Aboveground Biomass Estimation Using Spaceborne LiDAR in Managed Conifer Forests in South Central British Columbia

By

Laura Innice Duncanson
B.Sc. Queen's University, 2007

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Geography

Aboveground Biomass Estimation Using Spaceborne LiDAR in Managed Conifer Forests in South Central British Columbia

by

Laura Innice Duncanson
B.Sc. Queen's University, 2007

Supervisory Committee

Dr. K. Olaf Niemann, Supervisor
(Department of Geography, University of Victoria)

Dr. Michael A. Wulder, Outside Member
(Department of Geography, University of Victoria)

Dr. Dennis E. Jelinski, Departmental Member
(Department of Geography, University of Victoria)

ABSTRACT

Dr. K. Olaf Niemann, Supervisor
(Department of Geography, University of Victoria)

Dr. Michael A. Wulder, Outside Member
(Department of Geography, University of Victoria)

Dr. Dennis E. Jelinski, Departmental Member
(Department of Geography, University of Victoria)

In the context of growing concerns regarding global climatic change, developing methods to assess the carbon storage of various ecosystems has become important. This research attempts to develop low or no cost methods to estimate carbon stock in forests using satellite-based data. More specifically, this research explores the utility of spaceborne Light Detection and Ranging (LiDAR) data for forest canopy height and aboveground biomass estimation. High-resolution (sub meter) airborne LiDAR data were collected and validated for a 75 000 ha area near Clearwater, British Columbia. Airborne LiDAR has been widely demonstrated to yield accurate aboveground biomass estimates. 110 temporally coincident Geospatial Laser Altimeter System (GLAS) waveforms from the study site were used in this research. First, I demonstrate that airborne LiDAR can be manipulated to represent waveform curves with a high degree of similarity to GLAS waveform curves. Based on the relationship between the GLAS and simulated waveforms I am able to visualize the ground contribution to GLAS waveforms. Second, I calculate a suite of novel GLAS waveform metrics and develop models of terrain relief, canopy height, and terrain adjusted canopy height. These models compare favourably to other GLAS studies (terrain relief R^2=0.76, canopy height R^2= 0.75-0.88) and indicate that

terrain relief should be included in GLAS derived canopy height models. Third, I attempt to extrapolate the spatially discrete GLAS estimates to spatially continuous estimates using Landsat TM data. Landsat data have been used extensively for AGBM estimation, although they are known to have limitations for studies in high biomass or structurally complex forests. I develop models to predict GLAS AGBM estimates from Landsat bands and indices (R^2=0.6). I then use an airborne LiDAR derived AGBM map to generate a map of over and under prediction of AGBM, and evaluate the success of the model in areas of differing forest species and structure. I conclude that GLAS data is appropriate for AGBM estimation in forests over a wide range of biomass values, but that GLAS and Landsat integration for AGBM estimation should only be conducted in forests with less than approximately 120 Mg/ha of AGBM, 60 years of age, or 60% canopy cover.

TABLE OF CONTENTS

ABSTRACT ..... **III**

TABLE OF CONTENTS ..... **V**

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **IX**

ACKNOWELDGEMENTS ..... **X**

CO-AUTHORSHIP STATEMENT ..... **XI**

1.0 INTRODUCTION **1**

1.1 RESEARCH CONTEXT 1
1.2 RESEARCH GOALS AND OBJECTIVES 5
1.3 REFERENCES 7

2. ESTIMATING TERRAIN RELIEF AND CANOPY HEIGHT FROM GLAS WAVEFORM METRICS **12**

2.1 ABSTRACT 12
2.3 METHODS 20
2.3.1 Study Area *20*
2.3.2 Data *21*
2.3.2.1 SATELLITE-BASED LIDAR DATA 21
2.3.2.2 AIRBORNE LIDAR DATA 22
2.4 ANALYSIS 24
2.4.1 GLAS Data Processing *25*
2.4.2 Simulated Waveform Curve Construction *27*
2.4.3 Terrain and Canopy Metrics Calculation *28*
2.4.4 GLAS Waveform Metrics *29*
2.4.5 Modeling Terrain Relief and Canopy Height from GLAS Metrics *32*
2.4.6 Terrain Classification *33*
2.4.7 Canopy Height Models using Terrain Relief Classes or Dummy Variables *33*
2.4.8 Model Validation *34*
2.5 RESULTS/DISCUSSION 36
2.5.1 Curve Comparison *36*
2.5.2 Gaussian Decomposition *42*
2.5.3 Correlation Analysis *45*
2.5.4 Terrain Relief Model *47*
2.5.5 Preliminary Canopy Height Model *48*
2.5.6 Outliers *49*
2.5.7 Canopy Height Models with Terrain Relief Inputs *51*
2.5.8 Dummy Variable Models *55*
2.5.9 Relief Classification *56*
2.5.10 Model Validation *60*
2.6 CONCLUSIONS 62
2.7 REFERENCES 64

3. INTEGRATION OF GLAS AND LANDSAT TM DATA FOR ABOVEGROUND BIOMASS ESTIMATION **69**

3.1 ABSTRACT 69
3.2 INTRODUCTION 71
3.3 METHODS 76

3.2.1 Study Area ... *76*
3.2.2 Data ... *77*
3.2.2.1 AIRBORNE LIDAR DATA ... 77
3.2.2.2 FIELD DATA ... 78
3.2.2.3 AGBM MAP ... 78
3.2.2.4 SATELLITE-BASED LIDAR ... 79
3.2.2.5 SATELLITE MULTISPECTRAL ... 80
3.2.2.6 ANCILLARY DATA MAP ... 80
3.2.3 Analysis ... *81*
3.3 RESULTS ... 86
3.3.1 GLAS Model Results ... *86*
3.3.2 Landsat Model Results ... *87*
3.3.3 Error Image Decomposition ... *88*
3.4 DISCUSSION ... 94
3.4.1 GLAS AGBM Estimation ... *94*
3.4.2 Landsat AGBM Estimation ... *94*
3.4.3 Error Decomposition ... *95*
3.4.4 Canopy Height, Rugosity, DBH ... *97*
3.4.5 AGBM ... *97*
3.4.6 Age ... *98*
3.4.9 Per Cent Canopy Cover ... *98*
3.5 CONCLUSION ... 100
3.6 REFERENCES ... 102

4. CONCLUSION ... 109

4.1 DISCUSSION AND CONCLUSIONS ... 109
4.2 RESEARCH CONTRIBUTIONS ... 111
4.3 RESEARCH OPPORTUNITIES ... 113
4.4 REFERENCES ... 115

LIST OF FIGURES

Figure 2-1. Map of British Columbia and study area showing the GLAS transect. The center of the study is located at 51° 44′ N 120° 18′ W. 21
Figure 2-2. Analysis framework depicting data processing steps and model development. The shaded boxes correspond to the steps which fulfilled research objectives. 25
Figure 2-3. The top image shows the energy quarter divisions, the bottom image shows the elevation quarters. Waveform metrics calculate the proportion of energy in both energy and elevation quarters. 31
Figure 2-4. The *startpeak* and *peakend* metrics correspond to the distance between the location of the energy peak and the start and end of the waveform, respectively. 32
Figure 2-5. Histogram of correlation coefficients between simulated and waveform curves and cumulative curve. 37
Figure 2-6. Low correlation values between simulated and GLAS curves were found when the peak return from the simulated curve (blue) did not match the peak return from the GLAS waveform (red). The footprint on the left was sparsely vegetated, with the higher of the two peaks being partially attributed to ground return, partially to canopy return. The footprint on the right is a heavily vegetated area with a large, steep drop in elevation to the edge of the footprint. The two peaks are both attributed to ground return. 37
Figure 2-7a. Column 1 shows contour images of each footprint giving a visual representation of the canopy and underlying terrain. Column two shows a comparison of the simulated and waveform curves, with the simulated curves shown in blue and the waveform curves shown in red. Column three shows the simulated curve with the classified ground portion shaded yellow and column four shows the waveform curve with its 40
Figure 2-8. The left images show the GLAS return and corresponding Gaussian curves provided with the GLAS data products. The right images show the same waveforms with the corresponding Gaussian curves found in my Gaussian distribution. The top left image has five Gaussian curves, the top right has 14 Gaussian curves, the bottom left has three Gaussian curves and the bottom right has five Gaussian curves. 44
Figure 2-9. Histogram depicting the distribution of the number of Gaussian curves found per GLAS waveform from my Gaussian decomposition. 45
Figure 2-10. Left shows the results from the terrain maximum relief model. Right shows results from the preliminary canopy height model. The larger dots represent the outliers that were removed from the analysis. 48
Figure 2-11. Residual plot of 85th percentile hits height prediction against maximum relief. Disregarding the outliers there is an apparent increase in the variability of residuals for footprints with maximum relief <7 metres and >15 metres. 52
Figure 2-12. The relationship between maximum relief, 85^{th} percentile hits height and startpeak. It is apparent that for areas of low and high relief the relationship between canopy height and startpeak is less consistent. 55
Figure 2-13. Results from dummy variable models. 58

Figure 2-14. Results of classifying data and using classification as inputs to the three new forest height models for the three new relief classes. The dotted lines show the 1:1 line, the solid lines are the best fit lines. 59
Figure 2-15. Results from Dummy variable model using relief class inputs from discriminant analysis. 60
Figure 3-1 Study area map showing GLAS transect and elevation distribution across study area. 77
Figure 3-2. Analysis Framework. The results and discussion sections proceed through this framework sequentially. 82
Figure 3-3 AGBM distribution for the study area, as estimated from airborne LiDAR data. The subset shows the AGBM distribution from the GLAS transect. 85
Figure 3-4. Results from Models 1 and 3, GLAS and Landsat estimates of the square root of AGBM in GLAS footprints. 86
Figure 3-5 Model 3 spatial distribution of error (Model minus Observed AGBM per pixel). 88
Figure 3-6 Standard deviation of model 3 error, mean AGBM and standard deviation of AGBM for dominant species segmentation. 91
Figure 3-7. 10^{th} – 90^{th} percentile box and whiskers plots decomposing pixel-based error into various forest parameter classes. 92

LIST OF TABLES

Table 2-1. Waveform Metrics and Abbreviations 30
Table 2-2. Correlation matrix. The variables used in the terrain relief model or any of the canopy height models appear in bold. Only variables with correlations lower than r=0.5 with other input variables were included in models. 46
Table 2-3. Information pertaining to outliers found from the preliminary canopy height model 51
Table 2-4 Canopy height model information 56
Table 2-5. Model Validation Results for Maximum Terrain Relief Model. 60
Table 2-6 Model Validation Results for Dummy Variable Canopy Height Model. 61
Table 3-1. Waveform Metrics and Abbreviations. 79
Table 3-2. AGBM Model Information. 87
Table 3-3. Number of pixels in various forest classes, as used in Figure 3-7. 93

ACKNOWELDGEMENTS

First, I would like to thank my supervisor, Dr. K Olaf Niemann, who first introduced me to GLAS data and who has been incredibly supportive of my research throughout my masters program (financially and otherwise). He allowed me the flexibility to design my own research project and was very encouraging of my ideas, while always acting as a source of knowledge and solutions when I ran up against walls. Secondly, I would like to thank Dr. Michael A. Wulder, without whom I would not have been able to turn my research into publication quality papers. His incredible breadth of knowledge and experience, as well as his meticulous edits, were essential to the success of the papers. I would also like to thank Dr. Dennis Jelinski for acting as my third committee member and helping me prepare for conference presentations. I would also like to thank the members of the Hyperspectral LiDAR Research Group for all of their help, and particularly Rafael Loos and Roger Stevens for patiently answering so many questions about GIS and IDL programming. Also, thank you to the SPAR lab members, for keeping me sane and hydrated. Additionally, thank you to the Geography Department and the graduate students of 2007-2009 – you changed a mere degree into the experience of a lifetime. Thank you to the National Snow and Ice Data Center for distribution and assistance with GLAS products. For funding, thank you to the Natural Science and Engineering Council , the Derek Sewell fund, Graphic Office Interiors and the University of Victoria Fellowship. Finally, thank you to Ben and my family for your constant support and encouragement.

CO-AUTHORSHIP STATEMENT

This thesis was divided into two distinct but related manuscripts for which I was the lead author, followed by Dr. K. Olaf Niemann and Dr. Michael A. Wulder. Dr. Niemann first suggested aboveground biomass estimation using GLAS data, and provided me with airborne LiDAR data with which to analyze the GLAS data. For the two papers I was the lead researcher, meaning that I developed the research goals and objectives, developed and applied the methods, interpreted the results, and prepared the final manuscripts. Drs. Niemann and Wulder were both instrumental in influencing the direction of the research and in editing the final manuscripts. The first paper, entitled 'Estimating forest canopy height and terrain relief from GLAS waveform metrics' has been accepted for publication in the journal Remote Sensing of Environment. The second, entitled 'Integration of GLAS and Landsat TM Data for Aboveground Biomass Estimation' is intended for submission to the Canadian Journal of Remote Sensing.

1.0 INTRODUCTION

1.1 Research Context

Remote sensing technologies have developed at accelerating rates for the past few decades (Boyd & Danson, 2005). Current sensors have the ability to collect a wide array of data; hyperspectral sensors can break a surface's reflectance into hundreds of image bands at the nanometer level while airborne LiDAR sensors are collecting such high spatial resolution data that LiDAR data can be processed to visually simulate real world environments. Satellites are collecting data at increasingly high spatial and spectral resolutions for areas that have never been studied by humans on the ground (Potter et al., 2003; Houghton, 2005). This technological development has been closely followed by a development of data analysis; however, I argue that the data analysis has not kept up with the technological progression. There is a wealth of information hidden within remotely sensed data that researchers are working to uncover; remote sensing may hold the key to understanding the natural world in ways that have not been imagined before. One particularly important area in which remote sensing approaches are increasingly employed is that of carbon budget development for support of climatic change research and mitigation (Rosenqvist et al., 2003).

Accounting for baseline carbon stocks in forests has been noted as essential for the implementation of climate change policies (Rosenqvist et al., 2003), and as necessary for increasing the accuracy of global carbon cycle models (Goodale et al., 2002). Dry weight aboveground biomass (AGBM), which constitutes all living or dead material above the soil surface in a forested ecosystem, is made up of approximately 50% carbon (Drake et al., 1992). As such, AGBM is an important forest structural characteristic to study.

Traditional AGBM estimates have been developed using species-based allometric relationships between field measures of tree diameter at breast height (DBH) and AGBM (Jenkins et al., 2003). These approaches are useful for small area inventories, and for the validation of remote sensing data, but are inappropriate for regional or national inventories because of their time consuming nature and high associated costs (Patenaude et al., 2004). Estimates over larger areas have traditionally involved the extrapolation of field-based measurements to areas of similar land cover, as outlined by Hall et al. (2006). These extrapolations are, however, limited by ecological differences over large areas, variation between data collection dates, differences in inventory classification systems across large areas, geographically scattered biomass source data and equations, and the availability of inventory-based biomass estimates for only managed forest areas (Hall et al., 2006).

Optical, Synthetic Aperture Radar (SAR), and LiDAR are three types of remote sensing technologies have been used, in combination with various models, to estimate AGBM at various scales. Optical remote sensing has been used to estimate AGBM based on two assumptions: 1) the spectral reflectance from a forested area is correlated to the spectral reflectance from the surface elements (leaves, forest floor, etc.), and 2) there is a predictable relationship between the reflectance from surface elements and ratios between surface elements and forest structural characteristics (Wulder et al., 2004). The first assumption frequently fails in densely forested areas and structurally complex forested areas (i.e. old growth forests) where vertical information about the forest becomes more important (Drake et al., 2002). Optical sensors are fundamentally limited in biomass estimation because they integrate energy from a 2D plane, while biomass depends

strongly on 3D forest structure (Hudak et al., 2002).

Two active remote sensing technologies, SAR and LiDAR have been used to add a vertical dimension to AGBM models. SAR and LiDAR differ in that they use energy of different wavelengths; SAR uses microwave energy and LiDAR uses visible and near infrared wavelengths (Hudak et al., 2002). The use of SAR data to estimate AGBM is based on the assumption that the forest floor, large branches, and tree trunks will scatter all SAR frequencies similarly, but that the foliar components of the forest will primarily scatter higher frequencies (Dobson et al., 1992). SAR based biomass estimations are limited by canopy density because closed canopies often do not allow enough SAR penetration for biomass estimates to be accurate (Dobson et al., 1992).

LiDAR largely avoids biomass density threshold issues (Lefsky et al., 1999). LiDAR biomass estimations differ from optical estimations in that LiDAR takes measurements of physical canopy properties rather than spectral properties. For example, where optical sensors rely on the development of models to relate the reflectance from a pixel to reflectance from trees in the pixel, and subsequently to physical properties of the trees, LiDAR takes a measure of the tree heights themselves. Discrete return LiDAR measures the time elapsed between the emission and return of light pulses, producing a high-resolution (usually more than 1 pulse per square metre) three-dimensional point cloud. Each point represents the location at which the reflection of a pulse surpassed a defined threshold. Full waveform LiDAR sensors differ in that they record a continuous energy return from every emitted pulse. Typically these sensors have lower spatial resolutions, with a single pulse illuminating 5-25 meters on the ground. Both discrete return and full waveform airborne LiDAR data have been demonstrated as useful for

forest structural property estimation in a wide range of forested environments (Patenaude et al., 2004, Coops et al., 2007, Nelson et al., 2003, Means et a., 1999, Drake et al., 2002), and in denser canopies than those that can be accurately analyzed by optical or RADAR systems (Lefsky et al., 1999). However these data have high associated volumes and costs, and consequently are inappropriate for studying AGBM at the scales necessary for wall to wall national carbon budgeting. Satellite derived estimates are necessary to document current AGBM distributions and to analyze temporal changes (Houghton, 2005).

The Geospatial Laser Altimeter System (GLAS) is the only functioning spaceborne LiDAR system. Unlike more ubiquitous discrete return airborne LiDAR data, which are used to develop high resolution canopy height models and digital elevation models, the full waveform GLAS data only have one LiDAR return every 172 meters (Schutz et al., 2005). Each return corresponds to reflection off an approximately 65 meter diameter ellipsoid, or footprint. Each emitted LiDAR pulse is measured in amount of energy returned over time, and the elapsed time between the reflection off the ground and the reflection off the canopy can be used to estimate canopy height. The entire waveform return represents the vertical distribution of material within the GLAS footprint. In a relatively flat, homogeneous footprint the earliest reflection will correspond to the elevation of the top of the canopy, and the last reflection will correspond to the lowest ground elevation (Lefsky et al., 2005). As such, the breadth of GLAS waveforms has been demonstrated as useful for AGBM estimation in flat, relatively homogeneous forests (Rosette et al., 2008, Lefsky et al., 2005, Lefsky et al., 2007, Harding & Carabajal, 2005, Sun et al., 2008). However, the top of the canopy and lowest point on the ground may not

be useful for characterizing the footprint's surface - it is possible that a single tall tree or the lowest point in sloping terrain will result in these first and last signal reflections being inappropriate for canopy characterization. Terrain slope, in particular, has been shown to complicate methods for canopy height estimation (Lefsky et al., 2005). GLAS waveforms are as complex as the environments they represent, and therefore it is unreasonable to assume that a single waveform metric will be able to relate a variable such as canopy height or AGBM.

The utility of GLAS data is further limited by its spatial distribution. It is not currently possible to develop regional AGBM maps using GLAS data alone, because along a given GLAS transect there are 110 meters of terrain that fall outside of GLAS footprints and useful transects are not typically spatially contiguous. Spatially continuous data are needed to extend the utility of GLAS data in forested environments. One spatially continuous data source is from the Landsat TM sensor. Landsat sensors are multispectral, and have been used extensively in combination with inventory or airborne LiDAR to model forest attributes such as Leaf Area Index (LAI), cover type, and canopy height (Cohen & Spies, 1992, Hudak et al., 2002, Wulder & Seemann, 2003, Wulder et al., 2007). Extrapolating GLAS AGBM estimates using Landsat data may be an appropriate method for regional AGBM estimation.

1.2 Research Goals and Objectives

The goals of this research are to increase the understanding of GLAS waveforms and their relationship to terrain structure, and to increase the utility of GLAS data for

canopy height and AGBM estimation. This thesis has been separated into two distinct but related manuscripts that share the same research goals. The objectives of the first manuscript, Chapter 2, are four fold;

1. To validate the information content of GLAS waveforms by comparing coincident airborne LiDAR elevation profiles with GLAS waveforms;

2. To examine the utility of derived GLAS waveform metrics for the modeling of terrain relief and forest height;

3. To determine how terrain relief affects the utility of GLAS waveform metrics to model forest height; and

4. To develop a methodology to consistently model forest height directly from GLAS waveform metrics.

The objectives of the second manuscript, Chapter 3, are three fold;

1. Develop a method to model AGBM from GLAS and Landsat integration;

2. Determine the relationships between model error and forest cover properties, such as dominant species and stand age;

3. Establish reliable ranges of forest structural properties for which GLAS and Landsat data integration is appropriate for AGBM estimation.

1.3 References

Boyd, D.S. & Danson, F.M. 2005. Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography*. 29, 1, 1-26.

Cohen, W.B. & Spies, T.A. 1992. Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment.* 41, 1, 1-17.

Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., & Trofymow, J.A. 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees*, 21, 295-310.

Dobson, M.C., Ulaby, F.T., LeToan, T., Beaudoin, A., Kasischke, E.S., & Christensen, N. 1992. Dependence of Radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*. 30, 2, 412-415.

Drake, J.B., Dubayah, R.O., Know, R.G., Clark, D.B., & Blair, J.B. 2002. Sensitivity of large-footprint LiDAR to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, 81, 378-392.

Goodale, C.L., Apps, M.J., Birdsey, R.A., Field, C.B., Heath, L.S., Houghton, R.A., Jenkins, J.C., Kohlmaier, G.H., Kurz, W., Liu, S., Nabuurs, G.J., Nilson, S., &

Shvidenko, A.Z. 2002. Forest carbon sinks in the northern hemisphere. *Ecological Applications*. 12, 3, 891-899.

Hall, R.J., Skakun, R.S., Arsenault, E.J., & Case, B.S. 2006. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, 225, 378-390.

Harding, D.J., & Carabajal, C.C. 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32, L21S10.

Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterretche, M. 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82, 397-416.

Houghton, R.A. 2005. Aboveground forest biomass and the global carbon balance. *Global Change Biology*. 11, 945-958.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S. & Birdsey, R.A. 2003. National-scale biomass estimators for United States tree species. *Forest Science*. 49, 1, 12-35.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A. and D. Harding. 1999. LiDAR remote sensing of the canopy structure and biophysical properties of Douglas-Fir Western Hemlock forests. *Remote Sensing of Environment.* 70, 339-361.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B. Carabajal, C.C., Espirito-Santo, F.D.B., Hunter, M.O., & de Oliveira Jr., Raimundo. 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32, L22S02.

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E. & McKee, W.A. 1999. Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment.* 67, 298-308.

Nelson, R., Valenti, M.A., Short, A., & Keller, C. 2003. A multiple resource inventory of Delaware using airborne laser data. *BioScience*, 53, 10, 981-992.

Patenaude, G., Hill, R.A., Milne, R., Gaveau, D.L.A., Briggs, B.B.J., & Dawson, T.P. 2004. Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, 93, 368-380.

Potter, C., Klooster, S., Myneni, R., Genovese, V., Tan, P.N., & Kumar, V. 2003. Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982-1998. *Global and Planetary Change*. 39, 201-213.

Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., & Dobson, C. 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy*, 6, 441-455.

Rosette, J.A.B., North, P.R.J., & Suarez, J.C. 2008. Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing*, 29, 5, 1475-1493.

Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D., & DiMarzio, J.P. 2005. Overview of the ICESat Mission. *Geophysical Research Letters*, 32, L21S01.

Wulder, M.A., Hall, R.J., Coops, N.C. & Franklin, S.E. 2004. High spatial resolution remotely sensed data for ecosystem characterization. *BioScience.* 54, 6: 511-521.

Wulder, M.A. & Seemann, D. 2003. Forest inventory height update through the integration of lidar data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, 29, 5, 36-543.

Wulder, M.A., Han, T., White, J.C., Sweda, T., & Tsuzuki, H. 2007. Integrating profiling LiDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, 110, 123-137.

2. ESTIMATING TERRAIN RELIEF AND CANOPY HEIGHT FROM GLAS WAVEFORM METRICS

2.1 Abstract

Quantifying aboveground biomass in forest ecosystems is required for carbon stock estimation, aspects of forest management, and further developing a capacity for monitoring carbon stocks over time. Airborne Light Detection And Ranging (LiDAR) systems, of all remote sensing technologies, have been demonstrated to yield the most accurate estimates of aboveground biomass for forested areas over a wide range of biomass values. However, these systems are limited by considerations including large data volumes and high costs. Within the constraints imposed by the nature of the satellite mission, the GeoScience Laser Altimeter System (GLAS) aboard ICESat has provided data conferring information regarding forest vertical structure for large areas at a low end user cost. GLAS data have been demonstrated to accurately estimate forest height and aboveground biomass especially well in topographically smooth areas with homogeneous forested conditions. However in areas with dense forests, high relief, or heterogeneous vegetation cover, GLAS waveforms are more complex and difficult to consistently characterize.Iuse airborne discrete return LiDAR data to simulate GLAS waveforms and to subsequently deconstruct coregistered GLAS waveforms into vegetation and ground returns. A series of waveform metrics was calculated and compared to topography and vegetation information gleaned from the airborne data. A model to estimate maximum relief directly from waveform metrics was developed with an R^2 of 0.76 (n=110), and used for the classification of the maximum relief of the areas sensed by GLAS.

Discriminant analysis was also conducted as an alternative classification technique. A model was also developed estimating forest canopy height from waveform metrics for all of the data (R^2=0.81, n=110) and for the three separate relief classes; maximum relief 0-7 meters (R^2=0.83, n=44), maximum relief 7-15 meters (R^2=0.88, n=41) maximum relief >15 meters (R^2=0.75, n=25). The moderate relief class model yielded better predictions of forest height than the low relief class model which is attributed to the increasing variability of waveform metrics with terrain relief. The moderate relief class model also yielded better predictions than the high relief class model because of the mixing of vegetation and terrain signals in waveforms from high relief footprints. This research demonstrates that terrain can be accurately modeled directly from GLAS waveforms enabling the inclusion of terrain relief, on a waveform specific basis, as supplemental model input to improve

estimates of canopy height.

2.2 Introduction

Forests have been identified as important ecosystems in terms of capacity for carbon storage (Rosenqvist et al., 2003), as reservoirs of biodiversity (Turner et al. 2003), for an influence on the surrounding microclimate (Chen et al., 1999), and as drivers of a range of ecological processes (Parker et al., 2004). The physical structure of a forest largely determines the carbon storage capacity and related ecological functionality (Wulder et al., 2004). Several types of models have been developed in an attempt to relate forest structure to forest function. Models that characterize ecosystems (Potter, 1999), net primary production (Running et al., 2004), and climate (Hurtt et al., 1998), for example, require measures of forest structure. Remote sensing technologies have become widely used for forest structure characterization (Lefsky et al., 2002, Wulder et al. 2004). While optical imagery is widely available over large spatial scales, ready collection and generation of reliable measures of canopy heights in support of biomass estimation are more problematic (Patenaude, 2004). The major limitation to using optical remotely sensed data for measurements of vertical forest structure is the necessary reliance on the assumption that there is a predictable relationship between the two-dimensional structural properties of a forest that can be sensed by these systems, and the three-dimensional structural properties of a forest that are required for forest volume and aboveground biomass estimations (Lefsky et al., 1998). Although forest structural properties have been estimated from optical data, the accuracy of these estimates typically decreases with increasing biomass and LAI (Hudak et al., 2002; Foody et al., 2001).

Accounting for carbon stocks is a crucial element in the understanding of the global carbon cycle as well as for creating and updating national and regional (including state /

provincial) forest inventories (Hese et al., 2005; Wulder et al. 2004). Current remote sensing technologies provide an ability to acquire higher spatial resolution imagery than previously possible. While high spatial resolution measures may enable improved depictions of forest structure (Wulder et al., 2004, Zimble et al., 2003), there is often a trade-off between the scale of measurement and the spatial resolution and the spatial extent of a given image. As a result, detailed depictions of structure, whether based upon LiDAR or high spatial resolution imagery, can be made, yet often only over smaller areas. To provide estimates of carbon stocks representative of regional or global forest extents, different tools should be investigated to provide for accurate estimates of aboveground biomass over national, continental, and even global scales (Patenaude et al., 2004). LiDAR data have been demonstrated to produce accurate estimates of tree height, canopy closure, and aboveground biomass (Hyyppa et al, 2008, Lim et al., 2003), with the relationship with above ground biomass found to be non-asymptotic (Lefsky et al., 1998, 1999).

LiDAR systems function by emitting pulses of light energy towards a surface, and recording the elapsed time between emission and return of each pulse. In combination with accurate navigation and positioning systems aboard an aircraft, discrete return LiDAR systems yield three dimensional point clouds of forested areas, from which tree heights and vertical structural measures can be extracted (Lim et al., 2003). However, due to large data volumes and costs these technologies are typically deemed inappropriate for the large areas necessary for forest inventory (Ranson et al., 2007). Sampling schemes developed to utilize LiDAR data inputs are one means to capitalize upon the richness of LiDAR measures to provide a calibration data source for the characterization of large

areas (Lefsky et al., 2005; Nelson et al., 2003). Sampling may also be undertaken using space-borne LiDAR data transects in combination with optical data to generate regional biomass estimates (Boudreau et al., 2008), however these approaches remain constrained by the accuracy of space-borne LiDAR biomass models.

The Geoscience Laser Altimeter System (GLAS) aboard NASA's ICESat is the first space-borne LiDAR system capable of providing global datasets of the Earth's surface (Schutz et al., 2005). GLAS has been collecting data since 2003, with models under development to estimate forest structural properties from GLAS data for much of globe (Boudreau et al., 2008; Harding and Carabajal, 2005; Lefsky et al., 2007; Rosette et al., 2008; Sun et al., 2008). Of the three lasers onboard GLAS, the first laser failed approximately 38 days into the mission (March 29, 2003), resulting in an alteration in the design of the rest of the mission. In order to extend the life of the second and third laser, three 33 day operating periods per year replaced the original temporally continuous measurements (Sun et al., 2008). The third laser failed on October 10, 2008 and was replaced by Laser 2 which was operational as of July, 2009. GLAS was designed primarily to study relatively flat and homogeneous ice sheets, with low initial expectations for characterization of vegetated surfaces (Zwally et al. 2002); and consequently, considerable effort is required to apply GLAS data products to forested environments given their spatially variable nature (Harding and Carabajal, 2005).

GLAS is a full waveform sensor using a 1064 nanometer Laser operating at 40 Hz. The laser illuminates an elliptical area on Earth's surface with a diameter of approximately 65 meters, with footprint centroids located approximately 172 meters apart (Schutz et al., 2005). The sensor records the returned energy from these footprints over

time since pulse emission. Consequently the returned GLAS waveforms reflect the vertical distribution of the terrain and vegetation within each footprint. Canopy metrics, such as maximum canopy height, can be routinely extracted from flat, relatively homogeneous footprints that show distinct canopy and ground peaks within a waveform (Harding and Carabajal, 2005), and have been demonstrated as useful for aboveground biomass estimates (Harding and Carabajal, 2005, Boudreau et al., 2008) and tree height estimates (Lefsky et al., 2007, Rosette et al., 2008) in such areas. However, in areas of moderate to high relief, energy will strike lower elevation canopy tops at the same time as higher elevation ground, resulting in a more complex and difficult to interpret waveform (Harding and Carabajal, 2005).

Several studies have been undertaken to model forest characteristics from GLAS waveform metrics with varying levels of success. Lefsky et al. (2007), used the distance from the beginning to the end of the signal (called waveform extent), the distance from the start of the signal to one half of the waveform maximum power (leading edge extent), and the distance from the end of the signal to one half of the waveform maximum power (trailing edge extent) to estimate maximum canopy height (r^2 = 0.83, n=198). Boudreau et al., 2008, also used waveform extent but instead of the leading and trailing edges they used the slope between the signal start and first Gaussian peak, and a terrain index from SRTM data to model aboveground biomass (R^2=0.59, n=1325). Another study, presented by Sun et al. (2008), showed that a high correlation existed between airborne LiDAR measures of canopy height and GLAS measures of canopy height, based on quartile energies from both systems (R^2= 0.83). Although no methodology was presented to estimate canopy height or aboveground biomass in Sun et al. (2008), it was demonstrated

that energy quantile information from GLAS could be used the same way airborne LiDAR quantiles have been used to predict canopy heights.

Gaussian decomposition is another method used to extract information from GLAS waveforms. GLAS waveforms, like all full waveform LiDAR returns, can be thought of as the sum of individual Gaussian returns reflected from each element within the footprint (Blair and Hofton, 1999). A GLAS waveform from flat, homogenous terrain can be represented by a single Gaussian curve, while flat, homogeneously forested areas will yield bimodal returns that can be approximated from two Gaussian curves - the first representing the canopy, the second representing the underlying terrain (Harding and Carajabal, 2005). More discrete elements within the footprint (i.e., trees, fluctuations in terrain, canopy gaps) will increase the number of Gaussian curves that make up a given waveform.

The ICESat processing software fits up to six Gaussian curves to each waveform, and the location of the last Gaussian peak in a flat area will likely represent the ground return, and may represent the elevation of the underlying terrain (Boudreau et al., 2008). However it was demonstrated that the last Gaussian does not always represent the bulk of the ground signal, and in some situations the second lowest Gaussian peak is a better representation of ground elevations (Rosette et al., 2008). In areas of greater relief or more complex topography than those explored in these studies, it is unlikely that only one of the lowest Gaussian curves represents the bulk of the ground signal.

The number and distribution of Gaussian curves within a waveform are expected to correspond to canopy and terrain properties. Therefore, although setting a limit of six Gaussian curves allows for a more simplistic and comparable approximation of GLAS

waveforms, it also limits the utility of a Gaussian decomposition for terrain and canopy characterization.

Developing accurate methods to account for the influence of terrain in moderate or high relief areas is an integral step in the development of accurate models for characterization of forest structural properties (Lefsky et al., 2007).

To address the need to account for terrain in the development of approaches for forest structural characterizationsIhave defined four main objectives:

1. Validate GLAS information content of GLAS waveforms by comparing coincident airborne LiDAR elevation profiles with GLAS waveforms;

2. To examine the utility of derived GLAS waveform metrics for the modeling of terrain relief and forest height;

3. To determine how terrain relief affects the utility of GLAS waveform metrics to model forest height; and

4. To develop a methodology to consistently model forest height directly from GLAS waveform metrics.

2.3 Methods

2.3.1 Study Area

Airborne LiDAR data were acquired for 75,000 ha area of industrially managed forest near Clearwater, British Columbia, Canada. One GLAS transect was selected from the available GLAS data for the area, based on temporal proximity to the airborne LiDAR acquisition (Figure 2-1). Three main biogeoclimatic zones occur within the study area: the Sub-Boreal Spruce (SBS) zone, the Engelmann Spruce-Subalpine Fir (ESSF) zone and the Interior Cedar-Hemlock (ICH) zone (Pedersen and Forester, 2000). The predominant tree species in the area are Engelmann spruce (*Picea engelmannii*), white spruce (*Picea glauca*), lodgepole pine (*Pinus contorta*), balsam fir (*Abies balsamea*), Douglas fir (*Pseudotsuga*), western hemlock (*Tsuga heterophylla*) and western redcedar (*Thujia plicata*) (Pedersen and Forester, 2000). The area is characterized as a high elevation plateau of gently rolling terrain with an elevation range of approximately 800 meters (Pedersen and Forester, 2000). The GLAS data selected for this study represent a transect running almost north-south through the area, as seen in Figure 2-1. This transect covers all three biogeoclimatic zones present, as well as a variety of species compositions and cover types ranging from undisturbed mature forests to recently harvested areas. This variety of cover types and topography make the area ideal for exploration for better understanding the effects of each on waveforms.

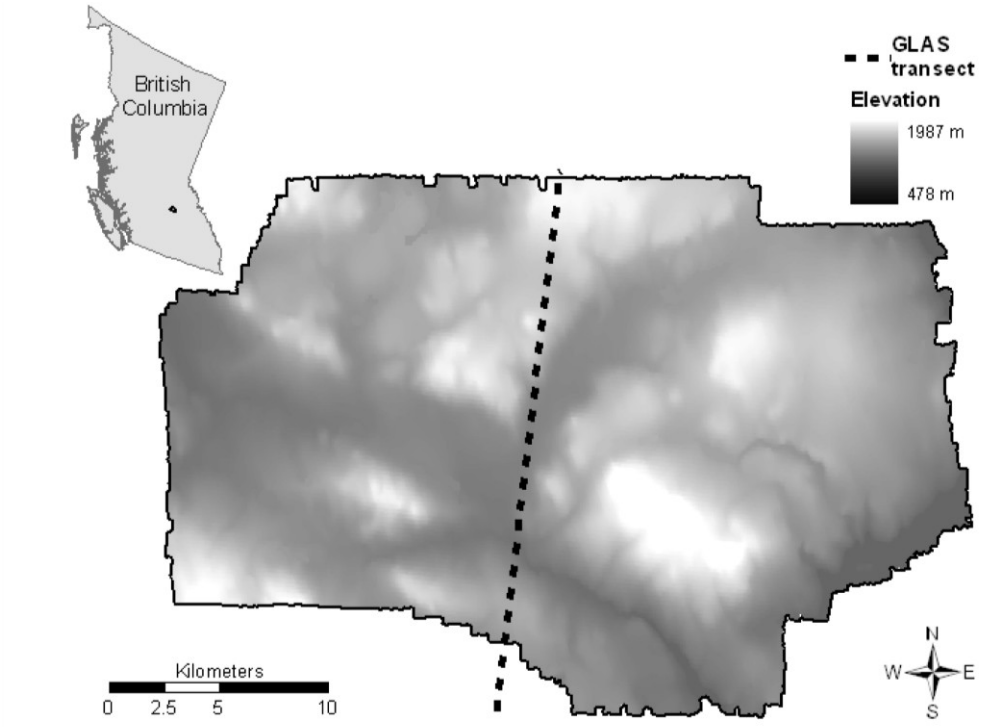

Figure 2-1. Map of British Columbia and study area showing the GLAS transect. The center of the study is located at 51° 44′ N 120° 18′ W.

2.3.2 Data

2.3.2.1 Satellite-Based LiDAR Data

The GLAS data used in this study were collected June 26, 2006. These data were distributed by the National Snow and Ice Data Center (http://nsidc.org/cgi-bin/snowi/search.pl). Raw waveform data were gleaned from data products GLA-01 and the locations of waveform centroids were gleaned from data product GLA-14. GLA-01 contains raw waveforms in volts against time since pulse emission, and GLA-14 contains the latitude and longitude position of the waveform centroid. The GLAS data flag i_FRir_qaFlag indicates the estimated atmospheric conditions over each GLAS footprint using a cloud detection algorithm. This atmospheric flag was used in combination with

coincident airborne LiDAR data to ensure that all 110 waveforms analyzed in this study were not affected by atmospheric conditions. These waveforms were from the F campaign of Laser 3. Laser 3 footprints have been found to be more circular than the elliptical footprints of Laser 2 (Abshire et al., 2005), rendering this data suitable for calibration with airborne LiDAR data because airborne data can be reliably clipped to coincident areas around each footprint centroid. The locational accuracy of GLAS footprints has been evaluated by matching GLAS waveforms to Shuttle Radar Topography Mission (SRTM) DEMs, and shown that the on-ground locational error is less than 60 meters and likely much smaller (Sun et al., 2008). My first objective was to assess the accuracy of GLAS data products, in order to further support the use of GLAS data in studies of forest structure.

2.3.2.2 Airborne LiDAR Data

Airborne LiDAR data were acquired August 16, 2006 by a first and last return 60 kHz instrument with a maximum 20° scan angle and 37 Hz scanning speed. The platform was flown at 1600 meters above ground resulting in approximately 2.25 hits per square meter. The data were filtered, using the Terrascan (Terrasolid, Helsinki, Finland, http://www.terrasolid.fi/en) to separate ground hits from those of vegetation, enabling development of a bare earth model and a canopy height model (CHM). The CHM was maintained as a set of points with heights being assigned as the difference between the elevation of that point and the elevation of the coincident bare earth model. Consequently a dataset of points classified as ground hits, a topographically normalized vegetation dataset, and the original set of all points co-located in each GLAS footprint are used in this study.

2.4 Analysis

There were 110 GLAS footprints with coincident airborne LiDAR found in the study area, and this study involves the generation of a suite of GLAS waveform metrics that relate to various terrain properties. Figure 2-2 shows the analysis framework for this project. The analysis will be discussed sequentially in terms of research objectives, beginning with the processing of GLAS and airborne LiDAR data and subsequent construction of simulated waveforms from airborne data to address objective 1. Objective 2 required the calculation of a suite of GLAS waveform metrics and airborne derived canopy and terrain metrics and the development of terrain relief and canopy height models using these metrics. Objectives 3 and 4 involved the classification of terrain relief from waveform metrics, and the subsequent inclusion of relief as an input to canopy height models from GLAS waveform metrics.

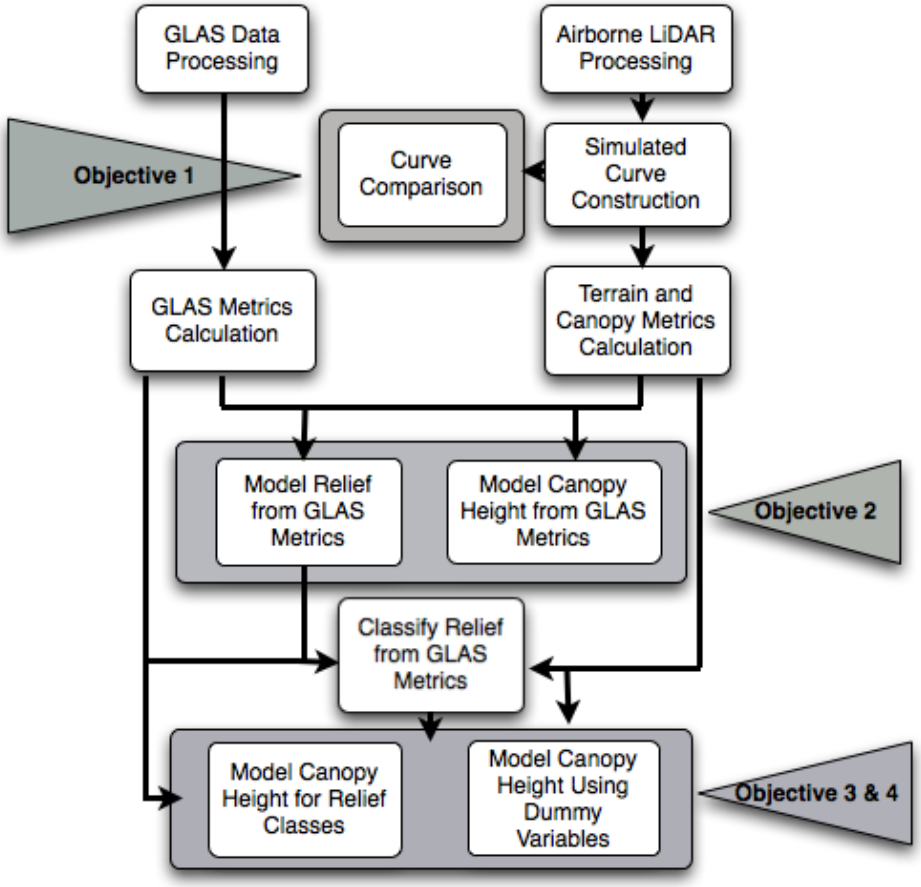

Figure 2-2. Analysis framework depicting data processing steps and model development. The shaded boxes correspond to the steps which fulfilled research objectives.

2.4.1 GLAS Data Processing

GLAS waveforms were filtered by clipping the beginning and end of waveforms below 4.5 times the standard deviation the noise of the waveform return, following the methodology presented by Lefsky et al. (2007). Waveforms were rescaled so the maximum return of the simulated curve and maximum return of the waveform curve were at the same elevation. This resulted in waveforms being measured in volts against elevation.

A Gaussian decomposition was applied in this study that did not constrain the number of Gaussian curves found in the waveforms. An iterative Gaussian decomposition was run on the GLAS data, using in-house IDL (Interactive Data Language, ITT Visualization) code. Gaussian decomposition is a technique based on the assumption that a waveform can be defined by the sum of a series of Gaussian curves, equation 1.

$$y = \sum_{i=1}^{n} ae^{-\frac{(x-b)^2}{2c^2}} \quad (1)$$

where y is the approximated waveform curve, n is the number of Gaussian peaks found, a is the amplitude of the nth Gaussian curve, x is elevation, b is the elevation position of the nth Gaussian peak, and c is the halfwidth of the nth Gaussian curve.

Except where specified, the Gaussian decomposition followed the methodology presented by Hofton et al. (2000). First, the waveforms were smoothed with a moving average filter of 4 nanoseconds (15 centimeters of height), which corresponds to the vertical resolution of the raw waveforms. In order to run the Gaussian decomposition iterative code an initial estimate of all coefficients in Equation 1 was required. Gaussian curves were initially detected by identifying inflection points in the waveform. The initial half width of a Gaussian was set to half the distance between two subsequent inflection points, the initial position was set to the elevation between two subsequent half widths, and the initial amplitude was set to the half of the energy return value at the initial position. This amplitude was selected as the initial parameter because amplitudes are always between zero and the maximum return at a position, but rarely approach either

zero or that maximum. The Levenburg-Marquardt Least Squares technique, presented in Markwardt (2008), was applied, which iteratively adjusts each of the initial parameters until a best fit is found between the result of equation 1 and the GLAS waveform curve.

The Gaussian decomposition was constrained so that half widths were greater than 0.1 meters, curve peaks were greater than one meter apart, and amplitudes were between zero and the maximum return. The minimum half width of 0.1 meters was applied as a smaller Gaussian curve is either the result of noise, or a relatively unimportant element within the footprint, and should be merged with a larger, adjacent Gaussian curve. Similarly, Gaussian curve peaks were forced to be at least one meter apart in elevation because two or more curves found within one meter of elevation likely represent the same element or two very similar elements within the footprint. Consequently, if two peaks were found within one meter of elevation the peaks were merged by adding the half widths, and averaging the amplitudes and positions.

2.4.2 Simulated Waveform Curve Construction

Simulated waveforms were created from the airborne LiDAR data by creating 15 cm height bins and summing the number of hits per height interval to match the vertical resolution of the GLAS waveforms. The height of maximum energy return in each GLAS waveform was set to the elevation of maximum energy return in the coincident simulated curve. The simulated curves and waveform curves were then compared by correlating the two curves once they were placed on the same elevation axis. Each elevation bin for the two curve sets was compared against a combined elevation axis, ensuring that the curves were vertically matched. When there was no waveform or airborne value for an elevation

bin a value of zero was set for that curve at that height. A Pearson's correlation coefficient was calculated between the energy returns of each simulated and GLAS waveform curve. It should be noted that althoughIrefer to the binned airborne LiDAR curves as 'simulated waveform curves' they do not represent true GLAS waveform simulations as they represent the number of airborne LiDAR hits rather than an amount of reflected energy per elevation bin.

The hits classified as ground hits in preprocessing were also divided into 15 centimeter elevation bins. This allowed for a visualization of ground contribution to the simulated curves. This ground contribution was translated to the GLAS waveforms by dividing the ground hits at an elevation by the total number of hits at that elevation and subsequently multiplying by the total waveform energy at that elevation to calculate the ground proportion of waveform energy at each elevation.

2.4.3 Terrain and Canopy Metrics Calculation

It has been demonstrated that the 80^{th} and 90^{th} percentiles of airborne LiDAR measured heights are useful for aboveground biomass and forest volume estimation (Naesset, 2004, Means et al., 2000). Consequently,Iused the 85^{th} percentile hits height as the canopy height metric for this analysis. Maximum terrain relief was calculated as the elevation difference between the highest and lowest ground hit. A mean terrain slope metric was also calculated by fitting a plane to the ground data using a least squares best fit approach and averaging the best fit slope in the east direction with the best fit slope in the west direction. However, maximum relief was selected as the sole terrain variable in this study because it is highly correlated (r=0.98) to mean terrain slope, and is more

physically meaningful for this research.

2.4.4 GLAS Waveform Metrics

In this study several waveform metrics were explored for utility to estimate both canopy and terrain characteristics (see Table 2-1). Waveform extent (*wf_extent*), was calculated after Lefsky et al. (2007). *Wf_max_e* was the highest energy value in the waveform, which should diminish with dense tall canopies and high terrain reliefs as the same amount of energy is spread over a greater vertical distance. *Wf_variance* should increase with landscape complexity within a footprint. *wf_skew* depends on the location of the bulk of the energy within the waveform, and therefore should be useful for terrain and canopy characterization.

The distribution of waveform energy both in terms of elevation and energy intensity is a function of the distribution of the terrain sensed by each GLAS pulse. As such, the proportion of energy in four equal elevation divisions and energy return divisions should act as useful descriptors of the waveform (Figure 2-3). It is expected that for flat areas with little to no vegetation, the greatest proportion of energy will be in the lowest elevation quarter representing the ground with energy being spread evenly between quarters. Similarly, in more highly vegetated areas with higher relief, energy will be spread more evenly across the elevation quarters (more in the higher quarters for heavily vegetated areas, more in the second and third lowest quarter in high relief areas) and more proportional energy will be in the mid energy quarters as the ground peaks will be more numerous and of a lower energy intensity.

Startpeak and *peakend* are similar to waveform extent but intended to represent

canopy height and terrain relief, respectively (shown in Figure 2-4). The number of Gaussian curves is expected to increase with terrain complexity. The number of curves found in *startpeak* should correspond to the complexity of the canopy and the number of curves found in *peakend* should correspond to the complexity of terrain. The location of the Gaussian curve peaks are shown as dots in Figures 2-3 and 2-4.

Table 0-1. Waveform Metrics and Abbreviations

Metric Abbreviation	Metric
85_hit	85th percentile hits height
max_relief	The difference between the highest and lowest LiDAR hit classified as ground
wf_extent	The difference between the beginning and end of the waveform signal
wf_max_e	The highest energy value in the waveform
wf_variance	The variance of the waveform
wf_skew	The skew of the waveform
e_44	Proportion of energy in highest elevation quarter
e_34	Proportion of energy in second highest elevation quarter
e_24	Proportion of energy in second lowest elevation quarter
e_14	Proportion of energy in lowest elevation quarter
Energy_highest	Proportion of energy in highest energy quarter
Energy_34	Proportion of energy in second highest energy quarter
Energy_24	Proportion of energy in second lowest energy quarter
Energy_14	Proportion of energy in lowest energy quarter
startpeak	The difference in elevation between the beginning of the signal and the position of wf_max_e
peakend	The difference in elevation between the end of the signal and the location of wf_max_e
wf_n_gs	The number of Gaussian curves found in the waveform
wf_n_gs_startpeak	The number of Gaussian curves found between the beginning of the waveform signal and the position of wf_max_e

Metric Abbreviation	Metric
wf_n_gs_endpeak	The number of Gaussian curves found between the position of wf_max_e and the end of the waveform signal

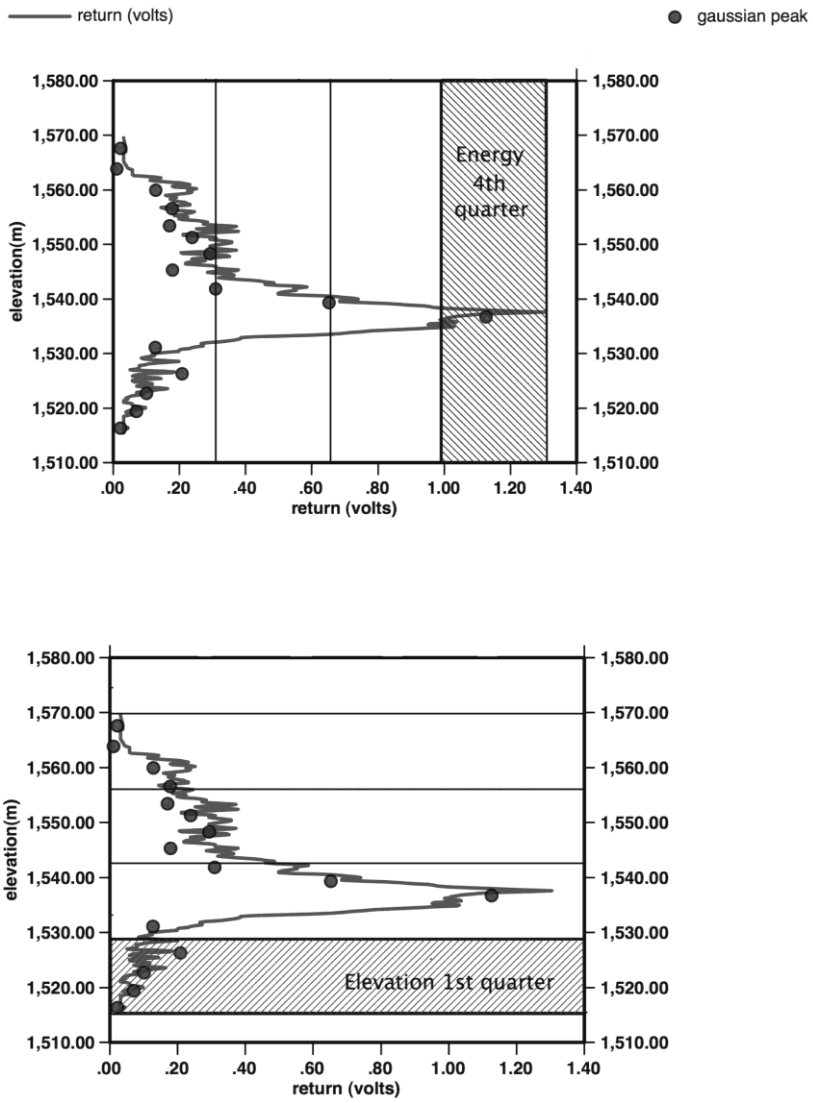

Figure 2-3. The top image shows the energy quarter divisions, the bottom image shows the elevation quarters. Waveform metrics calculate the proportion of energy in both energy and elevation quarters.

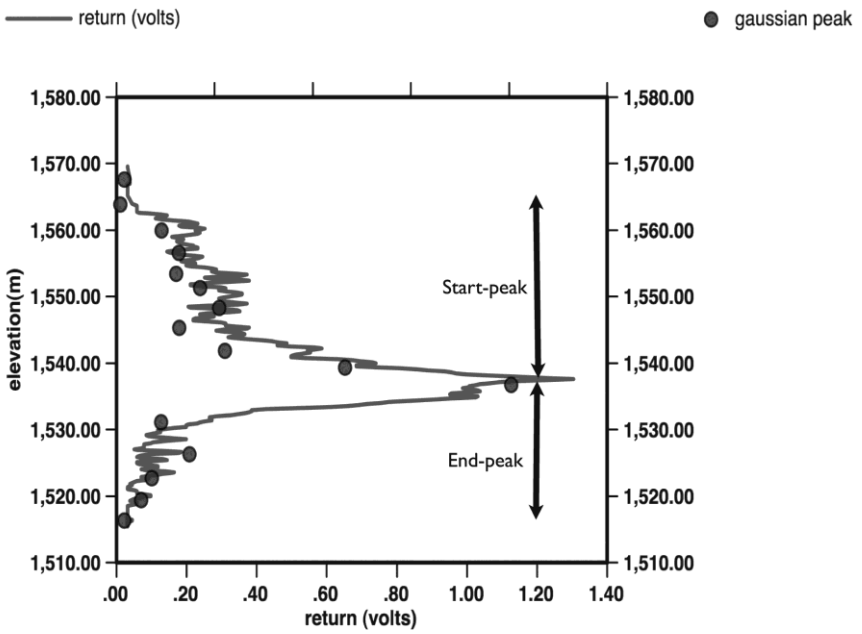

Figure 2-4. The *startpeak* and *peakend* metrics correspond to the distance between the location of the energy peak and the start and end of the waveform, respectively.

2.4.5 Modeling Terrain Relief and Canopy Height from GLAS Metrics

Our second objective was to examine the utility of derived GLAS waveform metrics for the modeling of terrain relief and forest height. A correlation analysis was conducted on all waveform metrics, maximum relief and 85th percentile hits height in order to determine which metrics would be appropriate to model relief. If two waveform metrics were correlated with a Pearson's R value greater than 0.5 the metric with a lower correlation to the variable of interest (maximum relief or 85th percentile hits height) was discarded from the analysis. The remaining GLAS variables were utilized as inputs to two stepwise regression models. This first canopy height model was used as a baseline for comparison with subsequent canopy height models that incorporate terrain relief, discussed later. Four outliers were found from the two models, two of which were

removed from the analysis. These outliers are discussed in the results/discussion section.

2.4.6 Terrain Classification

The third objective was to determine how terrain relief affects the utility of GLAS waveform metrics to model forest height. To explore this question, the residuals from the first canopy height model were plotted against terrain maximum relief. A visual analysis of the ground contribution to the simulated waveform curves was also conducted to assist in determining the maximum relief, if any, at which it becomes more difficult to visually interpret waveforms. Additionally, hierarchical cluster analysis was conducted to examine natural breaks in the dataset.

To incorporate terrain relief into canopy height models relief classes were developed. Two methods by which to classify maximum relief into these classes were developed and tested. The first used the maximum relief model to predict the maximum relief of each footprint and these predictions were subsequently assigned to the appropriate relief class. The second method used a stepwise discriminant analysis to classify maximum relief, with all non-autocorrelated waveform metrics used as inputs, assuming equal prior probabilities and using within-group covariance for classification.

2.4.7 Canopy Height Models using Terrain Relief Classes or Dummy Variables

The fourth and final research objective was to develop a methodology by which to consistently model forest canopy height directly from GLAS waveform metrics. Several models were developed to predict 85th percentile hits height with greater accuracy than the preliminary canopy height model. The dataset was divided into separate relief classes and distinct models were developed for each individual relief class using only footprints

with a maximum relief that fell within that class. However, this method required the separation of the 110 footprints, reducing the sample size for each model, as seen in table 2-4. To address this issue, binary coded dummy variables were used for each of the relief classes and the predictor variables selected for each set of individual relief classes were multiplied by the dummy variable corresponding to their class. The resulting variables were input into a stepwise multiple regression. The resulting model will be more robust in terms of sample size. Although dummy variable models rely on more input variables as a whole, for each individual footprint only 2-5 of the variables were actually used, depending on the relief class of a given footprint. As such, the dummy model represents a combination of the individual terrain relief class models. The results of the dummy variable model are also discussed in the results/discussion section.

The independent relief class models and dummy variable model depend on maximum terrain classification as an input. These model results assume 100% accuracy in the classification of terrain relief. To assess the accuracy of the methods outlined in this paper the accuracy of both the classification of terrain relief and subsequent canopy height prediction were taken into account. The classification methods were both run on the data and the resulting predicted relief classes were used as inputs into the various canopy height models. The predicted canopy heights from all of the models, including those run using classified rather than raw input data, were plotted against the observed airborne LiDAR 85th percentile hits height for each footprint.

2.4.8 Model Validation

We validated the maximum terrain relief model and the dummy variable canopy height

model by recreating each model using a random selection of 80% of the footprints and testing the model on the remaining 20% of the footprints. The same independent variables were used in the generation of the test regression models. The predictions of terrain relief or canopy height for the remaining 20% were regressed against the observed values.

2.5 RESULTS/DISCUSSION

2.5.1 Curve Comparison

The mean Pearson's R correlation coefficient value for the GLAS waveforms and simulated waveform curves was 0.77, with a standard deviation of 0.17. Although R values as low as 0.12 were found, the histogram R values (Figure 2-5) for the set of data between the simulated curves and real waveform curves, shows that the majority of values are greater than 0.8. The instances where the R value was below 0.5 usually correspond to situations when there are two pronounced peaks in the curves. In these instances the elevation of maximum energy return for the airborne data did not match the elevation of maximum return in the GLAS data (Figure 2-6). This was either due to differences in returns from discrete return and full waveform sensors, or due to a spatial disconnect between the two data sets. However, the majority of GLAS waveforms were well matched to the simulated curves, which relates to the geolocational accuracy and circular shape of GLAS Laser 3 data products, thus fulfilling research objective 1.

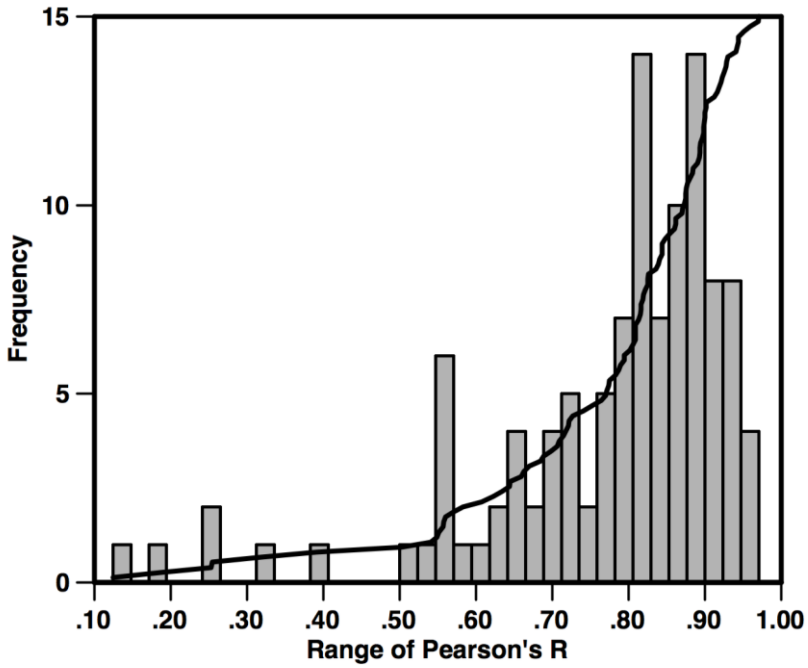

Figure 2-5. Histogram of correlation coefficients between simulated and waveform curves and cumulative curve.

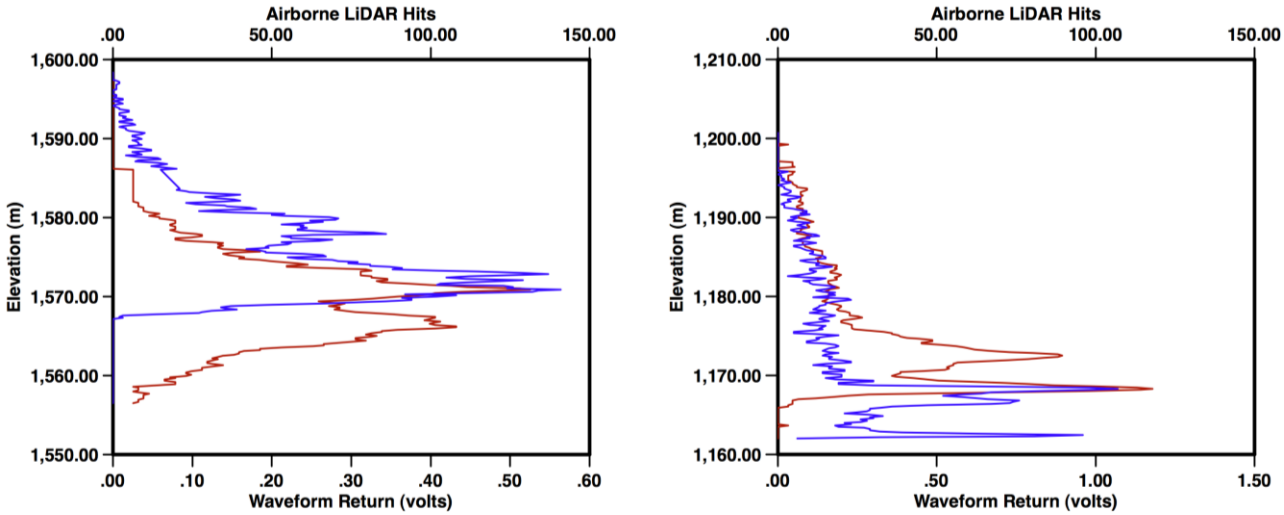

Figure 2-6. Low correlation values between simulated and GLAS curves were found when the peak return from the simulated curve (blue) did not match the peak return from the GLAS waveform (red). The footprint on the left was sparsely vegetated, with the higher of the two peaks being partially attributed to ground

return, partially to canopy return. The footprint on the right is a heavily vegetated area with a large, steep drop in elevation to the edge of the footprint. The two peaks are both attributed to ground return.

The high percentage (60%) of Pearson's R values greater than 0.8 allowed for the use of the proportion of the ground from the airborne dataset to calculate and visualize the proportion of ground return in the waveform dataset. Figures 2-7a and 2-7b show contour maps of the airborne hits, comparisons of waveform curves and simulated curves for several footprints, and the ground contributions to both the airborne and waveform curves. These images show that in many instances the ground contribution to the signal was as expected, exhibited as the final pronounced peak in the curve. However in this dataset the ground contribution was often not as simple.

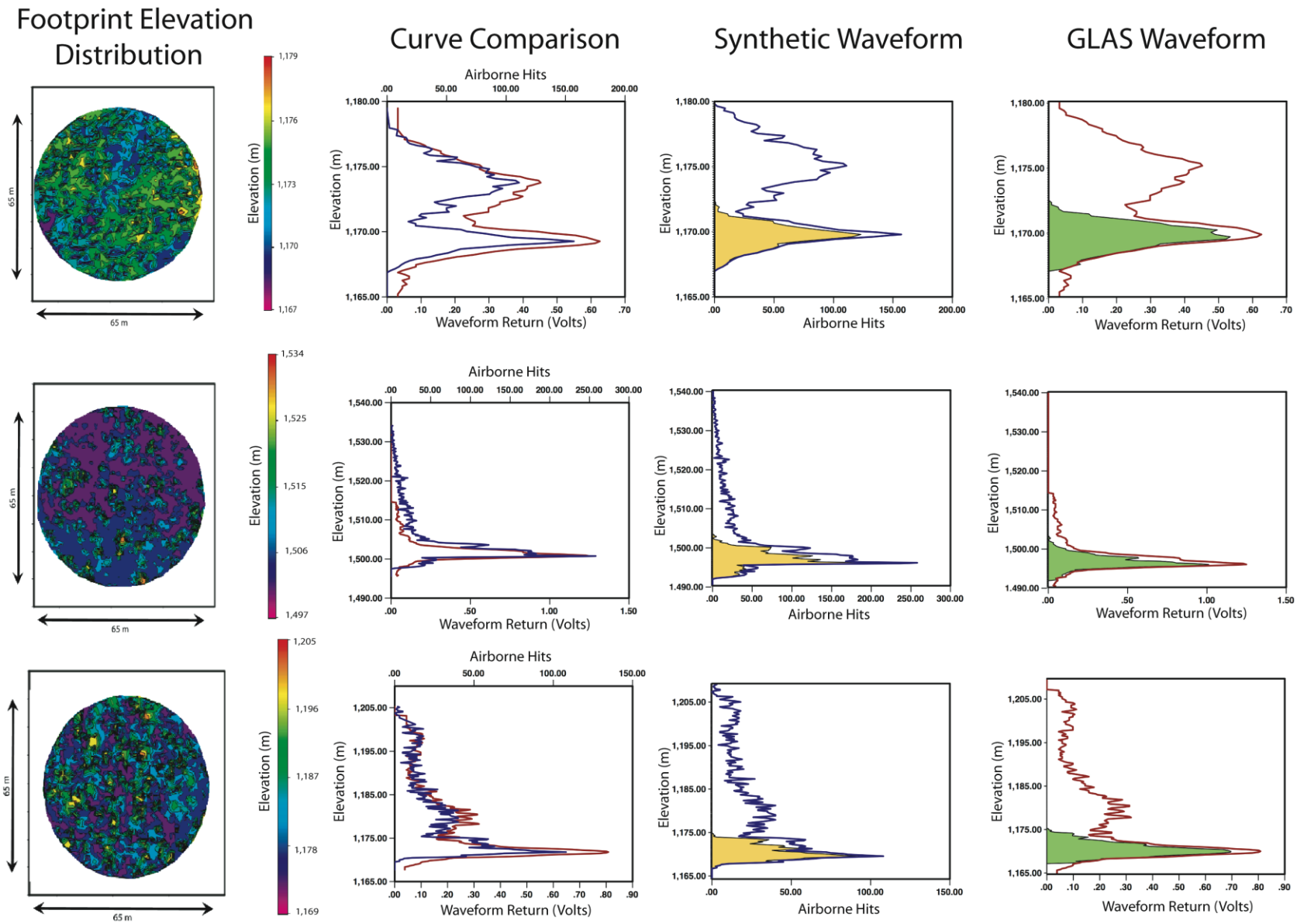
Footprint Elevation Distribution
Curve Comparison
Synthetic Waveform
GLAS Waveform
65 m
65 m
Elevation (m)
1,179
1,176
1,173
1,170
1,167
Airborne Hits
Waveform Return (Volts)
1,534
1,525
1,515
1,506
1,497
1,205
1,196
1,187
1,178
1,169

Figure 2-7a. Column 1 shows contour images of each footprint giving a visual representation of the canopy and underlying terrain. Column two shows a comparison of the simulated and waveform curves, with the simulated curves shown in blue and the waveform curves shown in red. Column three shows the simulated curve with the classified ground portion shaded yellow and column four shows the waveform curve with its simulated ground portion shaded green. Maximum relief ranges from 3.5 to 6.8 metres for these footprints.

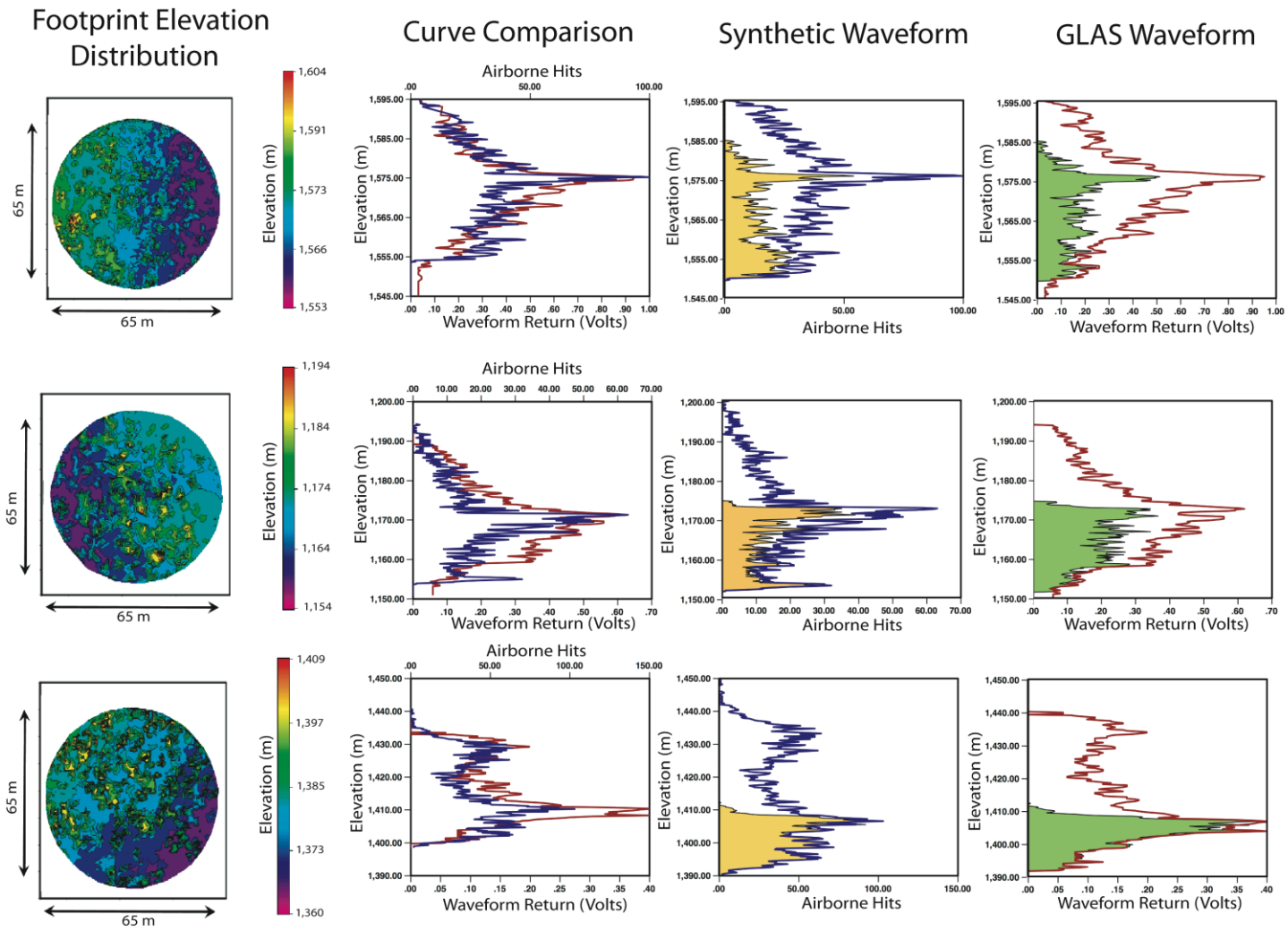

Figure 2-7b. Columns are the same as in Figure 2-7a. Maximum relief ranges from 14.3 metres to 28.4 metres.

The footprints depicted in Figure 2-7a show the expected ground contribution. To provide a footprint level explanation, the first, footprint 84, had a maximum relief across the footprint of 3.5 meters. The second, footprint 23 had a maximum relief of 6.8 meters. The third, footprint 74, had a maximum relief of 5.46 meters. The footprints depicted in Figure 2-7b did not exhibit the typical ground contribution seen in flat areas. The ground portions of these waveforms were more difficult to interpret, had a greater vertical extent and had many peaks within them representing complex topography. The first, footprint 32, had a maximum relief of 28.4 meters, the second, footprint 77, had a maximum relief of 17.8 meters, and the third, footprint 44, had a maximum relief of 14.3 meters. Although the examples presented in Figures 2-7a and 2-7b represent only a small sample of the total footprints analyzed in this study, they depicted the trend that over relatively flat areas the ground contribution was represented by a large peak near the end of the return. Areas of moderate to high relief were more difficult to visually interpret, and consequently, these results support the need for a new methodology to characterize terrain relief, such as presented in this study.

2.5.2 Gaussian Decomposition

Figure 2-8 shows a histogram of the number of Gaussian curves found per GLAS waveform. My Gaussian decomposition found a minimum of two and a maximum of 17 Gaussian curves per GLAS waveform. Figure 2-9 shows the Gaussian curves from my Gaussian decomposition in comparison to the Gaussian curves provided in GLAS data products for two of the waveforms shown in Figures 2-7a and 2-7b. The waveform shown in the bottom two images in Figure 2-8 is from a low relief area and has an almost bimodal return distribution. The waveform shown in the top two images, conversely,

shows the return from a high relief footprint. For curves where my Gaussian decomposition found less than six Gaussian curves the curves found were similar to the Gaussian curves provided in the GLAS data products. However, in more complex footprints associated with densely vegetated or high relief areas, the results of the two Gaussian decompositions appear to differ substantially. Approximating these complex distributions with a maximum of six Gaussian curves results in only the highest return portions of the waveform being associated with a Gaussian curve, as seen in the top left image in Figure 2-8. The highest and lowest portions of each waveform, which have been associated with terrain and canopy information (Lefsky et al., 2006), are frequently not represented by the six Gaussian curves provided with GLAS data products. My Gaussian decomposition, however, does not constrain the number of Gaussian curves used to approximate GLAS waveforms. Consequently metrics corresponding to the number and distribution of these Gaussian curves should be more useful for extracting terrain and canopy information.

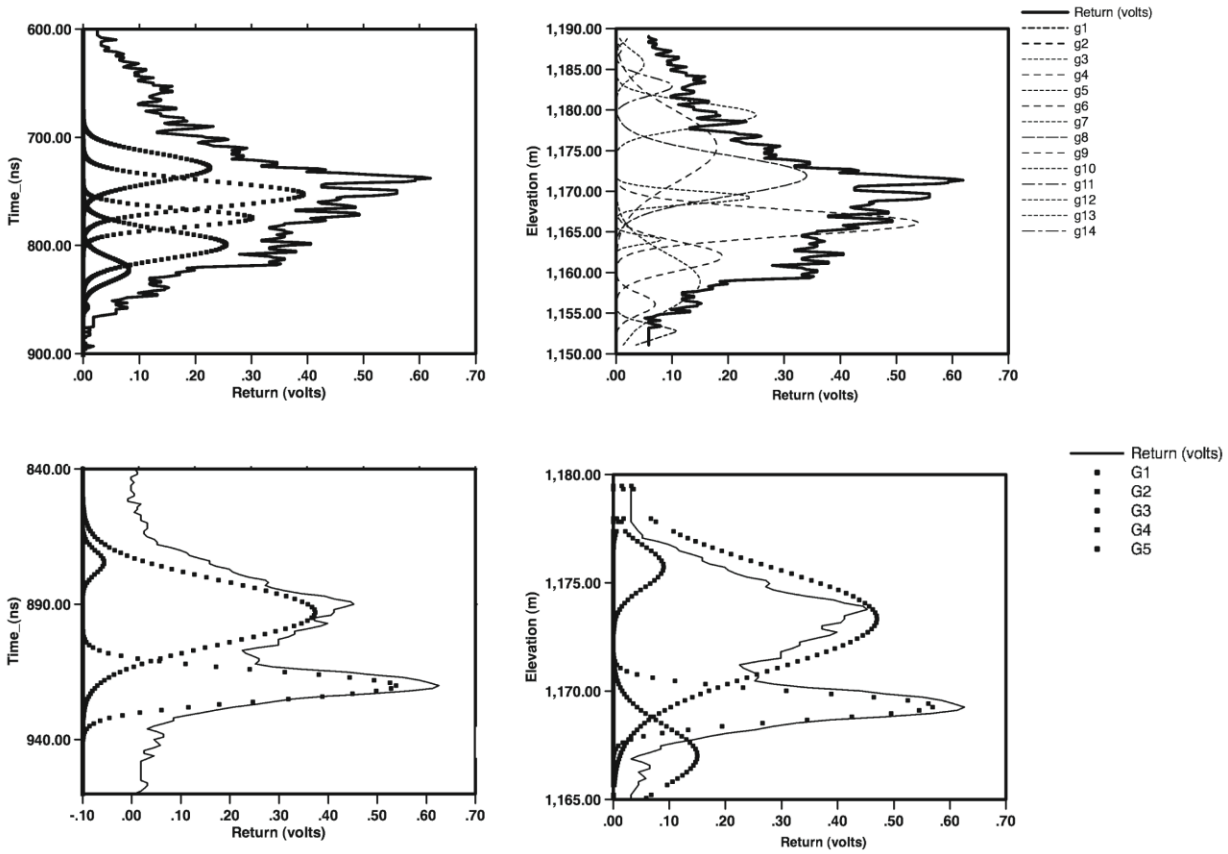

Figure 2-8. The left images show the GLAS return and corresponding Gaussian curves provided with the GLAS data products. The right images show the same waveforms with the corresponding Gaussian curves found in my Gaussian distribution. The top left image has five Gaussian curves, the top right has 14 Gaussian curves, the bottom left has three Gaussian curves and the bottom right has five Gaussian curves.

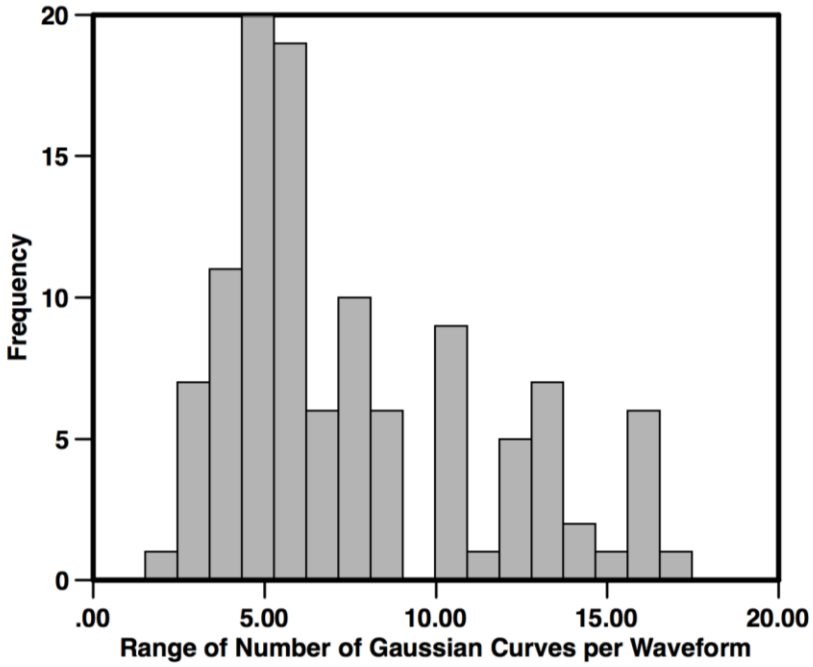

Figure 2-9. Histogram depicting the distribution of the number of Gaussian curves found per GLAS waveform from the Gaussian decomposition.

2.5.3 Correlation Analysis

Table 2-2 shows the results of the correlation analysis of all GLAS waveform metrics, maximum terrain relief and 85th percentile hits height. The correlation coefficient corresponding to the GLAS metrics used to model either of the two variables are presented in bold in the table. Note that although many values are shaded for 85th percentile hits height, these correspond to all of the canopy height models. The maximum number of these input variables used to model canopy height was five.

Table 2-0-2. Correlation matrix. The variables used in the terrain relief model or any of the canopy height models appear in bold. Only variables with correlations lower than r=0.5 with other input variables were included in models.

	85hits_ht	*wf_extent*	*wf_maxe*	*wf_var*	*wf_skew*	*e_4/4*	*3/4_e*	*2/4_e*	*1/4_e*	*wf_ngs*	*startpeak*	*ngs_peakend*	*ngs_startpeak*	*e_highest*	*energy_2/4*	*energy_1/4*	*peakend*	*max_relief*
85hits_ht	**1.00**	0.70	-0.35	-0.57	0.52	0.38	-0.50	-0.11	0.45	0.63	0.78	0.32	0.64	-0.30	0.33	0.52	0.24	0.35
wf_extent	0.70	**1.00**	-0.10	-0.33	0.27	0.25	-0.14	-0.05	0.10	0.88	0.80	0.68	0.72	0.08	0.59	0.66	0.70	0.69
wf_maxe	**-0.35**	-0.10	**1.00**	0.83	0.12	-0.46	0.06	0.10	0.00	-0.15	-0.02	-0.18	-0.06	0.62	0.04	0.35	-0.14	**-0.38**
wf_var	**-0.57**	-0.33	0.83	**1.00**	-0.25	-0.42	0.21	0.09	-0.15	-0.32	-0.31	-0.21	-0.29	0.76	-0.17	0.02	-0.18	-0.40
wf_skew	0.52	0.27	0.12	-0.25	**1.00**	-0.07	-0.47	0.00	0.48	0.12	0.55	-0.21	0.34	-0.31	-0.01	0.46	-0.21	**-0.19**
e_4/4	**0.38**	0.25	-0.46	-0.42	-0.07	**1.00**	0.19	-0.50	-0.06	0.26	0.02	0.39	0.05	-0.21	0.25	-0.03	0.40	0.36
e_3/4	**-0.50**	-0.14	0.06	0.21	-0.47	0.19	**1.00**	-0.38	-0.68	-0.12	-0.53	0.25	-0.37	0.18	0.00	-0.32	0.39	0.14
e_2/4	-0.11	-0.05	0.10	0.09	0.00	-0.50	-0.38	**1.00**	-0.37	-0.01	0.02	-0.12	0.08	0.06	0.00	0.08	-0.10	-0.05
e_1/4	**0.45**	0.10	0.00	-0.15	0.48	-0.06	-0.68	-0.37	**1.00**	0.04	0.49	-0.26	0.26	-0.16	-0.08	0.24	-0.42	-0.20
wf_n_gs	**0.63**	0.88	-0.15	-0.32	0.12	0.26	-0.12	-0.01	0.04	**1.00**	0.69	0.73	0.84	0.07	0.61	0.63	0.65	0.73
startpeak	**0.78**	0.80	-0.02	-0.31	0.55	0.02	-0.53	0.02	0.49	0.69	**1.00**	0.19	0.83	0.03	0.42	0.75	0.14	0.33
ngs_end	0.32	0.68	-0.18	-0.21	-0.21	0.39	0.25	-0.12	-0.26	0.73	0.19	**1.00**	0.24	0.06	0.47	0.24	0.90	0.73
ngs_start	0.64	0.72	-0.06	-0.27	0.34	0.05	-0.37	0.08	0.26	0.84	0.83	0.24	**1.00**	0.06	0.48	0.70	0.20	**0.45**
e_highest	**-0.30**	0.08	0.62	0.76	-0.31	-0.21	0.18	0.06	-0.16	0.07	0.03	0.06	0.06	**1.00**	0.13	0.29	0.10	-0.02
ener_2/4	**0.33**	0.59	0.04	-0.17	-0.01	0.25	0.00	0.00	-0.08	0.61	0.42	0.47	0.48	0.13	**1.00**	0.39	0.48	0.52
ener_1/4	0.52	0.66	0.35	0.02	0.46	-0.03	-0.32	0.08	0.24	0.63	0.75	0.24	0.70	0.29	0.39	**1.00**	0.20	0.27
peakend	0.24	0.70	-0.14	-0.18	-0.21	0.40	0.39	-0.10	-0.42	0.65	0.14	0.9	0.20	0.10	0.48	0.20	**1.00**	**0.76**
max_relief	0.35	0.69	-0.38	-0.40	-0.19	0.36	0.14	-0.05	-0.20	0.73	0.33	0.73	0.45	-0.02	0.52	0.27	0.76	**1.00**

2.5.4 Terrain Relief Model

Figure 2-10 shows the results of the terrain relief model. The stepwise multiple regression model for terrain relief found that the most useful waveform metrics to predict maximum relief were *wf_max_e*, *n_gs_start_peak*, *wf_skew*, and *peak_end*. The model for maximum relief was:

$$\text{Max\_relief} = 5.574 + 0.543 * \text{peak\_end} + .998 * \text{n\_gs\_start\_peak} - 4.395 * \text{wf\_max\_e} - 1.512 * \text{wf\_skew} \quad (2)$$

All of the variables were statistically significant (p<0.05). Of these variables, *peakend* had the highest correlation (r=0.70) with maximum terrain relief (see Table 2-2). This is expected as the distance between the end of the signal and the largest peak in the signal would be small in waveforms from flat footprints in which the ground peak will be the most distinct peak, and the last peak in the signal. As terrain relief increases the energy peak migrates up the signal.

The number of Gaussian peaks found in the top half of the waveform was also used as an indicator. Although it may seem illogical that information from the top half of the signal is correlated to terrain relief, the number of Gaussian peaks found in the top half of the waveform is indicative of the roughness, or number of total peaks, in the entire waveform. The three Gaussian waveform metrics are highly correlated to one another, and the number of Gaussian peaks found in the bottom half of the waveform depends on the breadth of the waveform itself and could not be used as an input variable. Consequently, *n_gs_startpeak* was the only Gaussian peak metric that could be used in the analysis without causing collinearity issues. The number of Gaussian peaks in a waveform increases with terrain complexity. *Wf_max_e* and *wf_skew* both decrease with

increasing terrain relief because of the changing shape of a waveform as the bulk of the return moves up the signal and the amplitude of the energy peak decreases.

These results show that terrain relief can be modeled directly from the GLAS waveform metrics used in this study. The R^2 of this model is 0.76 which compares favourably with other terrain extraction techniques for GLAS data.

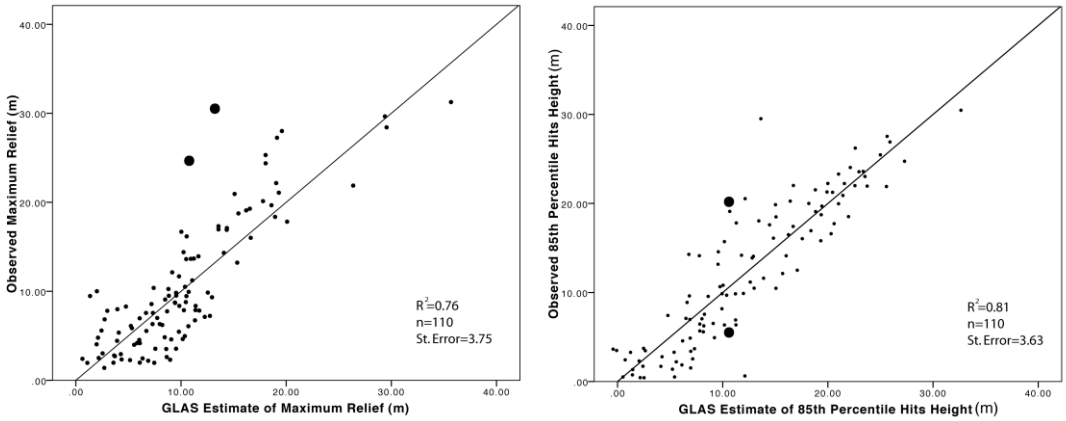

Figure 2-10. Left shows the results from the terrain maximum relief model. Right shows results from the preliminary canopy height model. The larger dots represent the outliers that were removed from the analysis.

2.5.5 Preliminary Canopy Height Model

The results from the preliminary canopy height model are also shown in Figure 2-10. The variables found to be most useful in the canopy height model were *startpeak*, *e_44* and *e_highest*. The model for canopy height was:

85th_ht=-0.544+0.665*startpeak+40.365*e_44-0.121*e_highest (3)

All of the variables were statistically significant ($p<0.05$) except the constant for the canopy height model, which had a *p* value of 0.60. These three input variables were

expected to be useful in a canopy height model; *startpeak* increases with canopy height, as the distance between the top of the canopy and a ground signal increases, *E_44* increases with canopy height and likely density as more relative energy is reflected by the upper canopy, and *e_highest* represents the portion of energy in high energy peaks, which are typically ground peaks. Thus, *e_highest* decreases with canopy height as more energy is reflected off the canopy, leaving less energy to reflect off the ground. Figure 2-11 shows the predicted canopy height residuals against maximum terrain relief. There appears to be more variability at lower values of terrain relief, with variability decreasing as terrain relief increases. There are also two obvious outliers in the dataset.

These results show that it is possible to predict canopy height directly from GLAS waveform metrics, and that the use of proportional energy metrics are useful for this purpose. However, the R^2 value of this model is not higher than previous models developed using GLAS data, and the constant in the model is not statistically significant, indicating that this model is not necessarily suitable for use without ancillary calibration data such as airborne LiDAR data.

2.5.6 Outliers

The four outliers found in the preliminary canopy height model and the maximum relief model pertained to footprints 102, 46, 85 and 101. Footprints 102 and 46 were outliers in the vegetation height model. While the maximum relief model for both of these footprints yielded acceptable estimates, the predicted canopy height value was too high for footprint 102 and too low for footprint 46, as shown in Table 2-3. Some of the predictor variables

for vegetation height also varied with increasing terrain relief. High relief and tall vegetation both have a broadening effect on GLAS waveforms and consequently waveform characteristics are occasionally misinterpreted. Footprint 102 represents an unvegetated area with moderate relief while footprint 46 represents a heavily forested area with high terrain relief and tall trees (maximum tree height approaching 40 meters). The model did not distinguish between the two broadening influences and attributed broadened waveform metrics to ground relief alone for footprint 46 and canopy height alone for footprint 102.

Information pertaining to the four outliers is shown in Table 2-3. The two outliers from the maximum relief model had predicted values far lower than the actual relief values. The discrepancy between the actual and predicted slope values in these cases appeared to be due to a locational disconnect between the airborne and satellite data. It was inferred from the airborne LiDAR data that these footprints were located at the edge of a road or flat harvested area. The area clipped from the airborne LiDAR data included some peripheral vegetation, while the GLAS instrument does not appear to have illuminated this vegetation. This may be explained by the lack of a weighting algorithm applied to the airborne hits, as this research did not account for the spatial Gaussian energy distribution across the footprint and the forested area in these footprints was towards the edge of the footprint. In footprints without distinct landscape changes toward their edges this would not meaningfully influence the utility of waveform metrics for prediction of terrain metrics. However in these cases, there was a significant disconnect between the two datasets. These two footprints were consequently removed from the analysis process.

Footprints 102 and 46, however, were included in the analysis because there was a good match between the airborne and satellite data and the difficulty in modeling these plots was due solely to issues with the canopy height model itself. The second and third research objectives for this paper are to examine how useful these GLAS waveform metrics are to predict and differentiate between terrain relief and canopy height, and to determine the extent of terrain influence on the utility of metrics to predict canopy height. Outliers 102 and 46 demonstrate that over moderate and high relief areas canopy height models may perform poorly. Dividing the footprints into relief classes prior to estimating canopy height was conducted in an attempt to minimize this problem.

Table 2-0-3. Information pertaining to outliers found from the preliminary canopy height model.

Footprint Number	*Max relief (m)*	*85th percentile hits height (m)*	*Predicted max relief (m)*	*Predicted hits height (m)*
102	14.40	0.62	10.52	13.08
46	31.25	29.52	34.27	13.63
85	31.89	6.77	13.80	12.06
101	26.4	20.05	11.12	10.83

2.5.7 Canopy Height Models with Terrain Relief Inputs

An analysis of the residuals from the preliminary canopy height model indicated that terrain relief influences the ability of a model based on these waveform metrics to predict canopy height (Figure 2-11). There is a decrease in the variability of the residuals as terrain increases, if the two model outliers are ignored. Contrary to expected findings, this

indicates that canopy height can be more reliably modeled if maximum terrain relief is greater than approximately 7 m.

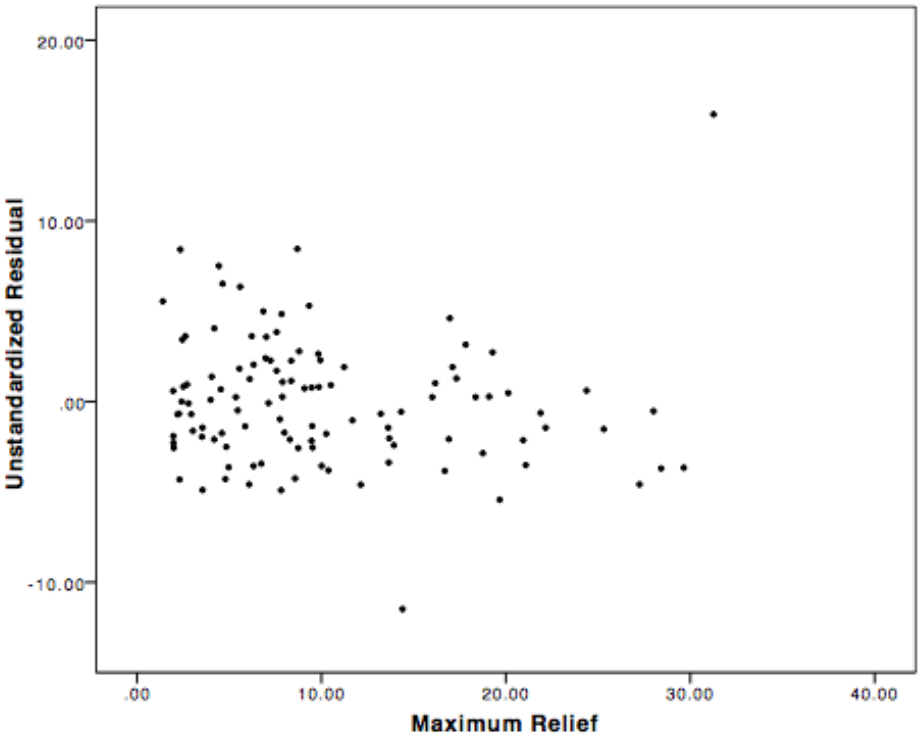

Figure 2-11. Residual plot of 85th percentile hits height prediction against maximum relief. Disregarding the outliers there is an apparent increase in the variability of residuals for footprints with maximum relief <7 metres and >15 metres.

Three relief classes were defined as 0-7 meters, >7-15 meters and >15 meters of maximum relief based on natural breaks in the data. The canopy height models for each individual relief class and the dummy variable model with associated R^2 values, sample sizes and standard errors are listed in Table 2-4. Other than the preliminary canopy height model, which has already been discussed, all of the variables in the models in Table 2-4, and their constants are statistically significant (p <0.05). Not only did terrain relief determine which variables best predicted canopy height but also how well the related

models performed. The model for the lowest maximum relief class had an R^2 of 0.83, which shows an improvement in performance over the original canopy height model. This model, however, requires five input variables. The moderate relief class' canopy height model predicted canopy height using only two independent variables, and had an R^2 of 0.88, indicating that it is empirically simpler to predict canopy height in areas of moderate relief rather than in low relief areas. The canopy height model for the third and highest relief class has an R^2 of 0.75, and three input variables. Although this is lower than the other relief classes, this model indicates that it is possible to glean canopy height information from waveforms even from areas of high relief.

In flat areas *startpeak* was the most important predictor variable because the highest peak in the waveform will almost always be from the ground, and consequently vegetation would be related to the distance between that peak and the beginning of the waveform signal. *Waveform_skew* was also more useful in flat areas as it varies based on where the ground signal was relative to the rest of the signal. *E_44*, the proportion of energy in the highest quarter which in flat areas will be entirely attributed to canopy, increasing as the bulk of biomass moves from the third quarter to the fourth.

The models for high relief replace *startpeak* as their most useful predictor variable with *wf_n_gs*. This was likely because the predominant ground peak typically apparent in flat areas and usually apparent in moderately sloped areas has been separated into several peaks and the distance from the peak that happens to have the highest maxima does not necessarily represent the average elevation of the underlying terrain. Consequently *startpeak* loses its usefulness in this class and was replaced by *wf_n_gs* as more peaks will be found in heavily vegetated areas than in unvegetated areas, regardless of the relief

of the terrain.

The canopy height models are found to perform best in areas of moderate relief. This was not an expected result, as typically it is assumed that the more complex waveforms associated with increasing relief are more difficult to interpret in terms of canopy structure (Lefsky et al., 2005). Figure 2-12 shows the relationship between *startpeak*, 85th percentile hits height and maximum terrain relief. This figure illustrates why areas of moderate relief yield more accurate predictions of canopy height. In the low and high relief classes *startpeak* does not always increase with increasing canopy height. In the moderate relief class, however, *startpeak* increases consistently with canopy height. In flat areas, startpeak will always be high for canopies over 25 meters, but for canopies between 7 and 25 meters startpeak varies considerably. One possible explanation is because of the denser understories associated with shorter, immature conifer canopies, the reflection off of the understory may be larger than the reflection off the ground. Over moderate relief areas, however, the ground signal is vertically stretched through the waveform as is the overlying understory return and canopy return. Consequently it is less likely that understorey reflection will be mistaken for ground reflection in areas of moderate relief, resulting in more accurate canopy height predictions.

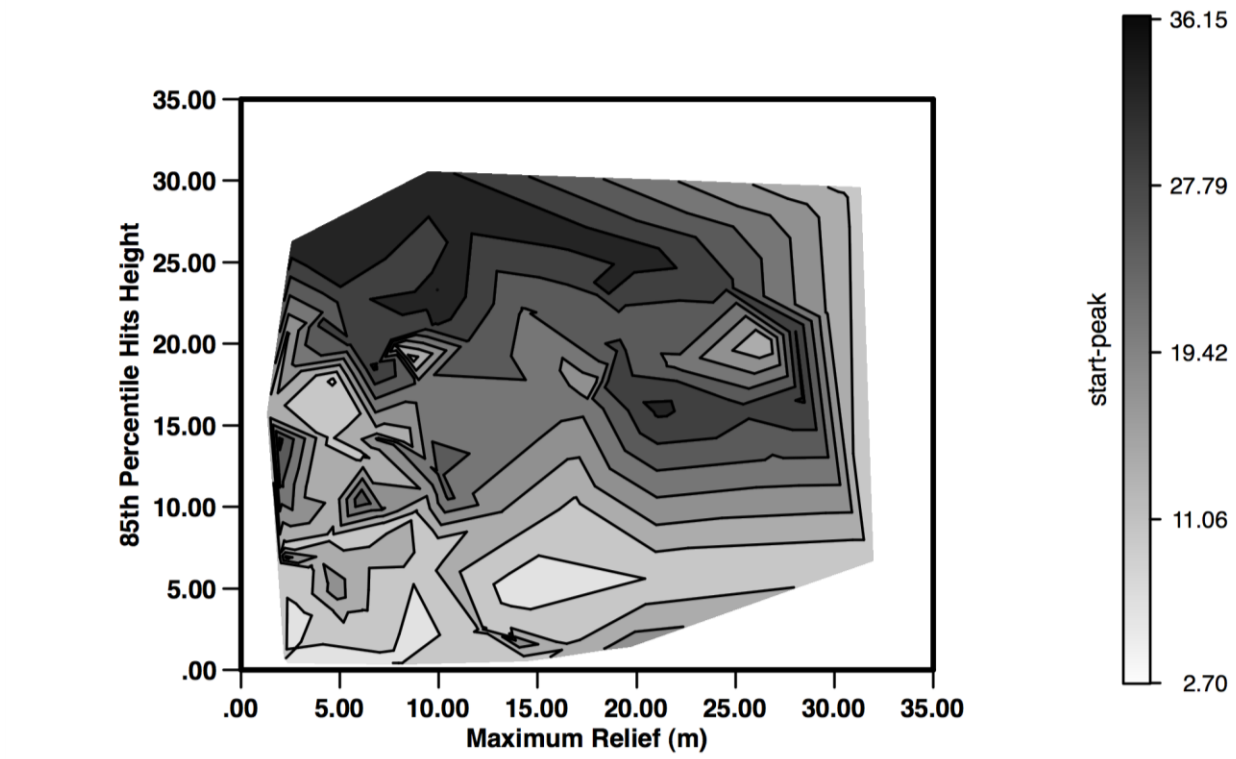

Figure 2-12. The relationship between maximum relief, 85th percentile hits height and startpeak. It is apparent that for areas of low and high relief the relationship between canopy height and startpeak is less consistent.

2.5.8 Dummy Variable Models

The results from the two dummy variable models are presented in Table 2-4, where the subscripts correspond to the relief classes. For the first dummy model a subscript of 0 corresponds to the maximum relief 0-7 meters class.

As with the individual relief class models, the dummy variable model's predictor variables and constants were statistically significant (p<0.05). This model was slightly more computationally demanding because dummy variables must be created after slope has been classified. Additionally, the number of input predictor variables in multiple

regressions should be minimized where possible. There is a trade-off between the sample size and the number of predictor variables used to create the model. However, it should be noted that the although there are numerous input variables in the dummy variable models, in the prediction of any given footprint only the input variables pertaining to the relief class in questions are actually used. The dummy models essentially represent a combination of the individual relief class models.

Table 2-0-4 Canopy height model information

Relief Class (m)	R^2	n	St. Error (m)	Model
All Data	0.81	110	3.63	85^{th}_ht=-0.544+0.665*startpeak+40.365*e_44-0.121*e_highest
0-7	0.83	44	3.04	$85^{th}_{0\text{-}7}$=1.758 + 0.609*startpeak + 75.192*e_44 - 9.280*e_34 - 0.084*e_highest - 0.318*energy_24
7-15	0.88	41	3.06	$85th_{7-15} = 5.348 + 0.808 * \text{startpeak} - 7.641 * \text{wf\_max\_e}$
>15	0.75	25	3.82	$85th_{>15} = -1.702 + 1.514 * \text{wf\_n\_gs} - 40.538 * \text{wf\_var} + 18.772 * \text{e\_14}$
Dummy Variable Model	0.85	110	3.28	$85th\_ht = 3.251 + 1.216 * \text{wf\_n\_gs}_2 + 0.843 * \text{startpeak}_1 + 0.562 * \text{startpeak}_0 - 6.418 * \text{wf\_max\_e}_1 - 46.843 * \text{wf\_var}_2 - 0.093 * \text{e\_highest}_0 + 73.755 * \text{e\_44}_0 - 10.754 * \text{e\_34}_0 + 13.758 * \text{e\_14}_2 - 0.320 * \text{energy\_24}_0$

2.5.9 Relief Classification

The utility of any of the models listed in Table 2-4 depends on the ability to successfully predict and/or classify maximum terrain relief. The discriminant analysis method correctly predicted the relief class (0-7m, 7-15m, >15m) of 71% of the footprints while

the regression-based method correctly predicted the relief class of 74% of the footprints.

These two classification methods were also compared by using the predicted classes from each method as inputs to the canopy height models. The results from classifying footprints into relief classes were used as inputs to the three related individual canopy height model and the dummy variable model. These plots are shown in Figures 2-13 and 2-14. Although the discriminant analysis method correctly classified a lower percentage of footprints, the predicted relief classes yielded a higher overall accuracy after inclusion in canopy height models, as seen in Figure 2-15. The three variables found to be useful to classify terrain relief into the specified relief classes by the discriminant analysis were *wf_variance*, *wf_skew* and *wf_n_gs*, all of which are logically related to terrain relief. Discriminant analysis is considered a more robust classification technique than multiple regression as it incorporates a measure of the covariance within each class as well as the distance to a class centroid for every case, which may explain its utility despite yielding a lower percentage of correctly classified footprints. Table 2-5 presents the classification results for the discriminant analysis classification. It is apparent from these classification results that it is rare for high relief footprints to be classified as low relief footprints, and vice versa. Unsurprisingly, the majority of misclassified footprints involved the moderate relief class.

Discriminant analysis classification may have misclassified more footprints than the prediction and classification method, but the misclassifications resulted in overall error in canopy height prediction when the classes were input into the canopy height models. As such, discriminant analysis was a more meaningful classification technique for this research.

The preliminary canopy height model is considerably less computationally demanding than the subsequent models which incorporate terrain relief and may seem more attractive. However, the presence of outliers in this model, the statistical insignificance of the constant the lower R^2 value and the higher standard error indicate that this model is less robust than the dummy variable model.

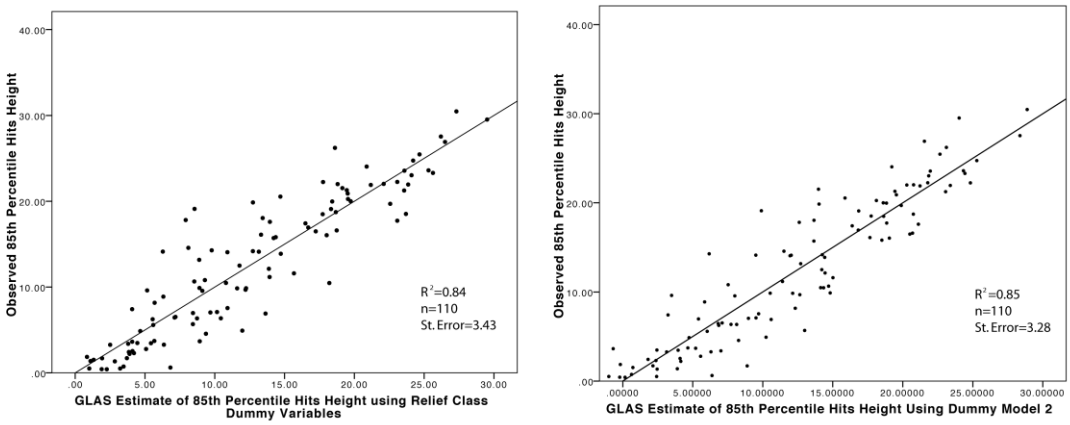

Figure 2-13. Results from dummy variable models.

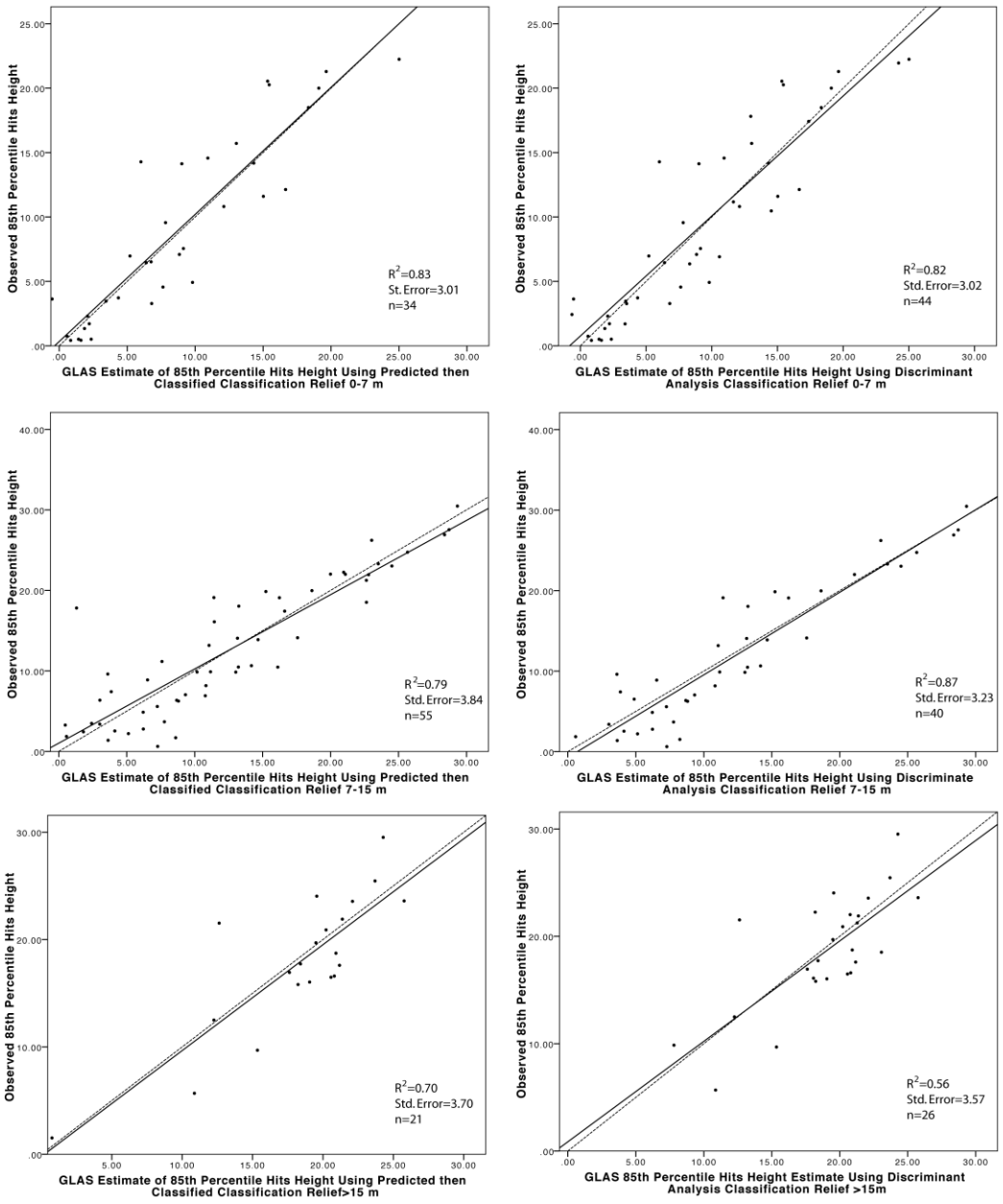

Figure 2-14. Results of classifying data and using classification as inputs to the three new forest height models for the three new relief classes. The dotted lines show the 1:1 line, the solid lines are the best fit lines.

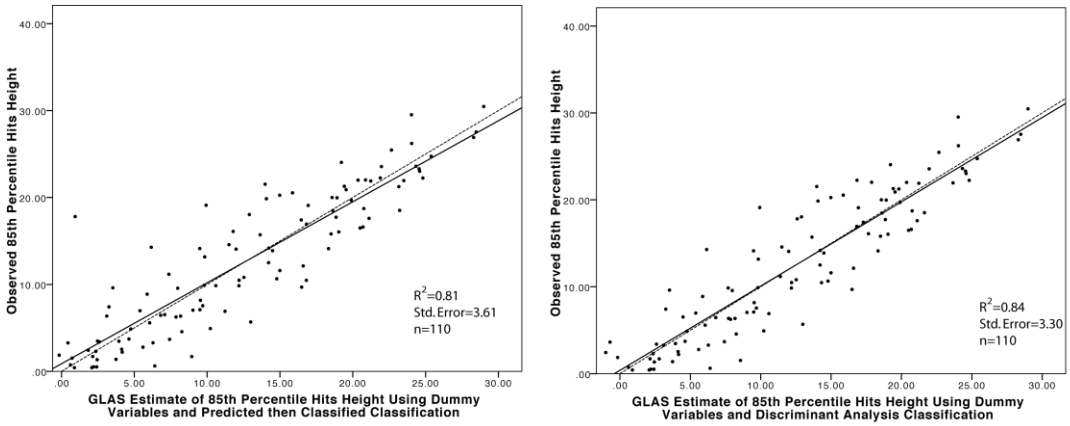

Figure 2-15. Results from Dummy variable model using relief class inputs from discriminant analysis.

2.5.10 Model Validation

Table 2-6 (maximum terrain relief) and 7 (canopy height) show the results of the validation of each of these models. The R^2_m corresponds to the model generated using a random selection of 80% of the cases. The R^2_v corresponds to a regression between the predicted and observed variables for the remaining 20% of the dataset. Both model validations show that the models generated using 80% of the cases can be used to predict terrain relief or canopy height for the remaining 20% of the cases with the expected accuracy.

Table 2-0-5. Model Validation Results for Maximum Terrain Relief Model.

Iteration	R^2_m	R^2_v
1	0.75	0.72
2	0.77	0.74
3	0.75	0.78
4	0.76	0.76
5	0.78	0.59
6	0.80	0.60
7	0.75	0.83

8	0.79	0.65
9	0.77	0.69
10	0.77	0.72

Table 2-0-6 Model Validation Results for Dummy Variable Canopy Height Model.

Iteration	R^2_m	R^2_v
1	0.86	0.75
2	0.85	0.84
3	0.86	0.83
4	0.84	0.92
5	0.86	0.87
6	0.86	0.85
7	0.84	0.88
8	0.85	0.86
9	0.87	0.80
10	0.85	0.86

2.6 Conclusions

In this studyIdemonstrated that maximum terrain relief can be directly predicted from GLAS waveform properties. Using an airborne scanning LiDAR dataset as a comparison,Ifound that 74% of footprints analyzed are correctly classified into relief classes of 0-7 meters, 7-15 meters and >15 meters of maximum relief. Terrain relief can also be classified using discriminant analysis and waveform metrics. Maximum terrain relief influenced how canopy height and structure affected other waveform characteristics, and consequently canopy height was modeled more accurately when terrain relief was incorporated into the models. The best methodology found in this paper was classify terrain relief using discriminant analysis and subsequently divide footprints into terrain classes that will have a dummy variable model applied to the entire dataset. I found that estimation of canopy height was more successful for footprints of moderate relief than for flat or high relief areas, likely due to my use of proportional energy metrics which require sufficient influence from vegetation on the waveform to be useful. In flat areas a ground peak tends to dominate waveforms, accounting for the majority of the energy and drowning vegetation signals. Waveforms in moderate relief areas (maximum relief 7-15 meters) have a more variable and complex energy return from which more information about both the terrain and the canopy can be gleaned when appropriate metrics are applied. However, as expected, in high relief areas the model was less successful due to the mixed nature of returns. When maximum relief is greater than 15 meters the individual class model explains less of the variability of canopy height, although 70.2% of the variability is explained which compares favourably with other

results from high relief terrain. The number of Gaussian peaks found in a Gaussian decomposition of waveforms was most useful in the prediction of canopy height in areas of high relief as the number of Gaussian peaks increases with the complexity and breadth of the signal.

In this paper I present a proof of concept that GLAS data can be used without ancillary data to characterize both terrain relief and canopy height. The finding that models developed for footprints over moderate slopes can be used to more reliably estimate canopy height than models developed for low relief areas was the first of its kind. One possible explanation was provided in this paper, however further research will address whether this is a function of the ecosystem investigated in this study, or whether this finding applies to other ecosystems. Similarly, tropical areas will be investigated to determine the influence of LAI and canopy structure on canopy height models.

2.7 References

Abshire, J.B., Sun, X., Riris, H., Sirota, J.M., McGarry, J.F., Palm, S., Yi, D., & Liiva, P. (2005). Geoscience Laser Altimeter System (GLAS) on the ICESat mission: on-orbit measurement performance. *Geophysical Research Letters*, 32, L21S02.

Blair, J.B., Hofton, M.A. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters*, 26, 2509-2512.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., & Kimes, D.S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, 112, 3876-3890.

Chen, J., Saunder, S.C., Crow, T.R., Naiman, R.J., Brosofske, K.D., Mroz, G.D., Brookshire, B.L., & Franklin, J.F. (1999). Microclimate in forest ecosystem and landscape ecology. *Bioscience*, 49, 288-297.

Foody, G.M., Cutler, M.E., McMorrow, J., Pelz, D., Tangki, H., Boyd, D.S., & Douglas, I. (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global Ecology & Biogeography*, 10, 379-387.

Harding, D.J., & Carabajal, C.C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32, L21S10.

Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., Neumann, K., Riedel, R., & Schroter, K. (2005). Global biomass mapping for an improved understanding of the CO_2 balance – the Earth observation mission Carbon-3D. *Remote Sensing of Environment*, 94, 94-104.

Hofton, M.A., Minster, J.B., & Blair, J.B. (2000). Decomposition of Laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing,* 38, 1989-1996.

Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment,* 82, 397-416.

Hurtt, G.C., Moorcroft, P.R., Pacala, S.W., & Levin, S.A. (1998). Terrestrial models and global change: challenges for the future. *Global Change Biology*, 4, 581-590.

Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29, 1339-1366.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Spies, T.A., Parker, G.G., & Harding, D. (1998). Lidar remote sensing of forest canopy structure and related biophysical parameters at the H.J. Andrews Experimental forest, Oregon, USA. *Geoscience and Remote Sensing Symposium Proceedings, IGARSS 1998*, 1252-1254.

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 Sensing of Environment*, 70, 339-361.

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding D.J. (2002). LiDAR remote sensing for ecosystem studies. *BioScience*, 52, 19-30.

Lefsky, M.A., Hudak, A.T., Cohen, W.B., & Acker, S.A. (2005). Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment*, 95, 517-531.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, 1, 013537.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27, 88-106.

Markwardt, C. (2008). Non-linear least squares fitting in IDL with MPFIT. *Astronomical Data Analysis Software and Systems XVIII.* Vol. TBD, p. TBD.

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., & Hendrix, C.J. (2000). Predicting forest stand characteristics with airborne scanning LiDAR. *Photogrammetric Engineering & Remote Sensing*, 66, 1367-1371.

Naesset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning Laser. *Scandinavian Journal of Forest Research*, 19, 164-179.

Nelson, R., Valenti, M.A., Short, A., & Keller, C. (2003). A multiple resource inventory of Deleware using airborne Laser data. *BioScience*, 53, 981-992.

Parker, G.G., Harmon, M.R., Lefsky, M.A., Chen, J., Van Palt, R., Weiss, 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*, 7, 440-453.

Patenaude, G., Hill, R.A., Milne, R., Gaveau, D.L.A., Briggs, B.B.J., & Dawson, T.P. (2004). Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, 93, 368-380.

Pedersen, L., & Forester, C. (2000). Tree Farm License 18: rationale for allowable annual cut (AAC) determination. *British Columbia Ministry of Forests.*

Potter, C. (1999). Terrestrial biomass and the effects of deforestation on the global carbon cycle. *BioScience*, 49, 769-778.

Ranson, K.J., Kimes, D., Sun, G., Nelson, R, Kharuk, V., & Montesano, P. (2007). Using MODIS and GLAS data to develop timber volume estimates in central Siberia. *Geoscience and Remote Sensing Symposium Proceedings, IGARSS 2007*, 2306-2309.

Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., & Dobson, C. (2003). A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science and Policy*, 6, 441-455.

Rosette, J.A.B., North, P.R.J., & Suarez, J.C. (2008). Vegetation height estimates for a mixed temperate forest using satellite Laser altimetry. *International Journal of Remote Sensing*, 29, 1475-1493.

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reevs, M. & H. Hashimoto. (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience*, 54, 547-560.

Schutz., B.E., Zwally, H.J., Shuman, C.A., Hancock, D., & DiMarzio, J.P. (2005). Overview of the ICESat Mission. *Geophysical Research Letters*, 32, L21S01.

Sun, G., Ranson, K.J., Kimes, D.S., Blair, J.B., & Kovacs, K. (2008). Forest vertical structure from GLAS: An evaluation using LVIS and ARTM data. *Remote Sensing of Environment*, 112, 107-117.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M.

(2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 18, 206-314.

Wulder, M.A., Hall, R.J., Coops, N.C. & Franklin, S.E. (2004). High spatial resolution remotely sensed data for ecosystem characterization. *BioScience*, 54, 511-521.

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., & Gerard, P.D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, 87, 171-182.

Zwally, H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., & Thomas, R. (2002). ICESat's Laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34, 405-445.

3. INTEGRATION OF GLAS AND LANDSAT TM DATA FOR ABOVEGROUND BIOMASS ESTIMATION

3.1 Abstract

Current regional aboveground biomass estimation techniques, such as those that require extensive field work or airborne LiDAR data for validation, are time and cost intensive. The use of freely available satellite-based data for carbon stock estimation mitigates both the cost and the spatial limitations of traditional, field based, techniques. Spaceborne LiDAR data have been demonstrated as useful for aboveground biomass (AGBM) estimation over a wide range of biomass values and forest types. However, the application of these data is limited because of their spatially discrete nature. Spaceborne multispectral sensors have been used extensively to estimate AGBM, but these methods have been demonstrated as inappropriate for forest structure characterization in high biomass, mature forests. This study uses an integration of ICESAT Geospatial Laser Altimeter System (GLAS) LiDAR and Landsat optical data to develop methods to estimate AGBM in an area of south central British Columbia, Canada. I compare estimates to a reliable AGBM map for the area derived from high resolution airborne LiDAR data to assess the accuracy of satellite-based AGBM estimates. Further, I use the airborne LiDAR dataset in combination with forest inventory data to explore the relationship between model error and canopy height, AGBM, stand age, canopy rugosity, mean DBH, canopy cover, terrain slope, and dominant species type. GLAS AGBM models were shown to reliably estimate AGBM (R^2=0.78) over a range of biomass

conditions. Inclusion of terrain relief in these AGBM models is found to improve model performance. A partial least squares AGBM model using Landsat input data to estimate predictions from the best GLAS AGBM model estimates had an R^2 of 0.6, and were found to underestimate AGBM by an average of 26 Mg/ha per pixel when applied to areas outside of the GLAS transect. This study demonstrated that Landsat and GLAS data integration are most useful for forests with less than 120 Mg/ha of AGBM, less than 60 years of age, and less than 60% canopy cover. These techniques have high associated error when applied to areas with greater than 200 Mg/ha of AGBM.

3.2 INTRODUCTION

The need exists to develop a systematic approach to inventory and monitor global forests, both for carbon stock evaluation and land use change analysis (Rosenqvist et al., 2003). Carbon stock estimation is also important given the development of international carbon credit trading schemes, which can be utilized to reduce deforestation and degradation, particularly in developing nations (Gibbs et al., 2007). Remote sensing provides the ability to collect data from remote forested areas that would otherwise be costly and logistically difficult to collect. Various remote sensing technologies have been demonstrated as useful for different aspects of forest research. Airborne hyperspectral remote sensing instruments have been demonstrated to provide highly accurate information pertaining to species type and composition (Clark et al., 2005, Van Aardt & Wynne, 2007), while airborne LiDAR instruments provide the ability to accurately estimate tree height (Means et al., 1999), volume and AGBM or carbon content (Patenaude et al., 2004, Nelson et al., 2003). As such, technologies are available that can be used to study both forest function and structure. However, these airborne technologies have high associated data volumes and costs (Wulder et al. 2008). Consequently they are best suited for studying forests over relatively small, localized areas, or for use in sampling strategies (Wulder & Seemann, 2003).

Satellite-based remote sensing technologies provide the ability to collect data over regional, national and global scales. Multispectral satellite data, such as collected by the Landsat Thematic Mapper, have been widely used for landscape classification and change detection in forested areas (Wulder, 1998). Studying forest structural properties

with multispectral data is more problematic, although many studies have attempted to model forest structural parameters using multispectral imagery with varying levels of success (Hall et al., 1995, Cohen and Spies, 1992, Cohen et al., 2003, Zheng et al., 2004). NDVI and Tassel Cap Transformations are two of the most common types of vegetation indices used in multispectral studies of forest structure (Rogan et al., 2002). Freitas et al., (2005) found that forest structural properties could be estimated using Landsat derived vegetation indices (R^2 = 0.49 - 0.72) for a tropical rain forest. Zheng et al. (2004) used Landsat 7 ETM+ data to estimate AGBM in managed forest in Wisconsin with an R^2 of 0.67. Hall et al. (2006) found that stand height (R^2 = 0.65) and crown closure (R^2 = 0.57) could also be estimated from Landsat ETM+ data in an area of Alberta with similar forest cover to the study area used in this research.

When using vegetation indices in combination with raw Landsat bands, collinearity is a common issue (Hensen & Schjoerring, 2003). One method to extract the most meaningful information from multispectral data is partial least squares (PLS) regression. PLS regression is a statistical technique used to reduce collinearity, and/or data dimensionality, which reduce the utility of ordinary least squares (OLS) regression (Næsset et al., 2004). PLS involves the extraction of linear combinations of the original input variables, and produces a new set of latent variables than can then be used to model the desired output, in this case AGBM. PLS has been demonstrated as useful for studies of biomass using optical data (Cai et al., 2009; Hensen & Schjoerring, 2003).

Despite attempts to model biomass directly from Landsat data, studies have demonstrated that models developed to estimate forest structure from multispectral data are sensitive to tree height (Donoghue & Watt, 2006), shadows (Hall et al., 1995), and

high biomass densities (Foody et al., 2001) Thus, although the spatially continuous nature of optical satellite imagery renders it ideal for studying phenomena over large areas, regional analyses of forest structural properties require information about the vertical distribution of biomass in a forest. Studies using optical imagery to model such vertical properties are limited because optical data integrate the vertical profile of a forest into a single reflectance value per pixel (Hall et al., 1995).

Satellite-based LiDAR provides an opportunity to accurately estimate AGBM from space. Recent studies have focused on using spaceborne LiDAR data to accurately model forest height and AGBM. NASA's GLAS instrument, on-board ICESat, is a full waveform satellite-based LiDAR system that collects data from 65 m diameter footprints on the Earth's surface (Schutz et al., 2005). The returned waveforms from these footprints can be used to estimate tree height, AGBM and basal area (Lefsky et al., 2007, Sun et al., 2008, Ranson et al., 2007, Rosette et al. 2008). The accuracy of these models was found to depend on both canopy density and terrain slope. Increasing density and slope increase the complexity of GLAS waveform footprints, and in high relief or high biomass areas it can be difficult to resolve ground reflection from canopy reflection within waveforms. This confusion can be reduced through the introduction of ground filtering techniques. Duncanson et al. (2009) developed a methodology to estimate and account for terrain relief in GLAS AGBM estimates. It was determined that adjusting for terrain relief increases the accuracy of GLAS AGBM models.

A number of issues render GLAS data less than ideal for large area forest inventories. The first is the limited nature of the data coverage. The design of the satellite orbit yields nearly contiguous coverage of the polar regions. The swaths diverge away

from the poles however so that other areas are covered by discrete orbits. Additionally, the GLAS laser samples along the flight path so that a 65 metre diameter footprint is spaced every 172 metres. This sampling scheme yields an incomplete coverage. A second issue involves atmospheric noise. As LiDAR is a line of sight sensor it is affected by clouds and only cloud free waveforms are appropriate for terrain analysis. Finally, due to technological issues early in the GLAS mission, GLAS does not continually sense the Earth's surface but instead operates in three annual 33-day campaigns (Schutz et al., 2005). Considering these limitations, the GLAS data available from the National Ice and Snow Data Distribution Centre (NISDC) is not spatially continuous but represents a series of transects. The distribution of footprints available from any region depends largely on the location of the region and the atmospheric conditions at the time of GLAS passage. Although GLAS data has the potential to be useful for the estimation of AGBM and carbon content of forests, these logistics limit the applicability of GLAS-derived models for regional biomass mapping.

Integrating the two-dimensional, spatially continuous data from Landsat TM with the three-dimensional, spatially discrete GLAS data may provide a systematic method to estimate AGBM over large areas, on regional, national and even global scales. Studies have integrated optical and LiDAR technology in attempts to either extrapolate LiDAR forest structure measurements to areas with similar spectral properties as those sensed by LiDAR (Hudak et al., 2002, Donoghue & Watt, 2006) or to more meaningfully interpret results from LiDAR studies (Wulder et al., 2007). This study employs a regression-based AGBM modeling approach, similar to those used in other Landsat AGBM studies. This approach was selected because it does not rely on field data, and can be applied using

only freely available satellite data. The availability of high-resolution airborne LiDAR data and a current forest inventory for the entire study area allows us to evaluate the accuracy of Landsat and GLAS integration AGBM estimates on a pixel by pixel basis. This allows the comparison of model error with terrain characteristics in order to improve my understanding of the utility of these methods. The objectives of this study are to:

1. Develop a method to model AGBM from GLAS and Landsat integration;
2. Determine the relationships between model error and forest cover properties, such as dominant species and stand age; and
3. Establish reliable ranges of forest structural properties for which GLAS and Landsat data integration is appropriate for AGBM estimation.

3.3 METHODS

3.2.1 Study Area

Airborne LiDAR data were acquired for a 75,000 ha area of industrially managed forest near Clearwater, British Columbia, Canada. One GLAS transect was selected from the available GLAS data for the area (Figure 3-1). The predominant species in the area are lodgepole pine (*Pinus* contorta*)*, balsam fir (*Abies balsamea*), Douglas-fir (*Pseudostuga menziesii*), western hemlock (*Tsuga heterophylla*) and western redcedar (*Thujia plicata*), Engelmann spruce (*Picea engelmannii*), white spruce (*Picea glauca*) (Pedersen and Forester, 2000). The area is characterized as a high elevation plateau of gently rolling terrain with an elevation range of approximately 800 metres (Pedersen and Forester, 2000). The GLAS data selected for this study represent the transect seen in Figure 3-1.

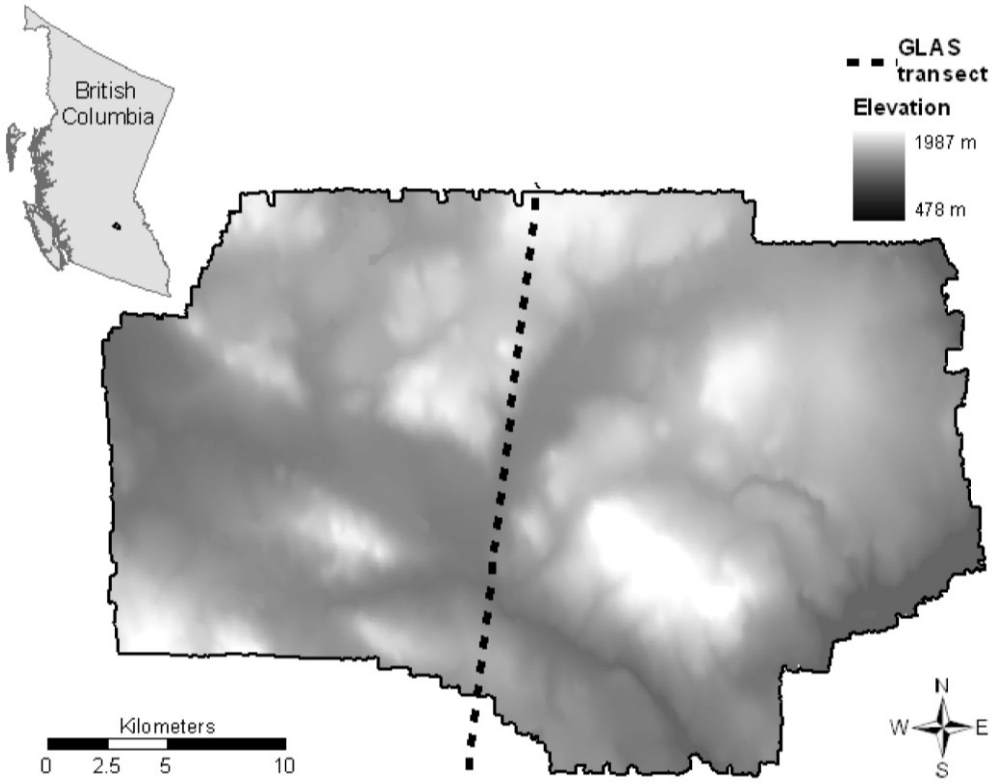

Figure 3-1 Study area map showing GLAS transect and elevation distribution across study area.

3.2.2 Data

3.2.2.1 Airborne LiDAR Data

All airborne data collection and processing has been reported in depth in Niemann (2009). Airborne scanning Laser LiDAR data were acquired August 16, 2006 using the UVic-TERRA Multisensor Airborne Platform (MAP). This system has an itegrated LiDAR and full range (400 to 2500 nm) hyperspectral imaging system. The data used for this project consists of discrete, first and last return 100 kHz LiDAR with a maximum +/- 20° scan angle and 37 Hz scanning speed. The Pulse Repetition Frequency (PRF) of the

dataset used was 60 KHz. The platform was flown at 1600 meters above ground which in combination with the PRF resulted in approximately 1.25 points per square metre at nadir, and 2.25 LiDAR returns per square meter when side overlap was included. Processing was conducted to separate ground returns from those of vegetation. A bare earth model and canopy height model (CHM) were subsequently generated; the CHM heights were calculated as the difference between the elevation of a return and the coincident elevation of the bare earth model. These data were also used to calculate terrain slope and rugosity (square root of variance) of hits height, which were used to explore model results.

3.2.2.2 Field Data

Field data were collected during a four day period during October, 2008. Diameter at Breast Height (DBH), tree species, and tree height were measured for every tree within a 10 meter diameter plot for 35 plots representative of the cover classes found in the study area. Plots were positioned to within 2 metres absolute accuracy using post processed differential GPS.

3.2.2.3 AGBM Map

A 20 m by 20 m AGBM map was generated for study area using ordinary-least squares regression to fit field measured AGBM to a series of airborne LiDAR metrics (for a complete discussion of the modeling please refer to Niemann (2009)). The natural logarithm of the total AGBM (dry weight) was found to be a function of the 55^{th} percentile of the CHM height and the natural logarithm of canopy cover. This model is shown in equation 1.

$$\ln TAGB = 0.0914 + 1.0496 \ln Lh_{0.55} + 0.5719 \ln CC \quad (1)$$

where TAGB is total AGBM (Dry Weight) in Mg/ha, $Lh_{0.55}$ is the 55^{th} percentile of the canopy height model, and CC is laser canopy cover (%). This model has an R^2 value of 0.89, with a RMSE of 46.07. This AGBM map was deemed sufficiently accurate to provide unique wide-area calibration/validation for AGBM estimates made using the GLAS and Landsat data.

3.2.2.4 Satellite-based LiDAR

GLAS is a full waveform LiDAR sensor using a 1064 nanometer Laser operating at 40 Hz. The GLAS data used in this study were acquired June 26, 2006 and represent a transect with 105 GLAS waveforms. This was the only temporally coincident transect available in the study site from GLAS Laser 3 data. Laser 2 transects were available, but where laser 3 footprints approximate 65 diameter circles, laser 2 footprints are more elliptical and the spatial pattern of light reflection varies considerably between footprints. As such, the airborne LiDAR data could not be reliably matched to Laser 2 footprints and these were not included in the analysis. Details on the processing of these data can be found in Duncanson et al. (2009). A series of waveform metrics were calculated describing the shape of each waveform in terms of both their vertical (temporal) distribution and energy return distribution. Table 3-1 provides a description of the waveform metrics used in this study.

Table 3-1. Waveform Metrics and Abbreviations.

Metric Abbreviation	Metric
wf_max_e	The highest energy value in the waveform
wf_variance	The variance of the waveform
wf_skew	The skew of the waveform

Metric Abbreviation	Metric
e_44	Proportion of energy in highest elevation quarter
e_34	Proportion of energy in second highest elevation quarter
startpeak	The difference in elevation between the beginning of the signal and the position of wf_max_e
wf_n_gs	The number of Gaussian curves found in the waveform

3.2.2.5 Satellite Multispectral

A single Landsat TM scene was selected for use in this study based on its temporal proximity to the acquisition of both the GLAS and airborne LiDAR data, and based on the atmospheric conditions of the scene (<1% cloud). The image was acquired on August 17th, 2006. It is unlikely that noteworthy change in AGBM occurred between the acquisition of the GLAS data and certainly very little in the day between the airborne LiDAR and Landsat acquisition. Top of atmosphere correction was applied using the methodology presented in Han et al. (2007).

3.2.2.6 Ancillary data Map

The British Columbia Vegetation Resources Inventory (VRI) (Ministry of Sustainable Resource Management, 2003) was used to explore the results found in this study. The variables explored using this inventory were canopy height, stand age, % canopy cover, dominant species and mean Diameter at Breast Height (DBH).

3.2.3 Analysis

This study involved the development of several models to estimate AGBM from various sets of input data. First, a relationship was developed between field data and airborne LiDAR data, and a reliable AGBM map for the study area was generated. This AGBM map acted as the control, and AGBM values from this map were the independent variables for subsequent models. Second, the AGBM map values were modeled from GLAS waveform metrics. Third, GLAS AGBM estimates were modeled from Landsat data. The GLAS estimates were modeled from Landsat rather than the AGBM map values in an attempt to eliminate the reliance on expensive data. The Landsat model was applied to the entire study area and the difference between the observed (AGBM map) values and the predicted (Landsat GLAS) values were evaluated for different species types and forest structural characteristics using forest inventory data and airborne LiDAR. Figure 3-2 provides a visual representation of this analysis framework.

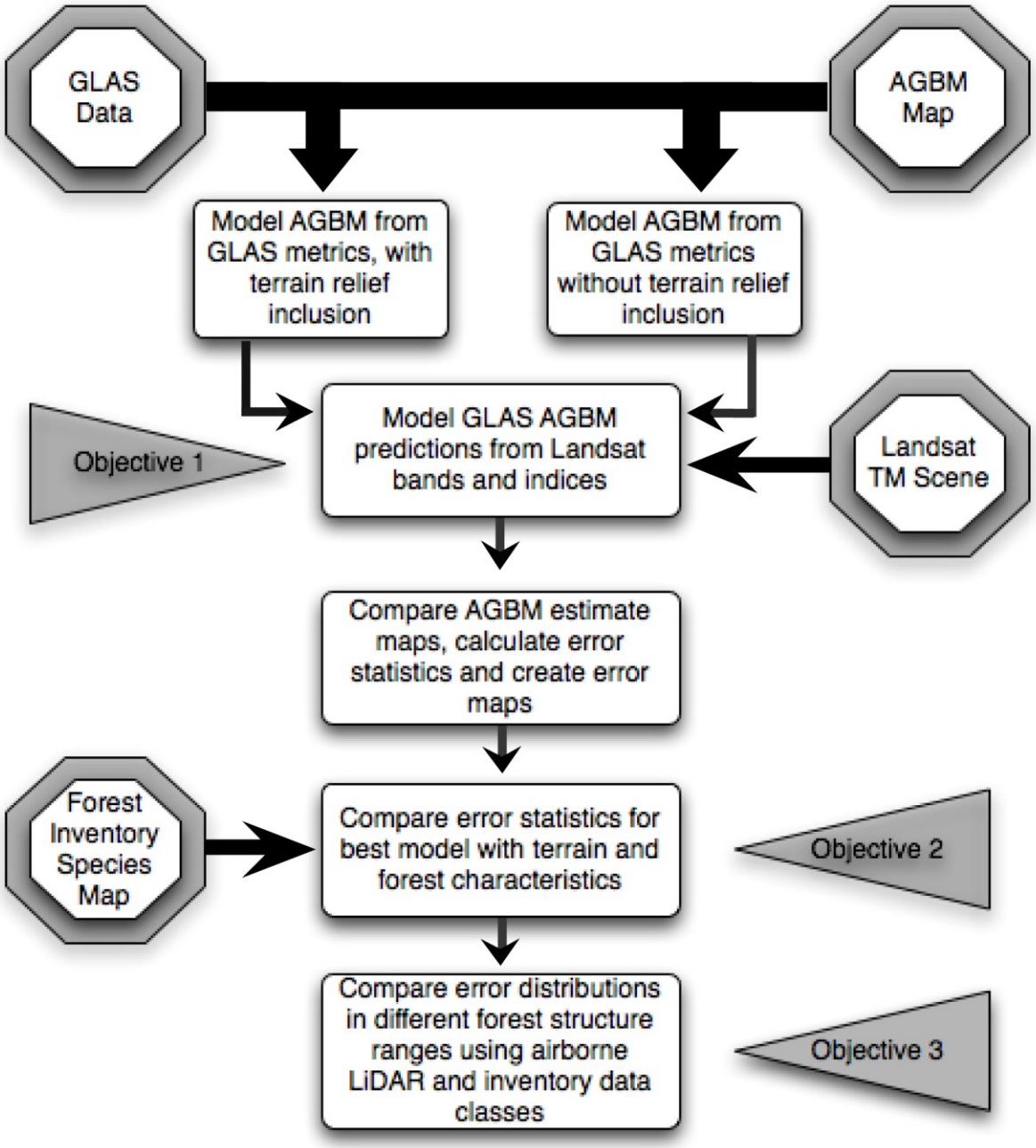

Figure 3-2. Analysis Framework. The results and discussion sections proceed through this framework sequentially.

Successfully modeling AGBM from GLAS waveforms relied on the assumption that the AGBM map could be manipulated to provide a single AGBM estimate for the 65 meter diameter circle illuminated by each GLAS pulse. To achieve this, 65 meter diameter circles, with centroids located at each GLAS centroid, were overlain over the AGBM map. The AGBM map was translated into polygons, which were then clipped to the overlaid GLAS footprints. An average of 12 AGBM map pixels fell within, or partially within each GLAS footprint. A weighted average was used to calculate the average AGBM in Mg/ha for each GLAS footprint by multiplying each polygon's AGBM value by the area within the GLAS footprint, summing the resulting values for each footprint and dividing by the area of the footprint. Thus, I generated a dataset representing a 'control' value of AGBM for the area corresponding to each GLAS pixel.

These AGBM values were then root transformed to reduce the positive skew of AGBM for the scene (Figure 3-3), and improve normality and linearity among variables. Following the methods presented in Duncanson et al. (2009) two ordinary least squares models were developed to predict these AGBM values from GLAS data; the first model used GLAS waveform metrics to directly model the square root of AGBM, and the second model incorporated terrain relief by using binary coded dummy variable GLAS waveform metrics. Only variables with less than a correlation of less than 0.5 with any other input variable were used in this model. The dummy variables were coded based on maximum terrain relief estimated from GLAS waveform metrics, without reliance on ancillary topographic data, as outlined in Duncanson et al. (2009). Dummy variables were used to adjust for different relationships between waveform metrics and AGBM in flat (0-7 metres), moderate (7-15 metres) or high (>15 metres) relief areas. The

predictions from each model were squared to produce predictions of AGBM in Mg/ha for each GLAS footprint. These two approaches yielded two sets of estimated AGBM for the 105 footprints, the latter adjusting for terrain relief.

Both ordinary least squares (OLS) and partial least squares (PLS) approaches were used to estimate the relationship of the Landsat data to each set of GLAS AGBM estimates. Due to the 65 meter spatial resolution of the GLAS data, the Landsat image, AGBM map, and forest inventory map were resampled to 60 meter pixels, using nearest neighbour resmpling. I rely on the assumption that the Landsat pixel coincident to a GLAS centroid location represents the multispectral response corresponding to a GLAS waveform. Landsat TM bands 1, 2, 3, 4, 5, and 7, as well as NDVI and three tassel cap transformation bands (brightness, greenness, wetness) were used as potential inputs into OLS and PLS regression models to estimate the square root of AGBM.

The airborne LiDAR AGBM (observed AGBM) map was subtracted from each of the Landsat estimate AGBM maps. To ensure that only areas forested at the time of Landsat and airborne LiDAR data acquisitions were included in the study, the airborne LiDAR were used to mask out zero AGBM pixels. The resulting error images, representing the over and under prediction of AGBM for each pixel, were therefore masked for zero biomass. A histogram of error was generated for each model, and the mean and standard deviation of error distributions were calculated.

The VRI maps of dominant species, canopy height, age, % canopy cover and mean DBH were used to decompose both the original AGBM map and the selected error map in order to explore differences in AGBM and error in areas of differing land cover. The mean observed AGBM (Mg/ha), standard deviation of observed AGBM (Mg/ha),

and standard deviation of error (Mg/ha) were calculated for the area associated with each dominant species. Box and whiskers plots were created to visualize trends in error over increasing canopy height, AGBM, age, rugosity, mean DBM, % canopy cover and terrain slope.

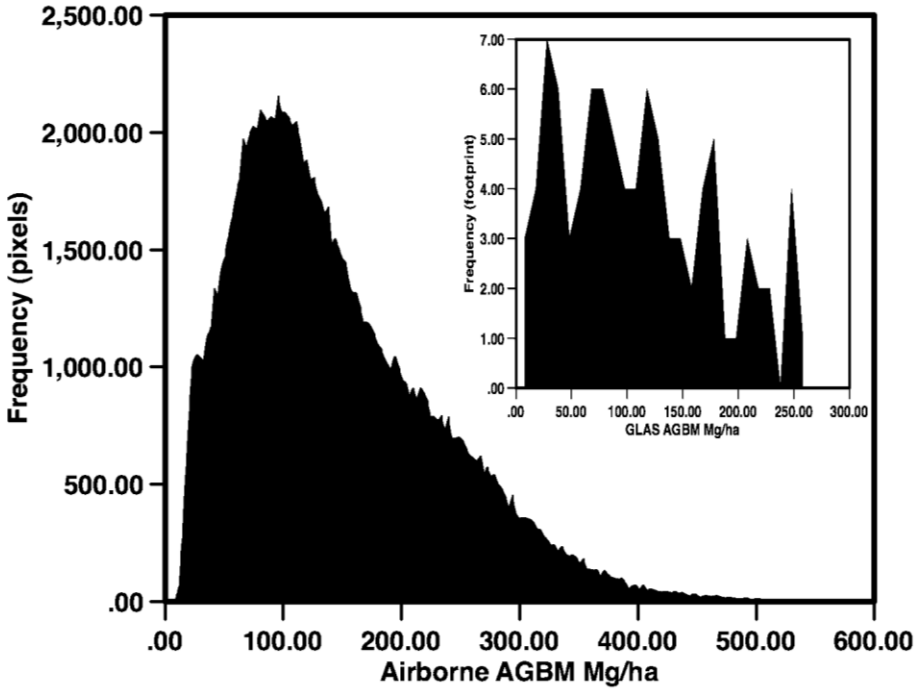

Figure 3-3 AGBM distribution for the study area, as estimated from airborne LiDAR data. The subset shows the AGBM distribution from the GLAS transect.

3.3 RESULTS

3.3.1 GLAS Model Results

The results of the comparison of the predicted versus the observed biomass estimates are summarized in Table 3-2. Figure 3-4 shows the results of the best GLAS model (model 1 in Table 3-2). Although the R^2 values and standard errors of models 1 and 2 are comparable, the estimates from model 1 are more evenly distributed around the best fit line.

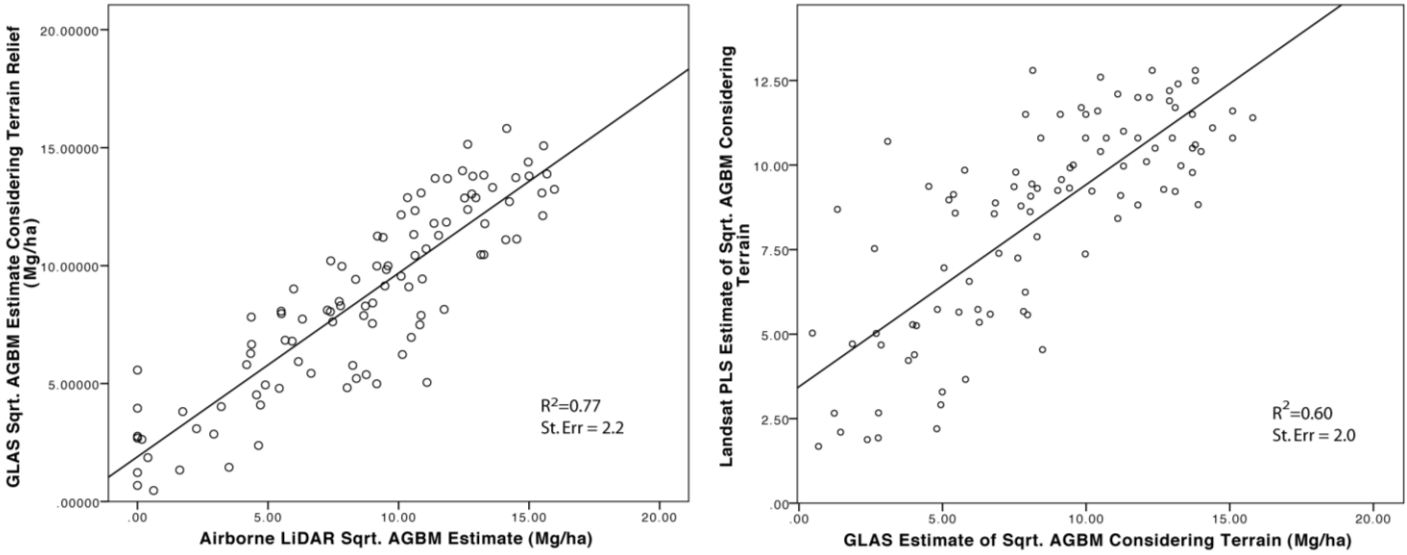

Figure 3-4. Results from Models 1 and 3, GLAS and Landsat estimates of the square root of AGBM in GLAS footprints.

Table 3-2. AGBM Model Information.

Input Data	Model Equation	R^2	St. Err
GLAS	(1) $AGBM^{0.5}_{\text{terrain inclusion}}$ $=4.21+0.13(\text{startpeak}_2)+3.86(\text{wf\_skew}_0)+0.42(\text{startpeak}_1)-22.41(\text{wf\_variance}_1)-32.05(\text{wf\_variance}_2)+0.51(\text{wf\_n\_gs}_2)-1.40(\text{wf\_max\_e}_0)$	0.77	2.2
GLAS	(2) $AGBM^{0.5}_{\text{no terrain inclusion}} = 4.94+0.27(\text{startpeak})-12.10(\text{wf\_variance})+24.23(\text{e\_44})-4.84(\text{e\_34})$	0.78	2.2
Landsat (PLS)	(3) $AGBM^{0.5}_{\text{terrain inclusion}} = 20.84-193.00(\text{Band7})+75.82(\text{Band5})-51.58(\text{Band4})+220.03(\text{Band3})-312.8(\text{Band2})131.28(\text{Band1})$	0.60	2.0
Landsat (PLS)	(4) $AGBM^{0.5}_{\text{no terrain inclusion}}= 11.41- 237.41(\text{Band1})+74.05(\text{Band2}) -56.19(\text{Band3}) +249.25(\text{Band4}) -275.20(\text{Band5})+280.26(\text{Band7})$	0.55	2.0
Landsat (OLS)	(5) $AGBM^{0.5}_{\text{high biomass inclusion}} = 9228.27-919.28(\text{T1})+0.373(\text{Band3})$	0.46	5.1

3.3.2 Landsat Model Results

Table 3-2 lists the results from the two PLS models (with and without terrain relief inclusion) and the OLS model. Models 3 and 5 estimated the GLAS estimates from model 1 using PLS or OLS, respectively. Model 4 estimated the GLAS estimates from model 2. Figure 3-4 shows the results from model 3, the Landsat model with the highest accuracy and therefore selected for error analysis. Figure 3-5 shows the spatial distribution of the model 3 estimates subtracting the observed airborne AGBM. For both PLS models 3 components, or latent variables, were used.

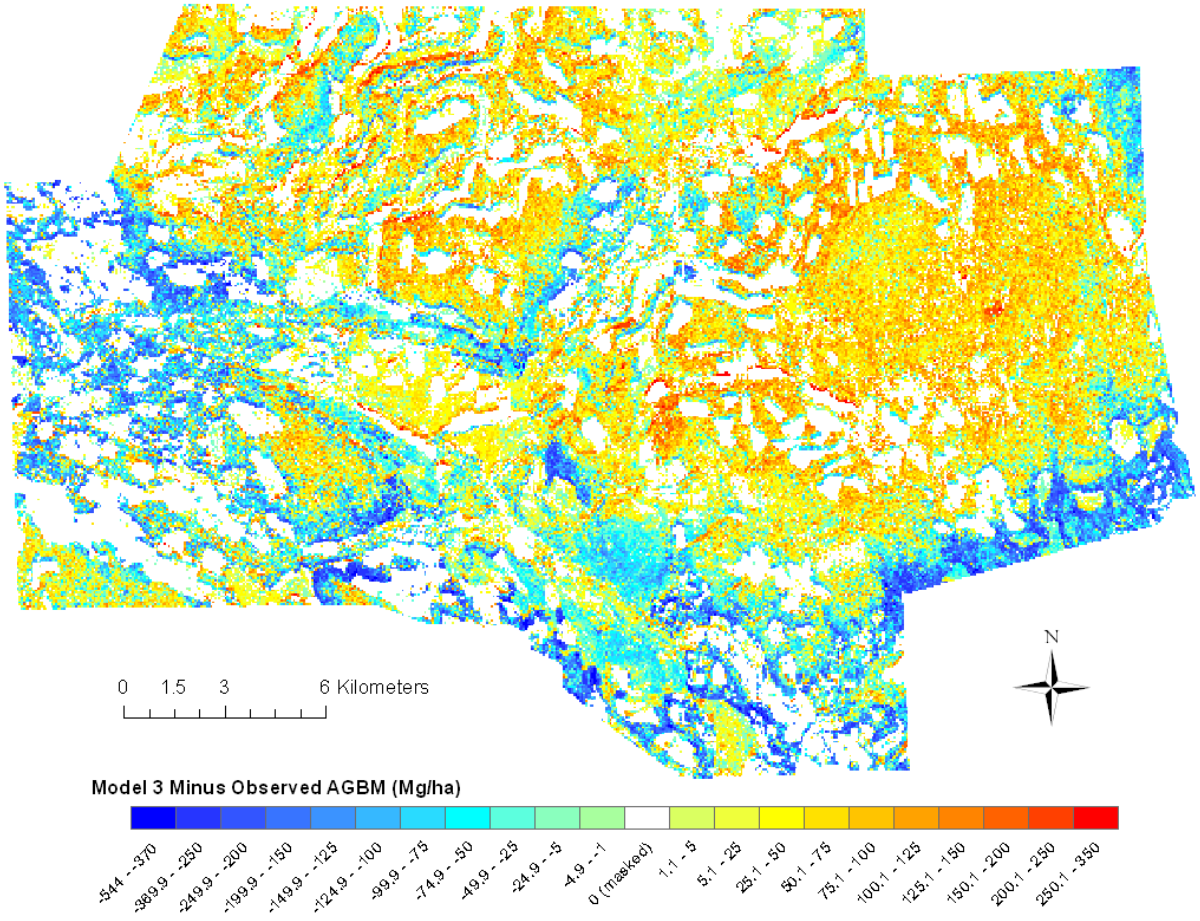

Figure 3-5 Model 3 spatial distribution of error (Model minus Observed AGBM per pixel).

3.3.3 Error Image Decomposition

The standard deviation for the study area was 89.3, the mean AGBM was 146.5 Mg/ha, and the standard deviation of AGBM was 85.9 Mg/ha. The error associated with model 3 was decomposed based on terrain and forest cover characteristics derived from the airborne LiDAR data or the forest inventory. The standard deviation, mean AGBM, and standard deviation of AGBM related to the dominant species are shown in Figure 3-6. Douglas-fir, immature lodgepole pine, trembling aspen, birch and paper birch are the

species for which the model performed poorly. Douglas-fir had a high average AGBM and standard deviation of AGBM. The areas of immature lodgepole pine tended to be overestimated. Balsam, subapline fir and white spruce were the species for which the model was most accurate.

Figure 3-7 shows how error mean, error median and 10^{th}-90^{th} percentiles of error change with increasing canopy height, AGBM, age, rugosity, mean DBH and % canopy cover. The masking of zero AGBM data from the analysis results in the lowest class of many of these parameters having no data. The class divisions and number of pixels within each class are shown in Table 3-3.

There is an overestimate of approximately 30 Mg/ha of AGBM when AGBM is less than approximately 120 Mg/ha. After 200 Mg/ha there is a large underestimation of AGBM, and the underestimation increases dramatically.

The relationship between age and model error is more complicated than many of the other relationships. Forests between 1 and 40 years of age have relatively small error distributions, and AGBM is overestimated. Between 41 and 60 years of age the model yields a good estimate of AGBM, but the distribution of error ranges from over predictions of 100 Mg/ha to under predictions of the same magnitude. Between 61 and 120 years of age the model under estimates AGBM, as do forests older than 250 years, and these have the widest range of errors. Interestingly, stands between 141 and 250 years of age show overestimates of AGBM and narrower error distributions.

The relationship between model error and % canopy cover is also non intuitive. Between 0 and 60% canopy cover, increasing canopy cover appears to cause increasing

over predictions of AGBM. After approximately 60%, however, the distribution of error increases.

The model overestimated AGBM when canopy height was less than 20 metres, or greater than 30 metres. The distribution of error increases from 11-15 m to 16-20 m and again for stands with greater than 20 meter trees. These methods become unreliable after approximately 20 meters of canopy height.

As mean DBH increases so does the distribution of error. Similarly, there appears to be a slight increase in the distribution as rugosity increases, but the results appear to be generally inconclusive.

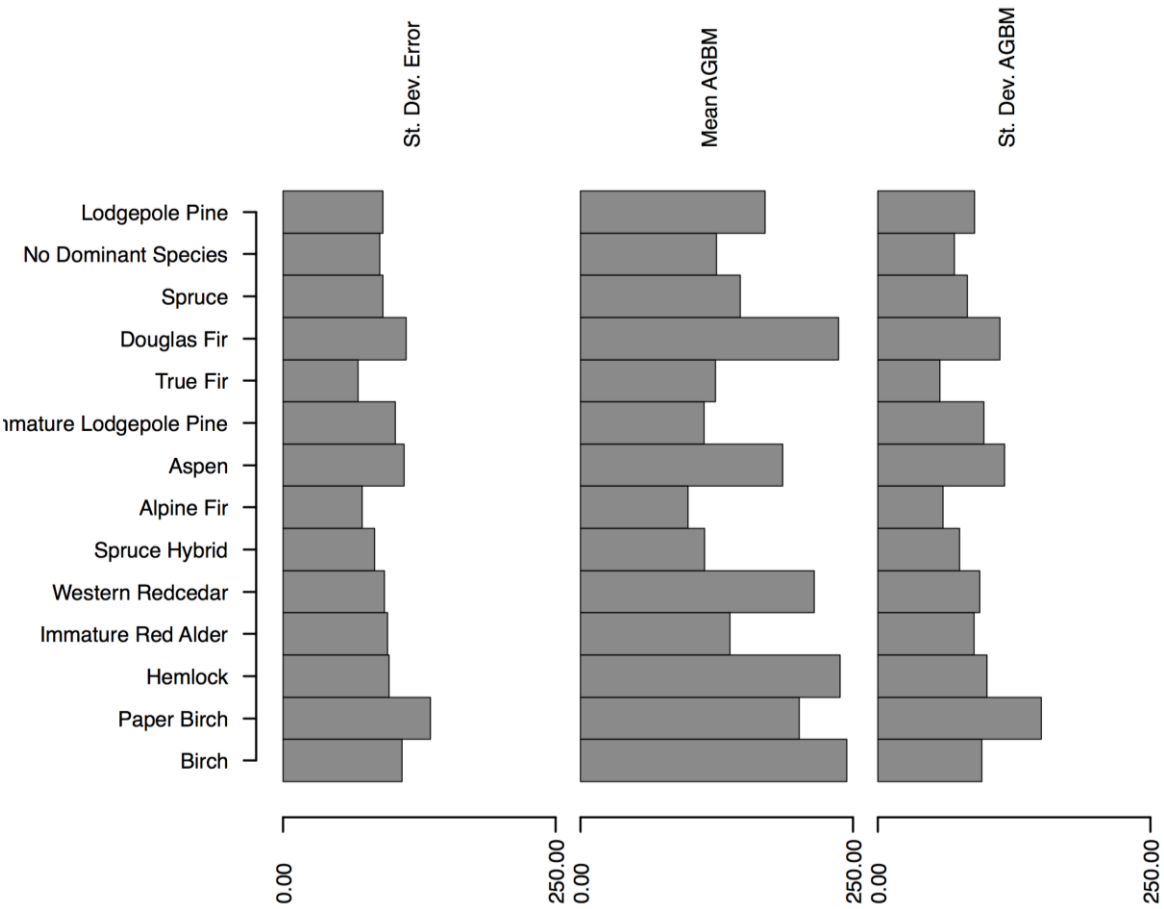

Figure 3-6 Standard deviation of model 3 error, mean AGBM and standard deviation of AGBM for dominant species segmentation.

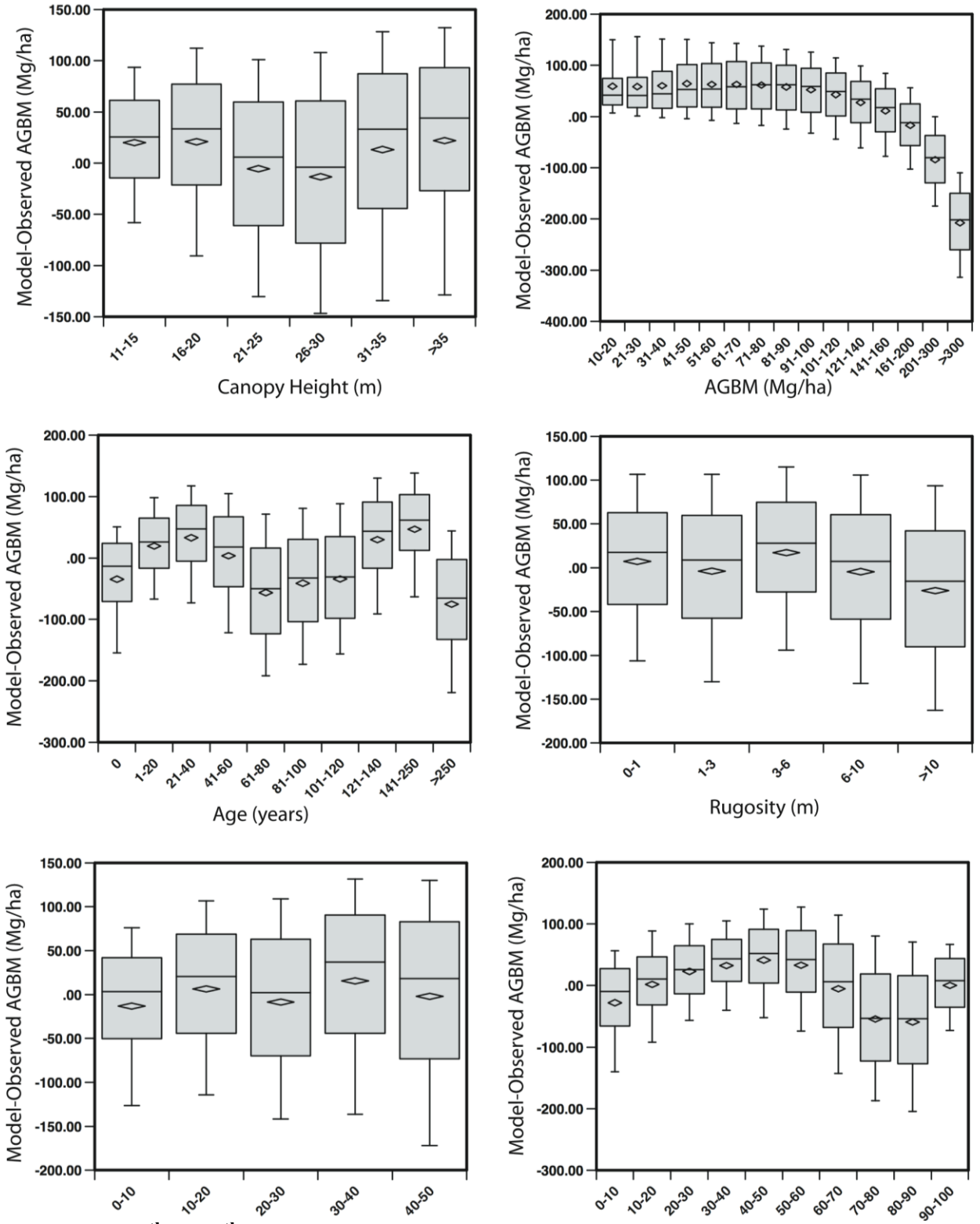

Figure 3-7. 10th – 90th percentile box and whiskers plots decomposing

pixel-based error into various forest parameter classes.

Table 3-3. Number of pixels in various forest classes, as used in Figure 3-7.

Variable	Class	# Pixels	Variable	Class	# Pixels
Rugosity (m)	0-1	1480	%CC	0-10%	5
	1.1-3	11868		10.1-20%	1000
	3.1-6	35254		20.1-30%	5080
	6.1-10	27026		30.1-40%	11701
	>10	2055		40.1-50%	18870
Mean DBH (cm)	0-10	35510		50.1-60%	18970
	10.1-20	15140		60.1-70%	11716
	21.1-30	35255		70.1-80%	5464
	30.1-40	26078		80.1-90%	3186
	>40	795		90.1-100%	1691
AGBM (Mg/ha)	0-20	910	Age (years)	0	23429
	21-30	3269		1-20	8871
	31-40	3620		21-40	6081
	41-50	4447		41-60	6919
	61-70	5224		61-80	7291
	71-80	6258		81-100	12518
	81-90	6652		101-120	7267
	91-100	6955		121-140	33304
	101-120	13439		141-250	7098
	121-140	11667		>250	369
	141-160	9738	Canopy Height (m)	0-15	3772
	161-200	14983		15.1-20	1769
	201-300	22733		20.1-25	15323
	>300	7454		25.1-30	21613
				30.1-35	16450
				>35	16437

3.4 DISCUSSION

3.4.1 GLAS AGBM Estimation

The GLAS estimates of AGBM compare favorably with other studies (Lefsky et al. 2005, Sun et al. 2007) that include high AGBM (>250 Mg/ha) areas. The relationship between GLAS waveform metrics and AGBM is linear and error does not increase with increasing AGBM. Model 1 performed better than model 2, indicating that terrain relief should be accounted for in GLAS AGBM models. Areas of low or high terrain relief, as seen in Duncanson et al. (2009) yield less accurate AGBM predictions in relation to areas of moderate relief. This likely causes the greater divergence of residuals from the line of best fit seen in model 2. These results reaffirm that GLAS is an appropriate tool for AGBM estimation.

3.4.2 Landsat AGBM Estimation

The utility of GLAS and Landsat integration depends on the characteristics of the landscape. The applicability of these methods depends on the assumption that the relationship between AGBM and vegetation spectral response is consistent between the area sensed by GLAS and the study area. The highest biomass areas fall outside of the GLAS transect, which may explain the trend to underestimate the AGBM in the study area by all of the Landsat models explored in this study. However, from deconstructing the scene into biomass classes, it was observed that the underestimation of AGBM began at approximately 140 Mg/ha, within the AGBM range sampled by the GLAS transect. Therefore I believe that the transect sampling may have empirically influenced the

AGBM estimates, but this influence was not sufficiently notable to alter the results or conclusions of this study.

Partial least squares regression was found to be more reliable for AGBM estimation than OLS Regression. This is attributed to the ability to include more input variables in PLS, regardless of collinearity amongst the variables. All six Landsat bands were used as inputs to the PLS models, although the inclusion of NDVI or any of the tassel cap bands did not improve the model. Only tassel cap band 1 and Landsat band 3 were useful for the OLS model. Other bands would have improved the model, but could not be included due to collinearity issues.

Even in the most successful areas of estimation, there were pixels with over and under predictions of 100 Mg/ha AGBM. This wide range of estimation is likely the result of lowered observable variance associated with satellite-based estimates that are incapable of documenting the variance observed with high resolution airborne data. Therefore signals related to areas with mixtures of high and low biomass are aggregated to moderate levels of biomass, resulting in large under and over predictions dependent on the resampling of imagery. Future satellite LiDAR missions will have smaller footprint sizes, which may eliminate the need to aggregate Landsat data to 60 meters. This may increase the utility of spaceborne LiDAR integration with Landsat.

3.4.3 Error Decomposition

The standard deviation of error in model 3 was the statistic deemed most indicative of model performance, as it describes the breadth of the error distribution. The

standard deviation for the study area was 89.3, the mean AGBM was 146.5 Mg/ha, and the standard deviation of AGBM was 85.9 Mg/ha. The range of AGBM values for this study area was likely a reason for the large distribution of errors. However, the range allows for the evaluation of this model over a wide range of AGBM values.

Figure 3-6 shows the standard deviation of model 3 error as well as the mean and standard deviation of AGBM. Douglas-fir, immature lodgepole pine, trembling aspen, birch and paper birch are the species for which the model performed poorly. Douglas-fir had a high average AGBM and standard deviation of AGBM. Douglas-fir were in some of the most under predicted areas. Douglas-fir trees in the study area are representative of old growth stands which are known to have variable AGBM in the study area. Immature lodgepole pine stands likely had similar spectral signatures to older, higher biomass members of their species. The areas of immature lodgepole pine tended to be overestimated, although some were also underestimated which likely caused the high standard deviation of error for this species.

The lower standard deviations of error associated with balsam, subapline fir and white spruce were likely related to the low mean and standard deviation of AGBM associated with these areas. Thus, GLAS estimates for these areas were likely successfully extrapolated. Additionally, the field validation of the airborne LiDAR AGBM estimates was conducted almost exclusively within conifer forests, which may have influenced the accuracy of AGBM estimates in broadleaf environments.

The relationships between forest properties and error show that modeling AGBM using Landsat is only appropriate for certain forested environments and that the conditions of a forest should be known prior to the application of these methods. Figure

3-7 shows a visualization of six relationships between increasing forest structural metrics and the distribution of error in each structure division.

3.4.4 Canopy Height, Rugosity, DBH

The models, as expected, were more successful for shorter canopies as these canopies are typically less complex and will have a lower proportion of understory and shaded trees, for which Landsat data is less useful. These methods become unreliable after approximately 20 meters of canopy height.

Rugosity is a measure of the vertical complexity of the canopy. I expected that vertical complexity played a significant role in the success of my models. However, neither the median nor the distribution of error showed a noteworthy change as rugosity increased. There appears to be a slight increase in the distribution as rugosity increases, but the results appear to be generally inconclusive.

As mean DBH increases so does the distribution of error. This was the expected trend, as large DBH values are related to high AGBM areas and it has been demonstrated that Landsat data is typically unsuitable for high AGBM plots.

3.4.5 AGBM

The trend between error and AGBM is the most noteworthy trend portrayed in Figure 3-7. The results indicate that these methods should not be used to estimate AGBM for forests greater than 200 Mg/ha, and should be used with caution for forests with between 120-200 Mg/ha. This depends on the information needs of an AGBM evaluation.

3.4.6 Age

The relationship between age and model error is more complicated than many of the other relationships. These age divisions were taken from the divisions provided in the forest inventory. Forests of age '0' are likely recently clear cut or planted stands, or non-forest environments that fell outside the zero AGBM mask and may not be appropriately comparable to the LiDAR data. The relationship between forest age and AGBM estimation error is likely related to the relationship between forest age and biomass. A comparison of the observed AGBM map and the inventory age map showed that biomass increased with age until approximately 61-80 years of age. After this age biomass no longer had a clear relationship with age; some of the highest biomass pixels were between 61-80 years of age, and some of the oldest stands had biomass values as low as 100 Mg/ha. These old, lower biomass stands may have been affected by disturbances such as wind damage. Consequently, if a forest stand is older than approximately 60 years, Landsat data may be useful for AGBM estimation but this is dependent on other forest structural properties, and not age in particular.

3.4.9 Per Cent Canopy Cover

The decrease in model accuracy after approximately 60% canopy cover is likely related to the relationships between age, canopy cover, and AGBM in the study area. An analysis of the associated maps shows that canopy covers greater than 60% typically correspond to stand ages 41-140 years of age. Older canopies typically have a canopy cover of 50-60% and younger stands have <50% canopy cover. Similarly, the AGBM over 60% canopy cover varies independently of canopy cover, while values of canopy

cover are related to AGBM; 60% canopy cover has an associated AGBM of approximately 140 Mg/ha. As a consequence, canopy cover may be a useful measure for determining whether Landsat is useful for AGBM estimation if an ancillary estimate of AGBM itself is not available.

3.5 CONCLUSION

Landsat and GLAS integration may be useful for biomass estimation when funding or logistical limitations prevent extensive data collection. The advantage of these data types is that they are freely available and have near global coverage. These methods offer an improvement over some carbon budget methods, such as biome based AGBM extrapolations that are currently in use for carbon credit generation in tropical forests (Gibbs et al., 2007).

The strength of this study was its ability to test empirically derived models in a spatially contiguous area outside of the dataset from which they were generated. My first objective was to develop a method to model AGBM from GLAS and Landsat integration. This was achieved, and it was determined that the best method was to develop a PLS model of GLAS estimated AGBM using Landsat bands 1-5 and 7.

Our second objective was to determine the relationships between model error and forest cover properties, and I conclude that model performance varies between dominant species classes. The model estimates are more accurate in areas with lower average AGBM and lower standard deviation of AGBM.

Our third objective was to establish ranges of forest structural properties for which these methods are most appropriate. I determined that AGBM was the best indicator of model applicability, with optimal model performance corresponding to areas with less than 120 Mg/ha AGBM. These models may also perform well in areas between 120-200 Mg/ha AGBM. Forest age and % canopy cover may also be useful indicators of model applicability, with 60% canopy cover and 60 years of age acting as minimum

thresholds under which the model performs well. These methods may be appropriate above these thresholds depending on the required accuracy of AGBM estimation, and other forest properties such as dominant species, canopy height, or mean DBH.

Additionally, the models presented in this paper are more accurate in conifer forests. This may be due to the favouring of conifer plots in the validation of the airborne LiDAR, and because the majority of the study site is forested by conifers. For forests outside the recommended range of conditions, I believe the use of satellite-based multispectral data may limit the accuracy of AGBM estimation. Future research will explore other techniques for extrapolating GLAS data for spatially continuous AGBM estimates. Radar data holds promise, particularly considering upcoming joint LiDAR-RADAR satellites such as DESDynI.

3.6 REFERENCES

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S. 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, 112, 3876-3890.

Cai, T., Ju, C., & Yang, X. 2009. Comparison of ridge regression and partial least squares regression for estimating above-ground biomass with Landsat images and terrain data in Mu US Sandy Land, China. *Arid Land Research and Management*, 23, 248-261.

Clark, M.L., Roberts, D.A., Clark, D.B. 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96, 375-398.

Cohen, W.B., Maiersperger, T.K., Gower, S.T., Turner, D.P. 2003. An improved strategy for regression of biophysical variables from Landsat ETM+ data. *Remote Sensing of Environment*, 84, 561-571.

Cohen, W.B. & Spies, T.A. 1992. Estimating structural attributes of Douglas-Fir/Western Hemlock Forest Stands from Landsat and SPOT Imagery. *Remote Sensing of Environment*, 41. 1-17.

Donoghue, D.N.M. & Watt, P.J. 2006. Using LiDAR to compare forest height estimates from IKONOS and Landsat ETM+ data in Sitka spruce plantation forests. *International Journal of Remote Sensing*. 27, 2161-2175.

Duncanson, L.I., Niemann, K.O., Wulder, M.A. 2009. Estimating foreset canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*, doi: 10.1016/j.rse.2009.08.018

Foody, G.M., Cutler, M.E., McMorrow, J., Pelz, D., Tangka, H., Boyd, D.S., & Douglas, I. 2001. Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global Ecology & Biogeography*, 10, 379-387.

Freitas, S.R., Mello, M.C.S., Cruz, C.B.M. 2005. Relationships between forest structure and vegetation indices in Atlantic Rainforest. *Forest Ecology and Management*. 218, 353-362.

Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A. 2007. Monitoring and estimation tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2, 045023.

Hall, R.J., Skakun, R.S., Arsenault, E.J., Case, B.S. 2006. Modeling forest stand structure attributes using Landsat ETM + data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, 225, 378-390.

Hall, F.G., Shimabukuro, Y.E., & Huemmrich, K.F. 1995. Remote sensing of forest biophysical structure using mixture decomposition and geometric reflectance models. *Ecological Applications*, 5, 993-1013.

Han, T., Wulder, M.A., White, J.C., Coops, N.C., Alvarez, M.F., & Butson, C. 2007. An efficient protocol to process Landsat images for change detection with tasseled cap transformation. *Geoscience and Remote Sensing Letters*, 4, 1, 147-151.

Hansen, P.M. & Schjoerring, J.K. 2003. Reflectance measusrement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86, 542-553.

Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterretche, M. 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82, 397-416.

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., Shugart, H.H. 1999. Surface LiDAR remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, 67, 83-98.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo, F.D.B., Hunter, M.O., & de Oliveira Jr., R. 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32, L22S02.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, 1, 013537.

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E., McKeww, W.A. 1999. Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, 67, 298-308.

Ministry of Sustainable Resource Management, 2003. Vegetation Resources Inventory: Quality Assurance Procedures for VRI Ground Sampling. *Resources Information Standards Committee*, British Columbia, ISBN 0-7726-433-7.

Næsset, E., Martin Bollandsas, O., & Gobakken, T. 2004. Comparing regression methods in estimation of biopysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment*, 28, 541-553.

Nelson, R., Valenti, M.A., Short, A., Keller, C. 2003. A multiple resource inventory of Delaware using airborne Laser data. *BioScience*, 53, 981-992.

Niemann K.O. 2009. Evaluation of Emerging Remote Sensing Technologies for the Assessment of Forest Resources. UVic Hyperspectral LiDAR Research Group Report 2009-1, 97.

Patenaude, G., Hill, R.A., Milne, R., Gaveau, D.L.A., Briggs, B.B.J., Dawson, T.P. 2004. Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, 93, 368-380.

Pedersen, L., & Forester, C. 2000. Tree Farm License 18: Rationale for Allowable Annual Cut (AAC) determination.. *British Columbia Ministry of Forests.*

Ranson, K.J., Kimes, D., Sun, G., Nelson, R., Khaaruk, V., & Montesano, P. 2007. Using MODIS and GLAS Data to Develop Timber Volume Estimates in Central Siberia. *IEEE IGARSS*, July 23-26, 2007, 2306-2309.

Rogan, J., Franklin, J., & Roberts, D.A. 2002. A comparison of methods for monitorint multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, 80, 143-156.

Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., Dobson, C. 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science and Policy*, 6, 441-455.

Rosette, J.A.B., North, P.R.J., & Suarez, J.C. 2008. Vegetation height estimates for a mixed temperate forest using satellite Laser altimetry. *International Journal of Remote Sensing*, 29, 1475-1493.

Schutz., B.E., Zwally, H.J., Shuman, C.A., Hancock, D., & DiMarzio, J.P. 2005. Overview of the ICESat Mission. *Geophysical Research Letters*, 32, L21S01.

Sun, G., Ranson, K.J., Masek, J., Fu, A., & Wang, D. 2007. Predicting tree height and biomass from GLAS data. In: *10th International Symposium on Physical Measurements and Signatures in Remote Sensing*, Davos, Switzerland.

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 Sensing of Environment*, 112, 107-117.

VanAardt, J.A.N. and Wynne, R.H. 2007. Examining pine spectral separability using hyperspectral data from an airborne sensor: an extension of field-based results. *International Journal of Remote Sensing*, 28, 431-436.

Wulder, M. 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography*, 22, 449-476.

Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A., & Franklin, S.E. 2003. Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The Forestry Chronicle*, 79, 1075-1083.

Wulder, M.A. & Seemann, D., 2003. Forest inventory height update through the integration of LiDAR data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, 29, 536-543.

Wulder, M.A., Han, T., White, J.C., Sweda, T., Tsuzuki, H. 2007. Integrating profiling LiDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresse, M., Le Moine, J., Ryu, S.R. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*, 93, 402-411.

Zwally, H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., &Thomas, R. 2002. ICESat's Laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34, 405-445.

4. CONCLUSION

4.1 Discussion and Conclusions

Technological development has fundamentally changed the nature of forest studies. Prior to satellite-based remote sensing data acquisition it was not possible to make spatially continuous measurements of forests beyond a local area of study. While areas of British Columbia's forests have been extensively documented by field campaigns, forest inventory cannot be documented for the entire province and satellite-based data are the only data available for many areas (Wulder et al., 2003). In countries that rely less on forestry for national income or lack the necessary funds to undertake extensive inventory campaigns, satellite data become even more important. However, generating reliable methods for modeling terrain characteristics from satellites will always be more problematic than field or even airborne data processing. Reduced spatial resolution, atmospheric influence, and a lack of appropriate calibration or validation data complicate the utility of satellite measurements (Turner et al., 2003).

My research focuses on increasing the understanding and utility of satellite-based LiDAR data for studies of forest structure. Studies have demonstrated that for flat, homogeneous areas data from the Geospatial Laser Altimeter System (GLAS) aboard NASA's ICESat is useful for canopy height and AGBM characterization (Lefsky et al., 2007, Sun et al., 2007, Rosette et al., 2008). However, these studies found that increasing terrain relief complicates GLAS signals and extracting canopy information from moderate or high relief waveforms is problematic.

In Chapter 2, I demonstrate that terrain relief complicates GLAS waveforms, and that increasing terrain relief increases the number of Gaussian curves found within waveforms, and influences other waveform metrics. I develop a method to classify terrain relief directly from GLAS waveform metrics (74% correctly classified), and use these relief classes as supplemental inputs to improve canopy height models (R^2=0.84). This chapter demonstrates that GLAS data are appropriate for canopy height modeling and stresses the influence of terrain relief on canopy height modeling.

Although GLAS data are useful for forest structure studies, as demonstrated in Chapter 2, these data do not represent a spatially continuous coverage of global forests. Instead, they represent series of transects, each composed of pulse centroids spaced 172 m along the Earth's surface. To generate spatially continuous estimates of forest structural properties, such as AGBM, supplemental data are required. Landsat TM data are selected as a supplemental data type because these data are spatially continuous in nature and can be freely available.

In Chapter 3, I develop a method to estimate AGBM from GLAS and Landsat integration by employing partial least squares (PLS) regression to predict GLAS AGBM estimates using Landsat bands 1-6 (R^2=0.6). I use a spatially continuous reliable measure of AGBM as a base map for comparison with my model estimates. Additionally, I use a current forest inventory including forest age, % canopy cover, mean DBH, and mean canopy height to decompose model error and assess ranges of forest structural properties over which these methods are appropriate. I determine that GLAS and Landsat integration are most useful for AGBM estimation for forests with less than 120 Mg/ha of AGBM, although they may be appropriate for forests with up to 200 Mg/ha depending on

other forest properties. Prior to employing GLAS and Landsat integration I recommend that researchers determine the nature of the forests within their study area. If AGBM information is not available, other forest structural properties can be used to determine how appropriate these methods are. These methods are more appropriate for coniferous forests with less than 60% canopy cover. Additionally, forests of 60-140 years of age appear to yield particularly poor results and should be avoided. In general, the AGBM of young forests with simple vertical structures and shorter, smaller trees will be easier to model with Landsat.

4.2 Research Contributions

The first major contribution of this research was the development of a novel suite of waveform metrics that are useful for terrain relief classification, and canopy height and AGBM estimation. Previous research has been unable to successfully account for terrain relief in canopy studies, likely because the appropriate waveform metrics had not been incorporated into models. My Gaussian decomposition technique is particularly noteworthy because other studies have used the Gaussian decomposition metrics provided with the GLAS data products without challenging the appropriateness of these metrics. The GLAS data products were designed for studies of ice sheets and not many Gaussian curves are expected in waveforms reflecting off these slowly undulating surfaces. Consequently, the GLAS processing software only fits a maximum of six curves to each waveform. However, in waveforms associated with forests, more than six Gaussian curves usually exist. My Gaussian decomposition does not constrain the number of Gaussian curves found in waveforms, and metrics related to the number and

location of these Gaussian curves are consequently more appropriate for canopy and terrain characterization in forested environments.

In addition to the Gaussian decomposition metrics, energy quantile metrics have not previously been used in studies of forest structure. Other studies have focused on metrics related directly to waveform breadth, such as the distance between the beginning of a waveform and the maximum energy return in the waveform, or the distance between the beginning and end of a waveform (Sun et al., 2007, Lefsky et al., 2007, Rosette et al., 2008). Although these metrics are useful I demonstrate in Chapter 2 that quantiles are also useful for modeling canopy height and AGBM. These metrics correspond to the distribution of energy returned in portions of the waveform – both in terms of relative energy and relative vertical distribution. Although forests with one tall tree yield waveforms with large breadths, the bulk of biomass may be held in understory trees, and models incorporating both measures of breadth and distribution of energy are less likely to overestimate AGBM.

The models and methods developed and explored in Chapter 2 can be used as guides by other researchers attempting to develop similar models in different forested environments. These models would likely not perform as well in broadleaf forests, or the coastal Douglas-Fir dominated forests found on Vancouver Island. They would also likely be inappropriate for dense tropical forests with significantly different structure than the forests for which the models were developed. However, the metrics used in these models may be applicable in other forested environments and other researchers are encouraged to adopt similar techniques in order to explore their applicability.

Although perhaps less novel than the contributions associated with Chapter 2, the evaluation of GLAS and Landsat integration also increases the understanding of the relationship between forests and satellite data. Although the models developed in Chapter 3 are less successful at estimating AGBM, the ability to evaluate the constraints on model performance enabled the development of a set of forest structural conditions for which these methods are most appropriate. The methods and associated caveats presented in Chapter 3 can be applied to generate an estimate of AGBM using data that have become freely available to researchers at zero cost. Although the methods have been demonstrated to be less accurate in high biomass (>200 Mg/ha), old (>60 years), or dense (canopy cover> 60%) forests, their utility can provide an estimate of base-line AGBM for areas with no ancillary data requirements. These methods are superior to traditional carbon budget methods that involve a single biome-based estimate of AGBM (Gibbs et al., 2007).

4.3 Research Opportunities

The findings in this research demonstrate that some of the problems associated with satellite studies of forest structure can be overcome. However, there remain challenges and opportunities for future research in this area. Evaluating the applicability of the methods and models presented in this thesis to other forested areas, and to other remote sensing data types, are two main areas that should be further explored.

Satellite-based LiDAR data, although useful, require ancillary data for spatially continuous studies. Landsat data are only useful for lower AGBM forests, and likely inappropriate for some of the more complex and higher biomass forests in the world.

Tropical wet, tropical dry and temperate rain forests are some of the richest ecosystems on the planet in terms of biomass and biodiversity (Houghton, 2005). They also represent some of the most unexplored and threatened areas. Research into tropical forest applications, in particular, is critical due to increasing rates of tropical deforestation (Olander et al., 2009). Carbon credit trading schemes are becoming popularized, and programs such as Reduce Emissions from Deforestation and Degradation of Forests (REDD) can provide funds to protect these forests by capitalizing on their carbon content (Gibbs et al., 2007). The carbon content of tropical forests, however, is largely unknown and the funds to generate highly accurate measurements through field campaigns or airborne data collection are usually not available in developing countries (Gibbs et al., 2007). Future research will attempt to develop reliable methods to generate accurate, large area carbon budgets of tropical forests using free or low cost satellite-based data. This research will likely involve research into integrating satellite-based LiDAR with other satellite sensor types, such as RADAR.

Research is currently being undertaken to test the integration of spaceborne LiDAR and RADAR data for the upcoming NASA DESDynl satellite. One of the research priorities of this mission will be aboveground biomass estimation, with the aim of using spaceborne LiDAR to estimate AGBM in forests past recognized thresholds of RADAR estimation (Dubayah, 2009). Evaluation of the GLAS metrics and GLAS AGBM models in higher biomass ecosystems is necessary to determine whether the conclusions drawn in this research are specific to the study area or more universally applicable.

4.4 References

Dubayah, R. 2009. Global vegetation structure dynamics from NASA's DESDynI Mission. *American Association of Geographers AAG*, Las Vegas, Nevada, 2009.

Gibbs, H.K., Brown, S., Niles, J.O., & Foley, J.A. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2, 045023, 1-13.

Houghton, R.A. 2005. Aboveground forest biomass and the global carbon balance. *Global Change Biology*. 11, 945-958.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, 1, 013537.

Olander, L.P., Boyd, W., Lawlor, K., Madeira, E.M., & Niles, J.O. 2009. International forest carbon and the climate change challenge: Issues and options. Report published by: *Nicholas Institute*, Duke University, North Carolina.

Rosette, J.A.B., North, P.R.J., & Suarez, J.C. 2008. Vegetation height estimates for a mixed temperate forest using satellite Laser altimetry. *International Journal of Remote Sensing*, 29, 1475-1493.

Sun, G., Ranson, K.J., Masek, J., Fu, A., & Wang, D. 2007. Predicting tree height and biomass from GLAS data. In: *10th International Symposium on Physical Measurements and Signatures in Remote Sensing*, Davos, Switzerland.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 18, 6 306-314.

Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A., & Franklin, S.E. 2003. Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The Forestry Chronicle*, 79, 1075-1083.