the AGB and Δ AGB estimates at the plot-scale. This is likely due to the consistent vertical pattern of point clouds (within the canopy structure) at the two points in time (Figure 6.1). Furthermore, due to the relatively large size of the sample plots (i.e., 30×30 m), the two LiDAR

datasets appear to be sufficiently comparable when aggregated into the size of 900m². The 1.93 points m⁻² mean point density in 2007 can be converted to a mean of approximately 1740 points per plot (900 m²), which could be sufficient to produce a stable canopy height distribution for calculating canopy height metrics (Hudak et al., 2012). In this study, the wavelength of LiDAR surveys in 2007 (1064nm with Optech) was slightly different from 2013 (1550nm with Riegl). The two near-infrared wavelengths are both commonly used by commercial LiDAR instruments because they can be transmitted through the atmosphere, are eye-safe, and vegetation typically has high reflectivity in these regions of the spectrum (Lewis and Hancock, 2007). Wichmann et al. (2015) evaluated the performance of multi-spectral LiDAR (with three wavelengths, i.e., 532, 1064 and 1550 nm) for land cover classification in the city of Oshawa (Ontario, Canada). They found a good agreement of the spatial pattern for vegetation between the two LiDAR channels (of 1064 and 1550 nm), and observed no significant difference of in the peaks and distributions of the histograms between these wavelengths for the vegetation classification.

In this study, canopy height metrics such as the delta value of mean height (Δh_{mean}) and some upper percentile heights (i.e., Δh_{75} and Δh_{95}) were chosen in the final AGB change estimation model (Table 6.4.Eq.(5)). Similarly, Næsset and Gobakken (2005) reported that upper percentiles of h_{50} and h_{90} are especially suitable for detecting forest growth, due to height distributions tending to be shifted upwards due to height growth, and the relative portion of the upper canopy increased by the elongation of the branches (St-Onge and Vepakomma, 2004). Maximum canopy height (h_{max}) may be less reliable to detect changes with multi-temporal LiDAR datasets because, due to the lower point density, LiDAR predicted heights in 2007 tend to have more bias (i.e., underestimated the height by missing the tree-tops) than in 2013. The metrics of canopy return densities were only selected in the AGB estimation models at two points in time (Table 6.4. Eq.(3) and (4)), but were not selected for estimating the change of AGB (Table 6.4. Eq.(5)). Similarly, Bollandsås and Næsset (2010) argued that although density metrics have proved useful for modeling biomass in previous studies, the delta value of the density metrics had little effect on the biomass change model. Hudak et al. (2012) also reported that compared with the metric of mean canopy height that has a linear relationship to the

biomass change, the relationship of the biomass change to either maximum canopy height or canopy densities is curvilinear and more scattered.

Canopy height distributions (calculated by LiDAR first returns) in 2007 and 2013 for some sample plots are shown in Figure 6.2. A marked vertical upward shift of the canopy height distributions can be found in broadleaved dominated plots (especially in the Sawtooth oak dominated plot) during the 6-year period. Furthermore, for these broadleaved dominated plots, ground returns decreased from 2007 and 2013, likely due to the horizontal growth of the crowns that reduced the canopy gaps (Næsset & Gobakken, 2005). These changes in pattern can also be found in some conifer dominated plots (especially in the Slash pine dominated plot); they were, however, much less marked than in the broadleaved dominated plots. This is probably because most coniferous forests in the study site are either mature or over-mature and tend to grow more slowly than productive young and middle-aged broadleaved forests (see Figure 6.7 and Table 2.4). In addition, the results in Figure 6.2 also confirm that the first returns are sensitive to the changes of canopy height distribution patterns in the forests. Næsset and Gobakken (2005) found that the first returns, representing the first part of the reflected signal, are more sensitive to changes on the canopy surface, and can therefore be used to explain the variation in forest growth. It should be noted that LiDAR-derived canopy height distributions are also sensitive to changes to factors other than forest growth, for example, thinning and harvesting activities, natural disturbance, seasonal change of undergrowth and the amount of foliage for deciduous trees. Therefore, results should be interpreted with caution if these effects are known to exist in the study area (Yu et al., 2008). In this study, the multi-temporal LiDAR technique used for growth estimation was based on the assumption that the canopy height distribution was only influenced by forest growth. No plots had evidence of significant mortality or silvicultural operations (e.g., thinning) being undertaken during the laser acquisition period.

For estimating the changes of AGB, both direct (i.e., directly predicted biomass change using the differences between LiDAR metrics) and indirect approaches (i.e., modeling biomass for each point in time and predicting the change as their differences) were used in this study. It was found that the model for estimating Δ AGB using the direct approach ($CV-R^2=0.63$, $CV-rRMSE= 25.64\%$) has a

higher accuracy than the indirect approach ($CV-R^2=0.59$, $CV-rRMSE= 28.35\%$), which is consistent with Bollandsås et al. (2013), who postulated that it was because the direct approach was only affected by errors from one model rather than two. Likewise, Næsset et al. (2013a) also reported that the direct approach resulted in more accurate estimates of biomass change in young and mature forest stands. However, these studies were all conducted in boreal forests. This result confirms these conclusions in another type of forests (i.e., the subtropical forest).

The spatial prediction of biomass change shows that, in general, most of the forest area saw an overall gain in biomass. Areas of biomass gain more commonly occurred on northern slopes (i.e., northern side of the central road) than on southern slopes, partially due to the better site quality on northern aspects. In these areas, conifer-dominated forests may contribute higher gains of biomass in areas with good site quality (e.g., stands dominated by Masson pine in some valleys with thick soil layers) and relatively flat areas (e.g., stands dominated by Slash pine at the foot of the mountain). In some of the areas, relatively high biomass gain was associated with broadleaved forests (e.g. stands dominated by middle-aged Sawtooth oak, see Figure 6.4, location: a, b, d e and g). There are two main areas of biomass loss; one along the central ridge of the mountain, reflecting the new development of Chinese fir/tea agroforestry systems by silvicultural treatments (see Figure 6.4, location: c, f). Secondly, some scattered losses (of biomass) can also be found within the pine dominated forests, likely due to selective thinning lowering the risk of pine caterpillar damage (SFMP, 2013).

In general, the middle-aged stands (typically dominated by Sawtooth oak between 41-60 years old) show marked growth in biomass (median $\Delta AGB = 13.76 \text{ Mg ha}^{-1}$ or $2.29 \text{ Mg ha}^{-1} \text{ yr}^{-1}$), followed by mature (typically dominated by Masson pine between 31-60 years old) (median $\Delta AGB = 10.31 \text{ Mg ha}^{-1}$ or $1.72 \text{ Mg ha}^{-1} \text{ yr}^{-1}$) and young stands (typically consist of regenerated broadleaved trees younger than 40 years old) (median $\Delta AGB = 8.86 \text{ Mg ha}^{-1}$ or $1.48 \text{ Mg ha}^{-1} \text{ yr}^{-1}$), whereas over-mature stands (typically dominated by Chinese fir older than 35 years) show the lowest growth (median $\Delta AGB = 4.58 \text{ Mg ha}^{-1}$ or $0.76 \text{ Mg ha}^{-1} \text{ yr}^{-1}$) (Figure 6.6 and Table 2.4). Dubayah et al. (2010) conducted the first LiDAR-based AGB change detection study in a tropical forest (in Costa Rica), and showed that secondary forests were gaining biomass ($3.6 \text{ Mg ha}^{-1} \text{ yr}^{-1}$) but the old growth forest was mostly

neutral ($0.29 \text{ Mg ha}^{-1} \text{ yr}^{-1}$). Their results are comparable in terms of the trends of ΔAGB (at landscape-level) with this study, i.e., the secondary broadleaved forests (at middle age, $\Delta\text{AGB}=2.29 \text{ Mg ha}^{-1} \text{ yr}^{-1}$) showed marked growth, while the productivity of the old forests (at mature and over-mature age, $\Delta\text{AGB}=0.76\text{-}1.47 \text{ Mg ha}^{-1} \text{ yr}^{-1}$) were declining. Meyer et al. (2013) also performed change detection study in the tropics (in central Panama); they found that secondary forests gained $+1.8 \text{ Mg ha}^{-1} \text{ yr}^{-1}$, but the old growth forests lost $-0.7 \text{ Mg ha}^{-1} \text{ yr}^{-1}$ across the entire study site. In this study, most of the losses of AGB were likely due to management activities (e.g. silvicultural treatments), not natural mortality. It should be noted that the over-mature stand of Chinese fir in this study site (Table 2.4) was much younger than the old growth forests in Panama (e.g., $> 400 \text{ yr}$ old, Meyer et al. 2013).

This study has some implications for forest management, both in terms of the development of new tools for management, as well as informing biomass management strategies. The LiDAR-derived data provide a more precise basis for managers to measure, monitor and respond to changes in biomass for the entire forest over relatively short periods of time (i.e. a 6 year period in this study). Also, the spatial analysis allows managers to identify individual stands and patches that may need prioritized management responses, or to alert managers to specific problems at an early stage (if monitoring is undertaken regularly). For example, the Yushan data suggests that managers might focus on replacing the older Masson Pine stands when growth declines (where appropriate) if they are concerned about forest productivity or optimizing carbon sequestration; also, they could consider intensive management (e.g. fertilization) to enhance productivity of suitable younger and middle-aged stands. Optimal management rotations should also be considered when carbon sequestration is an important objective with regard to forest type and age. This tool therefore may help managers to balance the multiple ecosystem services and outputs from the forest and adapt their management plans.

Overall, repeated small-footprint airborne LiDAR can be used to estimate efficiently the amount and distribution of biomass change in subtropical forests over a 6-year period. The direct approach of biomass change modeling improved performance relative to the indirect approach in these forests. The canopy height metrics of delta mean height (Δh_{mean}) and some upper percentile heights (i.e., Δh_{75} and

Δh_{95}), calculated from the LiDAR first returns, tend to be sensitive to the change of biomass, whereas the metrics of maximum canopy height or canopy densities had little effect on the biomass change model. Furthermore, the differences of point cloud densities between the two times did not significantly affect the AGB and Δ AGB estimations at the plot-scale. The map of biomass change showed that biomass gains accounted for the largest area of the study site. The middle-aged stands show marked growth of biomass, followed by mature and young stands, whereas over-mature stands show the lowest growth. This study provides further evidence of the benefits of LiDAR approaches in identifying specific areas for management interventions to enhance forest productivity, maintain biomass stocks and optimize carbon sequestration.

7 Conclusions

Airborne LiDAR data has become increasingly accessible and affordable over the past decade. As the technology continues to develop, it will provide valuable information for managers globally for a variety of forestry applications (Merrick et al., 2013). This dissertation has investigated novel techniques for airborne LiDAR-based assessment of tree species and forest biomass (and its dynamics) in a forest within the subtropical biome. Specifically, airborne LiDAR was evaluated for its ability to:

- discriminate tree species using small-footprint full-waveform LiDAR metrics,
- estimate forest biomass components by discrete-return and full-waveform LiDAR metrics,
- spatially extrapolated the estimation of forest biomass components, and
- predict and map biomass dynamics using multi-temporal LiDAR data.

Four research questions were addressed in this dissertation:

- 1) Can small-footprint full-waveform LiDAR data be used to classify tree species in the subtropical forest stands?
- 2) What are the advantages of estimating the forest biomass components by discrete-return and full-waveform LiDAR metrics alone or combined?
- 3) How can airborne LiDAR be used to provide spatially explicit maps of forest biomass components?
- 4) How can multi-temporal airborne LiDAR data be used to estimate and map biomass dynamics over time?

In Chapter 3, small-footprint full-waveform LiDAR data was used to classify tree species in subtropical forest stands. A local maxima algorithm was applied for tree detection and crown delineation, and full-waveform metrics (extracted from a voxel-based composite waveform) were examined. Using a Random Forests (RF) classifier, six subtropical tree species (i.e., Masson pine,

Chinese fir, Slash pines, Sawtooth oak and Chinese holly) were discriminated at three levels of discrimination. The result showed that the local maxima algorithm was appropriate for delineating individual crowns (detection rate=79%, relative RMSE=5.1% and 11.8% for tree height and crown diameter, respectively), and the use of full-waveform metrics in the Random Forests classifier showed relatively high accuracy (overall accuracy = 68.6% for six classes, 75.8% for four main species and 86.2% for conifers and broadleaved trees) for classifying tree species. The full-waveform metrics (height of median energy (HOME), waveform distance (WD) and number of waveform peaks (NP)) had high classification importance and were shown to be stable among various voxel sizes. These most important metrics can be used for species classification in the subtropics with high accuracy (overall accuracy = 66.1% for six classes, 73.2% for four main species and 84.4% for conifers and broadleaved trees).

In Chapter 4, small-footprint discrete (DR) and full-waveform (FWF) metrics were evaluated to predict biomass and its components. The LiDAR metrics derived from DR and FWF LiDAR data were used alone, and in combination, in stepwise regression models to estimate total as well as above-ground, root, foliage, branch and trunk biomass. Overall, the results indicated that three sets of predictive models performed well across the different subtropical forest types ($Adj-R^2 = 0.42-0.93$, excluding foliage biomass). Forest type-specific models ($Adj-R^2 = 0.18-0.93$) were generally more accurate than the general model ($Adj-R^2 = 0.07-0.79$) with the most accurate results obtained for coniferous stands ($Adj-R^2 = 0.50-0.93$). In addition, LiDAR metrics related to vegetation heights were the strongest predictors of total biomass and the components. For the DR metrics, metrics related to vegetation heights, i.e., mean height, upper height percentiles (i.e. h_{50} , h_{75} , h_{95}) were the strongest predictors of biomass and its components. The FWF metrics were sensitive to the trunk biomass (in coniferous forests) and foliage biomass (in broadleaved forests). This research also illustrates the potential for the synergistic use of DR and FWF LiDAR metrics to accurately assess biomass stocks in the study site.

In Chapter 5, small-footprint discrete-return LiDAR data was used to estimate and map above- and below-ground biomass components of subtropical forests. Foliage, branch, trunk, root,

above-ground and total biomass of 53 plots (30×30 m) were modeled using a range of LiDAR-derived metrics, with individual models for each of the three dominant forest types using stepwise multi-regression analysis. A regular grid covered the entire study site with cell size 30×30 m corresponding to the same size of the plots was used to generate maps for each biomass component. Overall, results indicate that biomass estimation was more accurate in coniferous forests, compared to mixed and broadleaved stands. The coefficient of determination (R^2) for the forest type-specific models was significantly improved compared with an overall generic, or common, model. Using independent stand-level data from existing forest inventory, the results indicated that overall model fit was significant for most of the biomass components, with relationships close to a 1:1 line, indicating no significant bias. This research illustrates the potential for LiDAR as a technology to assess forest carbon accurately and to provide a better understanding of how forest ecosystems function in this region.

In Chapter 6, the capacity of multi-temporal airborne LiDAR data to estimate forest biomass dynamics was assessed, and the magnitude and spatial patterns of AGB change were analyzed over a 6-year period. To do so, direct and indirect approaches to estimate change in AGB were evaluated. Once models were developed, the change in AGB was spatially extrapolated across the study site and examined in relation to forest type and age. The results demonstrate that the direct approach ($CV-R^2 = 0.63$, $CV-rRMSE = 25.64\%$) produced more accurate estimates in change of AGB than the indirect approach ($CV-R^2 = 0.59$, $CV-rRMSE = 28.35\%$). Canopy height metrics of delta mean height (Δh_{mean}) and a number of upper percentile heights (i.e., Δh_{75} and Δh_{95}), calculated from the LiDAR first returns, were found to be most sensitive to changes in biomass, whereas metrics based on maximum canopy height or canopy density had less predictive power. Furthermore, differences in point cloud densities did not significantly impact the estimations of AGB and change of AGB at the 30×30 m plot-scale (in these subtropical forest conditions). Spatial extrapolation of the AGB change indicated that, in general, most of the forest area had an overall gain in biomass. The middle-aged stands show marked growth of biomass (median $\Delta AGB = 2.29 \text{ Mg ha}^{-1} \text{ yr}^{-1}$), followed by mature (median $\Delta AGB = 1.72 \text{ Mg ha}^{-1} \text{ yr}^{-1}$) and young stands (median $\Delta AGB = 1.48 \text{ Mg ha}^{-1} \text{ yr}^{-1}$), whereas over-mature stands show the

lowest growth (median $\Delta\text{AGB} = 0.76 \text{ Mg ha}^{-1} \text{ yr}^{-1}$). This study demonstrates the benefits of using multi-temporal LiDAR data to help identify specific areas for management interventions, to enhance forest productivity, maintain biomass stocks and optimize carbon sequestration.

7.1 Unique outcomes

To the best of my knowledge, this is the first application of full-waveform LiDAR for tree species classification and biomass components estimation in subtropical forest stands, and the first application of multi-temporal LiDAR data to assess biomass dynamics in these forest environments. In particular, the unique outcomes of this dissertation are:

- A new approach to classify tree species by full-waveform metrics and investigate the optimal voxel size for modelling the LiDAR returns.
- A new application of a locally-adapted algorithm for tree detection and crown delineation in the subtropical forest stands.
- A new method to estimate biomass components by using the discrete-return and full-waveform metrics at the plot level.
- Demonstration of a voxel-based approach of LiDAR processing which avoids scan angle effects and to synthesize non-vertical waveforms from different flight trajectories;
- For the first time, the direct and indirect approaches were assessed for estimating biomass dynamics in the subtropical forest stands.

In combination, these LiDAR-based approaches allow forest managers to develop a more comprehensive and robust understanding of the structure, composition and spatial distribution of tree species in the subtropical forest stands. These approaches also provide methodologies and techniques for developing high resolution, spatially explicit estimates of forest biomass (and its dynamics), with applications to sustainable forest management, forest carbon cycling studies and forest carbon accounting projects.

7.2 Limitations

While this dissertation presents novel technologies for assessing tree species and forest biomass (and its dynamics) in subtropical forest stands, there are a number of limitations which can point the way to further research:

1) an incomplete understanding of the full potential of full-waveform (FWF) metrics, requiring, in future, a more direct (physical) interpretation of FWF metrics especially from small footprint systems;

2) a lack of comprehensive understanding on the use of voxel based approaches to derive vertical waveforms, including the generation of composite waveforms which do not have complete Gaussian shapes;

3) a need to refine estimates of root biomass due to less accurate allometric equations than above ground biomass components;

4) a limited number of change detection approaches were tested to estimate change in biomass;

5) a lack of ground truth data in 2007 for biomass change estimation.

In Chapter 3 and 4, most of the full-waveform metrics were adapted from previous large-footprint full-waveform analysis (i.e., SLICER, LVIS and GLAS), however, the potential of small-footprint full-waveform data (for classifying tree species and estimating biomass) remains to be fully explored. A key challenge of using the full-waveform metrics is understanding their direct interpretability (with the tree crown architectural characteristics), which can then be validated in the field at fine spatial scales (Neuenschwander, 2012). In the full-waveform based tree species classification research presented in Chapter 3, the composite waveforms were constructed by directly georeferencing each bin of the raw waveform within the voxels. As a result, “incomplete” composite waveforms were created if the resolution of the voxels were “very high”, or if they were located around the edge of tree boundaries. Although “incomplete” composite waveforms were not used for calculating final metrics, new waveform synthesis approaches (which can create composite waveforms with complete Gaussian

shape) still need exploration.

Species-specific allometric equations developed from tree inventory data from local or nearby provinces (referenced from Feng et al. (1999), Qian (2000), Wang and Shi (1990) and Xu et al. (2011)) were used for calculating plot-scale biomass (i.e., total and above-ground biomass) and biomass components (i.e., stem, root, branch and foliage) in this dissertation. However, it should also be noted that the root biomass of some less common tree species were calculated from generalized allometric equations (in Chapter 4 and 5). The use of generalized equations can lead to a bias in estimating root biomass for a particular species (Clark et al., 2001; Litton et al., 2006). In addition, in Chapter 5, the relationships of LiDAR metrics and field estimated biomass change may vary between forest types, however the impact of differences in forest type (i.e. coniferous, broadleaved and mixed dominated) for estimating biomass change has not been fully evaluated. Likewise only two approaches were compared for biomass change estimation, and the advantages (or disadvantages) of the transformations (e.g., log or squared transformations) for the (independent or dependent) variables (in the predictive models) on the biomass change estimation have not been evaluated.

In this study, as the sample plots were not established in 2007, biomass estimations for 2007 were predicted from the field data collected in 2013 and the estimated forest growth (derived from the DBH increments measured from extracted tree cores). This could introduce some errors into the ground reference data (in 2007). In addition, the height estimations for 2007 were predicted from the estimated DBH (in 2007) and the fitted species-specific diameter-height relationships, could also introduce some errors. Destructive sampling could be considered as a potential source of reference growth measurement techniques if historic data is unavailable and it would likely produce more accurate estimates of forest growth (for height and biomass). It should also be noted that the variability in stand structures present within the Yushan forest study area is not necessarily representative of structural variability within the subtropical forest stands. The regression models developed in this dissertation are designed to demonstrate the potential of this methodology for biomass and biomass change estimations, and should be tested before they are applied more broadly to other locations or subtropical forest types.

7.3 Future directions and recommended implications

Future research will focus on these areas:

Although this research demonstrates significant potential for full-waveform metrics to classify tree species and estimate biomass in subtropical forest biomes, more FWF metrics still need to be explored in order to maximize their potential for future forest studies. The tree crown-based geometric (e.g. echo height distributions, mean and standard deviation of echo widths) and radiometric (e.g. mean intensities and backscatter cross-sections) FWF metrics (following Gaussian decomposition) have been used to tree species classification, and achieved a relative high classification accuracy (overall accuracy = 83%) (Hollaus et al., 2009). Other metrics such as rise time of the leading edge (Neuenschwander, 2012), area under the third Gaussian peak (Nelson et al., 2009) and ground return ratio (calculated as the ground intensity divided by total canopy intensity, used in large footprint LiDAR study) (Drake et al., 2002), were found to be useful for predicting stand volume and biomass. These FWF metrics could also be explored, adapted (from large footprint LiDAR literatures) and implemented at the plot and individual tree scale in the subtropical forests. The exploration of spectral, textural or shape metrics from hyperspectral and high-resolution data (these data were also acquired simultaneously with the full-waveform LiDAR data by the LiCHy Airborne Observation System) could also further extend the range of preferred metrics and enhance the capability of species classification and biomass estimation in the subtropical forests.

Building links between the inherent architectural differences of tree species and how they are manifested in LiDAR metrics remains a critical endeavor for remote sensing scientists and ecologists alike. For example, is the height of median energy (HOME) correlated with tree crown volume density or penetration into the crown? Popescu et al. (2011) demonstrated that the metric of energy penetration index (EPI) (which is calculated as the ratio of ground energy to total waveform energy) was related to canopy cover in large footprint LiDAR systems (i.e., Geoscience Laser Altimeter System (GLAS) onboard the Ice Cloud and land Elevation Satellite (ICESat)), and is a good indicator for characterizing forest structural variables (e.g., mean canopy height) (Neuenschwander et al., 2008).

However, in small footprint LiDAR studies, one footprint (e.g., diameter = 30cm) only describes a portion of the tree structure. As a result, gathering field measurements (potentially, terrestrial laser scanning (TLS) data that can be used as a precise and efficient tool to directly measure detailed 3D models of vegetation elements) and developing relationships between crown structural characteristics and waveform metrics (aggregated or calculated variability within tree crown) is important to improve the direct interpretability of the full-waveform metrics (Neuenschwander, 2012).

Future work should also focus on how to improve the waveform synthesis approach in order to construct composite waveforms (with complete Gaussian shape in any cases). Jung and Crawford (2012) proposed a waveform reconstruction approach, which geo-references the decomposed components and then reconstructs waveforms by combining all Gaussian functions. Their approach included several steps: 1) waveform decomposition using an expectation–maximization (EM) algorithm; 2) georeferencing of the decomposed components using the estimated mean (of the components); 3) waveform reconstruction (using the geo-referenced components for each voxel); 4) and waveform normalization and shifting (using the DTM). After the preliminary FWF data processing, they found that compared with the ground truth data, the metrics (e.g., mean canopy height) extracted from the aggregated (or reconstructed) waveform ($R^2=0.85$) had a higher accuracy than from recorded (or raw) waveform ($R^2=0.68$). As a result, their approach could be adopted in subtropical forest types for creating improved composite waveforms, and may improve the tree species and estimate biomass in future studies.

For the calibration and validation of LiDAR derived biomass, accurate ground reference measurements are needed. In this dissertation, the ground biomass was estimated by allometric equations, which relate biomass to tree diameter and height, based on a theory that assumes that one or more parts of an organism are proportional to the size or growth of other tree parts (Komiyama et al., 2008). The allometric approach can be effective when applied within the productivity and species range of the calibration data, but the species-specific allometric relationships may change between regions and site conditions (Smith and Whelan, 2006), and the creation of site-specific allometric relationships needs intensive fieldwork and destructive sampling (therefore only be conducted on a

limited basis). TLS allows for measuring forest structure with a high degree of detail and accuracy, and has the potential to reduce uncertainties in estimating above ground biomass since it directly estimates tree structure parameters (Calders et al., 2015). Feliciano et al. (2014) used TLS data to estimate AGB of various mangrove plants in the coast of South Florida, United States. They found that the directly estimated AGB (by TLS) showed comparable results with published mangrove allometry, and suggested that TLS based methodology could almost be as accurate as destructive sampling and harvesting technology. Their methodology could be adopted (in Yushan Forest) for creating site-specific allometric equations of the subtropical tree species.

In relation to modelling approaches, studies have found differences in the relationships between LiDAR variables and field estimated biomass such as a study by Drake et al. (2003) who found differences between tropical wet and moist forest. In future studies, dummy variables could be added into the predictive models to assess the differences of forest types (i.e. coniferous, broadleaved and mixed dominated) for biomass change estimation (Kutner et al., 2004).

Srinivasan et al. (2014) assessed the multi-temporal TLS data to model tree biomass change (available for both 2009 and 2012) and field data (only available for 2009). In their approach, AGB for 2012 was estimated by the model developed from field and TLS data in 2009 and then applied to the 2012 TLS data. AGB change was then calculated as the difference between the estimated AGB from the two times. This approach was based on remote sensing paradigm of reducing field work by applying previously developed models to update remote sensing estimates, which also has potential to be adapted to airborne LiDAR studies for future investigation in the subtropical forest stands. Økseter et al. (2015) tested five different approaches for biomass change estimation in a boreal forest: 1) predicting change as the difference between separate AGB predictions; 2) predicting change using the differences of LiDAR metrics; 3) predicting change using the LiDAR metrics from each point in time; 4) predicting change using relative growth rate and 5) predicting change using relative reduction rate. In their research, several different transformations were applied to the independent variables, and the natural logarithm (LN) transformations used for the dependent variables in the indirect approaches. They found that the use of LN transformation (for dependent variables) increased the accuracy of

biomass estimation for each point in time (relative RMSE decreased from 14.8% to 18.0%). Future studies could test more approaches in biomass change estimation and evaluate the advantages (or disadvantages) of the transformations for the model variables on the biomass change estimation.

Previous studies have also shown that the relationships between biomass and LiDAR metrics are dependent on forest type (Drake et al., 2003; Asner et al., 2012; Naesset and Gobakken, 2008), which means that the biomass estimation models developed in this study may not be directly transferred to other study areas. Some previous studies have shown promising results transferring between forest types using power models based on mean canopy height (MCH) to predict biomass across large geographic extent. Lefsky et al. (2002) found that 84% of biomass variations across three sites in North America (a temperate conifer forest in Oregon, USA, a temperate deciduous forest in Maryland, USA, and a boreal conifer forest in Manitoba, Canada) can be explained using a model based on the squared MCH. Asner et al. (2012) found that plot-scale biomass across four tropical forest sites in Panama, Peru, Madagascar, and Hawaii can be estimated by LiDAR-based MCH, after accounting for the differences of wood density and basal area (BA) to MCH relationships for each site. More research is needed to find reliable and stable LiDAR metrics for creating transferable biomass estimation models as well as an ability to assess model uncertainties.

In addition to these future directions, this dissertation does provide some recommended implications for using LiDAR in forest management, particularly in the sub-tropical forest biome:

To support sustainable forest management, high spatial resolution forest structure information is required to accurately assess the species composition, structure, and distribution of forest stands that, in turn, can be used as base information for management decisions across a range of spatial and temporal scales (Wulder et al., 2008). Airborne LiDAR has emerged as one of the most promising remote sensing technologies, providing detailed, spatially explicit, three-dimensional information on forest structure, for operational applications in a wide range of disciplines related to the management of forest ecosystems (Vauhkonen et al., 2014). With increasing availability of LiDAR data, forest managers have seen opportunities for using LiDAR to meet a wider range of forest information needs

(Nelson et al., 2003). For instance, this dissertation has shown promising results that LiDAR can be effectively used to classify tree species and estimate biomass (and biomass change) with relatively high accuracy in the study site. The tree species information (predicted by LiDAR) can be helpful for forest management, because the development and use of species-specific growth and yield models or to inform a range of forest treatment schedules that are highly dependent on species (Vauhkonen et al., 2014). The LiDAR estimated biomass (and biomass change) can help identify specific areas for management interventions, to enhance forest productivity, maintain biomass stocks and optimize carbon sequestration.

The costs associated with LiDAR data will likely continue to be relatively high (compared to other remote sensing data), as the regular survey costs (e.g., fuel, mobilization and aircraft ferrying charges) need to be borne, and the instrument costs need to be amortized (Wulder et al., 2008). At this time, it may be advisable for forest managers to cooperate with other interested agencies to share the cost of LiDAR (for multiple use), making it feasible to apply LiDAR on a regular basis. Full-waveform (FWF) airborne LiDAR sensors and data have become increasingly available since 2004 (Mallet and Bretar, 2009). As there is no significant difference in the acquisition costs between discrete and FWF data, more contractors are promoting FWF data because of its perceived additional benefits (Hollaus et al., 2014). FWF systems have two important advantages with regard to the additional investigation possibilities, i.e., point cloud density can be enhanced by processing the full FWF return signal and additional metrics can be extracted by modelling the received waveforms. In the past decade, the additional information from FWF data has shown significant potential to forest management, for example in the estimation of forest structure parameters (Sumnall et al., 2012; Neuenschwander, 2012), single tree detection (Reitberger et al., 2008; Gupta et al., 2010) and tree species classification (Xu et al., 2012). This dissertation (Chapter 3 and 4) has also shown that the FWF metrics have strong capabilities for tree species classification and biomass estimation in the subtropical forest stands. Using appropriately designed surveys, multi-temporal LiDAR observations enable measurement of forest growth over time (Yu et al., 2008, 2006, 2004; Næsset and Gobakken, 2005). A change in tree height, however, can only be measured if the height increase is greater than any bias in the LiDAR

measurements. Bias can include instrument specifications, flying height, species architecture and so on. (Wulder et al., 2008). Hopkinson et al. (2008) analyzed the relationship between time interval and the standard deviation of error in growth estimates, and found that over a one-year period, the growth uncertainty was in the range of 0.3 m (around 100% of total growth), but reducing to less than 0.1 m (around 6% of total growth) after 5 years. For operational conifer plantation growth estimates, the acceptable uncertainty (10%) can be achieved within a three-year time interval. As a result this approach could be adopted (in the subtropics) for determining the minimum repeat acquisition time interval for an accurate and statistically significant estimate of forest growth in the future.

References

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi:10.1109/TAC.1974.1100705
- Allouis, T., Durrieu, S., Vega, C., Coueron, P., 2013. Stem volume and above-ground biomass estimation of individual pine trees from LiDAR data: Contribution of full-waveform signals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 6, 924–934. doi:10.1109/JSTARS.2012.2211863
- Andersen, H.E., McGaughey, R.J., Reutebuch, S., 2005. Estimating forest canopy fuel parameters using LIDAR data. *Remote Sens. Environ.* 94, 441–449. doi:10.1016/j.rse.2004.10.013
- Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J.S., van Breugel, M., 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* 168, 1147–1160. doi:10.1007/s00442-011-2165-z
- Asner, G.P., Powell, G.V.N., 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. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16738–16742. doi:10.1073/pnas.1004875107
- Bater, C., Wulder, M.A., Coops, N.C., Nelson, R., Hilker, T., Næsset, E., 2011. Stability of sample-based scanning-LiDAR-derived vegetation metrics for forest monitoring. *Trans. Geosci. Remote Sens.* 49, 2385–2392.
- Bater, C.W., Coops, N.C., 2009. Evaluating error associated with lidar-derived DEM interpolation. *Comput. Geosci.* 35, 289–300. doi:10.1016/j.cageo.2008.09.001
- Biging, G.S., Dobbertin, M., 1992. A comparison of distance dependent competition measures for height and basal area growth of individual conifer trees. *For. Sci.* 695–720.
- Bollandsås, O.M., Gregoire, T.G., Næsset, E., Øyen, B.-H., 2013. Detection of biomass change in a Norwegian mountain forest area using small footprint airborne laser scanner data. *Stat. Methods Appl.* 22, 113–129. doi:10.1007/s10260-012-0220-5
- Bollandsås, O.M., Næsset, E., 2010. Using delta values of multi-temporal first-return small footprint airborne laser scanner data to predict change of tree biomass in mountain spruce forests ., in: *Silvilaser 2010*. Freiburg, Germany, pp. 411–417.
- Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leafon conditions using airborne lidar. *ISPRS J. Photogramm. Remote Sens.* 61, 325–340.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.

- Bretar, F., Pierrot-Deseilligny, M., Roux, M., 2004. Solving the Strip Adjustment Problem of 3D Airborne Lidar Data, in: IEEE International Geoscience and Remote Sensing Symposium. Anchorage, AK, pp. 4734–4737.
- Calders, K., Newnham, G., Burt, A., Murphy, S., Raunonen, P., Herold, M., Culvenor, D., Avitabile, V., Disney, M., Armston, J., Kaasalainen, M., 2015. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods Ecol. Evol.* 198–208.
- Cao, L., Coops, N.C., Innes, J.L., Dai, J., She, G., 2014. Mapping Above- and Below-Ground Biomass Components in Subtropical Forests Using Small-Footprint LiDAR. *Forests* 5, 1356–1373. doi:10.3390/f5061356
- Carle, J., Holmgren, P., 2003. *Definitions Related to Planted Forests*. Rome.
- Chan, J.C.W., Paelinckx, D., 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* 112, 2999–3011.
- Chauve, a., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Pierrot Deseilligny, M., Puech, W., 2009. Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest. *Int. J. Remote Sens.* 30, 5211–5228. doi:10.1080/01431160903023009
- Clark, D. a, Sandra, B., David, K.W., Chambers, J.Q., Thomlinson, J.R., Ni, J., 2001. *Measuring Net Primary Production in Forests : Concepts and Field Methods*. *Ecol. Appl.* 11, 356–370.
- Clark, M., Roberts, D., Ewel, J., Clark, D., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sens. Environ.* 2931–2942.
- Clark, M.L., Clark, D.B., Roberts, D. a, 2004. Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sens. Environ.* 91, 68–89. doi:10.1016/j.rse.2004.02.008
- Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G.J., Siggins, A., Trofymow, J.A.T. a (Tony), 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees-Structure Funct.* 21, 295–310. doi:10.1007/s00468-006-0119-6
- Dixon, R.K., Brown, S., Houghton, R.A., Solomon, A.M., Trexler, M.C., Wisniewski, J., 1994. Carbon pools and flux of global forest ecosystems. *Science (80-)*. 263, 185–190. doi:10.1126/science.263.5144.185
- Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M. a, Chazdon, R.L., Weishampel, J.F., Prince, S., 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sens. Environ.* 79, 305–319. doi:10.1016/S0034-4257(01)00281-4
- Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., Hofton, M., 2003. Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Glob. Ecol. Biogeogr.* 12, 147–159. doi:10.1046/j.1466-822X.2003.00010.x
- Du, J., 1999. Modeling Individual Tree Growth of Larix by Using ForestManagement Inventory Plots. *For. Res.* 12, 160–164.

- Du, L., Zhou, T., Zou, Z., Zhao, X., Huang, K., Wu, H., 2014. Mapping Forest Biomass Using Remote Sensing and National Forest Inventory in China. *Forests* 5, 1267–1283. doi:10.3390/f5061267
- Dubayah, R., Drake, J., 2000. Lidar remote sensing for forestry. *J. For.* 98, 44–46.
- Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M. a., Blair, J.B., Hurtt, G.C., Chazdon, R.L., 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J. Geophys. Res.* 115, G00E09. doi:10.1029/2009JG000933
- Duong, H. Van, 2010. Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data. Delft University of Technology.
- Englhart, S., Jubanski, J., Siegert, F., 2013. Quantifying Dynamics in Tropical Peat Swamp Forest Biomass with Multi-Temporal LiDAR Datasets. *Remote Sens.* 5, 2368–2388. doi:10.3390/rs5052368
- Falkowski, M., Hudak, A., Crookston, N., Gessler, P., Smith, A., 2010. Landscape-scale parameterization of a tree-level forest growth model: A k-NN imputation approach incorporating LiDAR data. *Can. J. Remote Sens.* 40, 184–199.
- Fang, J., Chen, A., 2001. Dynamic forest biomass carbon pools in China and their significance. *Acta Bot. Sin.* 43, 967–973.
- FAO, 2001. Global Forest Resources Assessment 2000 - Main report. Rome.
- Feliciano, E.A., Wdowinski, S., Potts, M.D., 2014. Assessing Mangrove Above-Ground Biomass and Structure using Terrestrial Laser Scanning: A Case Study in the Everglades National Park. *Wetlands*. doi:10.1007/s13157-014-0558-6
- Feng, Z.W., Wang, X.K., Wu, G., 1999. Biomass and Productivity of Chinese Forest Ecosystem, 1st ed. Science Press, Beijing.
- Ferster, C.J., Coops, N.C., Trofymow, J.A., 2009. Aboveground large tree mass estimation in a coastal forest in British Columbia using plot-level metrics and individual tree detection from lidar. *Can. J. Remote Sens.* 35, 270–275.
- Fleming, R., Kanowski, P., Brown, N., Jenik, J., Kahumbu, P., Plesnik, J., 2011. Emerging Perspectives on Forest Biodiversity, in: *UNEP Year Book 2011*. pp. 46–59.
- Franco, M., Kelly, C.K., 1998. The interspecific mass-density relationship and plant geometry. *Proc. Natl. Acad. Sci. USA* 7830–7835.
- Franklin, S.E., 1994. Discrimination of subalpine forest species and canopy density using CASI, SPOT, PLA, and Landsat TM data. *Photogramm. Eng. Remote Sens.* 60.
- Franklin, S.E., Hall, R.J., Moskal, L.M., Maudie, A.J., Lavigne, M.B., 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. *Intern. J. Remote Sens.* 21, 61–79.

- Fu, T., Pang, Y., Huang, Q.F., Liu, Q.W., Xu, G.C., 2011. Prediction of subtropical forest parameters using airborne laser scanner. *J. Remote Sens.* 15, 1092–1104.
- Gatziolis, D., Andersen, H.E., 2008. A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest, General Technical Report. United States Department of Agriculture, Forest Service, Portland.
- Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environ. Res. Lett.* 2, 1–13.
- Givnish, T.J., 1986. Biomechanical constraints on self-thinning in plant populations. *J. Theor. Biol.* 139–146.
- Gupta, S., Weinacker, H., Koch, B., 2010. Single tree detection using full waveform laser scanner data, in: *Silvilaser 2010*. Freiburg, Germany, pp. 368–376.
- Harding, D.J., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophys. Res. Lett.* 32, L21S10. doi:10.1029/2005GL023471
- Hauglin, M., Dibdiakova, J., Gobakken, T., Næsset, E., 2013. Estimating single-tree branch biomass of Norway spruce by airborne laser scanning. *ISPRS J. Photogramm. Remote Sens.* 79, 147–156. doi:10.1016/j.isprsjprs.2013.02.013
- Heathington, K., Isibor, E., 1972. The use of dummy variables in trip generation analysis. *Transpn. Res.* 6, 131-142.
- Heinzel, J., Koch, B., 2011. Exploring full-waveform LiDAR parameters for tree species classification. *Int. J. Appl. Earth Obs. Geoinformation* 13, 152–160.
- Hermosilla, T., Coops, N.C., Ruiz, L.A., Moskal, L.M., 2014. Deriving pseudo-vertical waveforms from small-footprint full-waveform LiDAR data. *Remote Sens. Lett.* 5, 332–341.
- Hermosilla, T., Ruiz, L.A., Kazakova, A.N., Coops, N.C., Moskal, M., 2014. Estimation of forest structure and canopy fuel parameters from small-footprint full-waveform LiDAR data. *Int. J. Wildl. Fire* 23, 224–233.
- Höfle, B., Hollaus, M., Hagenauer, J., 2012. Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* 67, 134–147. doi:10.1016/j.isprsjprs.2011.12.003
- Höfle, B., Hollaus, M., Lehner, H., Pfeifer, N., Wagner, W., 2008. Area-based parameterization of forest structure using full-waveform airborne laser scanning data, in: *SilviLaser 2008, the 8th International Conference on LiDAR Applications in Forest Assessment and Inventory*. Edinburgh, Scotland, UK, pp. 229–235.
- Hollaus, M., Mücke, W., Höfle, B., Dorigo, W., Pfeifer, N., Wagner, W., 2009. Tree species classification based on full-waveform airborne laser scanning data, in: *SilviLaser 2009*. College Station, Texas, USA, pp. 54–62.

- Hollaus, M., Mücke, W., Roncat, A., Pfeifer, N., Briese, C., 2014. Full-Waveform Airborne Laser Scanning Systems and Their Possibilities in Forest Applications, in: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), *Forestry Applications of Airborne Laser Scanning Concepts and Case Studies*. Springer, New York London, pp. 43–61.
- Holmgrena, J., Persson, A. sa, 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sens. Environ.* 90, 415–423.
- Hopkinson, C., Chasmer, L., Hall, R.J., 2008. The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sens. Environ.* 112, 1168–1180. doi:10.1016/j.rse.2007.07.020
- Hopkinson, C., Chasmer, L.E., Sass, G., Creed, I.F., Sitar, M., Kalbfleisch, W., Treitz, P., 2005. Vegetation class dependent errors in lidar ground elevation and canopy height estimates in a boreal wetland environment. *Can. J. Remote Sens.* 31, 191–206. doi:10.5589/m05-007
- Hudak, A.T., Strand, E.K., Vierling, L. a., Byrne, J.C., Eitel, J.U.H., Martinuzzi, S., Falkowski, M.J., 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sens. Environ.* 123, 25–40. doi:10.1016/j.rse.2012.02.023
- Hyypä, J., Yu, X., Hyypä, H., Vastaranta, M., Holopainen, M., Kukko, A., Kaartinen, H., Jaakkola, A., Vaaja, M., Koskinen, J., Alho, P., 2012. Advances in Forest Inventory Using Airborne Laser Scanning. *Remote Sens.* 4, 1190–1207. doi:10.3390/rs4051190
- Jones, T.G., Coops, N.C., Sharma, T., 2010. No Title. *Remote Sens. Environ.* 114, 2841–2852.
- Jung, J., Crawford, M.M., 2012. Extraction of features from LIDAR waveform data for characterizing forest structure. *IEEE Geosci. Remote Sens. Lett.* 9, 492–496. doi:10.1109/LGRS.2011.2172769
- Kankare, V., Rätty, M., Yu, X., Holopainen, M., Vastaranta, M., Kantola, T., Hyypä, J., Hyypä, H., Alho, P., Viitala, R., 2013. Single tree biomass modelling using airborne laser scanning. *ISPRS J. Photogramm. Remote Sens.* 85, 66–73.
- Kankare, V., Vastaranta, M., Holopainen, M., Rätty, M., Yu, X., Hyypä, J., Hyypä, H., Alho, P., Viitala, R., 2013. Retrieval of Forest Aboveground Biomass and Stem Volume with Airborne Scanning LiDAR. *Remote Sens.* 5, 2257–2274. doi:10.3390/rs5052257
- Kant, S., Berry, R.A., 2005. *Institutions, Sustainability, and Natural Resources: Institutions for Sustainable Forest Management*. Springer Netherlands.
- Key, T., Warner, T.A., McGraw, J.B., Fajvan, M.A., 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. *Remote Sens. Environ.* 75, 100–112.
- Kim, Y., Yang, Z.Q., Cohen, W.B., Pflugmacher, D., Lauver, C.L., Vankat, J.L., 2009. Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sens. Environ.* 113, 2499–2510. doi:10.1016/j.rse.2009.07.010
- King, D., Loucks, O.L., 1978. The theory of tree bole and branch form. *Radiat. Environ. Biophys.* 141–165.

- Koch, B., Kattenborn, T., Straub, C., Vauhkonen, J., 2014. Segmentation of Forest to Tree Objects, in: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), *Forestry Applications of Airborne Laser Scanning*. Springer, Dordrecht Heidelberg New York London, p. 460.
- Komiyama, A., Ong, J., Pongpan, S., 2008. Allometry, biomass, and productivity of mangrove forests: a review. *Aquat. Bot.* 89, 128–137.
- Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* 53, 193–203. doi:10.1016/s0924-2716(98)00009-4
- Kutner, M., Nachtsheim, C., Neter, J., Li, W., 2004. *Applied Linear Statistical Models*, 5th ed. McGraw-Hill/Irwin, New York, NY, USA.
- Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *For. Ecol. Manage.* 95, 209–228.
- Larson, P.R., 1963. *Stem Form Development of Forest Trees*. The Society of American Foresters, Washington D.C.
- Lefsky, M. a., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *Bioscience* 52, 19–30. doi:10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sens. Environ.* 70, 339–361. doi:10.1016/s0034-4257(99)00052-8
- Lefsky, M.A., Cohen, W.B., Harding, D.J., 2002. Lidar remote sensing of above-ground biomass in three biomes. *Glob. Ecol. Biogeogr.* 393–399.
- Lefsky, M.A., Hudak, A.T., Cohen, W.B.W.B., Acker, S.A., 2005. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sens. Environ.* 95, 517–531. doi:10.1016/j.rse.2005.01.004
- Li, D.L., Li J., R., Wang, B.S., Wang, X.H., 2006. Analysis of vegetation types and woody plants flora in Yushan of Changshu. *J. Zhejiang For. Coll.* 23, 46–51.
- Li, W., Guo, Q., Jakubowski, M.K., Kelly, M., 2012. A new method for segmenting individual trees from the LiDAR point cloud. *Photogramm. Eng. Remote Sensing* 78, 75–84.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2, 18–22.
- Lieth, H., Whittaker, R.H., 1975. *Primary Productivity of the Biosphere*. Springer, New York.
- Lim, K., Treitz, P., Baldwin, K., Morrison, I., Green, J., 2003. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Can. J. Remote Sens.* 29, 658–678.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. *Prog. Phys. Geogr.* 27, 88–106. doi:10.1191/0309133303pp360ra

- Lim, K.S., Treitz, P.M., 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scand. J. For. Res.* 19, 558–570.
doi:10.1080/02827580410019490
- Lindberg, E., Olofsson, K., Holmgren, J., Olsson, H., 2012. Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data. *Remote Sens. Environ.* 118, 151–161.
doi:10.1016/j.rse.2011.11.015
- Litton, C.M., Sandquist, D.R., Cordell, S., 2006. Effects of non-native grass invasion on aboveground carbon pools and tree population structure in a tropical dry forest of Hawaii. *For. Ecol. Manage.* 231, 105–113.
doi:10.1016/j.foreco.2006.05.008
- Lovell, J.J.L., Jupp, D.L.B., Culvenor, D.D.S., Coops, N.C.C., 2003. Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Can. J. Remote Sens.* 29, 1–16.
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G., Saah, D., 2012. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *Int. J. For. Res.* 2012, 1–16. doi:10.1155/2012/436537
- Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: State-of-the-art. *ISPRS J. Photogramm. Remote Sens.* 64, 1–16. doi:10.1016/j.isprsjprs.2008.09.007
- Martin, M.E., Newman, S.D., Aber, J.D., Congalton, R.G., 1998. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sens. Environ.* 65, 249–254.
- McGaughey, R.J., 2014. FUSION/LDV: Software for LiDAR Data Analysis and Visualization. Seattle, WA, USA.
- McRoberts, R., Bollandsås, O., Næsset, E., 2014. Modeling and Estimating Change, in: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*. Springer Science, New York, p. 294.
- Merrick, M.J., Koprowski, J.L., Wilcox, C., 2013. Into the third dimension: benefits of incorporating LiDAR data in wildlife habitat models, in: *USDA Forest Service Proceedings RMRS-P-67*. pp. 389–395.
- Meyer, V., Saatchi, S.S., Chave, J., Dalling, J.W., Bohlman, S., Fricker, G. a., Robinson, C., Neumann, M., Hubbell, S., 2013. Detecting tropical forest biomass dynamics from repeated airborne lidar measurements. *Biogeosciences* 10, 5421–5438. doi:10.5194/bg-10-5421-2013
- Naesset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* 80, 88–99.
- Næsset, E., 2004a. Estimation of above-and below-ground biomass in boreal forest ecosystems, in: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI-8/W2. Freiburg, Germany, pp. 145–148.

- Naesset, E., 2004. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sens. Environ.* 91, 243–255. doi:10.1016/j.rse.2004.03.009
- Næsset, E., Bollandsås, O.M., Gobakken, T., Gregoire, T.G., Ståhl, G., 2013a. Model-assisted estimation of change in forest biomass over an 11year period in a sample survey supported by airborne LiDAR: A case study with post-stratification to provide “activity data.” *Remote Sens. Environ.* 128, 299–314. doi:10.1016/j.rse.2012.10.008
- Næsset, E., Gobakken, T., 2005. Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sens. Environ.* 96, 453–465. doi:10.1016/j.rse.2005.04.001
- Næsset, E., Gobakken, T., Bollandsås, O.M., Gregoire, T.G., Nelson, R., Ståhl, G., 2013b. Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. *Remote Sens. Environ.* 130, 108–120. doi:10.1016/j.rse.2012.11.010
- Næsset, E., Gobakken, T., Naesset, E., Gobakken, T., Næsset, E., Gobakken, T., 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* 112, 3079–3090. doi:10.1016/j.rse.2008.03.004
- Nelson, R., Ranson, K.J., Sun, G., Kimes, D.S., Kharuk, V., Montesano, P., 2009. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sens. Environ.* 113, 691–701. doi:10.1016/j.rse.2008.11.010
- Nelson, R., Valenti, M.A., Short, A., Keller, C., 2003. A multiple resource inventory of Delaware using airborne laser data. *Bioscience* 53, 981–992. doi:10.1641/0006-3568(2003)053[0981:amriod]2.0.co;2
- Neuenschwander, A., 2012. Mapping vegetation structure in a wooded savanna at Freeman Ranch, TX using airborne waveform lidar, in: 2012 Silvilaser International Conference on Lidar Applications for Assessing Forest Ecosystems. Vancouver, BC, Canada, pp. 42–53.
- Neuenschwander, A.L., 2008. Evaluation of Waveform Deconvolution and Decomposition Retrieval Algorithms for ICESat/GLAS Data. *Can. J. Remote Sens.* 34, SS240–SS246.
- Neuenschwander, A.L., Magruder, L.A., Tyler, M., 2009. Landcover classification of small-footprint, full-waveform lidar data. *J. Appl. Remote Sens.* 3, 1–13. doi:10.1117/1.3229944
- Neuenschwander, A.L., Urban, T.J., Gutierrez, R., Schutz, B.E., 2008. Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: Implications for vegetation mapping. *J. Geophys. Res. Biogeosciences* 113, 1–18. doi:10.1029/2007JG000557
- Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* 56, 1–7. doi:10.1016/0034-4257(95)00224-3
- Økseter, R., Bollandsås, O.M., Gobakken, T., Næsset, E., 2015. Modeling and predicting aboveground biomass change in young forest using multi-temporal airborne laser scanner data. *Scand. J. For. Res.* 30, 458–469. doi:10.1080/02827581.2015.1024733

- Ørka, H.O., Næsset, E., Bollandsås, O.M., 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens. Environ.* 113, 1163–1174.
- Pancel, L., 1993. *Tropical forestry handbook*, Springer-Verlag. Springer Verlag, Berlin, Germany.
- Pang, Y., Lefsky, M., Sun, G., Ranson, J., 2011. Impact of Footprint Diameter and Off-Nadir Pointing on the Precision of Canopy Height Estimates from Spaceborne Lidar. *Remote Sens. Environ.* 115, 2798–2809.
- Pang, Y., Li, Z.Y., 2012. Inversion of biomass components of the temperate forest using airborne Lidar technology in Xiaoxing'an Mountains, Northeastern of China. *Chinese J. Plant Ecol.* 36, 1095–1105.
- Parker, G.G., Harmon, M.E., Lefsky, M.A., Chen, J.Q., Van Pelt, R., Weis, S.B., Thomas, S.C., Winner, W.E., Shaw, D.C., Franklin, J.F., 2004. Three-dimensional structure of an old-growth *Pseudotsuga-Tsuga* canopy and its implications for radiation balance microclimate and gas exchange. *Ecosystems* 440–453.
- Parrish, C.E., Jeong, I., Nowak, R.D., Smith, R.B., 2011. Empirical Comparison of Full-Waveform Lidar Algorithms : Range Extraction and Discrimination Performance. *Photogramm. Eng. Remote Sens.* 77, 825–838.
- Piao, S., Fang, J., Zhu, B., Tan, K., 2005. Forest biomass carbon stocks in China over the past 2 decades: Estimation based on integrated inventory and satellite data. *J. Geophys. Res.* 110. doi:10.1029/2005JG000014
- Pirotti, F., Laurin, G., Vettore, A., Masiero, A., Valentini, R., 2014. Small Footprint Full-Waveform Metrics Contribution to the Prediction of Biomass in Tropical Forests. *Remote Sens.* 9576–9599. doi:10.3390/rs6109576
- Plourde, L.C., Ollinger, S. V., Smith, M., Martin, M.E., 2007. Estimating species abundance in a northern temperate forest using spectral mixture analysis. *Photogramm. Eng. Remote Sensing* 73, 829–840.
- Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy* 31, 646–655. doi:10.1016/j.biombioe.2007.06.022
- Popescu, S.C., Hauglin, M., 2014. Estimation of Biomass Components by Airborne Laser Scanning, in: *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*. Springer Science, New York, p. 294.
- Popescu, S.C., Wynne, R.H., 2004. Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogramm. Eng. Remote* 70, 589–604.
- Popescu, S.C., Wynne, R.H., Nelson, R.H., 2002. Estimating plot-level tree heights with LIDAR: Local filtering with a canopy-height based variable window size. *Comput. Electron. Agric.* 37, 71–95.
- Popescu, S.C., Wynne, R.H., Nelson, R.H., 2003. Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. *Can. J. Remote Sens.* 29, 564–577.

- Popescu, S.C., Zhao, K., Neuenschwander, A., Lin, C., 2011. Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sens. Environ.* 115, 2786–2797. doi:10.1016/j.rse.2011.01.026
- Qian, G., 2000. Studies on the dynamic change of the net production quantity of liquidambar formosana plantation. *Acta Agric. Univ. Jiangxiensis* 22, 400–404.
- Ranson, K.J., Sun, G., Kovacs, K., Kharuk, V.I., 2004. Landcover attributes from ICESat GLAS data in central Siberia, in: *International Geoscience and Remote Sensing Symposium, IGARSS 2004: IEEE International Geoscience and Remote Sensing Symposium Meetings*. Anchorage, AK, pp. 753–756.
- Reitberger, J., Krzystek, P., Stilla, U., 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *Int. J. Remote Sens.* 29, 1407–1431. doi:10.1080/01431160701736448
- Reitberger, J., Peter, K., Stilla, U., 2006. Analysis of full waveform LiDAR data for tree species classification, in: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (Part 3). pp. 228–233.
- Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS J. Photogramm. Remote Sens.* 64, 561–574. doi:10.1016/j.isprsjprs.2009.04.002
- Rosenqvist, Å., Milne, A., Lucas, R., Imhoff, M., Dobson, C., 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environ. Sci. Policy* 6, 441–455. doi:10.1016/S1462-9011(03)00070-4
- Ross, J., 1981. *The radiation regime and architecture of plant stands*. Springer Netherlands.
- Sands, R., 2005. *Forestry in a global context*. CABI Publishing, London.
- Schmitt, C.B., Burgess, N.D., Coad, L., Belokurov, A., Besançon, C., Boisrobert, L., Campbell, A., Fish, L., Gliddon, D., Humphries, K., Kapos, V., Loucks, C., Lysenko, I., Miles, L., Mills, C., Minnemeyer, S., Pistorius, T., Ravilious, C., Steininger, M., Winkel, G., 2009. Global analysis of the protection status of the world's forests. *Biol. Conserv.* 142, 2122–2130. doi:10.1016/j.biocon.2009.04.012
- Schuckman, K., Renslow, M., 2014. *Topographic Mapping with Lidar (GEOG 481) [WWW Document]*. Pennsylvania State Univ. URL https://www.e-education.psu.edu/geog481/l3_p4.html (accessed 8.3.15).
- SFA, 2011. *Regulations for age-class and age-group division of main tree-species [WWW Document]*. URL http://www.forestry.gov.cn/portal/main/govfile/13/govfile_1883.htm (accessed 8.29.15).
- SFMP, 2013. *The Sustainable Management Plan for Yushan Forest (2013–2020)*. Suzhou.
- Smith, T.I., Whelan, K., 2006. Development of allometric relations for three mangrove species in south Florida for use in the greater everglades ecosystem restoration. *Wetl. Ecol. Manag.* 14, 409–419.
- Song, Y.B., Ding, Y.P., Shen, F., 2009. The development and latest progress of JSCORS. *Bulltin Surv. Mapp.* 73–74.

- Stephens, P.R., Kimberley, M.O., Beets, P.N., Paul, T.S.H., Searles, N., Bell, A., Brack, C., Broadley, J., 2012. Airborne scanning LiDAR in a double sampling forest carbon inventory. *Remote Sens. Environ.* 117, 348–357. doi:10.1016/j.rse.2011.10.009
- St-Onge, B., Vepakomma, U., 2004. Assessing forest gap dynamics and growth using multi-temporal laser-scanner data, in: *Proceedings of the ISPRS Working Group VIII/2. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 8/W2.* Germany University of Freiburg, pp. 173–178.
- Sumnall, M.J., Hill, R.A., Hinsley, S.A., 2012. The estimation of forest inventory parameters from small-footprint waveform and discrete return airborne LiDAR data, in: *SilviLaser 2012.* Vancouver, pp. SL2012–020.
- Sun, G., Ranson, K.J., Kimes, D.S., Blair, J.B., Kovacs, K., 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sens. Environ.* 112, 107–117. doi:http://dx.doi.org/10.1016/j.rse.2006.09.036
- Suratno, A., Seielstad, C., Queen, L., 2009. Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS J. Photogramm. Remote Sens.* 64, 683–693.
- Turner, D.P., Cohen, W.B., Kennedy, R.E., Fassnacht, K.S., Briggs, J.M., 1999. Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sens. Environ.* 70, 52–68.
- Vaughn, N.R., Moskal, L.M., Turnblom, E.C., 2012. Tree Species Detection Accuracies Using Discrete Point Lidar and Airborne Waveform Lidar. *Remote Sens.* 4, 377–403.
- Vauhkonen, J., Maltamo, M., McRoberts, R.E., Næsset, E., 2014. Introduction to Forestry Applications of Airborne Laser Scanning, in: *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies.* Springer, New York London, p. 294.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS J. Photogramm. Remote Sens.* 60, 100–112.
- Wang, Q., Shi, Y., 1990. A preliminary study on the biomass and production of slash pine plantation in Jiangsu province. *Acta Phytoecol. Geobot. Sin.* 14, 2–12.
- Wang, X., Feng, Z., Ouyang, Z., 2001. The impact of human disturbance on vegetative carbon storage in forest ecosystems in China. *For. Ecol. Manage.* 148, 117–123. doi:10.1016/S0378-1127(00)00482-5
- Warren Wilson, J., 1965. Stand Structure and Light Penetration. I. Analysis by Point Quadrats. *J. Appl. Ecol.* 2, 383–390.
- Weisberg, S., 1985. *Applied linear regression*, 2nd ed. Wiley, New York, USA.

- White, J.C., Wulder, M. a, Varhola, A., Vastaranta, M., Coops, N.C., Cook, B.D., Pitt, D., Woods, M., 2013. A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. Victoria, British Columbia. doi:10.5558/tfc2013-132
- Wichmann, V., Bremer, M., Lindenberger, J., Rutzinger, M., Georges, C., Petrini-Monteferri, F., 2015. Evaluating the potential of multispectral airborne LiDAR for topographic mapping and land cover classification. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. II-3/W5:113-119.
- Wu, J., Aardt, J. Van, Asner, G.P., Knapp, D., Erasmus, B.F.N., 2009. LiDAR Waveform-based woody and foliar biomass estimation in savanna environments, in: *Silvilaser*. pp. 20–29.
- Wu, J., van Aardt, J. a N., Asner, G.P., 2011. A Comparison of Signal Deconvolution Algorithms Based on Small-Footprint LiDAR Waveform Simulation. *IEEE Trans. Geosci. Remote Sens.* 49, 2402–2414. doi:10.1109/TGRS.2010.2103080
- Wulder, M.A., Bater, C., Coops, N.C., Hilker, T., White, J.C., 2008. The role of LiDAR in sustainable forest management. *For. Chron.* 84, 807–826.
- Wulder, M.A., White, J.C., Stinson, G., Hilker, T., Kurz, W. a, Coops, N.C., St-Onge, B., Trofymow, J. a T., 2010. Implications of differing input data sources and approaches upon forest carbon stock estimation. *Environ. Monit. Assess.* 166, 543–61. doi:10.1007/s10661-009-1022-6
- Wyoff, W.R., 1990. A basal area increment model for individual conifers in the northern rocky mountains. *For. Sci.* 36, 1077–1104.
- Xie, X., Wang, Q., Dai, L., Su, D., Wang, X., Qi, G., Ye, Y., 2011. Application of China's National Forest Continuous Inventory database. *Environmental management.* 48:1095-1108.
- Xu, G., Pang, Y., Li, Z., Zhao, D., Liu, L., 2012. Individual trees species classification using relative calibrated fullwaveform LiDAR data, in: *Silvilaser2012*. Vancouver, pp. SL2012–084.
- Xu, J., Wang, M., Huang, Q., Gong, S., 2011. Study on aboveground biomass model of natural individual trees of *Quercus acutissima*. *Anhui For. Sci. Technol.* 37, 3–6.
- Yang, W., Ni-Meister, W., Lee, S., 2011. Assessment of the Impacts of Surface Topography, Off- Nadir Pointing and Vegetation Structure on Vegetation Lidar Waveforms using an Extended Geometric Optical and Radiative Transfer Model. *Remote Sens. Environ.* 115, 2810–2822.
- Yao, W., Krzystek, P., Heurich, M., 2012. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sens. Environ.* 123, 368–380.
- Yu, X., Hyypä, J., Kaartinen, H., Maltamo, M., 2004. Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sens. Environ.* 90, 451–462.

- Yu, X., Hyypä, J., Kaartinen, H., Maltamo, M., Hyypä, H., 2008. Obtaining plotwise mean height and volume growth in boreal forests using multi - temporal laser surveys and various change detection techniques. *Int. J. Remote Sens.* 29, 1367–1386. doi:10.1080/01431160701736356
- Yu, X., Hyypä, J., Kukko, A., Maltamo, M., Kaartinen, H., 2006. Change Detection Techniques for Canopy Height Growth Measurements Using Airborne Laser Scanner Data. *Photogramm. Eng. Remote Sens.* 72, 1339–1348. doi:10.14358/PERS.72.12.1339
- Yu, X., Litkey, P., Hyypä, J., Holopainen, M., Vastaranta, M., 2014. Assessment of Low Density Full-Waveform Airborne Laser Scanning for Individual Tree Detection and Tree Species Classification. *forests* 5, 1011–1031.
- Zhang, K., Cui, Z., 2007. LIDAR Technology [WWW Document]. URL <http://lidar.ihr.c.fiu.edu/aboutlidar.html> (accessed 8.3.15).
- Zhao, K., Popescu, S.C., Nelson, R.F., 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.* 113, 182–196.