A systematic approach for using lidar intensity to detect forest structure

by

Jaden Orion Langford
B.Sc., University of Victoria, 2003

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

MASTER OF SCIENCE

in the Department of Geography

A systematic approach for using lidar intensity to detect forest structure

by

Jaden Orion Langford
B.Sc., University of Victoria, 2003

Supervisory Committee

Dr. K. O. Niemann, (Department of Geography)

Supervisor

Dr. M. A. Wulder, (Department of Geography & Natural Resources Canada)

Departmental Member

Dr. M. S. Flaherty, (Department of Geography)

Departmental Member

Supervisory Committee

Dr. K. O. Niemann, (Department of Geography)

Supervisor

Dr. M. A. Wulder, (Department of Geography & Natural Resources Canada)

Departmental Member

Dr. M. S. Flaherty, (Department of Geography)

Departmental Member

Abstract

Lidar intensity, a quantity analogous to backscatter, has yet to be fully exploited as an information source in the characterization of coniferous forests. Intensity images appear noisy due to (1) dynamic survey geometry, and (2) complex laser interactions in a forested environment. The nature of these issues are explored, and a systematic procedure for processing, visualizing, and normalizing the intensity data is presented. Despite high variability among neighbouring intensity values, the data are inherently spatially structured. Results from an investigation into the spatial pattern of intensity demonstrate that (1) the scale and variability of global estimates of spatial autocorrelation derived from raw intensity (point) data were markedly different between stands of different age, and these differences were driven by the canopy and gap structure within each individual stand, and (2) the magnitude of local estimates of spatial autocorrelation varied with canopy height, and, particularly in old growth stands, these magnitudes are linked to compositional factors such as species.

Table of Contents

Abstract iii
Table of Contents iv
List of Tables vii
List of Figures viii
Acknowledgments x

Chapter 1 Introduction 1
1.1 Context 1
1.2 Rationale 2
1.3 Objectives and research questions 3
1.4 Thesis outline 3

Chapter 2 Literature Review 5
2.1 Lidar remote sensing 5
2.2 Intensity issues 10
2.2.1 Physics of intensity 11
2.2.2 Radiometric influences 16
2.2.3 Geometric influences 16
2.3 Utility of intensity 20
2.3.1 Forestry 20
2.3.2 Urban features 21
2.3.3 Glacial surfaces 22
2.4 Summary 23

Chapter 3 Data processing and visualization 24
3.1 Introduction 24
3.2 Lidar data acquisition 24
3.3 Data processing 25
3.4 Canopy height model 27
3.4.1 Digital terrain (ground) model 27
3.4.2 Subtraction 28
3.4.3 Filtering 28
3.4.4 Outputs 28
3.5 Lidar data structure 29
3.6 Visualization of intensity data 30
3.6.1 Interpolation 31
3.6.2 Enhancement 33
3.6.3 Filtering 34
3.7 38
3.7 Summary 38

Chapter 4 Normalization of intensity data 39
4.1 Introduction 39
4.1.1 Return type 41
4.1.2 Footprint area 41
4.2 Objectives 43
4.3 Methods 43
4.3.1 Remotely sensed data 43
4.3.2 Sample areas 43
4.3.3 Return type 45
4.3.4 Range 45
4.3.5 Validation 46
4.4 Results 47
4.4.1 Return type 47
4.4.2 Range 48
4.4.3 Validation 51
4.5 Discussion 54
4.6 Conclusion 55

Chapter 5 Intensity in a forested environment 57
5.1 Introduction 57
5.1.1 The nature of intensity data 57
5.1.2 Spatial pattern 58
5.1.3 Global spatial autocorrelation 59
5.1.4 Local spatial autocorrelation 61
5.2 Objectives 63
5.3 Methods 64
5.3.1 Study area 64
5.3.2 65
5.3.2 Field data 65
5.3.3 Remotely sensed data 69
5.3.4 Canopy vs. ground intensity analysis 71
5.3.5 Global spatial autocorrelation 72
5.3.6 Local spatial autocorrelation 72
5.4 Results 74
5.4.1 Canopy vs. ground intensity analysis 74
5.4.2 Global spatial autocorrelation 76
5.4.3 Local spatial autocorrelation 80
5.5 Discussion 100
5.6 Conclusion 103

Chapter 6 Conclusions 105
6.1 Thesis summary 105
6.2 Suggestions for future research 107

Bibliography 109

Appendix A – Generation of canopy height model 114
Appendix B – Laser range calculation and normalization 117
Appendix C – Height prediction 119
Appendix D – G_i* statistic computation 129
Appendix E – Raster sampling routine 130
Appendix F – Vector (point) sampling routine 132

List of Tables

Table 2-1: Reflectivity of diffusely reflecting surfaces and materials at 900nm..... 12

Table 3-1: Lidar technical specifications..... 25
Table 3-2: Job specifications 25

Table 4-1: Descriptive statistics for intensity by return type..... 47
Table 4-2: Mean differences between single (1/1) and second of double returns (2/2).... 47
Table 4-3: Range regression models 51
Table 4-4: Sensor and intensity data by plot and flightline..... 53
Table 4-5: Flightline differences in intensity..... 53

Table 5-1: Variation of single return (1/1) intensity data before and after range standardization for homogenous sample areas 58
Table 5-2: Field plot attributes..... 65
Table 5-3: Attributes collected for each tree 68
Table 5-4: Mean differences in *raw* intensity for canopy and ground classes 75
Table 5-5: Mean differences in *normalized* intensity for canopy and non-canopy 76
Table 5-6: Distances of maximum spatial autocorrelation..... 77
Table 5-7: Scale of spatial autocorrelation by plot age class..... 78
Table 5-8: Field data relationships with the scale of spatial autocorrelation of intensity . 79
Table 5-9: Direction and strength of association between canopy height, intensity, and G_i* 88
Table 5-10: Differences in G_i* for species, crown class and mortality..... 99

List of Figures

Equations

Equation 2-1: Intensity expressed as the ratio of received to transmitted power 11
Equation 2-2: Range-standardized intensity 17
Equation 2-3: Intensity standardized by angle of incidence 20

Equation 3-1: Linear stretch 34

Equation 4-1: Range-standardized intensity 51

Equation 5-1: Moran's *I* 59
Equation 5-2: Standardized Moran's *I* 61
Equation 5-3: Getis statistic 62
Equation 5-4: Standardized Getis statistic 62

Figures

Figure 2-1: The electromagnetic spectrum 6
Figure 2-2: Typical reflectance of vegetation, soil, and water 6
Figure 2-3: Lidar schematic 7
Figure 2-4: Comparison of the two types of lidar systems 9
Figure 2-5: Laser speckle 14
Figure 2-6: Phase angle 15
Figure 2-7: Hot spot effect for targets of known reflectivity 15
Figure 2-8: Range-standardized intensity 18
Figure 2-9: The effects of scan angle and terrain slope on angle of incidence 19

Figure 3-1: Example of a lidar data file 26
Figure 3-2: Elevation and canopy height images derived from lidar data 29
Figure 3-3: Influence of scan, roll, pitch, and yaw on lidar point spacing 31
Figure 3-4: Intensity image of forest stands 32
Figure 3-5: Intensity image histogram 33
Figure 3-6: Linear 2% stretch 34
Figure 3-7: Raw and 3x3 filtered intensity images: median, bit error, local sigma 37

Figure 4-1: Controls of footprint area: beam divergence, range, and angle of incidence 40
Figure 4-2: Normalization sample areas over canopy height 44
Figure 4-3: Range in 1m bins as a function of intensity mean and standard deviation 49
Figure 4-4: Intensity images created using: last returns , single returns, and single range-standardized returns 52

Figure 5-1: Example of various lag distances from a central point 60
Figure 5-2: Field plot locations over lidar canopy height model..... 66
Figure 5-3: Intensity image of the study area. 67
Figure 5-4: Ground (<=2m) point locations in plot 2..... 75
Figure 5-5: Behaviour of standardized Moran's *I* with increasing lag distance 77
Figure 5-6: Scatterplot of spatial scale of autocorrelation of intensity related to field-measured plot attributes (crown diameter and stem position)..... 79
Figure 5-7: Canopy height, intensity and G_i* images for regenerating, immature, and old growth plots..... 81
Figure 5-8: Lidar canopy height image with field data overlaid (plot 8) 84
Figure 5-9: G_i^* image with field data overlaid (plot 8) 85
Figure 5-10: G_i^* image draped over canopy height surface (plot 8) 87
Figure 5-11: Intensity and G_i* vs. canopy height for *regenerating* plots..... 89
Figure 5-12: Intensity and G_i* vs. canopy height for *immature* plots..... 90
Figure 5-13: Intensity and G_i* vs. canopy height for *old growth* plots..... 91
Figure 5-14: *Unfiltered* scatterplots of intensity and G_i* vs. height by crown class and species..... 93
Figure 5-15: Filtered intensity (top) and G_i* (bottom) scatterplots for *old growth* 95
Figure 5-16: Filtered intensity (top) and G_i* (bottom) scatterplots for *immature* 96
Figure 5-17: Filtered intensity (top) G_i* (bottom) scatterplots for *regenerating* 97
Figure 5-18: Interaction effect between species and crown class (old growth)..... 100

Acknowledgments

I thank Dr. Olaf Niemann for providing me with the opportunity to pursue graduate work in remote sensing. I was well on my way to becoming a high school teacher when I went to Olaf's office on a whim, and he mentioned the possibility of working with lidar and doing fieldwork in Haida Guai. My research did not take me to Haida Guai, but Olaf came through with his promise that the master's degree would provide me with "a unique set of experiences". It has been a pleasure.

Completion of this work would not have been possible without financial support from the University of Victoria, the Natural Sciences and Engineering Research Council, the Canadian Forest Service, and Coast Capital Savings. I have come to believe that graduate school can only be completed within prescribed timelines when students receive adequate funding with which to live. I am fortunate to have had this benefit.

I found my project to be extremely challenging in the analysis phase. It required constant perseverance, and ongoing brainstorming for new approaches. I would not have been able to complete the work that follows without help from Gordon Frazer. Gordon was my primary sounding board, and he generously offered me his expertise in forest ecology, lidar, fieldwork, and statistics.

I thank Drs. Mark Flaherty and Trisalyn Nelson for statistics consultations. I also acknowledge Dr. Michael Wulder for sharing his ideas and assisting with final editing.

Finally, I recognize Rafael Loos, Fabio Visintini, and Ryan Powers who assisted with field data collection. Special thanks to Francois Beaudet, who, in addition to helping in the field, produced several schematic figures that appear in this thesis.

Chapter 1 Introduction

1.1 Context

Forests cover nearly half of Canada's landmass and are a major component of the global economy and environment. Canada exports $44.6 billion of forest products per year, which accounts for 19% of the global market (Natural Resources Canada 2006). Ecologically, forests play an important role in producing oxygen and removing carbon dioxide from the atmosphere. Forests also help to purify water and stabilize soil. As a significant component of global carbon and hydrological cycles, forests play a central role in regulating climate. Canada contains 10% of the world's forests, and up-to-date inventory and structural data are needed to support sustainable forest management.

Light detection and ranging (lidar) is an active remote sensing technology that uses a pulsed near-infrared laser to determine the distance, or range to objects. Lidar data are well suited to the capture of forest structure because of the ability of laser pulses to penetrate through canopy gaps. High laser pulse rates result in near continuous sampling of forest elements from the canopy top to the ground. Depending on the survey and sensor configuration, lidar has the unique capability to detect the fine scale (<2m) spatial arrangement of canopy material (foliage, branches, and boles) and gaps, both key components of forest canopy structure (Parker 1995). In addition to capturing the vertical and horizontal distribution of canopy material, some pulses reach the ground surface. Generating a digital terrain model from these points enables standardization of all points to vertical heights above a locally measured ground surface.

The ability of lidar to estimate forest canopy height has led to significant interest by the forest science and management community. Several forest inventory attributes (e.g.,

stand height, crown closure, and timber volume) can be derived from lidar canopy height information. Height-based metrics have received considerable attention in the literature (Maclean & Krabill 1986; Magnussen & Boudewyn 1998; Maltamo *et al.* 2004; Nelson *et al.* 1988; Patenaude *et al.* 2004). In addition to height above ground, each lidar point contains a backscatter value (termed intensity), which has yet to be fully exploited as an information source in forest characterization. Intensity is expected to contain additional information about forest composition and structure, such as species, moisture content, leaf display, arrangement, and density (Schreier *et al.* 1985).

1.2 Rationale

Exploring the utility of lidar intensity is a timely research opportunity, however significant challenges exist. There are numerous technical issues, such as calibration, that must be considered *a priori*—before interpreting intensity data, or exploring the linkages between intensity and forest structure. The *first* goal of this thesis is to develop systematic routines for processing, visualizing, and normalizing intensity data.

Over the past 20 years, research has demonstrated that lidar data can successfully capture forest stand attributes of commercial interest, such as canopy height and timber volume. Traditionally, a forest stand is defined as "a contiguous group of trees sufficiently uniform age-class distribution, composition and structure…to be a distinguishable unit" (Franklin & Van Pelt 2004: 27-28). This definition implies homogeneity within the stand, and that stand-level attributes, such as canopy height, are sufficient descriptors of forest structure. Yet, some forest scientists suggest that forests, especially old-growth stands, are not homogenous structures but rather diverse mosaics of various structural components. Complex structures include "a broad range of sizes and

conditions of live trees, standing dead trees (snags), and boles on the forest floor." (Franklin & Van Pelt 2004: 22). Recent research has focused on using lidar data to characterize fine scale spatial structure inside the forest canopy (Frazer *et al.* 2005). Detecting structural complexity is of interest because spatial arrangements, and related compositional differences, result in varied microclimates, stores of energy, moisture and nutrients, all of which serve to create diverse habitat. The *second* goal of this thesis is to explore the utility of intensity data in the capture of forest structure, especially the components of structure (such as compositional differences) not expressed by canopy height alone.

1.3 Objectives and research questions

There are two key objectives for this thesis and research questions nested within each.

- Objective 1 – Process, visualize, and normalize intensity data
  - How can intensity data be visualized?
  - What are the sensor and survey geometry influences on intensity and how can these effects be accounted for such that intensity values are normalized?
- Objective 2 – Use intensity data to capture forest structure
  - What spatial patterns and trends exist within intensity data and how do these relate to forest structure?
  - Which aspects of forest structure does intensity respond to?

1.4 Thesis outline

This thesis consists of six chapters. Chapter 2 reviews the literature covering the key concepts of lidar remote sensing, summarizes pertinent issues to consider when working with laser backscatter (intensity) data, provides examples from previous studies using lidar intensity. Chapter 3 explains how the remote data were acquired and outlines the

procedures used to process and visualize the lidar data. Chapter 4 documents the intensity normalization and validation procedure, presents results, and discusses the outcomes. Chapter 5 deals with the second objective, using intensity data to capture forest structure. This chapter describes field data collection, statistical analysis methods, results, and discusses the findings. Chapter 6 summarizes the key results, and provides recommendations for future research.

Chapter 2 Literature Review

2.1 Lidar remote sensing

In the previous chapter, light detection and ranging (lidar) was briefly introduced as a remote sensing technique that is well suited to forest research. This section provides an overview of the basic principles of lidar technology, types of lidar systems, and applications for lidar data.

Remote sensing is defined as the "science and art of obtaining information about an object, area, or phenomenon through analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation" (Lillesand & Kiefer 2000: 1). There are two basic types of remote sensing, active and passive. Passive sensors record energy reflected off a surface that originates from an external source, such as the sun. Examples of passive remote sensing products include aerial photographs and earth observing satellite data, such as *Landsat* images. Active sensors, in contrast, do not rely on an external source of illumination; instead they supply their own energy and detect its interaction with the target surface. Lidar is an example of active remote sensing, as are radar and sonar (Jensen 2000).

A lidar system's source of energy is a laser, which is an acronym for light amplification by the stimulated emission of radiation. Lasers produce a very narrow, intense beam of energy at a specific wavelength. Lasers differ from common light sources (e.g., the sun), which emit a continuous spectrum of wavelengths in a much wider beam. Lasers used in terrestrial lidar systems typically operate at a wavelength between 0.9 and 1.5 μm, which is in the near-infrared region of the electromagnetic spectrum (Figure 2-1). A near infrared wavelength is chosen for two reasons: (1) reflectivity from vegetation is relatively high at these wavelengths (Figure 2-2), and (2) eye safety is a concern for lasers using shorter wavelengths, such as visible lasers (Wehr & Lohr 1999).

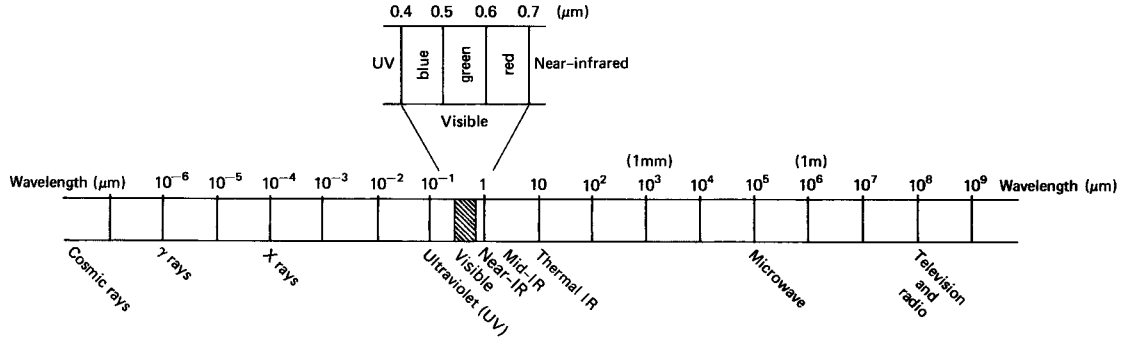


Figure 2-1: The electromagnetic spectrum (reproduced from Lillesand & Kiefer 2000: p.5)

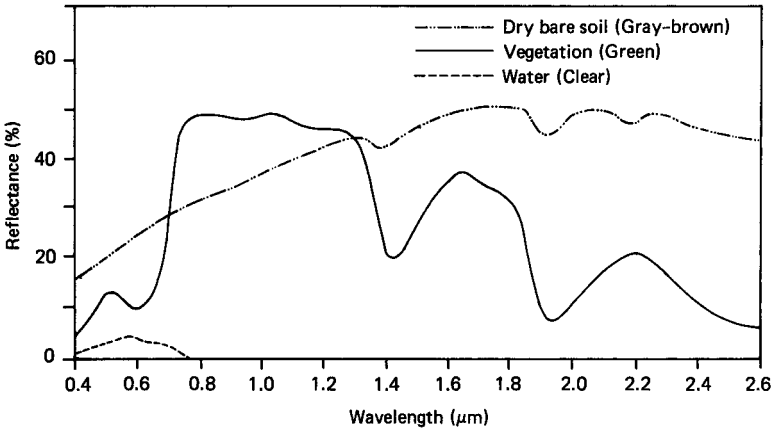


Figure 2-2: Typical reflectance of vegetation, soil, and water (reproduced from Lillesand & Kiefer 2000: p.17)

Lidar systems rely on reflection of laser pulses to determine the distance (range) to objects. The main assumption of laser ranging is that the speed of light in the atmosphere is constant. Range is then calculated using the travel time for a transmitted pulse to reach an object and return to the sensor, divided by two. The lidar sensor is typically mounted on an airborne platform (Figure 2-3). The range data it collects are individually georeferenced using time-matched positional data from differential GPS, and aircraft orientation data from an inertial measurement unit (IMU).

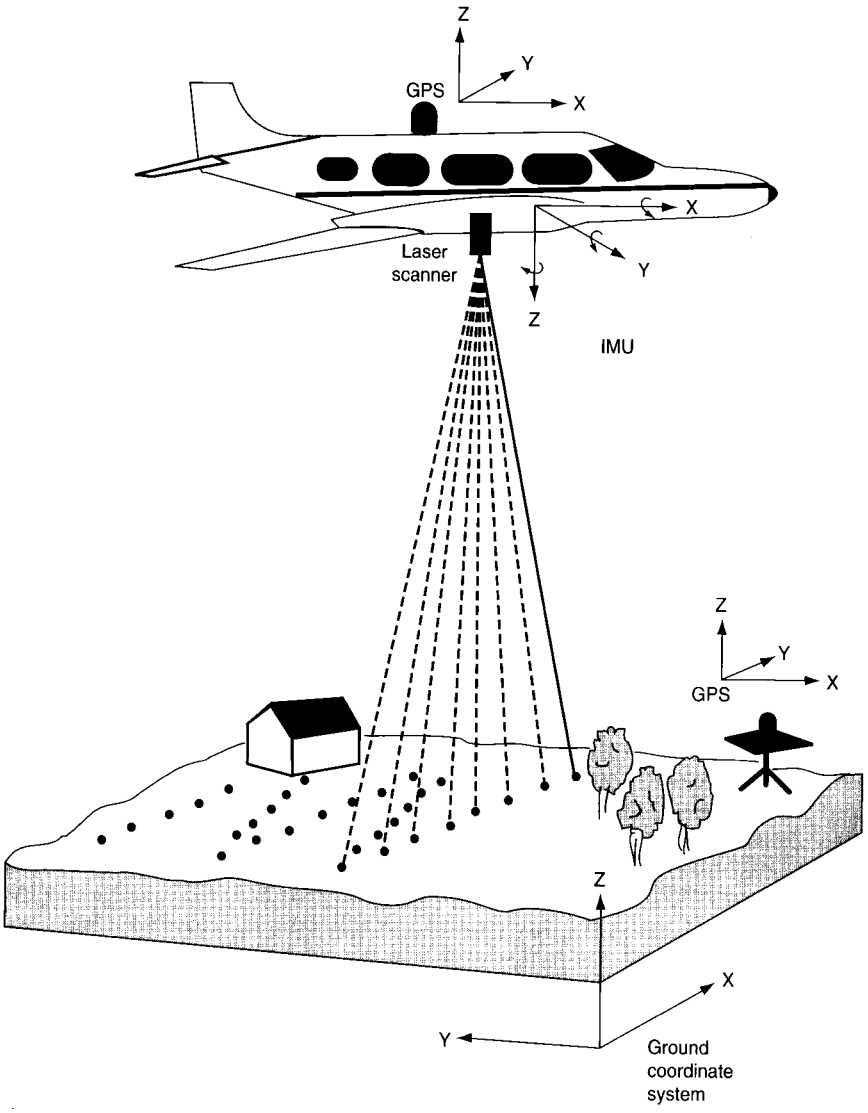


Figure 2-3: Lidar schematic (reproduced from Lillesand & Kiefer 2000: p.703)

There are two general types of lidar systems used for terrain modeling and forest characterization: “discrete return” and “full waveform” (Lim *et al.* 2003a). Discrete return systems detect location-specific peaks in backscattered energy above a threshold. Thus, if the pulse interacts with more than one surface (e.g., multiple layers of a vegetation canopy) multiple returns are recorded. A discrete return system, registers each return in *x*-*y*-*z* space and there are typically one to three returns per pulse. In addition to the return number, the intensity of each return is also recorded. The second type is full waveform lidar. These systems record the returning energy from each pulse at fixed intervals of time (which are equivalent to fixed height classes), resulting in intensity waveforms for each pulse (Lefsky *et al.* 1999). Figure 2-4 illustrates the difference between these two types of lidar. The lidar beam, shown in grey, propagates down toward the ground and the area it illuminates is called the footprint. The full waveform system, shown on the left side, records the amount of energy reflected as the beam encounters foliage on the tree and eventually the ground. In contrast, the discrete return system records only four reflections, one near the top of the tree, two in its lower foliage, and a final one from the ground.

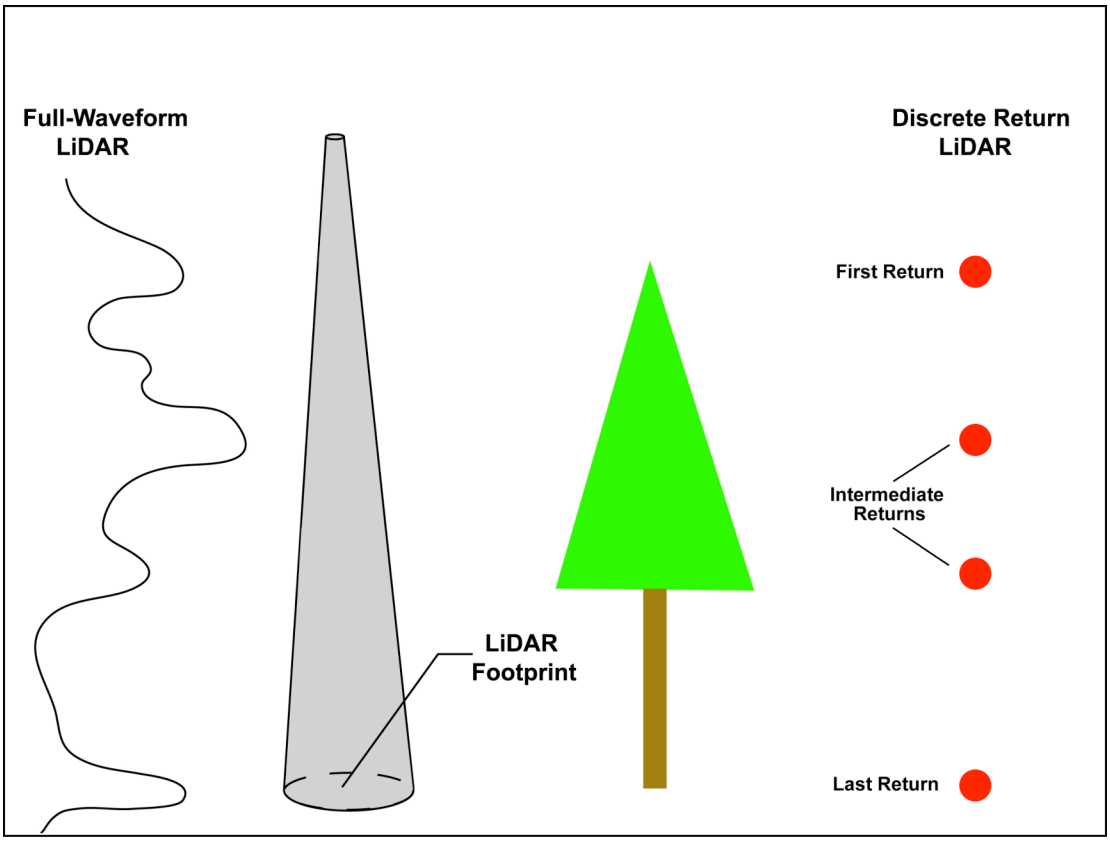


Figure 2-4: Comparison of the two types of lidar systems (after Lim *et al.* 2003a)

The two types of lidar systems described above explain differences in vertical sampling. Another important consideration is horizontal sampling area, or "footprint" of the lidar system. The footprint diameter is directly related to the laser beam divergence and the altitude of the platform (Baltsavias 1999). Due to the heavy storage requirements of full waveform lidar, these systems tend to have large footprints (8-70m in diameter). In contrast, discrete return systems have smaller footprints (0.2-0.9m in diameter) (Lim *et al.* 2003a). Few full waveform systems are available, and they are primarily used for research. In contrast, a wide variety of discrete return lidar systems, such as the one used in this study, are available for research and commercial applications. For the purpose of

this thesis, the term lidar refers to a discrete-return small-footprint system, unless otherwise stated.

The main advantage of lidar compared to other remote sensing techniques is the ability of some of the laser pulses to penetrate through gaps in vegetation canopies. This makes lidar an effective tool for creating terrain models of the ground surface (Krabill *et al.* 1984; Ritchie 1995). The topographic mapping application of lidar has been particularly useful in dense tropical forest regions (Arp & Griesbach 1982; Hofton *et al.* 2002). As lidar pulses proceed to the ground, part or all of the energy may be reflected by vegetation. In open canopies, high laser pulse rates result in near continuous sampling of forest elements from the canopy top to the ground. This vertical sampling capability makes lidar highly effective in remote sensing of forest structure, in addition to ground topography. Furthermore, the proportion of reflected energy (intensity), is also useful in analyzing the vegetation encountered (Schreier *et al.* 1985).

Three forest attributes can be directly retrieved from lidar data: canopy height, sub-canopy topography, and the vertical distribution of canopy material (Lim *et al.* 2003a). Other characteristics such as: biomass, volume, and above ground carbon content can be derived through empirical models developed using both lidar and field-based data (Lefsky *et al.* 1999; Lim *et al.* 2003b; Means *et al.* 2000; Naesset 1997; Patenaude *et al.* 2004; Riaño *et al.* 2003). Developing a better understanding of the presently unused lidar intensity data may be helpful as these predictive models are refined and improved.

2.2 Intensity issues

The ability of lidar to estimate forest canopy height has led to significant interest by the forest science community. To date, however, little attention has been paid to the possible

use of intensity data to retrieve information about forest composition and structure. The following sections describe some significant challenges that must be considered, and, wherever possible, rectified before beginning to analyse intensity data.

2.2.1 Physics of intensity

Intensity is the ratio of energy received by the detector after reflecting off a target surface, to the energy transmitted by the laser. Received energy is controlled by several parameters defined in Equation 2-1 (Baltsavias 1999).

$$\frac{P_R}{P_T} = \rho \frac{M^2 A_r}{\pi R^2} \tag{2-1}$$

Where: P_R = power received
P_T = power transmitted
ρ = reflectivity
M = atmospheric transmissivity
A_r = area of receiving optics
R = range

Equation 2-1: Intensity expressed as the ratio of received to transmitted power

It can be assumed for lidar surveys that the area of receiving optics (A_r) is constant. Range (R) is known, and intensity should be normalized to a standard range, especially if the mission was flown at different altitudes or over high relief areas (see Section 2.2.3.1). Therefore, under stable atmospheric conditions (constant M), intensity variability is directly linked to the reflectivity (ρ) of the surface encountered. The reflectivity of some materials at 900nm is presented in Table 2-1. Subsequent paragraphs will explain, however, that the precise relationship between intensity and reflectivity is complex.

Table 2-1: Reflectivity of diffusely reflecting surfaces and materials at 900nm (reproduced from Riegl Laser Measurement Systems, http://www.rieglusa.com/principles/e_gi004.htm)

MATERIAL	REFLECTIVITY
White paper	up to 100%
Dimension lumber (pine, clean, dry)	94%
Snow	80-90%
Beer foam	88%
White masonry	85%
Limestone, clay	up to 75%
Newspaper with print	69%
Tissue paper, two ply	60%
Deciduous trees	typ. 60%
Coniferous trees	typ. 30%
Carbonate sand (dry)	57%
Carbonate sand (wet)	41%
Beach sands, bare areas in dessert	typ. 50%
Rough wood pallet (clean)	25%
Concrete, smooth	24%
Asphalt with pebbles	17%
Lava	8%
Black neoprene	5%
Black rubber tire wall	2%

The absence of known calibration between intensity and reflectivity is one of the main limitations of intensity data. In one study, portable brightness tarps were used to investigate the relationship between reflectivity and intensity (Kaasalainen *et al.* 2005). The authors identified three main optical physics influences on intensity: laser polarization, laser speckle, and phase angle. Although researchers have discussed these influences, current lidar systems do not fully accommodate for them. Each influence is outlined in the following sub-sections to provide context for the challenges involved in exploring and interpreting intensity data.

2.2.1.1 Polarization

Lasers generate electromagnetic radiation at a specific wavelength. One of the benefits of laser remote sensing is that the receiving detector is sensitive to a narrow range of wavelengths, which diminishes the influence of background radiation (i.e., reflected sunlight) (Wehr & Lohr 1999). The electromagnetic radiation waves emitted by the laser can also be polarized, causing the waves to propagate through space with a pre-determined orientation. Polarimetric sensors provide more information about the texture of the surface detected. Due to multiple scattering, rough discontinuous surfaces tend to depolarize the signal, while smooth continuous surfaces do not (Lillesand & Kiefer 2000). Linear polarization (e.g., vertical and horizontal) is a common feature in radar systems but it is technically more difficult to implement in lidar systems (Tan & Narayanan 2004). As a result, nearly all lidar systems including the one used in this study are *not* polarimetric. Thus, the intensity data contain less information about the nature of the interaction, and consequently, unpolarized intensity is expected to contain more noise.

2.2.1.2 Laser speckle

Another source of noise is laser speckle (Figure 2-5). When the coherent beam/pulse of electromagnetic waves are incident on a surface with roughness larger than their wavelength, the reflected wavelets will interact with one another and produce an interference pattern (Ennos 1996). This is analogous to the irregular pattern of constructive and destructive interference that causes speckle in radar backscatter imagery. The effect of laser speckle has not yet been investigated for lidar systems. However, since

forests consist of rough, discontinuous surfaces, the effect of laser speckle is expected to contribute noise to the intensity data.

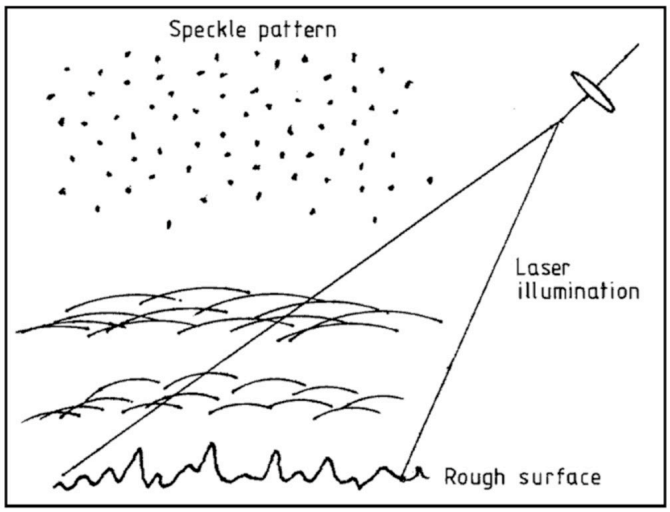


Figure 2-5: Laser speckle (reproduced from Ennos 1996: p.138)

2.2.1.3 Phase Angle

Although little is known about the impact of laser speckle on the data collected by lidar systems, a phenomenon related to the phase angle has been investigated as part of an attempt to calibrate intensity data with actual surface reflectivity (Kaasalainen *et al.* 2005). Lidar reflections are characterized by a central "hot spot" of high intensity around which the intensity decreases. The peak of the hot spot is defined by a phase angle of zero degrees. The phase angle is the angle between the backscattered wave and the observer or detector (Figure 2-6). When calibrated using targets of known reflectivity, the intensity of the hot spot is up to 1.5 times more than expected (

Figure 2-7). As the phase angle increases, moving away from the hot spot, intensity decreases exponentially. It has also been demonstrated that the higher the reflectivity of the surface, the more amplified the hot spot intensity is. In other words, if hot spot

intensity measurements are used, which is generally the case, the intensity scale will be non-linear with respect to actual surface reflectivity.

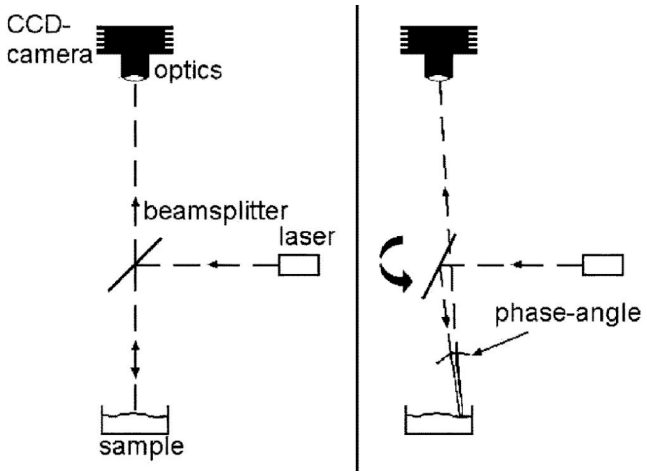


Figure 2-6: Phase angle (reproduced from Kaasalainen *et al.* 2005: p.256)

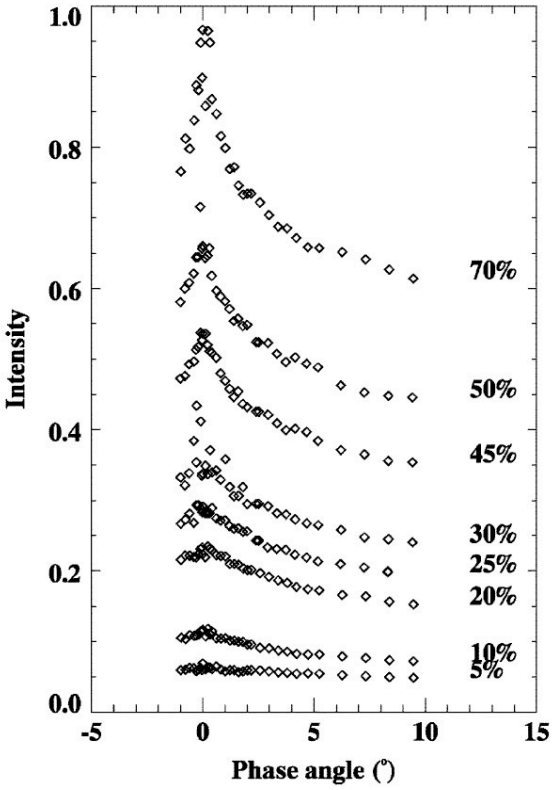


Figure 2-7: Hot spot effect for targets of known reflectivity (in %) (reproduced from Kaasalainen *et al.* 2005: p.257)

Although calibrating intensity data with known energy units is beyond the scope of the current investigation, these issues are important to consider if intensity is to be used as a quantitative measure of surface reflectivity in the future.

2.2.2 Radiometric influences

The relationship between reflected energy and the intensity values is also influenced by detector settings. The laser reflections are recorded and measured using a detector sensitive to the laser's wavelength. The laser power used depends on the altitude flown and atmospheric conditions. The returning signal is normalized as the ratio of the received to emitted energy (Jason Campbell, Terra Remote Sensing Inc., pers. comm. 26 Oct 2004). The sensor used in this study only recorded intensity on the last return from each pulse. Each intensity measurement was quantized on a 9-bit scale (digital numbers ranging from 0-511) for computer storage.

2.2.3 Geometric influences

In addition to physical and radiometric influences on the intensity values, two geometric influences must also be considered: range, and angle of incidence. Both of these quantities determine the footprint area, which is expected to affect the magnitude of the intensity measurements.

2.2.3.1 Range

The range, or path length, is the distance between the instrument and the surface being sensed. It is a function of flying height, scan angle, and ground topography. As range increases, so does the footprint area over which the laser energy is spread (Baltsavias 1999). Energy is not lost due to larger footprints, but it is less concentrated. However,

since the detector aperture is fixed in size, the amount of reflected energy received decreases as the footprint becomes larger. For surfaces of similar reflectivity, longer ranges weaken the intensity signal, while shorter ranges strengthen it (Figure 2-8). Intensity data can be normalized to a standard range using Equation 2-2 (Luzum *et al.* 2004). A standard range can be chosen by identifying the most frequent range distance from a histogram.

$$\text{intensity}_{new} = \text{intensity}_{old} * (\text{range}^2 / \text{standard range}^2) \quad (2\text{-}2)$$

Equation 2-2: Range-standardized intensity

The influence of range is substantial if: 1) the mission was flown at different altitudes, 2) wide scan angles were used, or 3) there were drastic changes in ground topography within the survey area (e.g., a deep valley) (Luzum *et al.* 2004).

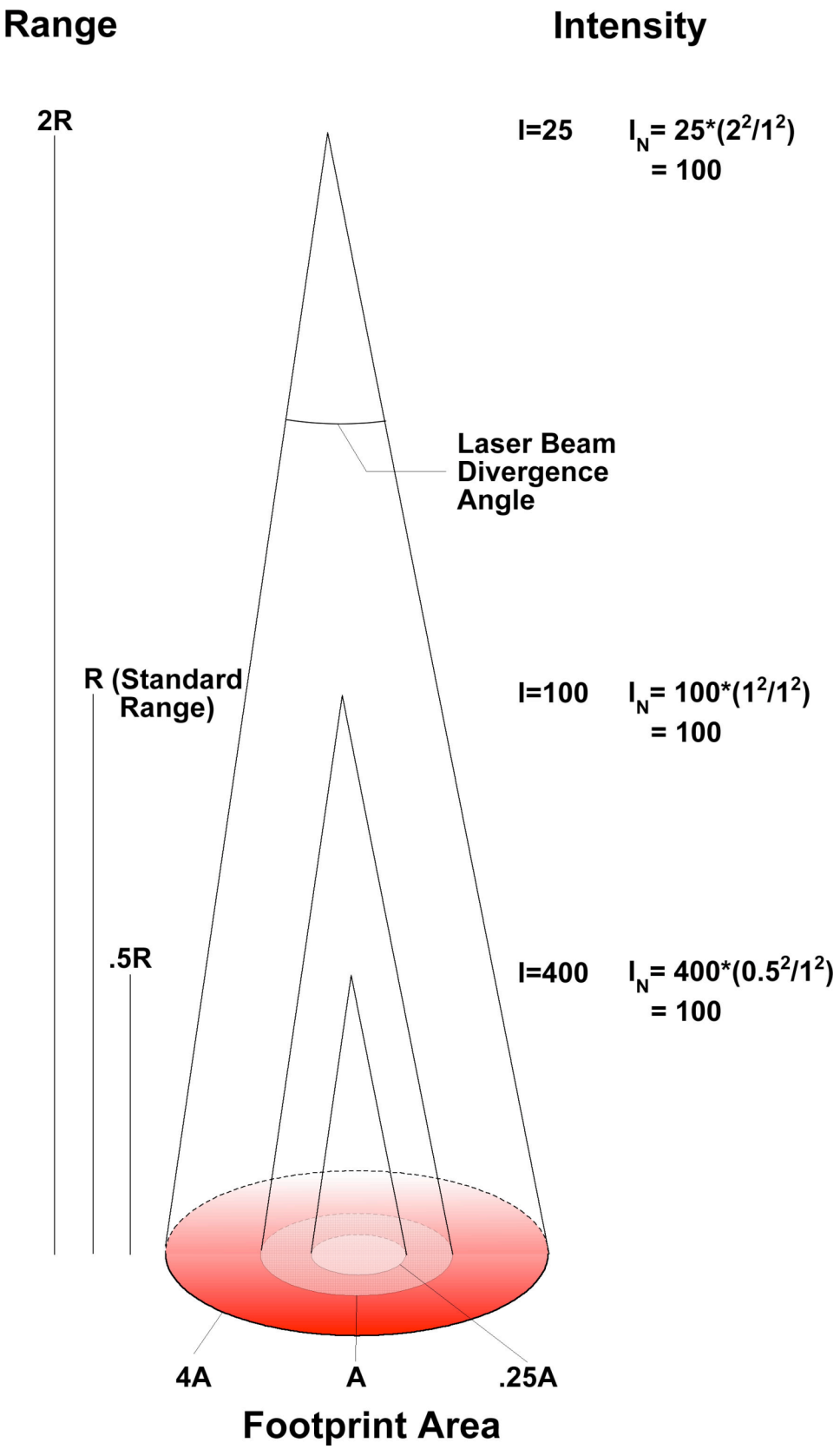


Figure 2-8: Range-standardized intensity

2.2.3.2 Angle of incidence

The angle of incidence is the angle between the surface normal and the incoming laser beam (Figure 2-9). On featureless (non-vegetated) terrain, incidence angle is controlled by the scan angle and terrain slope. As the angle of incidence increases, the footprint changes shape becoming more oblong and energy is spread over a larger area.

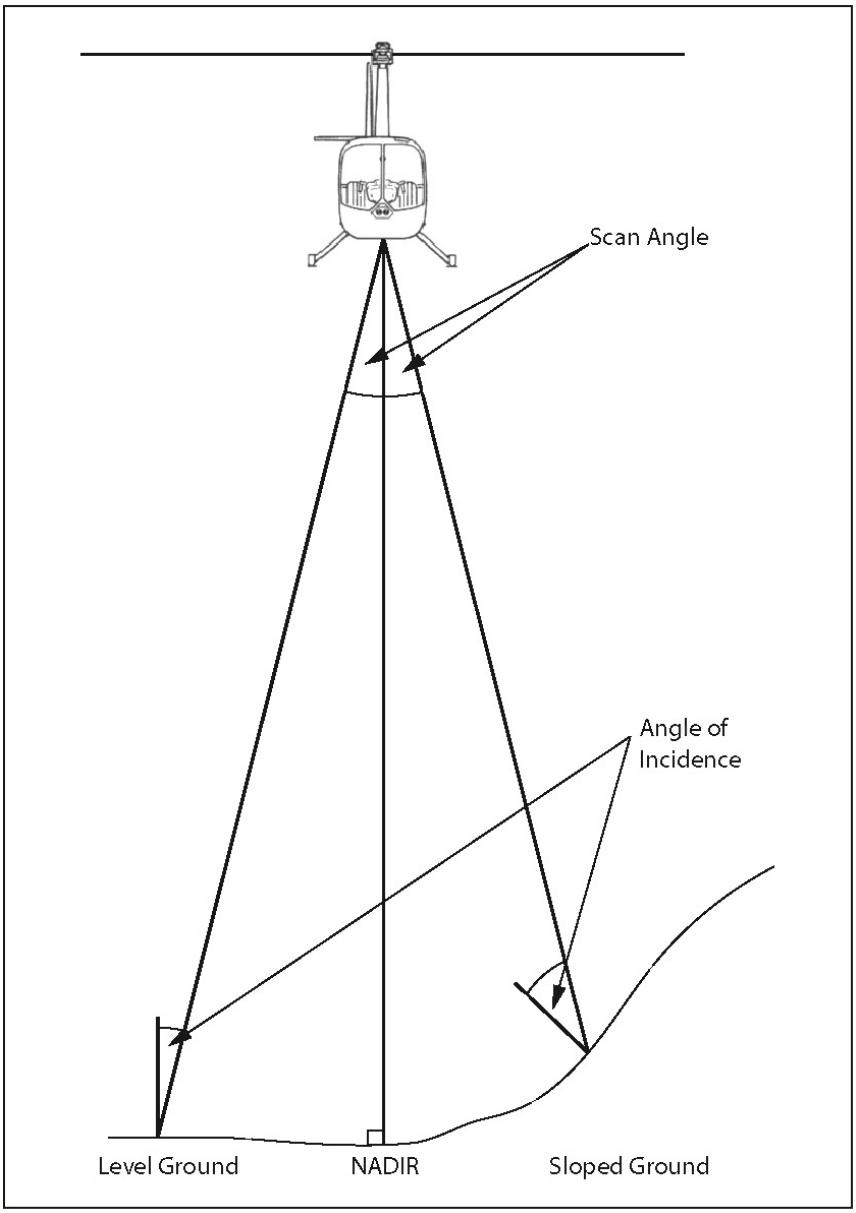


Figure 2-9: The effects of scan angle and terrain slope on angle of incidence

If a smooth "Lambertian" surface is assumed, Lambert's Cosine Law can be applied to model the apparent energy falloff (Jensen 2005). The intensity data can be normalized to zero degrees incidence using Equation 2-3:

$$intensity_{new} = intensity_{old} * COS (\text{angle of incidence}) \quad (2\text{-}3)$$

Equation 2-3: Intensity standardized by angle of incidence

This cosine correction factor has been used extensively in passive optical remote sensing, where it is necessary to correct for solar angle and terrain slope (Bishop *et al.* 2003). More complex sun-canopy-sensor correction models have been developed for forested environments (Soenen *et al.* 2005).

2.3 Utility of intensity

Having considered the issues associated with using intensity, including physical, radiometric and geometric factors, this section describes some practical applications for the information contained in intensity data. It begins with examples from forestry, followed by feature extraction in urban settings, and glacial surface classification. The latter two examples demonstrate how image processing and geometric modelling have attempted to accommodate some intensity issues.

2.3.1 Forestry

While lidar-derived canopy height is useful for predicting stand attributes such as volume and biomass, intensity data can provide information about vegetation type. Using mean intensity and intensity variability, it is possible to differentiate between coniferous and deciduous forests (Schreier *et al.* 1985). This study also found that intensity is

unaffected by differences in solar radiation incident on vegetated terrain. In another study, a minimum intensity threshold was used to reduce the influence of understory vegetation when modelling canopy height. Height estimates using data with intensity greater than 200, were better estimators of basal area, biomass and volume than estimates based on mean or maximum laser height metrics (Lim *et al.* 2003b). Intensity data have also been used to classify individual trees by species. In a supervised classification of Norway spruce and Scots pine trees in Sweden, the mean and standard deviation intensity of canopy surface returns produced classification accuracies of 82.3% and 83.6%, respectively (Holmgren & Persson 2004). In another species classification study of leaf-off deciduous forests in eastern USA, the maximum intensity, and intensity kurtosis resulted in accuracies of 41% and 43%, respectively. The authors of this study concluded that intensity responded to different colours of bark (Brandtberg *et al.* 2003).

2.3.2 Urban features

Surface reflectivity and lidar intensity have also been compared in the urban environment. In a study evaluating the separability of intensity values into classes of asphalt road, grass, house roof, and tree, it was found that although lidar intensity did not precisely match the theoretical reflectivity for each class it did follow the general magnitudes and permitted separability (Song *et al.* 2002). Lidar intensity data have also been used to identify roads. In a roadway feature extraction study, the maximum likelihood classification using lidar intensity data had producer's and user's accuracies of 71% and 81% respectively (Yu *et al.* 2002). Intensity was also used in a model for the detection and categorization of vehicles traveling on highways. The theory was that the different window orientation for sedans, multi-purpose vehicles, and trucks would result

in distinct intensity values. However, the study found that the addition of intensity into the principal components analysis resulted in poorer categorization of vehicles (Lovas *et al.* 2004).

Intensity image processing has yet to be fully investigated in the literature. One road extraction study applied triangular interconnected network (TIN), inverse distance weighted, and natural neighbour interpolation methods for rasterizing (irregularly spaced) intensity point data, and states that the optimum method will depend on the function of the imagery (Yu *et al.* 2002). After the intensity data are rasterized to a regular grid, a "salt and pepper" appearance is evident. The nature of the speckle is attributed to both sensor and target factors. A simple median convolution filter can be used to smooth the image for visualization and classification, and was found to improve separability of features in road extraction studies (Song *et al.* 2002; Yu *et al.* 2002). A special set of adaptive filters has been developed for radar imagery to reduce speckle, and improve classification (Durrand *et al.* 1987). These may be useful for intensity imagery as well.

2.3.3 Glacial surfaces

When interpreting intensity data it is necessary to consider the influence of laser geometry. In an investigation of lidar intensity and glacial surfaces, a geometric model was built to predict the laser angle of incidence (angle between the laser pulse and the surface normal), footprint size (area illuminated by the laser) and range (slant distance between the platform and surface) for each pixel of the intensity image. Correlations between all laser geometry variables and intensity were negative ($-0.63 < r < -0.80$), indicating that intensity decreases as the angle of incidence, footprint size, and range increase. The investigation also found differentiable intensity characteristics for classes

of snow, ice, water, and rock. An interesting finding was that although the intensity values for water were very low, water was found to be a specular reflector when the angle of incidence was close to zero, resulting in "intensity hot spots" (Lutz *et al.* 2003).

2.4 Summary

A large body of literature exists on lidar applications in forestry that focus on canopy height, yet little attention has been paid to the possible use of intensity data to retrieve information about forest composition and structure. There are several unique challenges to the investigation of lidar intensity data. These issues include: laser polarization, laser speckle, the phase angle used in detection, radiometric settings of the detector, and the footprint area (through its geometric controls: range and angle of incidence).

Although intensity data have yet to be fully analyzed with respect to forest surveys, results from forest, urban, and glacial classification studies suggest that intensity data may indeed contain important, added information related to forest composition and structure.

Another question is whether analysis should use raster images generated by interpolating intensity to a regular grid, or the raw *x*-*y*-*z* intensity data points collected by the lidar system itself. If raster analysis is chosen, image-processing issues that need to be addressed include: finding a suitable image interpolation method, and reducing speckle. The next chapter addresses these questions, as it outlines the data processing and visualization routines used.

Chapter 3 Data processing and visualization

3.1 Introduction

Lidar data present unique processing and visualization challenges. One of these challenges is the sheer volume of data collected by the lidar system, which typically pulses 40,000 times for every second of operation. The data are collected in vector (point) format, defined by *x*-*y*-*z* coordinates, and do not conform to a regular image-like grid pattern. The high volume, three-dimensional nature, and irregularly spaced point structure of lidar data make it problematic to visualize. Although point data can be converted to raster images, fine-scale detail is lost in the conversion. This is concerning because intensity values are variable over fine scales (i.e., speckled).

One goal of this chapter is to document how the lidar data were acquired and processed for this project. Another goal is to explain how the vertical (*z*) dimension was transformed from absolute elevations into canopy heights, creating a canopy height model. Finally, it will outline the steps used to determine a suitable visualization strategy for the intensity data.

3.2 Lidar data acquisition

Lidar data were acquired over 1300 hectares of forested land within the Greater Victoria (Sooke Lake) Watershed on southern Vancouver Island, British Columbia, Canada on July 24, 2004. The study area, a valley surrounding Rithet Creek, ranges in elevation from 180m to 750m above mean sea level.

The lidar system's technical specifications are presented in Table 3-1 and the specifications for this project are contained in Table 3-2. The data provider, Terra

Remote Sensing Inc. (Sidney, BC, Canada), classified the data into ground and non-ground (i.e., vegetation) classes using a proprietary ground-finding algorithm (Terrasolid Ltd., Jyvaskyla, Finland). Automated ground classification was inspected for the entire area by technicians at Terra Remote Sensing Inc.

Table 3-1: Lidar technical specifications

PARAMETER	SPECIFICATION
Laser	Nd-YAG
Wavelength	1089 nm
Beam divergence	0.45 milliradians
Pulse duration (length)	8 nanosecons
Pulse rate	40 kHz
Scan frequency	30 Hz
Scan angle	±24°

Table 3-2: Job specifications

PARAMETER	SPECIFICATION
Projection	UTM Zone 10N
Horizontal Datum	NAD83
Geoid	Canada 2000
Vertical Datum	CGVD
Number of flightlines	11
Flying altitude	855-1600m (flightline dependant)
Mean point density (ground)	0.49 points/m^2
Mean point density (non-ground)	3.14 points/m^2

3.3 Data processing

The lidar data were provided as comma separated ASCII files based on Format 1 of the ASPRS Lidar Data Exchange Format Standard (LAS) (http://www.lasformat.org). The fields included: x, y, z, intensity, return number, number of returns (given pulse), scan direction flag, edge of flight line, classification, scan angle, and GPS time (Figure 3-1). The total size of the dataset is 3.06 gigabytes. For improved processing performance the dataset was divided into 28 files, each representing a 1000 by 1000 metre sub-area. The

largest file is 250 megabytes in size, and it contains 3.7 million individual lidar data records (points).

```
-----------------------------------------------------------------------
X        |Y          |Z        |I  |Ret#|D±|Ed|Cla|Ang|Time
-----------------------------------------------------------------------
444997.54, 5384000.40,  278.55, 169, 1/1, 0, 0, 2, 0, 588531.995800
444999.16, 5384001.00,  278.38, 183, 1/1, 0, 0, 2, 0, 588531.995800
444997.81, 5384000.48,  294.75, 000, 1/2, 0, 0, 1, 0, 588531.996000
444999.17, 5384000.97,  296.67, 179, 1/1, 0, 0, 1, 0, 588531.996000
444998.33, 5384000.65,  306.30, 196, 1/1, 0, 0, 1, 0, 588531.996000
444999.81, 5384001.19,  306.40, 212, 1/1, 0, 0, 1, 0, 588531.996000
444998.85, 5384001.25,  306.02, 197, 1/1, 0, 0, 1, 0, 588532.012000
444998.49, 5384001.13,  304.44, 221, 1/1, 0, 0, 1, 0, 588532.012000
444998.55, 5384001.16,  294.72, 000, 1/2, 0, 0, 1, 0, 588532.012000
```

Figure 3-1: Example of a lidar data file

The volume of data and the size of the computer files requires a lidar data user to have advanced processing software and/or computer programming skills. Although there are software packages available specifically for lidar data processing, these applications are expensive, and do not offer the researcher complete control over the processing routines.

Interactive Data Language (IDL) (Research Systems Inc., Boulder, Colorado) is a computer scripting environment that is well suited to lidar data processing and analysis. IDL is packaged with a comprehensive library of data processing and visualization scripts. The user also has the ability to write custom procedures. A basic knowledge of computer programming is needed to use IDL, but it is considered a high-level language, so coding is concise and intuitive.

Each script developed to process the lidar data used 28 loops, reading one file at a time into memory, processing its contents, and writing output before moving on to the next file. Two fields in the original lidar dataset, return number and number of returns, were separated by a slash rather than a comma so it was first necessary to read all rows as

strings and replace the slashes with commas. After this process was complete, each file could be easily read into a numeric double precision IDL array (analogous to a spreadsheet table). Numerous IDL scripts were developed specifically for this project. These scripts are presented as appendices and will be referred to at various times throughout this thesis.

3.4 Canopy height model

The first major processing step was to convert the absolute elevation of all points classified as vegetation into heights above ground. This is an important procedure because height is a fundamental forest attribute. A canopy height model is created using three steps:

(1) A raster digital terrain model of the ground surface was created by interpolating all ground-class points to a regular grid.

(2) The elevation of each ground point was subtracted from the elevation of all vegetation points within its respective grid cell.

(3) All points greater than 100m above the ground were filtered and removed.

The following paragraphs describe each step in greater detail. Refer to Appendix A for the complete IDL code.

3.4.1 Digital terrain (ground) model

Although on average there is one ground point for every 2 square meters, it is necessary to create an interpolated ground surface model. Some areas, especially those with thick canopies and/or multiple vegetation layers, have a lower density of ground points due to low frequency of laser penetration to the ground. Rather than generating a separate digital terrain model for each file, the entire ground point network was used.

This mitigated edge effects that were encountered when file-specific models were used. The ground point network for the entire study area (all 28 files) consists of 6,361,726 points. IDL's TRIGRID function, linear interpolation using a Delaney triangulation (RSI 2005), was used to create the interpolated surface. A grid size of 2x2 meters was chosen to mimic the average resolution of the raw ground point network.

3.4.2 Subtraction

Vegetation point data were segmented into 2x2m grid cells, with a ground point at the centre of each cell. The ground elevation was then subtracted from all vegetation point elevations within the cell. The procedure was repeated for each grid cell in the file.

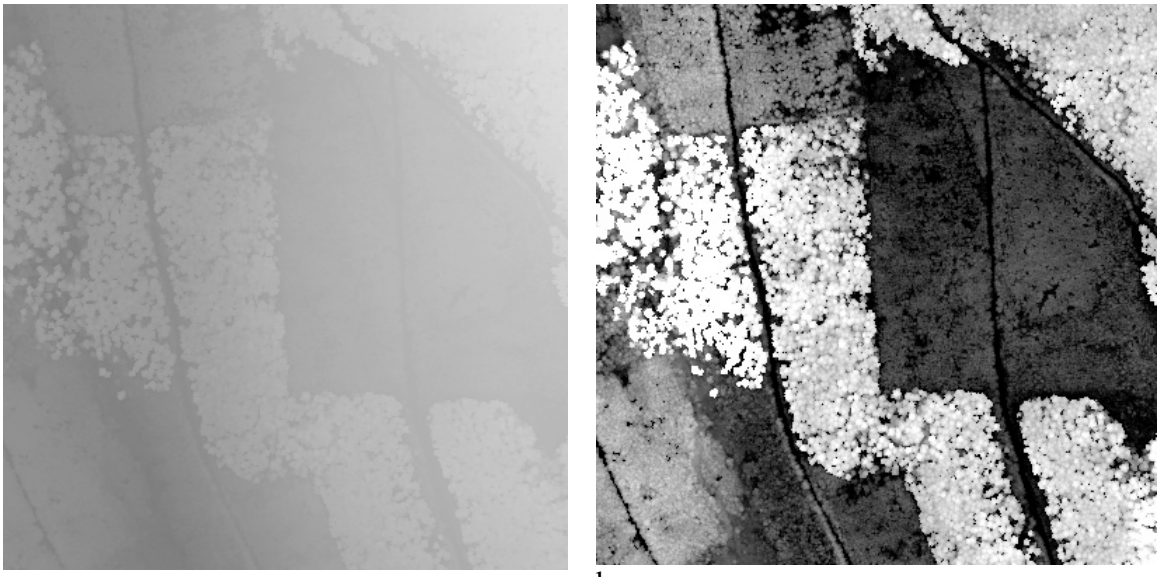
3.4.3 Filtering

Two unexpected outcomes can occur after subtraction: negative height values, and extremely high ones. The former are due to error in the ground classification, while the latter are due to extraneous reflections (e.g., birds), or sensor error. Minimum and maximum thresholds can be established to filter and remove these anomalies. In this study, a maximum threshold of 100m was used, but a minimum threshold was not applied. Negative height values were retained as they are assumed to represent either ground or ground covering vegetation, which may be of interest later.

3.4.4 Outputs

The outputs of the canopy height model procedure are (1) an ASCII text file for all vegetation points with the elevation (z) field replaced by height above ground, and (2) a 2x2m GEOTIFF raster image of canopy height (Figure 3-2). The text file is used for all

further analyses, as it contains the original point data (except elevation). The raster canopy height image, created using IDL's TRIGRID function, is used for visualization.



a b

Figure 3-2: Elevation (a) and canopy height (b) images derived from lidar data (2m spatial resolution). Roads and three distinct forest stand heights can be seen in this 64-ha area east of Rithet Creek.

3.5 Lidar data structure

As mentioned earlier, lidar data can be represented as independent points (vector data structure) or as an image (raster data structure). Lidar data are defined spatially by *x*-*y*-*z* triplets and thus are vector in origin. The original vector points can be visualized using three-dimensional rendering software, such as computer aided design (CAD) packages. Alternatively, the data can be rasterized into two-dimensional grid cells (pixels) and displayed as an image.

The vector approach provides the highest possible resolution, and maintains the integrity of the original data. This is especially important for the intensity data because of its fine-scale variability (the speckle characteristic of intensity imagery). As such a vector

approach will be used to *analyse* intensity data. However, it is difficult to visualize intensity in vector format because it adds a fourth dimension (*x*-*y*-*z*-*i*). Therefore, the two-dimensional raster approach will be used to *visualize* intensity data. By rasterizing intensity data, a suite of image processing techniques that have already been developed for other remote sensing and geographic applications can be applied to optimize visualization.

3.6 Visualization of intensity data

Raster images consist of a contiguous array of regularly spaced square grid cells (pixels). Due to scanning, aircraft motion, and platform attitude (e.g., roll, yaw, pitch), lidar data points do not conform to a regular grid pattern (Figure 3-3). The combination of these factors leads to a sample spacing that is difficult to predict. This creates issues for analysis in that some objects (i.e., trees) may be over-sampled while others are under-sampled. Irregularity of the sample positions together with the variability of intensity (speckle) present some unique challenges when visualizing intensity data. This section will describe the procedures used to create intensity images, and to reduce speckle in them.

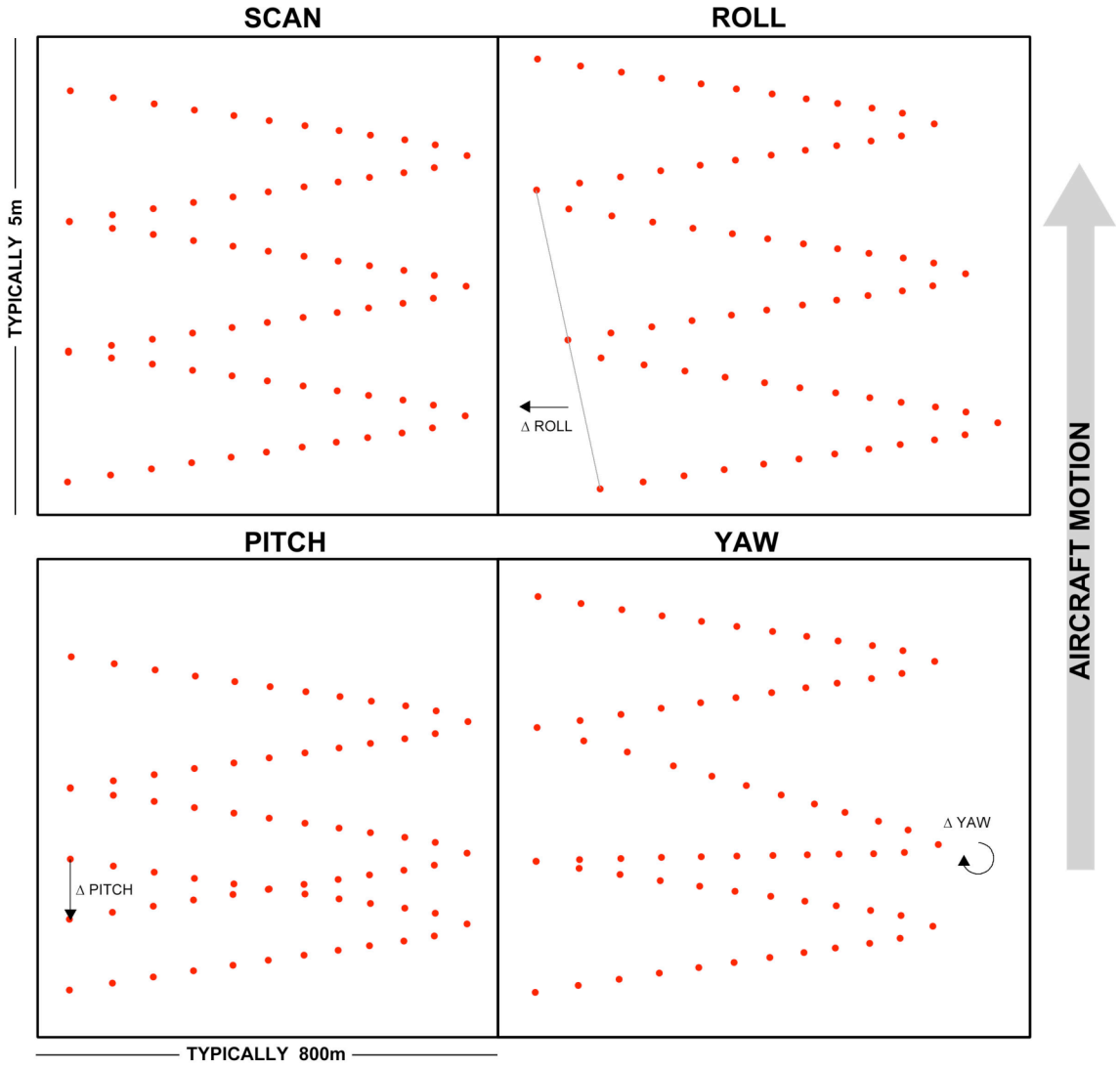


Figure 3-3: Influence of scan, roll, pitch, and yaw on lidar point spacing. The width has been compressed and point density reduced in these plan view schematics to emphasize the z-shaped scan pattern.

3.6.1 Interpolation

As mentioned in the introduction to this section, lidar datasets consist of point data. A wide variety of algorithms exist for interpolating point data to grid cells (e.g., nearest neighbour, inverse distance weighted, and kriging). IDL's TRIGRID function (linear interpolation from a TIN, see Section 3.4.1) was chosen because of its speed and its

ability to exclude edges of the images where no data exist due to the diagonal orientation of the study area. IDL has functions to interpolate point data using other algorithms, however, they cannot handle the large volume and high density of lidar point data. An attempt was made to use Surfer (Golden Software Inc., Golden, Colorado), which offers a wide variety of interpolation features. There were two problems with this software: (1) it interpolated to the edges of the files, not to the edge of the study area, and (2) the "image maps" it creates could not be exported as georeferenced images. Figure 3-4 shows and intensity image created using TRIGRID and enhanced with a linear 2% function (see next section).

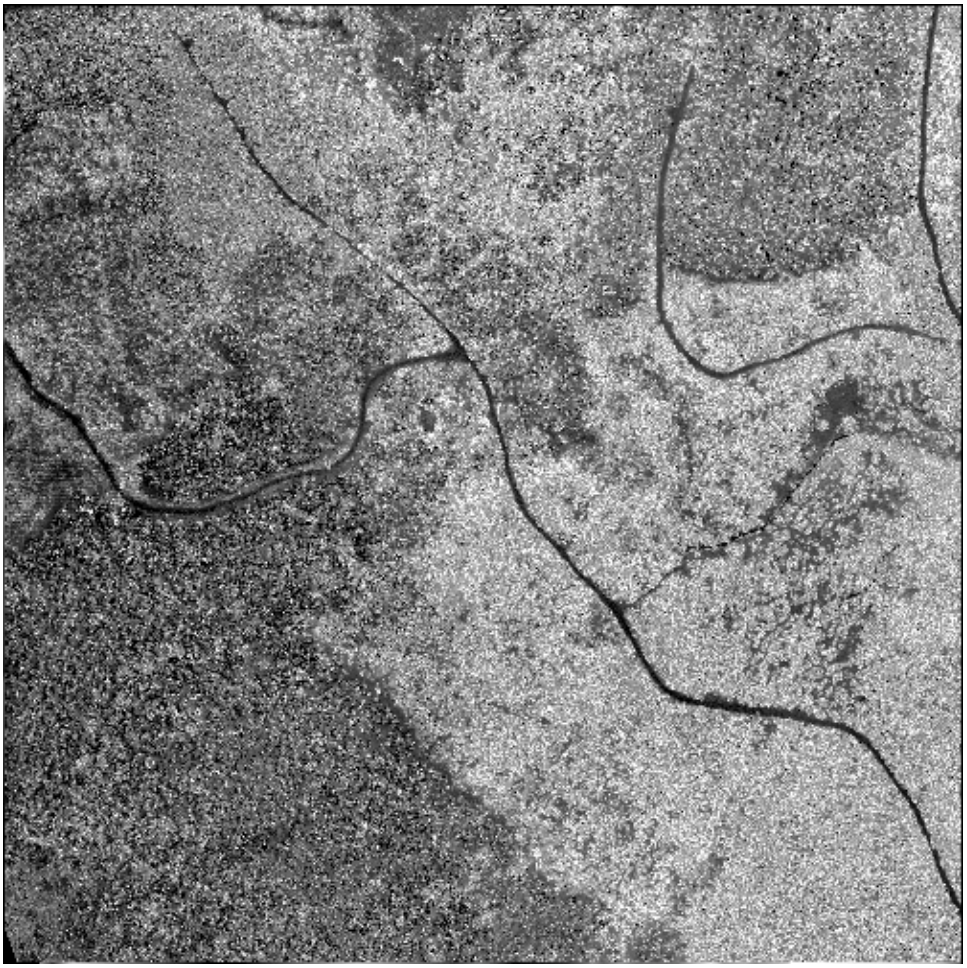


Figure 3-4: Intensity image of forest stands (100ha area; 2m spatial resolution)

3.6.2 Enhancement

The intensity values are approximately normally distributed, but highly leptokurtic (Figure 3-5). Note that the range of intensity values is confined by the radiometric resolution of 512 integer values. To improve visualization, contrast enhancement was applied to the intensity images.

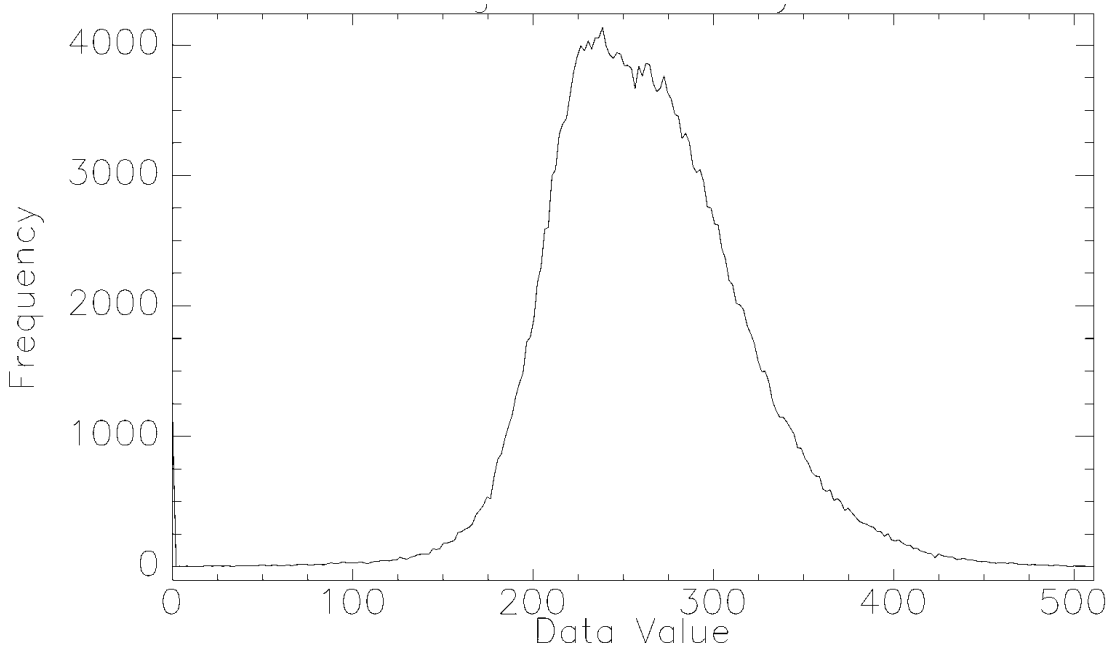


Figure 3-5: Intensity image histogram

Contrast enhancement, or histogram stretching, is a standard remote sensing and image processing procedure. The goal is to improve visual distinction between brightness values (BVs) by stretching the histogram over the full range of possible BVs (usually 256). It is important to note that enhancement is for visualization only. The original data values remain unchanged, but the displayed values are optimized for viewing. There are many different types of enhancements, each suited to different situations. The formula for one of the most common enhancements, the linear stretch, is presented as Equation 3-1.

$$BV_i = [(BV_i - MIN) / (MAX-MIN)] * 255 \quad (3\text{-}1)$$

Equation 3-1: Linear stretch

In the case of intensity images, a linear 2% enhancement produces optimum results. The 2% means that BVs falling in the upper and lower 2% of the histogram are saturated at 0 and 255, respectively, while the central 96% of BVs are stretched in a linear fashion over the full range of BVs (Figure 3-6). Another strategy to improve visualization, filtering, is discussed next.

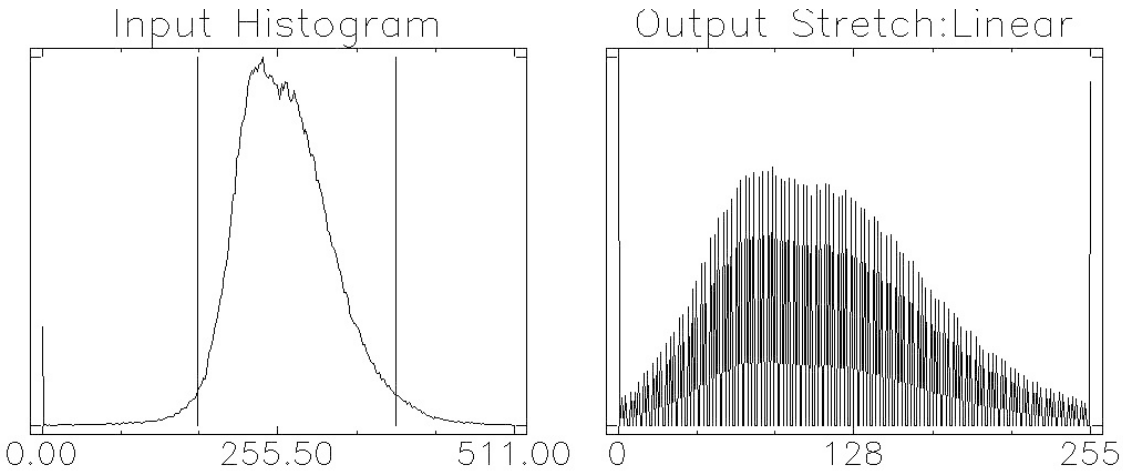


Figure 3-6: Linear 2% stretch

3.6.3 Filtering

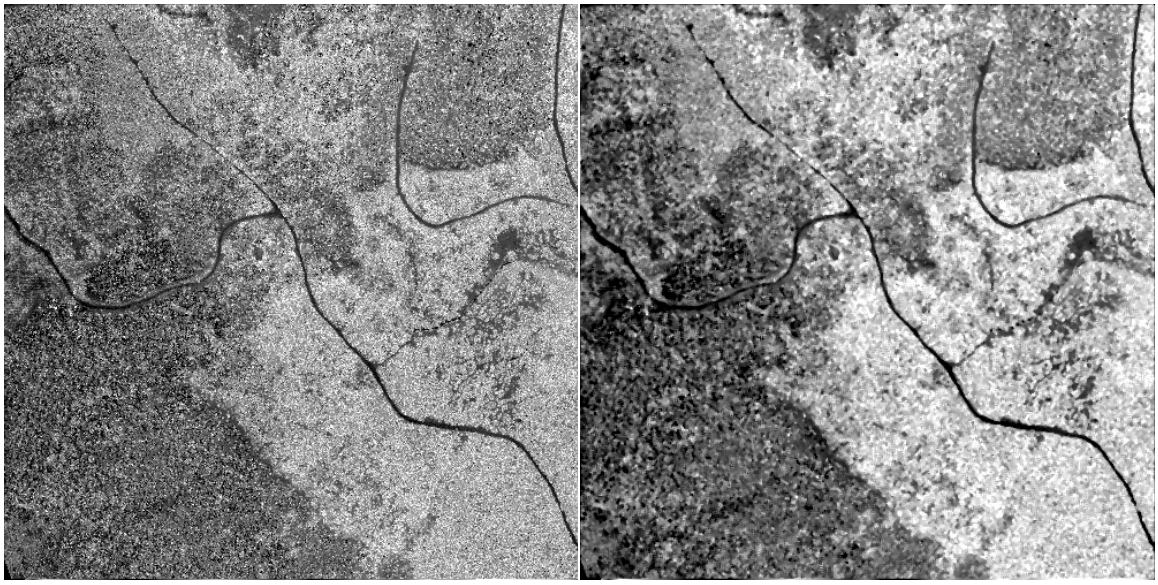
A common observation made about lidar intensity imagery is its fine scale speckling. It is hypothesized that the speckle is due to micro-scale (<1m) intrinsic surface complexity (variability in exposure angle, composition, moisture content, etc.). If one data point has an extremely high or low value, its intensity pixel will deviate from the general trend for the area. The nature of the speckle will be discussed further in Chapters 4 and 5. Filtering can be used to reduce speckle and improve visualization.

Spatial filtering is an image processing procedure, where a "window" (usually 3x3 pixels in size) is passed over an image systematically so that the window is centred on each pixel once. The central pixel value is modified based on statistics drawn from the entire window. In filtering the intensity images there are three goals: (1) reduce speckle, (2) improve visualization, and (3) emphasize overall trends.

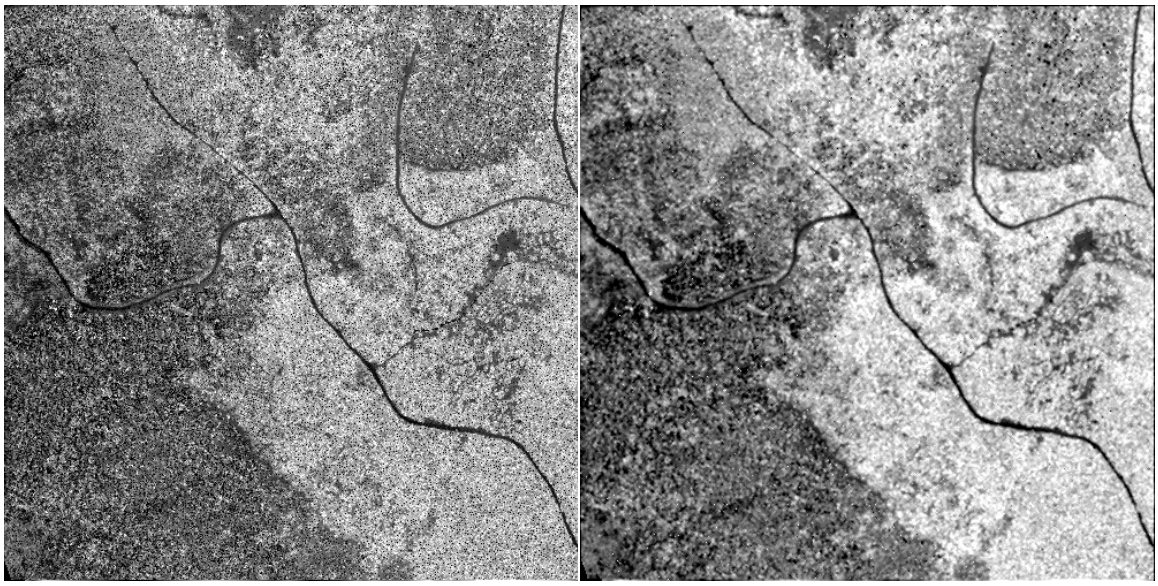
There are a plethora of options when choosing a filter. In light of the identified goals, a smoothing filter is needed. The simplest of these is a median filter. This filter replaces the central pixel with the median value of pixels in the window, thereby reducing the influence of extreme outliers. This filter is effective in reducing speckle within intensity images; however, the images are fuzzy and lack fine texture (Figure 3-7b). An alternative to the simple median is adaptive filtering. This category includes a variety of different filters that use the localized standard deviation of the window to compute central values. Adaptive filters are frequently used to reduce speckle in radar images without "oversmoothing".

Adaptive filters can achieve two goals: removing bit errors (extreme single pixels), and reducing speckle (Jensen 2005). Bit error filters compare the standard deviation of pixels in the window to the value of the central pixel. If the central pixel is outside of three standard deviations, it is declared a bit error and replaced by the mean of the window. Speckle reduction filters, such as the local sigma filter, compute the mean and standard deviation of the entire window and replace the central pixel with a mean calculated by including only pixels within one standard deviation of the original window mean. The advantage of speckle reduction filters is the retention of localized textures, since the standard deviation is computed separately for each window.

The local sigma filter was found to be more effective than the median filter, and much more effective than the bit error filter (Figure 3-7). Other adaptive filters available in remote sensing software are specifically designed for the visualization of radar imagery, and proved to be ineffective for lidar intensity imagery. Filtering improves the visual appearance of the intensity images, and, by reducing speckle, enables better observation of the intensity trends in forest stands.



a b



c d

Figure 3-7: Raw (a) and 3x3 filtered intensity images: (b) median, (c) bit error [sigma=2, tolerance=3], (d) local sigma [sigma=2.6]

3.7 Summary

In this chapter a few of the challenges faced in processing and visualizing lidar data were addressed. Due to its ability to handle large volumes of data, IDL was used for importing and processing lidar data. Standard IDL scripts were used, and custom ones were created to automate various processing routines. In order to investigate forest structure, absolute elevation values from the lidar data were converted to heights above ground by subtracting all points from a digital terrain model comprised of lidar ground points. Lidar height and intensity data are stored in vector (point) data structure, but can also be converted to raster images for visualization. By rasterizing intensity data, it is easier to observe general trends. Images can be enhanced and filtered using image processing techniques that have already been developed for other remote sensing applications. The resulting images improve visual interpretation of by emphasizing intensity trends. The next chapter deals with normalization of intensity data, the next major stage in lidar data processing.

Chapter 4 Normalization of intensity data

4.1 Introduction

Lidar intensity is recorded digitally as the ratio of the received to transmitted laser pulse energy. Chapter 2 identified several issues that must be considered before interpreting lidar intensity data. Some issues are related to laser physics and sensor engineering, while others are driven by dynamic survey geometry. Engineering issues, particularly the phase angle used in detection, need to be resolved before it is possible to calibrate intensity with known energy units (i.e., W/m^2). It is, however, possible to assess the impact of survey geometry issues on intensity values. In a previously mentioned study of lidar intensity over glacial surfaces, intensity was found to be directly related to range, angle of incidence, and footprint area (Lutz *et al.* 2003). These three factors can be reduced to one, because footprint area (the area illuminated by the laser) is a function of both range and angle of incidence (Figure 4-1). Differences in attenuation of laser energy through the atmosphere are assumed to be negligible given the wavelength and rages involved and are not considered here. This normalization procedure considers one sensor-specific issue (return type), in addition to footprint area.

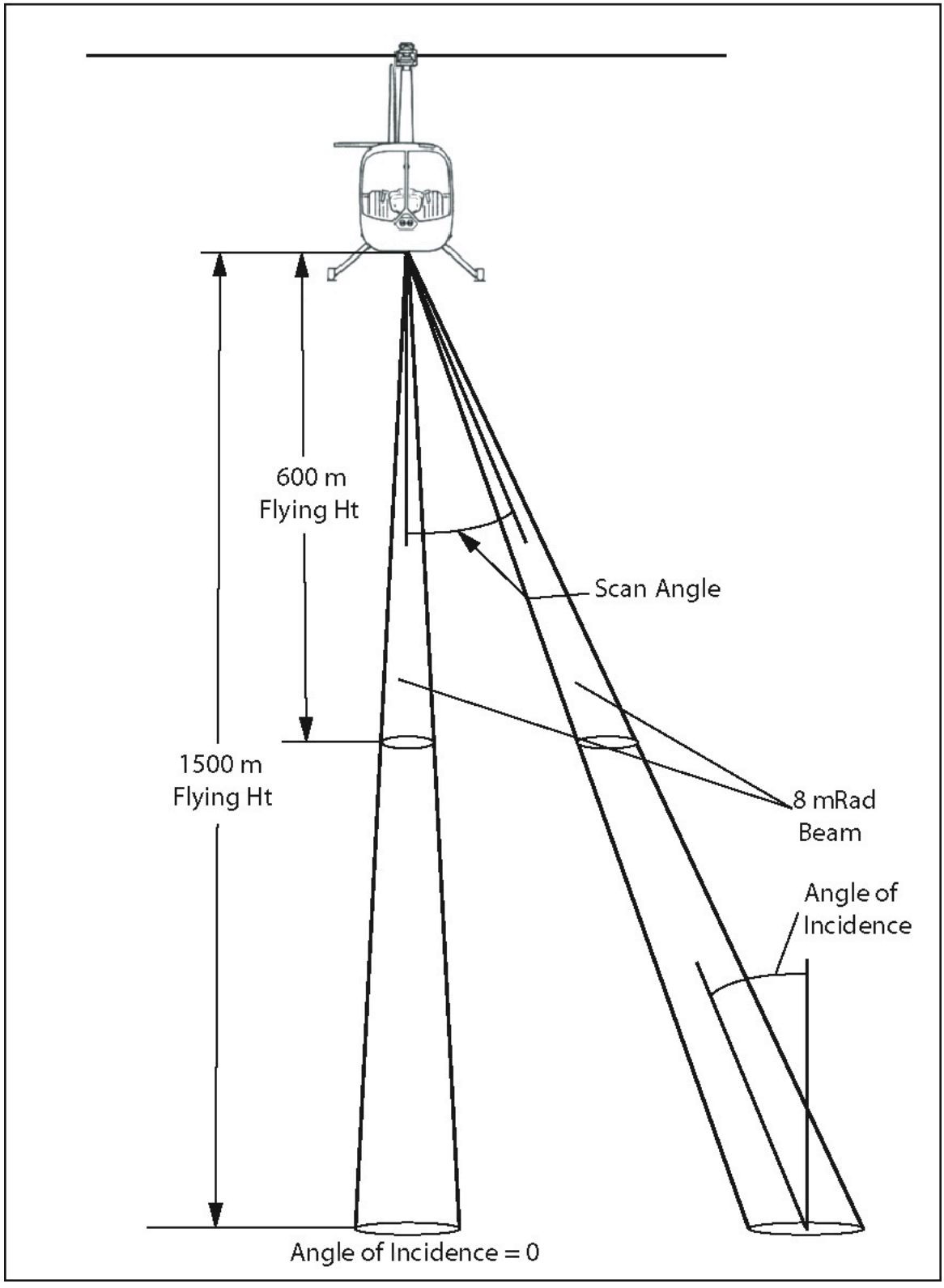


Figure 4-1: Controls of footprint area: beam divergence, range, and angle of incidence

4.1.1 Return type

The lidar system used in this study is capable of registering multiple returns for a given pulse. Due to sensor limitations, intensity was only recorded on the last return. In other words, intensity values are available for all singular returns (1/1), the second of double returns (2/2), the third of triple returns (3/3), and so on. Although it would be advantageous to have intensity values for all returns, the system does not have that capability. Nevertheless, there are still plenty of (last return) intensity data representing the forest cover, however, a question of comparability arises. The intensity of the second of a double return (2/2) is presumably reduced by the intensity of its first return (1/2), a value that is not recorded. Similarly, the third of a triple return (3/3) would be reduced by the intensity of both its first and second return (1/3 and 2/3).

4.1.2 Footprint area

The footprint is the area over which the laser energy is spread when it reaches the target (i.e., vegetation or ground). Since the laser beam divergence angle is constant, footprint area is directly controlled by range (the distance between the aircraft and the target), and angle of incidence (the angle between the incoming laser pulse and the surface normal). As the range increases, the laser beam spreads out and the footprint area increases. Similarly, as the angle of incidence increases, the shape of the footprint becomes more oblong and its area increases (Figure 4-1). Differences in footprint area are expected to influence intensity values, in addition to the intrinsic target (i.e., forest) reflection characteristics that we wish to examine later on.

4.1.2.1 Range

Range is the distance between the aircraft and the target. It varies primarily due to changes in aircraft altitude, but also due to the scan angle and terrain slope. Range and intensity are in fact the only two quantities measured directly by the laser. These original laser range data were not available, however, so range was calculated for each lidar data point using the point's coordinates, and time-matched aircraft trajectory data (Appendix B).

Normalization based on the inverse square of the range (see Section 2.2.3.1 on p.16) was attempted, but it grossly overcompensated for intensity differences due to range variation. This suggests that the relationship could be sensor-dependent.

4.1.2.2 Angle of incidence

Angle of incidence is a function of scan angle and terrain slope angle (see Figure 2-9 on p.19). However, determining angle of incidence using these two factors alone assumes a featureless (plane) surface. To determine angle of incidence, a digital elevation model (DEM) could be used to generalize the terrain slope. In the context of a forested landscape, however, the angle of incidence of incoming laser pulses on vegetation is much more complex. At the scale of the laser footprints (<30 cm in diameter) the forest materials (branches, shoots, leaves, needles, etc.) encountered have unique spatial orientations, which are different from the underlying terrain slope. Variability in angle of incidence is expected to add noise (speckle) to the intensity data; however, there is no practical way to normalize for it. As a result, the normalization will proceed based on range alone.

4.2 Objectives

The purpose of this work is to assess the impact of return type and range on intensity, and to determine the extent to which those impacts can be reduced. The goal, in the context of the overall thesis, is to enable meaningful comparison of intensity values occurring at different geographic locations and collected from different flightlines.

The objectives are as follows:

(1) Assess the impact of return type and range on intensity.

(2) Develop and implement a processing routine to normalize intensity data based on relationships identified in the previous objective.

(3) Validate the normalized intensity output.

4.3 Methods

4.3.1 Remotely sensed data

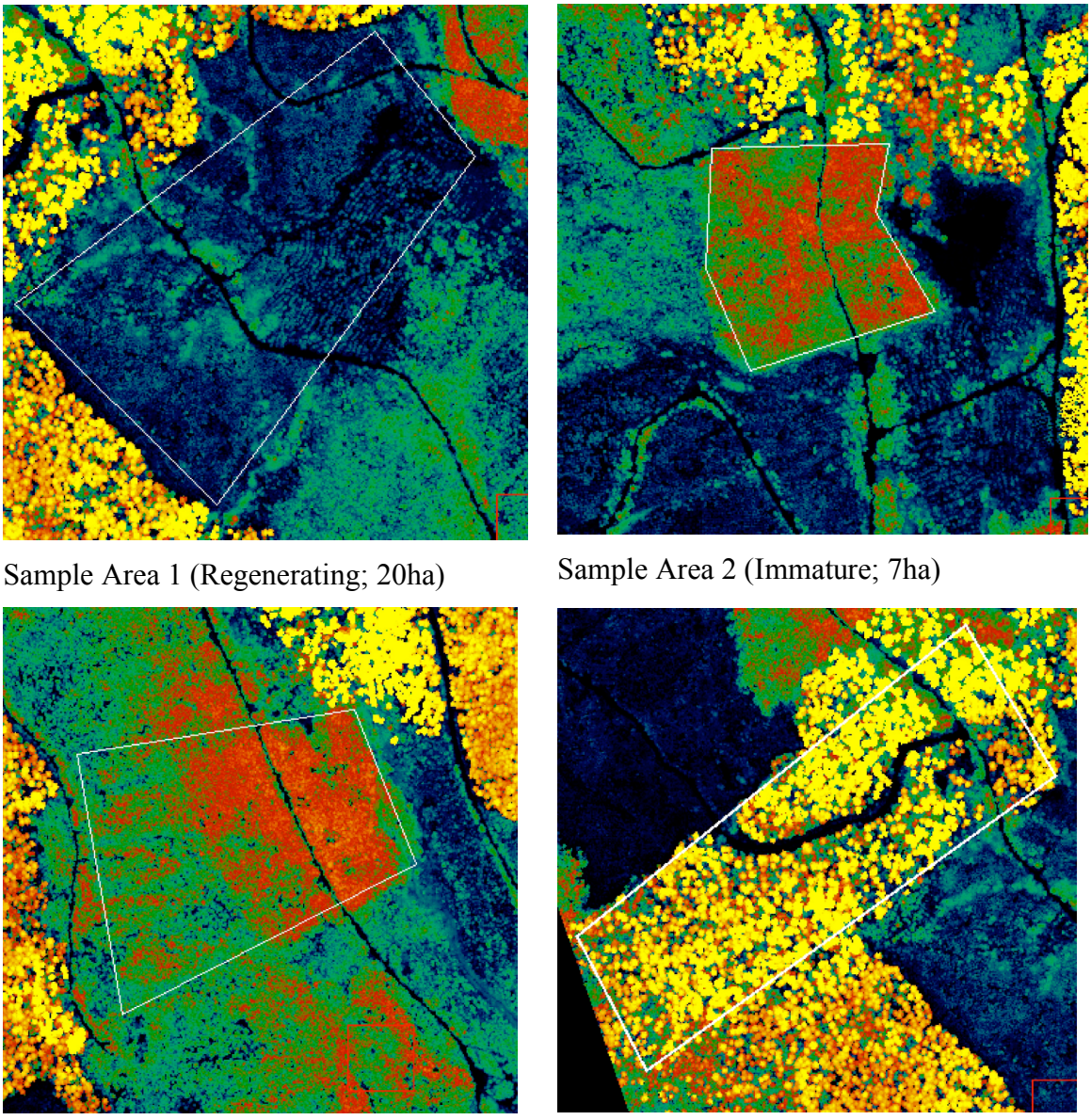
For details of the lidar instrument, technical specifications for this project, and data processing routines, refer to Section 3.2 on p.24. Data were acquired using 11 flightlines, flown at altitudes ranging from 855-1600m, with a maximum scan angle of ±24°.

4.3.2 Sample areas

In order to explore the influences of return type and range on intensity, the first step was to delineate sample areas of relatively homogenous forest cover. The purpose of this is to minimize the variation of intensity values due to changes in target composition so that other trends (return type and range) can be examined as clearly as possible. The samples should also be large enough to include data collected from several flightlines with a range of flying altitudes, scan angles, and terrain slopes, as all these factors contribute to the effective laser range.

Forest inventory maps, provided by the Capital Regional District, were used to select four sample areas representing three distinct and regionally representative stand age classes: one regenerating (<=20 years), two immature (>20 and <45 years), and one old growth (>200 years). These areas are outlined on Figure 4-2 over canopy height images.

Figure 4-2: Normalization sample areas (white lines) over canopy height (shaded)



Sample Area 1 (Regenerating; 20ha)

Sample Area 2 (Immature; 7ha)

Sample Area 3 (Immature; 14ha)

Sample Area 4 (Old growth; 17ha)

4.3.3 Return type

In order to demonstrate the impact of return type on intensity, the mean and standard deviation of intensity for returns of each possible type (1/1, 2/2, and 3/3) were calculated in each sample area. Because of the way energy is partitioned in multiple return scenarios (e.g., 1/2, 2/2), mean intensity is expected to decrease and variation is expected to increase as the number of returns for a given pulse increases.

To demonstrate statistically significant differences, a two-sample *t*-test was used to test for mean intensity differences between single (1/1), and second of double (2/2) returns. If significant differences exist, there is no need to test third of triple (3/3) returns, as it would already be confirmed that single and multiple returns are not comparable as expected. In such a case, only single (1/1) returns should be used for the next stage of this analysis, and for the remainder of the study.

4.3.4 Range

To explore the impact of range on intensity, mean and standard deviation of intensity values for 1-m binned increments of range were plotted for each sample. Once plotted, a suitable regression model was developed to explain the relationship statistically. This model was then used to create an equation to normalize intensity to a standard range. The standard range is used to scale all intensity values to the intensity value that would have resulted if the laser measurement were made at the standard range. An 800m standard range was chosen, because it represents the mean range for sample plots used in subsequent components of this study.

4.3.5 Validation

4.3.5.1 Visual

Three raster intensity images were generated for qualitative comparison. The first image is created from raw intensity data, the second from return-standardized intensity data, and the third from range and return standardized intensity data. The procedure used to rasterize, and enhance the images is documented in Section 3.6.

4.3.5.2 Statistical

Four 40 x 40 m (0.16ha) and two 20 x 20 m (0.04ha) validation plots representing the three distinct forest stand ages classes were used to validate range standardization. Locations of the validation plots correspond with field plots that will be used in later stages of this study. There are actually nine field plots, but three were excluded from validation because they are located inside the sample areas used to develop the normalization model. Due to overlapping flightline coverage, each validation plot contains lidar points originating from two or more flightlines. The occurrence of multiple flightlines contributes to a variety of range distances at which lidar data were acquired for each validation plot (see Table 4-4 on p.53). Areas covered by multiple flightlines provide an opportunity to assess the validity of the normalization. If intensity has been properly normalized, there should be no significant differences in intensity values between flightlines.

To ensure that only overlapping flightline coverage was included in such an analysis, the lidar points contained within each plot were segmented into 1x1m square grid cells. Cells containing points originating from all flightlines present within the plot were flagged as overlap cells. Any flightline representing less than 10% of the total population

of points was excluded. If these low-contribution flightlines were included, the quantity and representativeness of overlap cells would be poor.

Analysis tables were built for each plot, so that the intensity value and flightline number of each point within an overlap cell were included. Paired sample *t*-tests (for plots with only two unique flightlines), and repeated measures tests (for plots with more than two unique flightlines) were used to test for statistically significant differences between intensity grouped by flightline using SPSS (SPSS Inc., Chicago, Illinois).

4.4 Results

4.4.1 Return type

The results show that mean intensity decreases and variation increases with the number of returns (Table 4-1). Two sample *t*-tests confirm that the single (1/1) and second of double (2/2) mean intensity values are significantly different ($p<0.05$) for all samples (Table 4-2). These results confirm findings from another study that return type does indeed have an impact on intensity (Moffiet *et al.* 2005).

Table 4-1: Descriptive statistics for intensity by return type

Return type	1/1			2/2			3/3		
Statistic	Mean	StDev	COV	Mean	StDev	COV	Mean	StDev	COV
Sample 1 (R)	285	57	.20	181	78	.43	n/a	n/a	n/a
Sample 2 (I)	289	51	.18	176	76	.43	132	58	.44
Sample 3 (I)	282	54	.19	170	71	.42	129	53	.41
Sample 4 (O)	244	59	.24	150	64	.43	n/a	n/a	n/a

Table 4-2: Mean differences between single (1/1) and second of double returns (2/2)

Sample	*t*	*p*
1	244.781*	0.00000
2	247.944*	0.00000
3	411.651*	0.00000
4	431.846*	0.00000

* unequal variances assumed

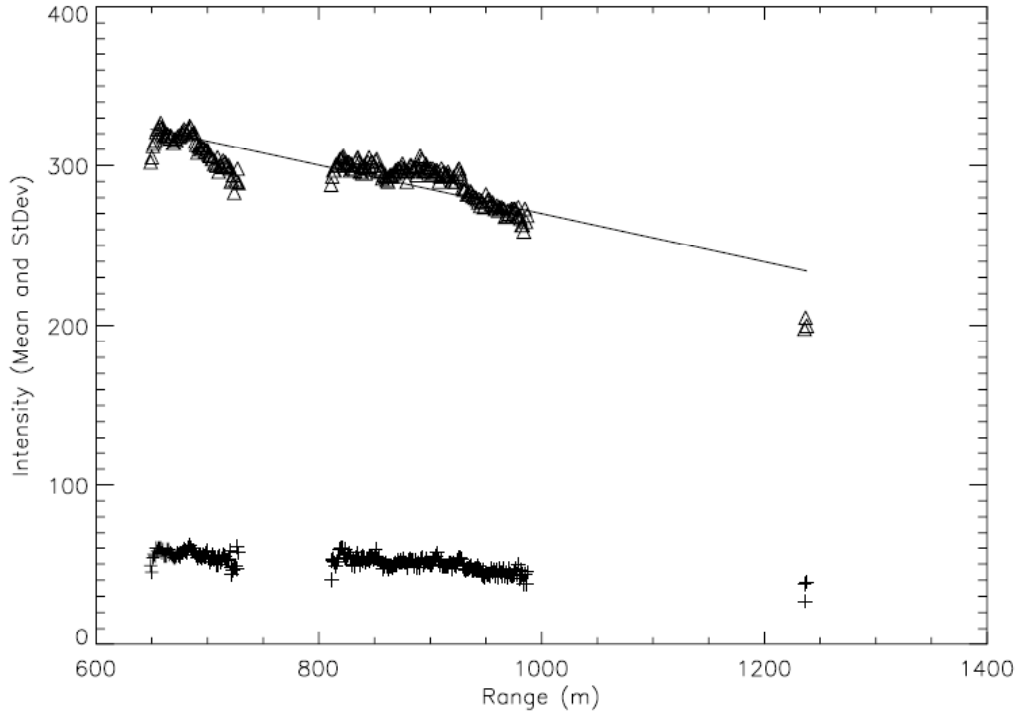
4.4.1.1 Implications

Knowing that single return intensity values are not comparable with double returns the next question that arises is what can be done about this. Since the variation also increases with return number, it is definitely not appropriate to scale the second of double (2/2) and third of triple (3/3) return intensity values. However, single returns comprise a large proportion of the available intensity data. In the regenerating sample area, single returns make up 93% of the data, while in the immature and old growth samples, single returns represent 76% and 74% respectively. With such a large portion of the dataset consisting of single returns, it was decided to extract single returns and work exclusively with them for the remainder of the study. From this point on, intensity refers to single return (1/1) values only unless otherwise specified.

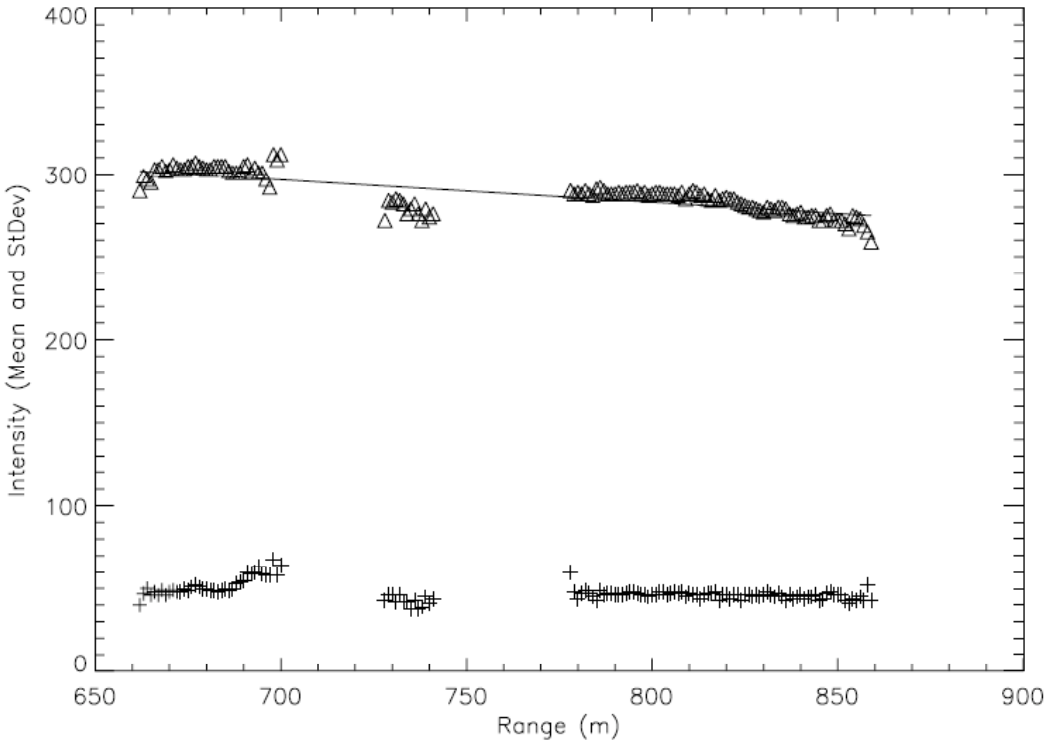
4.4.2 Range

To visualize the relationship, plots of range vs. mean and standard deviation of intensity were produced for all sample areas (Figure 4-3). As expected, a negative relationship exists between range and mean intensity. In contrast, the standard deviation of intensity appears to be fairly constant across a wide variety of range distances.

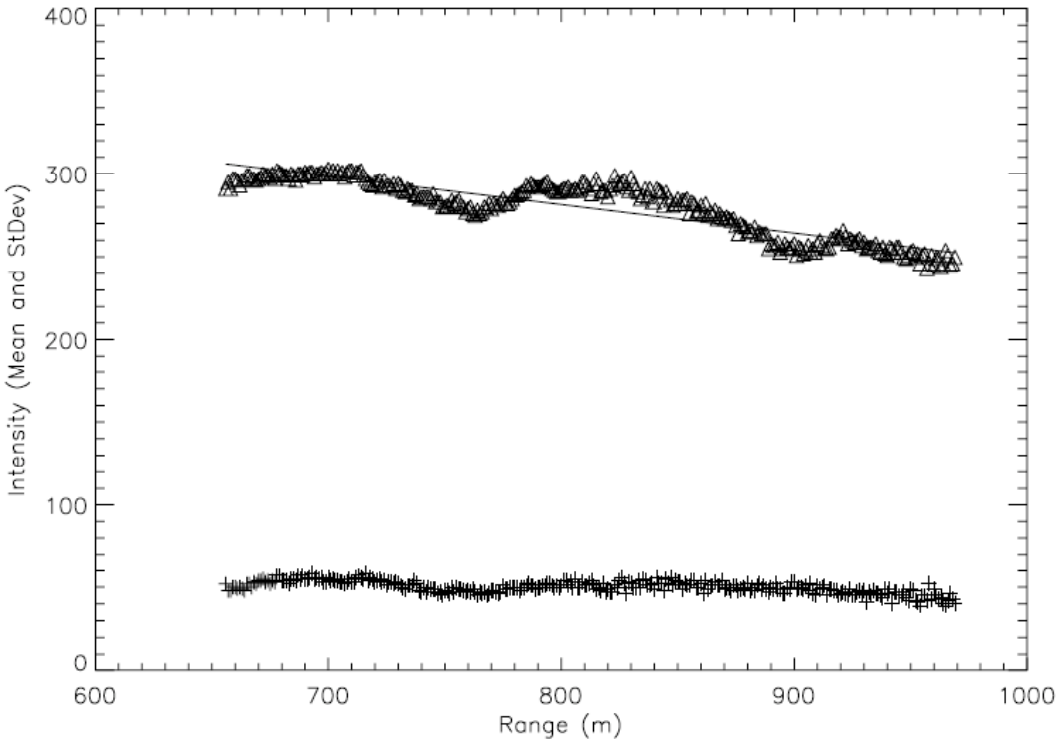
Figure 4-3: Range in 1m bins as a function of intensity mean (Δ) and standard deviation (+). The horizontal breaks represent flightline edges.



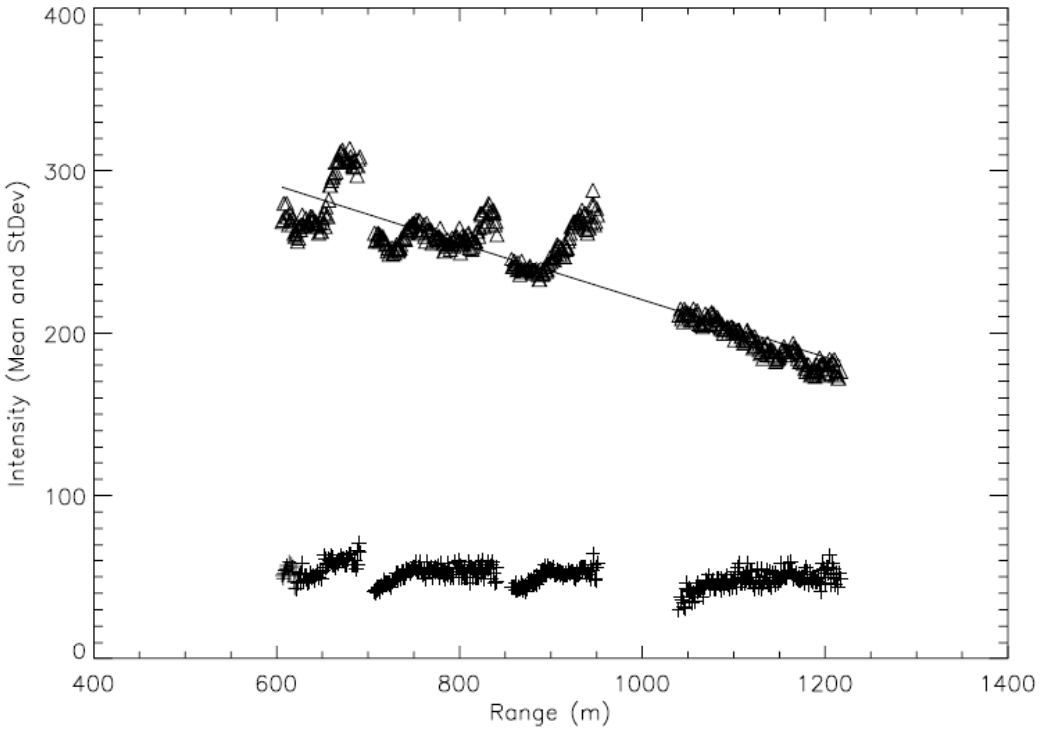
(a) Sample #1 (regenerating)



(b) Sample #2 (immature)



(c) Sample #3 (immature)



(d) Sample area #4 (old growth)

A simple linear regression was applied to determine the slope of the trend between range and intensity (Table 4-3). Similarity in slope for each of the sample areas offers some confidence that the relationship between range and intensity is consistent throughout the study area. The average slope (-.1567) was used in range standardization (Equation 4-1).

Table 4-3: Range regression models

Sample	Slope	r^2	p	Standard Error of the Estimate	Range (m)
1	-.1342	.744	.000	9.08	646-1269
2	-.1448	.669	.000	6.66	658-873
3	-.1683	.790	.000	7.89	651-992
4	-.1796	.829	.000	15.5	593-1248
Average	-.1567				

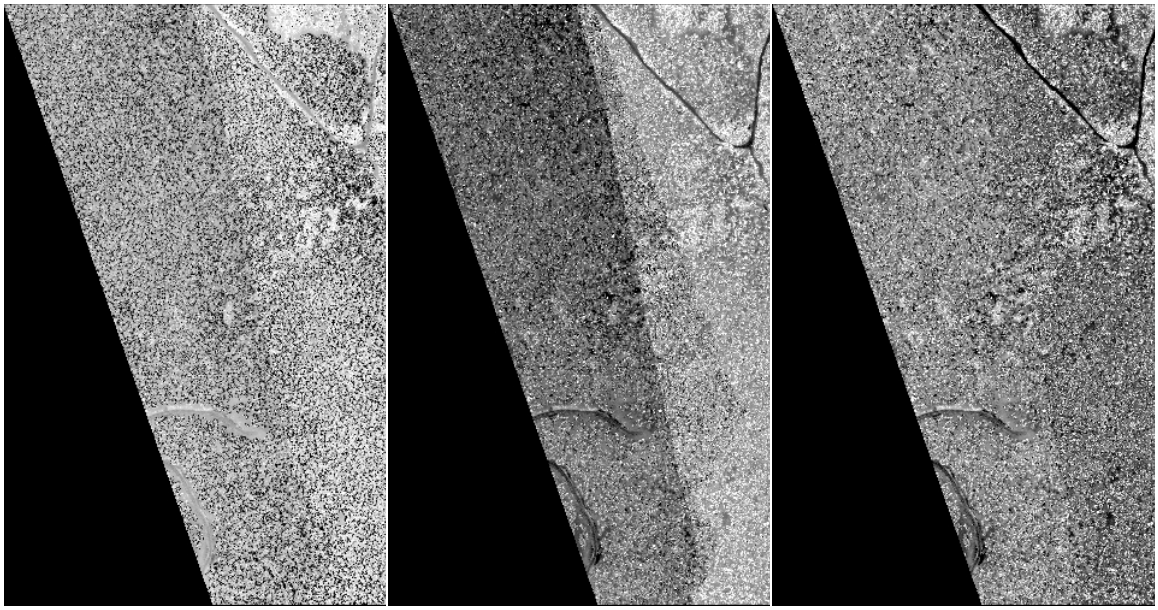
$$intensity_{standardized} = intensity_{raw} + [\ -0.1567 * (\ range - 800\ )\ ] \quad (4\text{-}1)$$

Equation 4-1: Range-standardized intensity

4.4.3 Validation

4.4.3.1 Visual

Raster intensity images created using (1) raw intensity data, (2) single return intensity data, and (3) range-standardized single return intensity data are presented in Figure 4-4.



a b c

Figure 4-4: Intensity images created using: last returns (a), single returns (b), and single range-standardized returns (c). These images represent the same 46ha of forested land and have a 2m spatial resolution. The black section is outside the lidar acquisition area.

4.4.3.2 Statistical

Summary statistics for raw and range-standardized intensity values are presented by plot and flightline in Table 4-4. Range standardization appears to improve the correspondence of means without affecting variation. Note that plots 1, 3, and 5 were removed from the validation because they are located within the sample areas used to build the model.

Table 4-4: Sensor and intensity data by plot and flightline

Plot ID	Flightline	Range (m)		Raw single return intensity			Range-standardized single return intensity			*n*	%
		Min	Max	Mean	StDev	CoV	Mean	StDev	CoV		
2	3	748	821	273	52	0.19	270	53	0.20	1097	38
2	9	611	681	293	59	0.20	268	60	0.22	1818	62
4	2	932	969	272	48	0.18	294	47	0.16	1802	28
4	10	661	696	307	54	0.18	286	54	0.19	1448	22
4	11	798	836	277	47	0.17	278	47	0.17	3217	50
6	4	827	842	308	54	0.18	313	54	0.17	557	49
6	8	752	771	320	57	0.18	313	57	0.18	570	51
7	1	1269	1313	170	48	0.28	245	46	0.19	373	9
7	2	800	834	283	51	0.18	285	51	0.18	1580	36
7	10	669	689	284	48	0.17	264	48	0.18	1149	26
7	11	833	873	257	45	0.18	264	45	0.17	1280	29
8	2	784	838	254	45	0.18	255	45	0.18	2233	47
8	10	634	674	249	39	0.16	227	38	0.17	570	12
8	11	805	869	234	45	0.19	239	45	0.19	1903	40
9	4	781	793	265	50	0.19	263	50	0.19	1211	55
9	8	658	671	297	57	0.19	276	57	0.21	1011	45

The results from the group difference tests (Table 4-5) show that there are in fact significant differences ($p < 0.05$) in five of the six plots. In these five cases, range-standardization has not been entirely effective in reducing intensity discrepancies due to flightline geometry issues.

Table 4-5: Flightline differences in intensity

Plot ID	Flightlines	GLM repeated measures test			Paired samples *t*-test			*n*
		F	*df*	*p*	*t*	*df*	*p*	
2	3,9				2.068	508	.039	509
4	2,10,11	20.994	2	.000				386
6	4,8				-.868	322	**.386**	323
7	2,10,11*	10.822	2	.000				190
8	2,10,11	13.198	2	.000				85
9	4,8				-2.051	392	.041	393

*flightline 4 was excluded because it represented <10% of the plot

4.5 Discussion

Since intensity values captured in multiple return situations are significantly lower than those from single returns, comparing intensity values of different return types is not appropriate. Fortunately, single returns comprise a large percentage (>74%) of the available data. The single return subset is considered to be the most representative of the target surfaces because these values are the least affected by earlier returns partitioning the laser pulse energy. As such, single return intensity values were extracted from the dataset for all future analyses.

Results from range-standardization of intensity are inconclusive. Normalized intensity images show marked improvement over images created from raw intensity data. It is also apparent (from Table 4-4) that, even where differences exist, the range-standardized intensity values for the flightlines within each plot have closer means than the raw intensity values do. Still, the statistical tests report that in five out of six plots there are significant differences between flightlines when there should not have been if the normalization were successful.

There are three potential factors explaining these unexpected significant differences:

(1) The normalization procedure was incomplete (e.g., it did not include angle of incidence, which is expected to add variation).

(2) The four sample areas were not sufficiently homogenous to capture the precise trend between range and intensity.

(3) The 1x1m grid cells used (in validation plots) to identify areas of flightline overlap were not homogenous, or were not sampled in the same way by all flightlines due to scan geometry (e.g., at nadir the laser detects a gap, while at higher scan angles it detects the side of a tree crown) and angle of incidence.

Range-standardization closes the gap between intensity differences caused by sensor geometry, but in some plots flightline-induced differences persist and are statistically significant.

To investigate the apparent normalization error, the range of slope coefficients estimated in the range standardization models (0.1342-0.1796) was considered. Even for the most extreme correction (at minimum and maximum range) this was not expected to contribute more error than ±9 intensity units. This could explain the 2-unit found in plot 2, however, it exceeds the mean differences found in other plots, which range from 13 to 28 units. An attempt was made to build range standardization models separately for each flighline, but the limited variation in range distances for individual flightlines resulted in inconsistent trends (i.e., the slope coefficients were inconsistent).

4.6 Conclusion

For lidar systems that are incapable of recording intensity for all returns, only single (1/1) returns should be used when interpreting intensity values. Range, a key contributor to the footprint area, has a strong negative relationship with intensity ($.669 < r^2 < .829$). The validation results suggest that while the range-standardization improved correspondence between different flightlines, it did not result in perfect matching of the intensity values between flightlines for a majority of validation plots.

To improve precision of the normalization model, a more uniform sample area (e.g., a paved tarmac surveyed at a variety of altitudes and scan angles) is suggested. In addition, a complex model needs to be developed to predict angle of incidence of the laser pulses within a forested environment.

The inconclusive outcome of this normalization procedure has consequences on the way intensity data can be used for further analyses. The range-standardized intensity values should be used with caution. Normalization of these values was not statistically validated for the majority of plots, and therefore is not considered to be reliable. On the other hand, even more caution should be taken when using raw (single return) intensity values. The normalization and validation processes have demonstrated that raw intensity values are flightline dependant, due to range differences. There are, however, a few options for using raw data. One is to use single flightlines, where the flight altitude is relatively stable, in order to minimize the variation in range. Another option is to use value-independent derivatives of the intensity data, such as the coefficient of variation, or spatial pattern statistics (e.g., Moran's *I*).

Chapter 5 Intensity in a forested environment

5.1 Introduction

Using lidar intensity data to characterize forests is a new area of study. While the range-based (height) data have been utilized for over 20 years, intensity data have largely been ignored. There are several reasons for its omission; most likely it is because of a daunting list of calibration and normalization issues, which have been documented in earlier chapters. The overall goal of this thesis is to better understand the meaning of the intensity data, and to learn what it can tell us about forest structure.

Due to a normalization procedure that remains unsolved, and the inherent variability of intensity generally, it is difficult to interpret the intensity values individually. There is, however, an opportunity to look for patterns in the intensity data to better understand its spatial structure (i.e., agglomerations of high and low values). Using spatial statistics to estimate the scale and magnitude of these spatial structures, an improved understanding of the stand (e.g., crown and gap structure), and tree-level attributes (e.g., height, crown class, species) that drive the intensity patterns can be developed. Relationships between spatial pattern and forest structure have been previously demonstrated with optical imagery (Wulder & Boots 1998).

5.1.1 The nature of intensity data

As mentioned in previous chapters, there are many issues concerning the reliability of intensity data in a forested environment. While range-standardization resulted in greater consistency of the intensity data, its inherent variability persists. Even in areas of relatively homogenous forest cover the variability of the intensity data is high. For

example, in the sample areas used for normalization, the coefficient of variation ranged from 0.18 to 0.24 for single (1/1) returns before range-standardization and from 0.17 to 0.20 after (Table 5-1).

Table 5-1: Variation of single return (1/1) intensity data before and after range standardization for homogenous sample areas defined in Figure 4-2 on p.44

Sample Area	Raw intensity			Range-standardized intensity		
	Mean	StDev	CoV	Mean	StdDev	CoV
1 (Regenerating)	285	57	.20	289	57	.20
2 (Immature)	289	51	.18	286	50	.17
4 (Immature)	282	54	.19	282	53	.19
3 (Old growth)	244	59	.24	260	52	.20

This variability is attributed to the complex interaction of the laser footprint with many different kinds of forest elements, including: needles, leaves, bark, and gaps. These unique canopy microstructures are composed of rough, discontinuous surfaces (shoots and branches) that vary in size, shape, density, and orientation with respect to the incident laser pulse. Another contributor to the variability is incomplete normalization. As mentioned in Chapter 4, the normalization procedure does not consider a complex factor, angle of incidence, which is highly dynamic in the forest context because of varying branch and foliage orientations.

5.1.2 Spatial pattern

Variability in the intensity values results in fine scale speckling of intensity imagery (see Figure 5-3 on p.67), however, coarser patterns are also evident. The presence of multi-scale spatial structure within the intensity data suggests that they contain important, added information about forest composition and structure. Essentially, spatial structure refers to clustering of related attribute values (i.e., intensity) in space. This structure can

be explored using spatial statistics. While lidar provides attribute data (i.e., intensity) with three spatial dimensions (x, y, and height), all known measures of spatial association accommodate only one (i.e., transect analysis), or two (i.e., spatial dependence) spatial dimensions. To take advantage of the three-dimensional spatial depth of lidar data, two-dimensional spatial dependence analysis (x, y, attribute) was selected.

Two measures of spatial dependence are global and local spatial autocorrelation. Global spatial autocorrelation detects spatial association in an entire scene. Global measures identify the average or typical dependence for a region. Local measures, in contrast, can be used to characterize variation in the local nature of spatial association (Boots 2002).

5.1.3 Global spatial autocorrelation

One common measure of spatial autocorrelation is Moran's I (Moran 1950). Moran's I considers the nature of association in the attributes of points as a function of distance (Equation 5-1).

$$I = \left(\frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \right) \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{5-1}$$

Equation 5-1: Moran's *I*

Where x is the attribute value, and w_{ij} is the spatial weighting matrix. For fixed-distance estimates w_{ij} is a binary factor equal to one for all points within the specified distance, and to zero for all points beyond. This binary spatial weighting matrix relates to a spatial statistics concept called the lag distance. This is the distance, or search radius,

surrounding each point that the statistic will look for autocorrelated neighbours (Figure 5-1). By computing Moran's *I* for several distances and comparing the Moran's *I* statistics, it is possible to determine the distance (i.e., scale) at which spatial associations are the strongest. This information tells us about the spatial scale of the overall intensity pattern. One of the main assumptions of the Moran's *I* is stationarity of the spatial pattern. By using a global measure to describe spatial autocorrelation, it must be assumed that the process(es) leading to spatial dependence (e.g., crown size, gap diameter, etc.) are the same (i.e., stationary) throughout the area for which the global statistic is computed (Boots 2002).

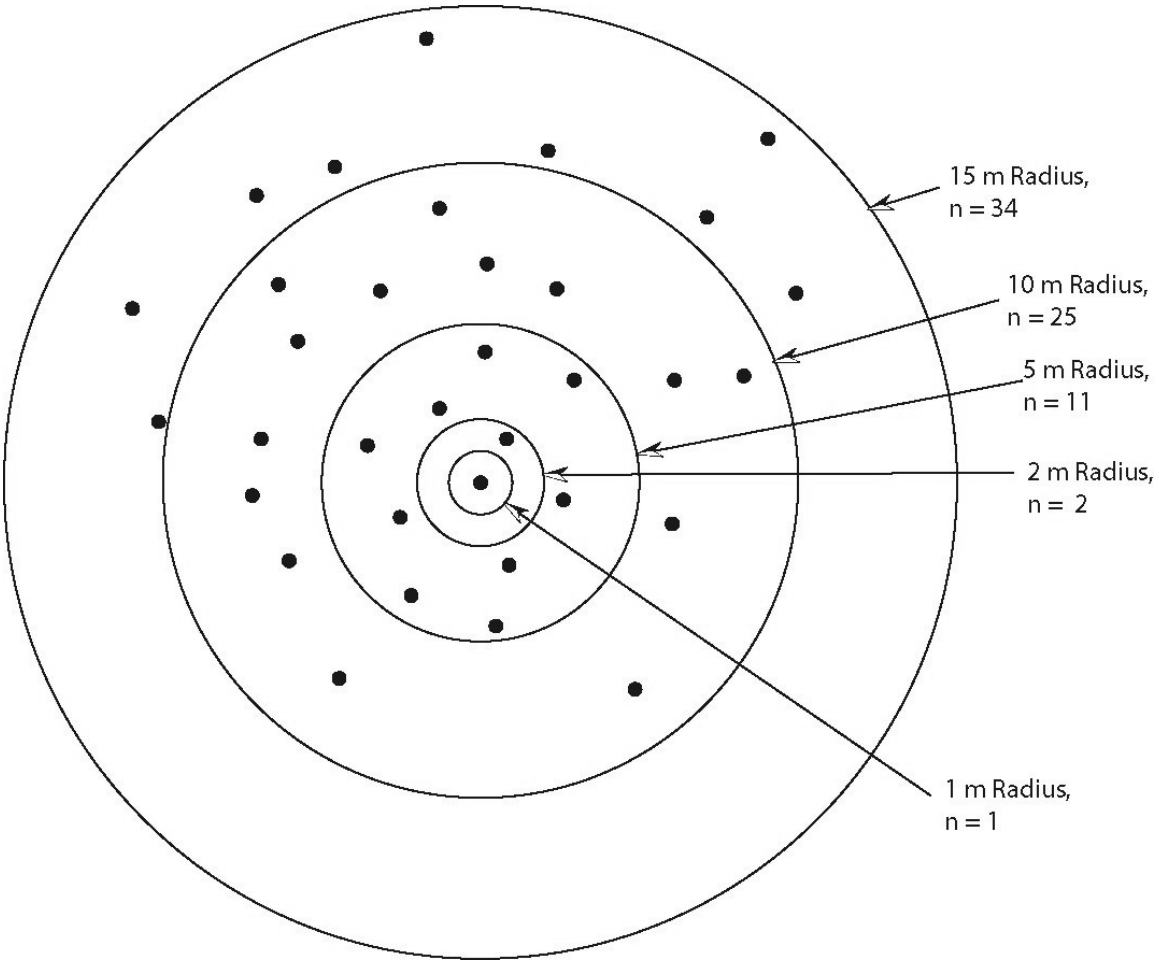


Figure 5-1: Example of various lag distances from a central point

The standardized version of Moran's *I* (Equation 5-2), is based on the expected distribution of Moran's *I* values (Cliff & Ord 1981). With the null hypothesis of complete spatial randomness, the standard values give information about the probability that the data are in fact spatially autocorrelated. Thus, the standardized Moran's *I* values are *z*-scores or standard deviations from the expected distribution. By definition, if the standardized value of Moran's *I* exceeds 1.0, there is 68% confidence that spatial autocorrelation exists, and if it exceeds 2.0 there is 95% confidence.

Standardized Moran's *I* is calculated as follows:

$$\mathrm{var}_N(I) = \frac{1}{(n-1)(n+1)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2} \times \left[n^2 S_1 - nS_2 + 3\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2 \right] - \frac{1}{(n-1)^2}$$

where $S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2$ and $S_2 = \sum_{i=1}^{n}(w_{ij}w_{ji})^2$

Whereas the expected is $E_N(I) = -(n-1)^{-1}$

and the z-score is calculated as $z = \frac{I - E(I)}{\sqrt{\mathrm{var}(I)}}$ (5-2)

Equation 5-2: Standardized Moran's *I*

5.1.4 Local spatial autocorrelation

A local measure of spatial autocorrelation, the Getis G_i*, is particularly suited for lidar intensity as it identifies spatial clusters of similar values that are extreme (high and low) relative to the mean (Ord 1992). The G_i* statistic is computed for each point, as a function of distance, using Equation 5-3.

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \tag{5-3}$$

Equation 5-3: Getis statistic

The standardized form of G_i* is computed using Equation 5-4. G_i* values are assumed to be normally distributed if there are eight or more points in the spatial neighbourhood defined by distance *d* (Griffith *et al.* 1996). G_i* values are also more likely to be significant in the presence global spatial autocorrelation (i.e., significant Moran's *I*). Thus G_i* is best used as an exploratory tool, rather than for hypothesis testing.

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - W_i^* x}{s[W_i^*(n - W_i^*)/(n-1)]^{1/2}} \tag{5-4}$$

Where: $x = \frac{\sum_j x_j}{n}$, and $s^2 = \frac{\sum_j (x_j - x)^2}{n}$, and $W_i^* = \sum_j w_{ij}(d)$

Equation 5-4: Standardized Getis statistic

From this point on, G_i* will refer to the standardized form. Negative G_i* values indicate the presence of spatial association among intensity values that are low relative to the mean, while positive G_i* values indicate spatial association of intensity values that are high relative to the mean. G_i* values near zero represent intensity values that are close to the mean. Thus, G_i* contains information about the strength of association, and the relative magnitude of the associated values. Near zero values can mean two things: 1) there is no significant spatial autocorrelation, or 2) there may be spatial autocorrelation but intensity values are near the mean.

Getis statistics have been used to examine spatial structure in raster images, including forest stands, snow cover, deserts, and coral reefs. The spatial characteristics of conifer stands, measured by Getis statistics, have been linked to ecological processes resulting from different disturbance histories (e.g., wildfires, and forest management) (Wells & Getis 1999). In a study using Landsat TM images, the G_i^* statistic applied to one spectral band identified forest clusters on par with classification algorithms requiring multiple band inputs (Wulder & Boots 2001). With passive microwave radiometry data, studies have used Getis statistics to link "brightness temperatures" of snow pack to atmospheric airflow (Derksen *et al.* 1998), and to map the intensity of desertification resulting from a sandstorm in China (Jin & Yan 2004). At Lunar Lake Playa (Nevada), an area deemed sufficiently uniform to calibrate satellite image radiometry, Getis statistics have been employed to identify subtle heterogeneities used to assess radiometric calibration of the SPOT satellite (Bannari *et al.* 2005). Getis statistics have also been applied to SPOT images of the marine environment to identify stress in coral reefs (LeDrew *et al.* 2004).

5.2 Objectives

The goal of this work is to better understand the nature of the intensity data in a forested environment. This will be accomplished through analysis of intensity values, and through global and local measures of spatial autocorrelation.

The objectives are as follows:

(1) Confirm that significant differences exist between intensity values representing the forest canopy, and those representing the ground beneath it.

(2) Estimate and identify trends in the spatial scale of the intensity pattern for plots representing each stand age-class.

(3) Explore relationships between intensity and *lidar*-measured canopy height, using intensity values, and local measures of spatial autocorrelation.

(4) Explore relationships between intensity and *field*-measured tree attributes (crown class, species, and mortality), using intensity values, and local measures of spatial autocorrelation.

5.3 Methods

5.3.1 Study area

The study area is the Rithet Creek sub-basin of the Greater Victoria (Sooke Lake) Watershed. This valley is approximately seven kilometres long and two kilometres wide, and it ranges in elevation from 180m to 750m above mean sea level. The ecosystem is part of the Coastal Western Hemlock very dry maritime (CWHxm) biogeoclimatic subzone. Overstorey canopies are generally dominated by Douglas-fir with lesser components of more shade-tolerant species, such as Western hemlock, and Western redcedar (Trofymow *et al.* 1997). Shrubs include salal, Oregon-grape, huckleberry, baldhip rose, ocean spray, false azalea, salmonberry, devil's club, and Labrador tea (CRD 2006).

5.3.2 Field data

A field campaign was conducted between May 26 and July 26, 2005 in order to make systematic observations and to collect tree-level attribute data. Nine plots were established to include three samples from each of the three distinct and regionally representative forest stand age classes in the study area: regenerating (<=20 years), immature (>20 and < 45 years), and old growth (>200 years). Stand age was determined using forest inventory maps maintained by the Capital Regional District. Regenerating plots were 20 x 20 metres (0.04 ha) in size and all other plots were 40 x 40m (0.16 ha). Table 5-2 presents a summary of field plot attributes and Figure 5-2 and Figure 5-3 show their locations in the Rithet Creek study area.

Table 5-2: Field plot attributes

PLOT ID	MEAN STAND AGE (YEARS)	SIZE (HECTARES)	STEM DENSITY (PER HECTARE)	BASAL AREA (PER HECTARE)
1	43	0.16	2325	47.95
2	217	0.16	2188	73.41
3	17	0.04	1800	26.41
4	38	0.16	1306	51.73
5	217	0.16	725	52.19
6	19	0.04	6225	30.36
7	39	0.16	1013	29.63
8	217	0.16	781	76.47
9	20	0.04	2050	11.27

Figure 5-2: Field plot locations (red) over lidar canopy height model (shaded)

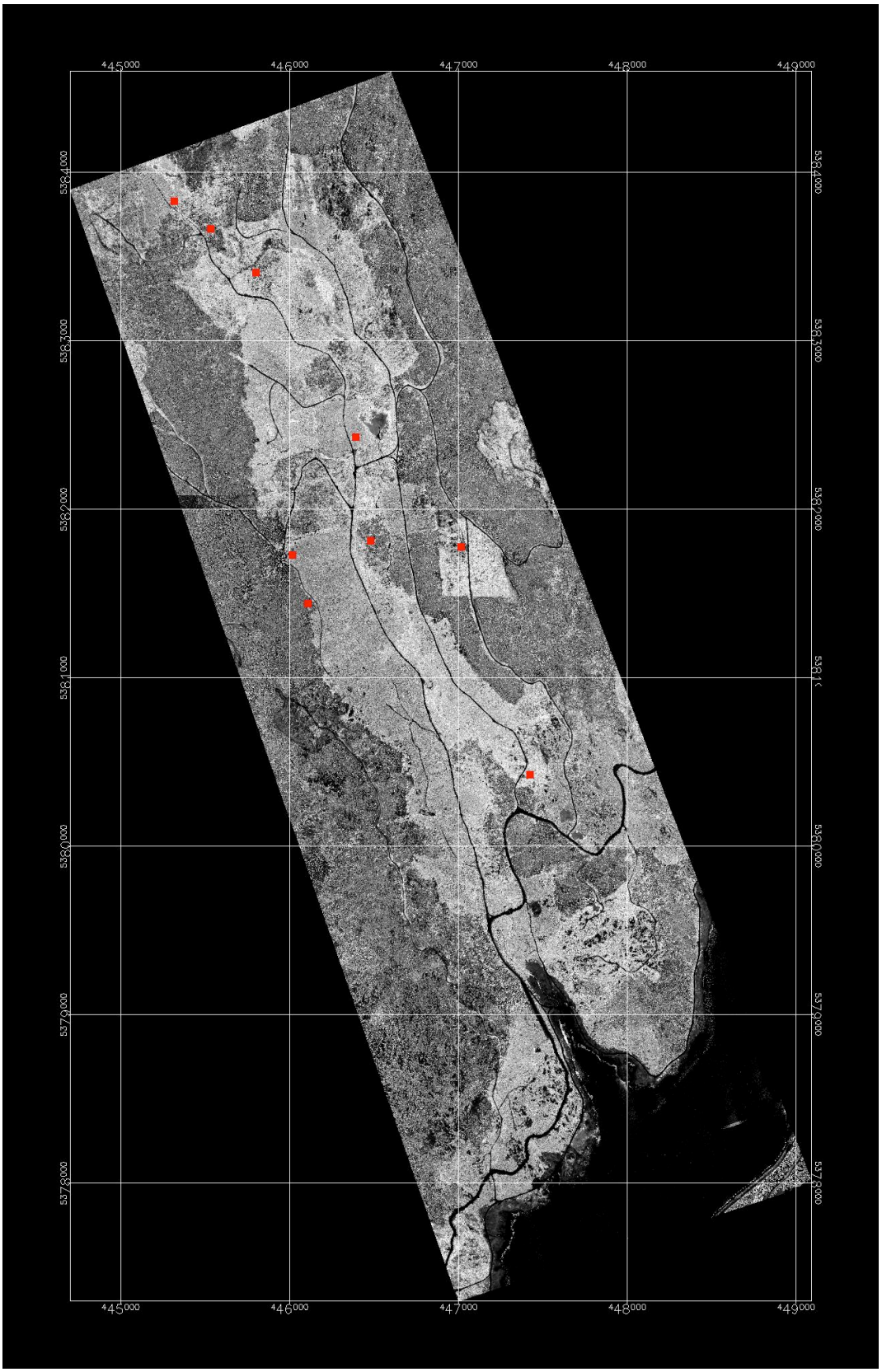


Figure 5-3: Intensity image of the study area.

Plot control points (centre and corner posts) were surveyed using a GeoXT GPS and differentially corrected using Pathfinder Office software (Trimble Navigation Ltd., Sunnyvale, California.) Within each plot, the coordinates of all trees higher than 2m were surveyed using an Impulse laser range finder and MapStar angle encoder (LaserTech Inc., Centennial, Colorado). A 2m threshold was chosen because lidar hits below this height could be confused with the ground surface during classification (see Section 3.4.3 on p.28). Tree height was measured, using the range and inclinometer capabilities of the Impulse instrument, for a sub-sample of trees with the intention of using height-diameter equations to predict heights for the remainder (see Appendix C). Attribute data including diameter at breast height (1.3m), species, crown radius, and mortality were also recorded for each tree. Table 5-3 summarizes the attributes collected in the field.

Table 5-3: Attributes collected for each tree

ATTRIBUTE	DESCRIPTION	TECHNIQUE
x,*y*,*z*	surveyed coordinates of the tree stem	GPS, laser rangefinder and angle encoder
height	tree height in metres (sub-sample only)	laser rangefinder and inclinometer
dbh	tree diameter at breast height (1.3m)	callipers (for plots 1-3 circumference at breast height was measured using tape and converted to diameter, assuming a circular bole)
species	tree species	tree book (Parish & Thomson 1994)
crown class	vertical position within the stand	observation
crown radius	upper crown dimensions measured in 4 cardinal directions (N,E,S,W) from stem	observation and measuring tape
mortality	living or dead	observation

Crown classes were defined as follows (Nigh & Love 2004):

> (1) Dominant — A tree whose crown extends above the general level of the main canopy of even-aged stands or, in uneven-aged stands, above the crowns of the tree's immediate neighbours and receiving full light from above and partial light from the sides.
>
> (2) Co-dominant — A tree whose crown helps to form the general level of the main canopy in even-aged stands or, in uneven-aged stands, the main canopy of the tree's immediate neighbours and receiving full light from above and comparatively little from the sides.
>
> (3) Intermediate — A tree whose crown extends into the lower portion of the main canopy of even-aged stands or, in uneven-aged stands, into the lower portion of the canopy formed by the tree's immediate neighbours, but shorter in height than the co-dominants and receiving little direct light from above and none from the sides.
>
> (4) Suppressed — A tree whose crown is completely overtopped by the crowns of one or more neighbouring trees.

5.3.3 Remotely sensed data

Lidar data were acquired for the study area on July 24, 2004. For details about the lidar instrument, technical specifications for this project, and data processing routines, refer to Section 3.2 on p.24. Data were acquired using 11 flightlines, flown at altitudes ranging from 855-1600m, with a maximum scan angle of ±24°.

5.3.3.1 Selection of lidar data for analysis

As there are a large quantity (>3000) of lidar data points within the bounds of each field plot, the goal is to use the best, most reliable points for analysis. This is particularly

important considering the normalization issues known to exist in intensity data. The following paragraphs outline how data were selected for use in the plot-level analyses.

Firstly, only singular (1/1) returns will be used. It was demonstrated in Section 4.4.1 that these are the most consistent values because the returning energy is not partitioned by multiple reflections.

Secondly, only points collected from the most nadir (lowest scan angle) flightline for each plot will be used. There are issues arising when using multiple view angles, especially in older stands. A nadir view may detect gaps, but a high-angle view may encounter the sides of crowns. While this is potentially an advantage in the capture of forest structure, the superposition (one object on top of another) problem is a complication when using two-dimensional spatial statistics techniques. Another issue with high-angle data is that the points tend to "pile up" on the sides of tree crowns after a discontinuity (e.g., a cut block, or a road). This results in an over-sampling in one specific area (i.e., crowns at the edge of the discontinuity) and under-sampling in other locations. These problems need to be avoided for appropriate application of local spatial statistics. By using the most nadir flightline for each plot, maximum effective scan angle will be limited to eight degrees off nadir.

Finally, raw intensity values will be used instead of normalized ones. Since the plot-level validation of the normalization procedure was inconclusive (effective for some plots but not others), it is prudent to use original, unaltered data for analysis. Since only one flightline for each plot will be used, points within small plots are comparable with the following qualifications:

(1) Due to the impact of range on footprint area (see Section 4.4.2), intensity is expected to decline through the height of the canopy. In old growth stands, where the canopy height nears 70m, this could result in up to 12 units difference in intensity between the canopy top and the ground due to footprint size.

(2) Raw intensity data are not appropriate for comparisons *between* different plots. This is due, primarily, to substantial differences in lidar ranges for the plots (see Table 4-4 on p.53). This does not, however, preclude comparison of plots using measures of spatial pattern (autocorrelation) derived from raw intensity values assuming stationarity at the plot level.

An IDL script was used to subset the lidar data based on these criteria, and a 10m buffer around the plot boundaries was added to mitigate edge effects. Having discussed the data that will be used, the following sections outline the analytical methods that will be employed to address each objective.

5.3.4 Canopy vs. ground intensity analysis

Before exploring links between intensity and forest attributes, it is prudent to first determine that there are in fact intensity differences between canopy and ground classes in each plot. This was accomplished in two steps: (1) each canopy height model was segmented into canopy and ground classes using a 2m height threshold introduced by (Brokaw 1982) to define canopy gaps; (2) a two-sample t-test was used to check for statistically significant mean differences between the two intensity distributions for each of the nine plots. If significant differences are found between the two classes, this would suggest that intensity data do indeed contain information about the forest canopy.

5.3.5 Global spatial autocorrelation

Moran's *I* was calculated using the "RookCase" visual basic add-in for MS Excel (Sawada 1999). This script computed the *I* statistic, and its standardized version, *zI*, for 1m cumulative lag increments (*d*) from 1m, out to 20m: ($0<d<=1$, $0<d<=2$, $0<d<=3$ … $0<d<=20$). Stationarity, one of the main assumptions of the Moran's *I*, is assumed on the basis that small (0.04-0.16 ha) individual plots were selected in relatively homogenous parts of each sampled stand. The scale of spatial autocorrelation will be determined by the lag distance at the first maximum of standardized Moran's *I*.

Trends in the scale of spatial autocorrelation of intensity values are expected to relate to the mean crown diameters, and to mean nearest neighbour distances between tree stems. Using crown radius measurements collected in the field, arithmetic and gemometric means of crown diameter were computed for each plot (including only dominant and co-dominant trees, and excluding dead trees). Mean nearest neighbour distance was computed using stem-map coordinates of live dominant and co-dominant trees in each plot. Finally, mean crown radius and mean nearest neighbour distance were compared to the spatial scale of the intensity pattern using regression analysis.

5.3.6 Local spatial autocorrelation

G_i* for intensity was computed (IDL code in Appendix D) using the plot-specific distance at which global spatial autocorrelation is maximized for Moran's *I*. Two approaches were used to explore relationships in the G_i^* data: (1) a raster sampling approach to compare intensity and G_i* with lidar-measured canopy heights, and (2) a vector (point) sampling approach to compare intensity and G_i* with field-measured tree attributes.

5.3.6.1 Raster sampling approach

To explore statistical relationships canopy height, intensity, and G_i* data were rasterized using a grid size equal to the plot-specific scale of global autocorrelation. Rather than using an interpolation strategy to create gridded surfaces from the lidar points, a custom IDL function was used to populate each grid cell with height, intensity and G_i* values belonging to the *highest* lidar point within the cell (Appendix E). The highest point was chosen to capture the purest (least mixed) canopy and understory reflections. Each pixel from the three resulting surfaces (height, intensity, and G_i*) represents one case in the plot-level analysis. Two scatterplots (intensity vs. height and G_i* vs. height) were created for each plot. Correlation analysis was used to assess the strength and direction of linear relationship.

5.3.6.2 Vector (point) sampling approach

In order to integrate field data, a conditional approach was used to sample lidar points. Field attributes, including stem-map tree coordinates, mean crown radius, crown class and height were each used to decide which lidar point to sample for each tree. An IDL procedure was written, taking into account both horizontal and vertical domains of the lidar point cloud within the plot (see Appendix F). The algorithm searched for all lidar points within the crown radius of each tree and selected the highest point (for dominant and co-dominant classes), or, the point nearest in height to the tree (for intermediate and suppressed classes). A minimum crown radius of 1m was used throughout, in an attempt to sample lidar points for snags (dead trees) with no crown. Once a lidar point was sampled it was flagged so that sample independence was maintained (i.e., no point was sampled twice). The lidar data record for each selected point was then fused with its

respective field data record. The new records were coded by plot and tree identification numbers, and written to an ASCII text file.

The first step in the analysis of these tree-level data was to create scatterplots of intensity vs. lidar height, and G_i* vs. lidar height. Colours and symbols were used to represent the qualitative data (e.g., species, crown class, and mortality). While the scatterplots are effective for qualitative observation, a more in-depth analysis is needed to demonstrate quantitative relationships.

Three-way analysis of variance (ANOVA) was used to test for significant differences in G_i* based on three categorical tree attributes (species, crown class and mortality). ANOVA assumes that variance of the dependant variable (G_i*) is equal for all categories. This assumption will be verified using Levene's test for the equality of variances.

5.4 Results

5.4.1 Canopy vs. ground intensity analysis

Two-sample *t*-tests confirm that there are significant mean differences between intensity values originating from canopy and ground classes for eight out of nine plots (Table 5-4). In the non-significant plot, only 2% of the points were recorded within 2m of the ground. Upon inspection, the majority of these reflections were clustered in one corner of the plot (Figure 5-4). It is likely that this limited sample does not contain a representative range of intensity values needed to detect a discernible difference from the canopy. How can lidar estimate canopy heights properly with such a poor sample of ground points? Remember that only single (1/1) returns from one flightline are considered here. More ground points are captured when using multiple returns.

Table 5-4: Mean differences in *raw* intensity for canopy and ground classes

Plot ID	Canopy (>2m) raw intensity, nadir flightline				Ground (<=2m) raw intensity, nadir flightline				Difference of means*	
	Mean	StDev	CoV	*n*	Mean	StDev	CoV	*n*	*t*	*p*
1	306	51	0.17	1808	277	41	0.15	66	5.58	**0.000**
2	274	53	0.19	1073	270	45	0.17	24	0.37	0.712
3	293	52	0.18	232	268	40	0.15	83	4.51	**0.000**
4	273	48	0.18	1769	223	29	0.13	33	9.85	**0.000**
5	279	65	0.23	1116	262	38	0.15	423	6.25	**0.000**
6	313	55	0.18	455	288	49	0.17	102	4.60	**0.000**
7	287	50	0.17	1485	227	24	0.11	95	21.27	**0.000**
8	259	49	0.19	1630	243	29	0.12	603	9.09	**0.000**
9	335	54	0.16	416	272	45	0.17	595	19.35	**0.000**

*unequal variances assumed

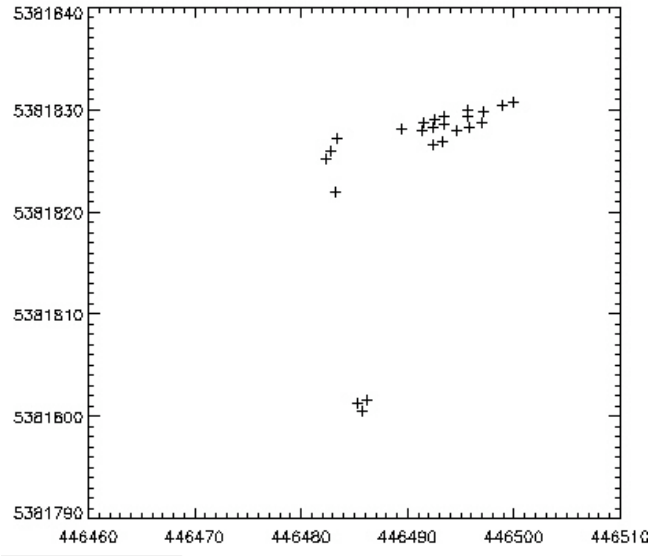


Figure 5-4: Ground (<=2m) point locations in plot 2

These results demonstrate that intensity data can differentiate between canopy and ground classes. But it was acknowledged earlier (in Section 5.3.3.1) that lower intensity near the ground could be due to the difference in footprint size (larger at the ground than at the canopy top). Although normalized intensity values are not used in any analysis that follows, to ensure that the canopy versus ground differences are not merely due to

footprint size the *t*-tests were re-run using range-standardized intensity values. The results from the second set of tests confirm that significant differences exist for the same eight plots (Table 5-5). As such, the observed differences between canopy and ground classes appear to be due to intrinsic properties of those classes, not footprint size.

Table 5-5: Mean differences in *normalized* intensity for canopy and non-canopy classes

Plot ID	Canopy (>2m) range-standardized intensity, nadir flightline				Ground (<=2m) range-standardized intensity, nadir flightline				Difference of means*	
	Mean	StDev	CoV	*n*	Mean	StDev	CoV	*n*	*t*	*p*
1	285	51	0.18	1808	260	41	0.16	66	4.922	**0.000**
2	270	54	0.20	1073	273	45	0.17	24	-0.316	0.755
3	304	52	0.17	232	280	40	0.14	83	4.314	**0.000**
4	298	48	0.16	1769	251	29	0.12	33	9.061	**0.000**
5	252	65	0.26	1116	240	38	0.16	423	4.499	**0.000**
6	318	55	0.17	455	294	49	0.17	102	4.447	**0.000**
7	289	50	0.17	1485	232	24	0.10	95	20.513	**0.000**
8	259	50	0.19	1630	249	29	0.12	603	6.012	**0.000**
9	310	54	0.17	416	249	44	0.18	595	19.216	**0.000**

* unequal variances assumed

5.4.2 Global spatial autocorrelation

Standardized Moran's *I* values indicate that spatial autocorrelation in the intensity data is highly significant ($zI > 2$) for all plots (Table 5-6). Standard estimates based on the expectation of normality were very similar to those computed using randomizations. This suggests that the *I* statistic is normally distributed for the lidar intensity data. Plot-level trends of Moran's *I* over the range of lag distances are presented in Figure 5-5.

Table 5-6: Distances of maximum spatial autocorrelation

Plot ID	Age	Lag distance (m) at first maximum *I*	Standardized I_N	Standardized I_R
1	I	4	6.45	6.45
2	O	7	11.42	11.42
3	R	5	7.28	7.29
4	I	6	10.37	10.38
5	O	8	43.33	43.35
6	R	5	13.92	13.93
7	I	6	17.83	17.84
8	O	4	34.46	34.47
9	R	3	57.37	57.37

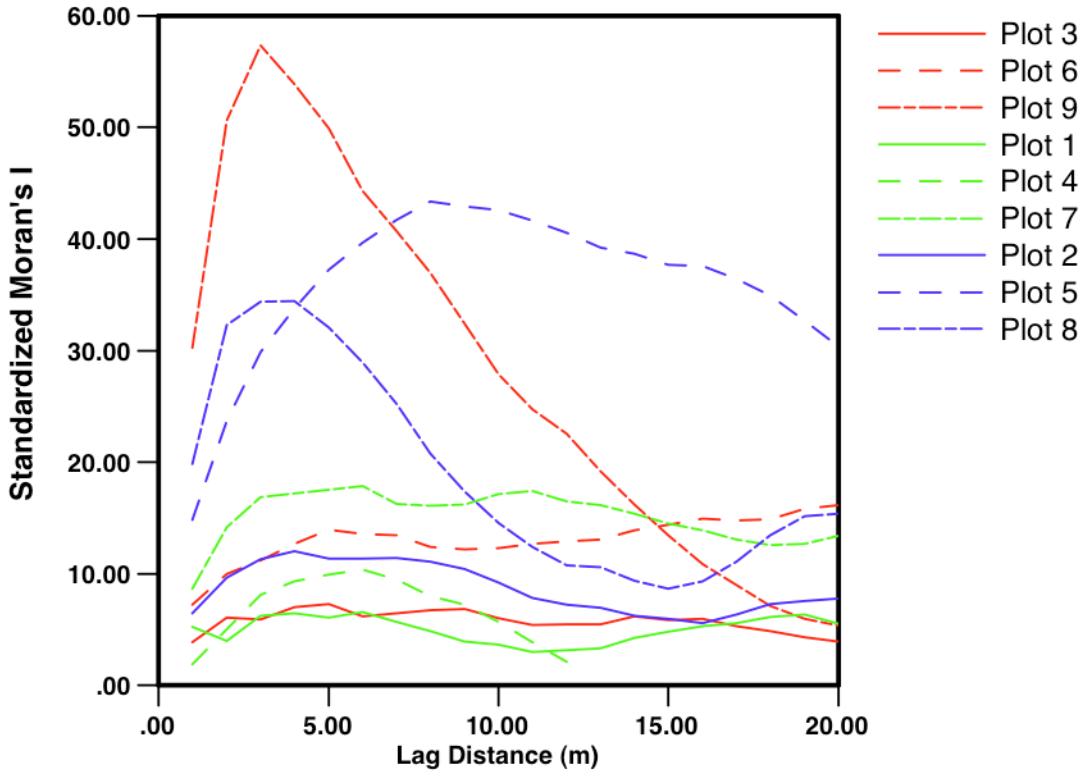


Figure 5-5: Behaviour of standardized Moran's *I* with increasing lag distance: regenerating <=20 years (red); immature <45 years (green); old growth >200 years (blue)

The spatial scale of intensity (i.e., standardized Moran's *I*), and its variability, is expected to increase as stands mature towards the later stages of forest development. Results shown in Figure 5-5 do not illustrate this trend as clearly as the plot-specific distances to the first maximum of standardized Moran's *I* (Table 5-7). Regenerating and immature plots have spatial clustering of intensity values at the smallest and most consistent scale, while old growth plots exhibit clustering at larger and more variable spatial scales. Although the trend is as expected, it is important to remember that only three replicates for each age-class are considered. As such, the descriptive statistics presented below should be interpreted with caution.

Table 5-7: Scale of spatial autocorrelation by plot age class

	Regenerating (<=20 years)	Immature (20-45 years)	Old growth (>200 years)
Plots 3, 1, 2	5	4	7
Plots 6, 4, 5	5	6	8
Plots 9, 7, 8	3	6	4
Mean	4.33	5.33	6.33
StDev	1.15	1.15	2.08
CoV	0.267	0.217	0.329

The spatial scale of the intensity pattern is expected to relate to mean crown diameter and mean nearest neighbour distance (Table 5-8). Both these quantities increase with stand age. Figure 5-6 shows that the scale of spatial autocorrelation of intensity relates well to the mean of crown diameter and mean nearest neighbour distance, but there are a few outliers.

Table 5-8: Field data relationships with the scale of spatial autocorrelation of intensity

Plot ID	Age	Scale of spatial autocorrelation of intensity (m)	Mean crown diameter[1] (m)		Mean nearest neighbour distance of tree stems[1] (m)
			Arithmetic	Geometric	
1	I	4	4.46	4.34	1.90
2	O	7	6.19	6.01	2.73
3	R	5	5.08	4.77	2.29
4	I	6	4.99	4.84	2.71
5	O	8	7.18	7.06	4.03
6	R	5	3.30	3.17	1.45
7	I	6	5.40	5.29	2.69
8	O	4	7.50	7.27	3.80
9	R	3	3.65	3.61	2.38
r^2			.2490	.2543	.1858

[1] Dominant and co-dominant trees only
[2] Independent variable is the scale of spatial autocorrelation

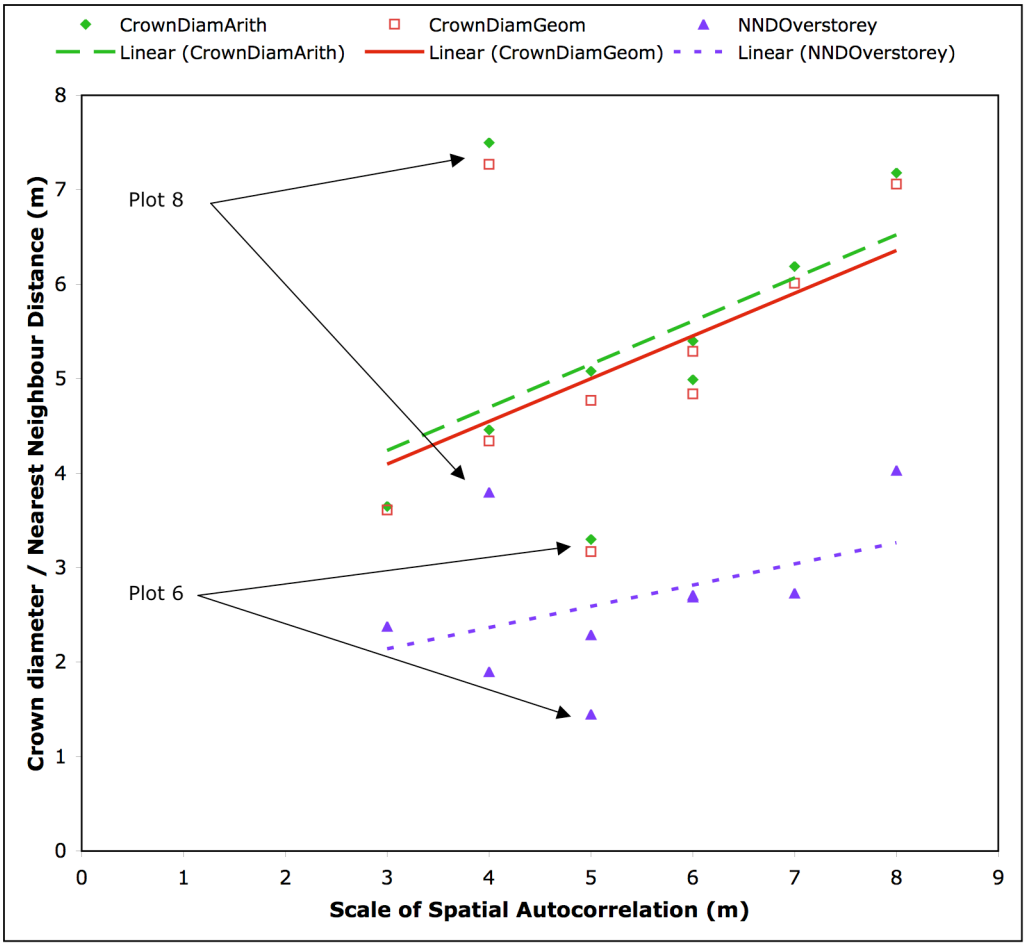
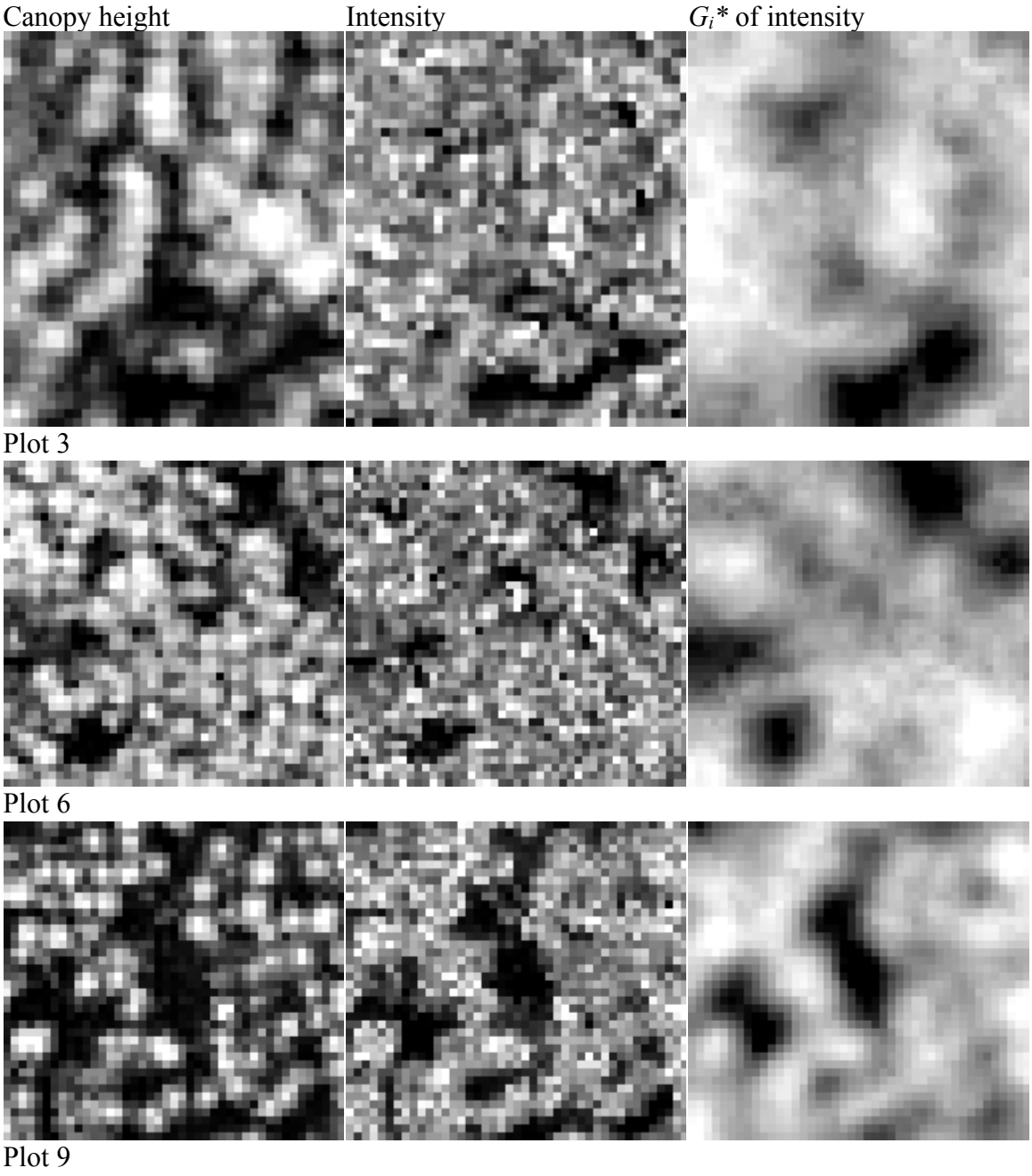


Figure 5-6: Scatterplot of spatial scale of autocorrelation of intensity related to field-measured plot attributes (crown diameter and stem position)

Two anomalies are apparent in Figure 5-6, plots 6 and 8. Plot 6 is an extremely dense regenerating plot with a stem density exceeding 6000 stems per hectare. Given that the trees in this plot, similar in size and composition, are so tightly packed they may resemble larger objects (multiple crowns agglomerated together) in the intensity pattern. Plot 8 (old growth) has a large number of small gaps with patches of Western hemlock emerging below, which contribute to a finer scale intensity pattern. Mean gap diameter measurements are not available, however, the nearest neighbour distance between dominant and co-dominant trees is fairly large, suggesting that there is more space between the overstorey trees (and perhaps more gaps). Greater variability in the scale of the intensity pattern for old growth is also a finding in itself. The spatial and compositional structure of old growth forests is highly complex and heterogeneous (Franklin & Van Pelt 2004).

5.4.3 Local spatial autocorrelation

While global measures of autocorrelation indicate differences between plots, standardized G_i^* images (Figure 5-7) reveal a distinct spatial structure of the intensity values within each plot that is much more difficult to see using raw intensity data (shown along side for comparison). In addition, the well defined patterns in the G_i^* images are quite distinct from canopy height patterns, suggesting that intensity does indeed contain important added information about forest structure. For growth plots (Figure 5-7c) there appears to be a link between G_i^* and gaps (higher intensity in gaps). Field data overlaid on canopy height (Figure 5-8) and G_i^* imagery (Figure 5-9) illustrates that trends at the individual tree level are not clear and that the relationship is complex.



(a) Regenerating plots

Figure 5-7: Canopy height, intensity and G_i* images for regenerating (a), immature (b), and old growth (c) plots. Images were rasterized to 1m pixels and enhanced using a linear 2% stretch.

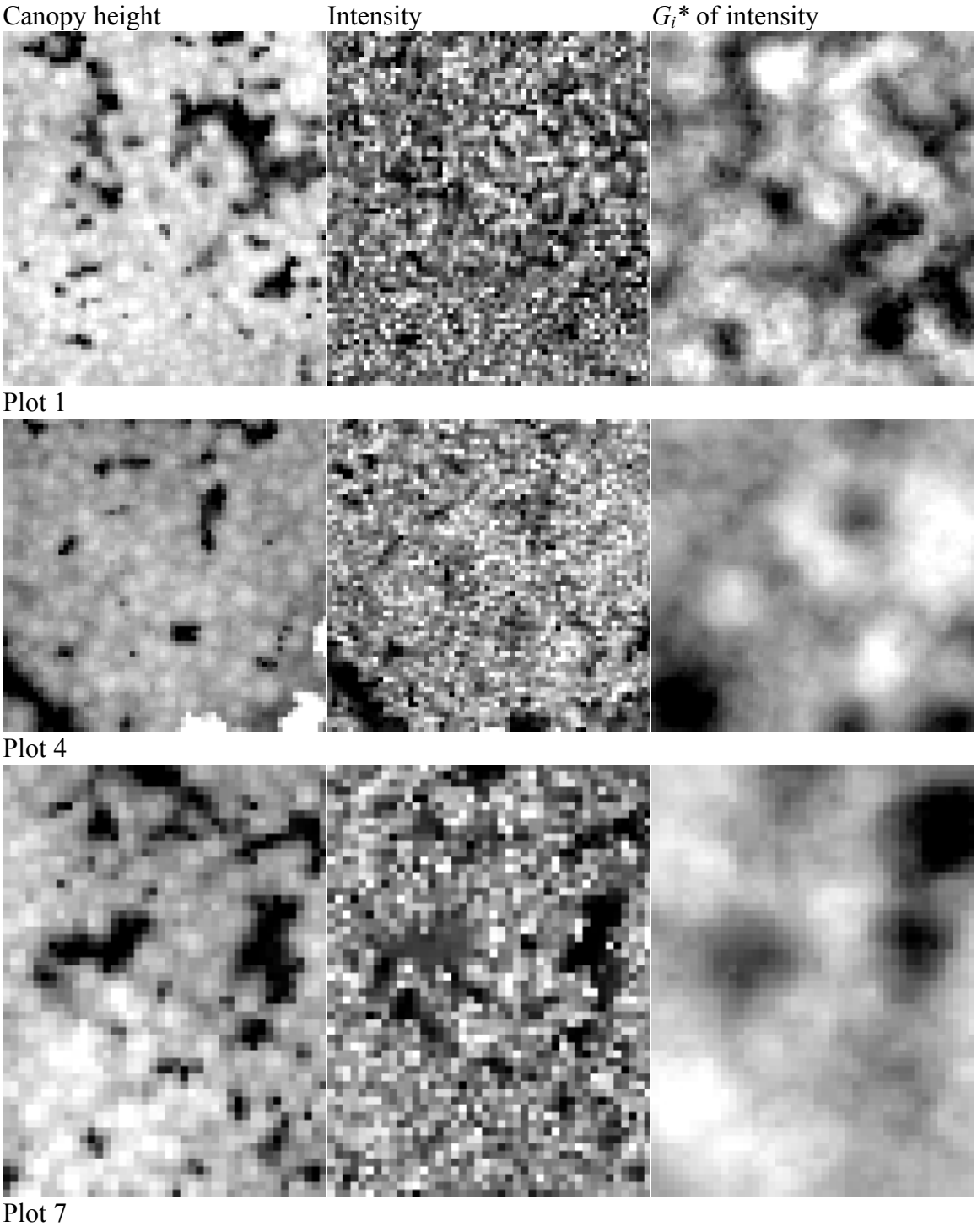


Figure 5-7(b) Immature plots

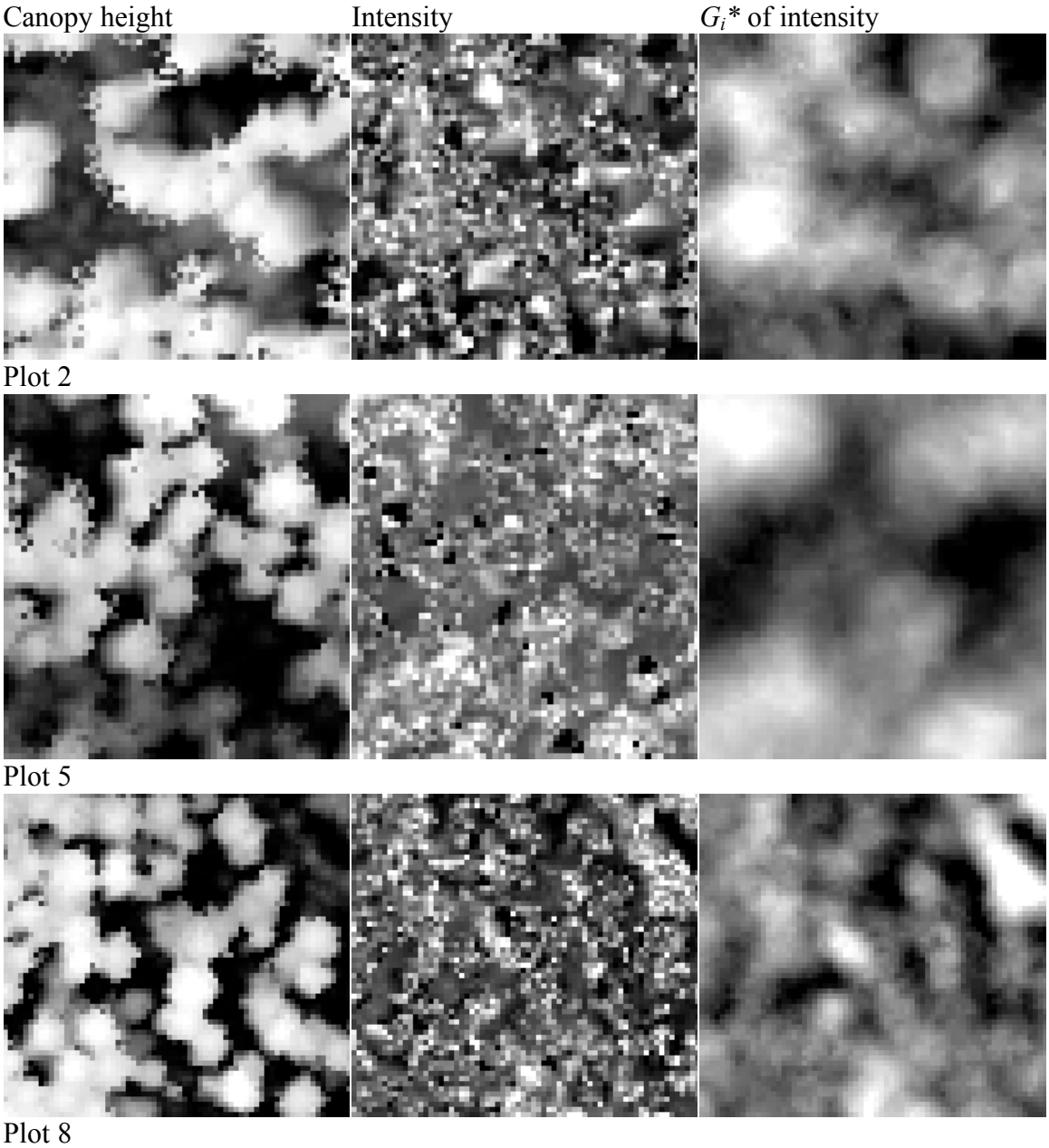


Figure 5-7(c) Old growth plots

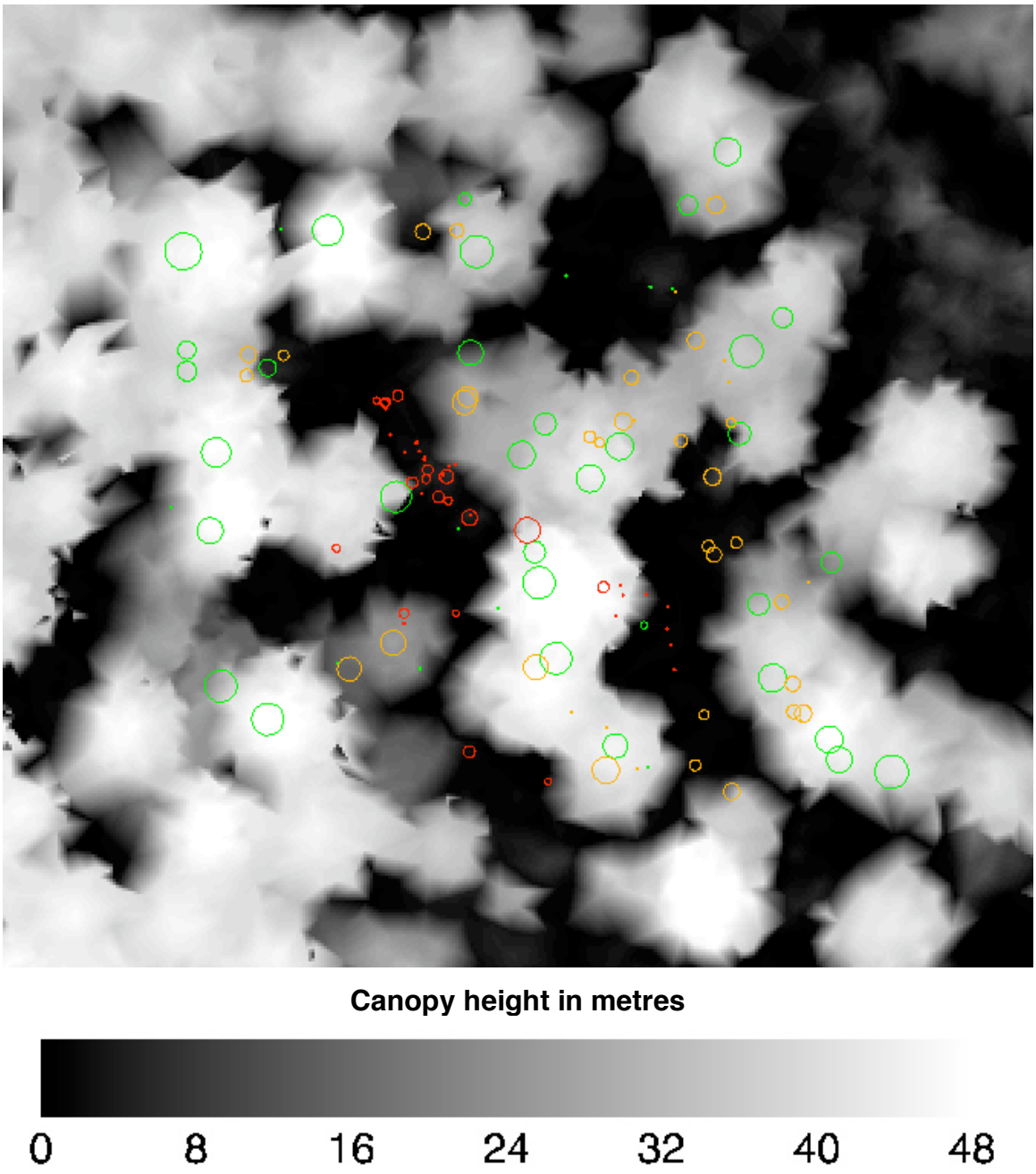


Figure 5-8: Lidar canopy height image with field data overlaid (plot 8). Circles are proportional to mean crown radius. Colours represent tree species: Douglas-fir (green), Western hemlock (red), Western redcedar (gold).

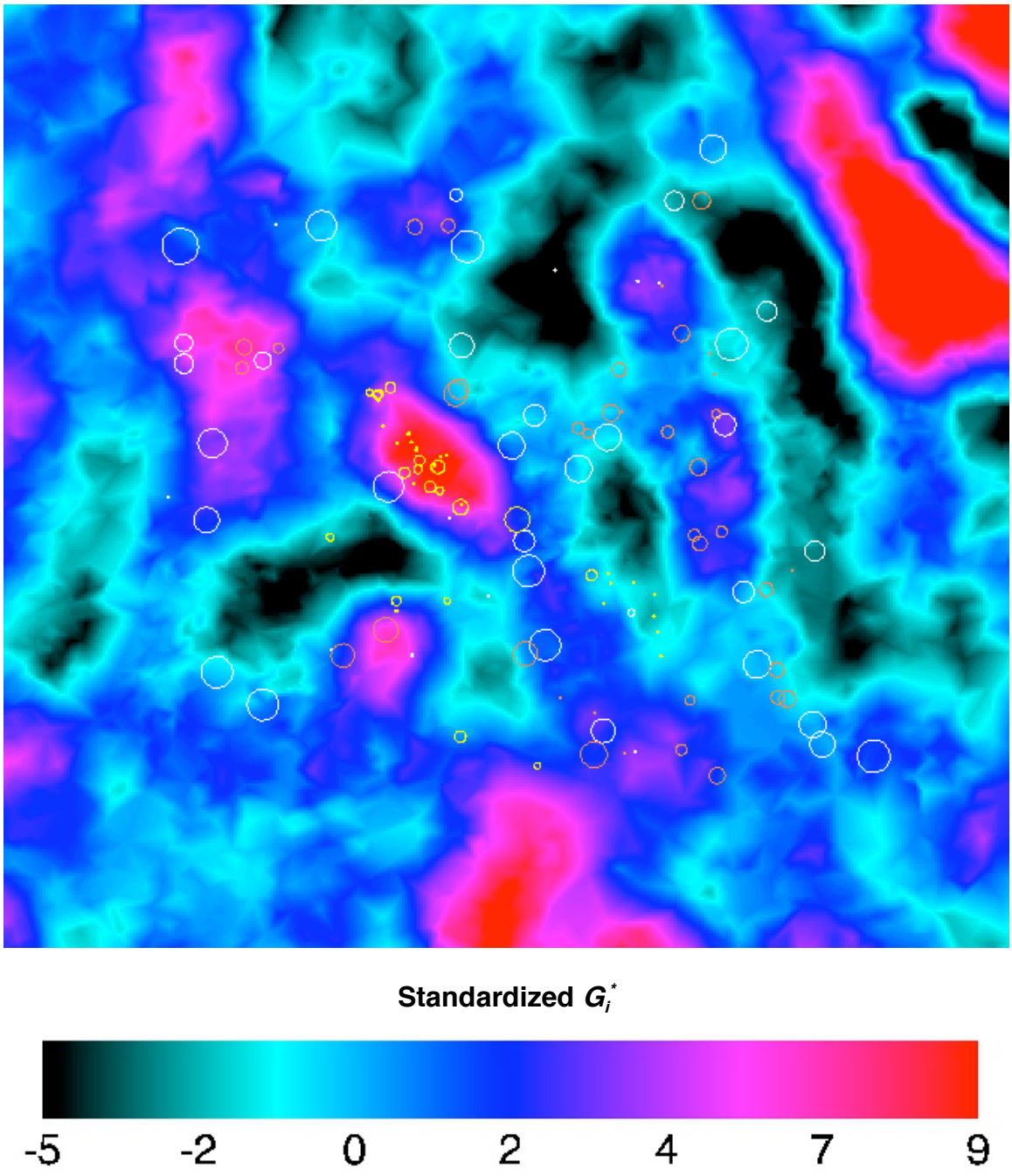
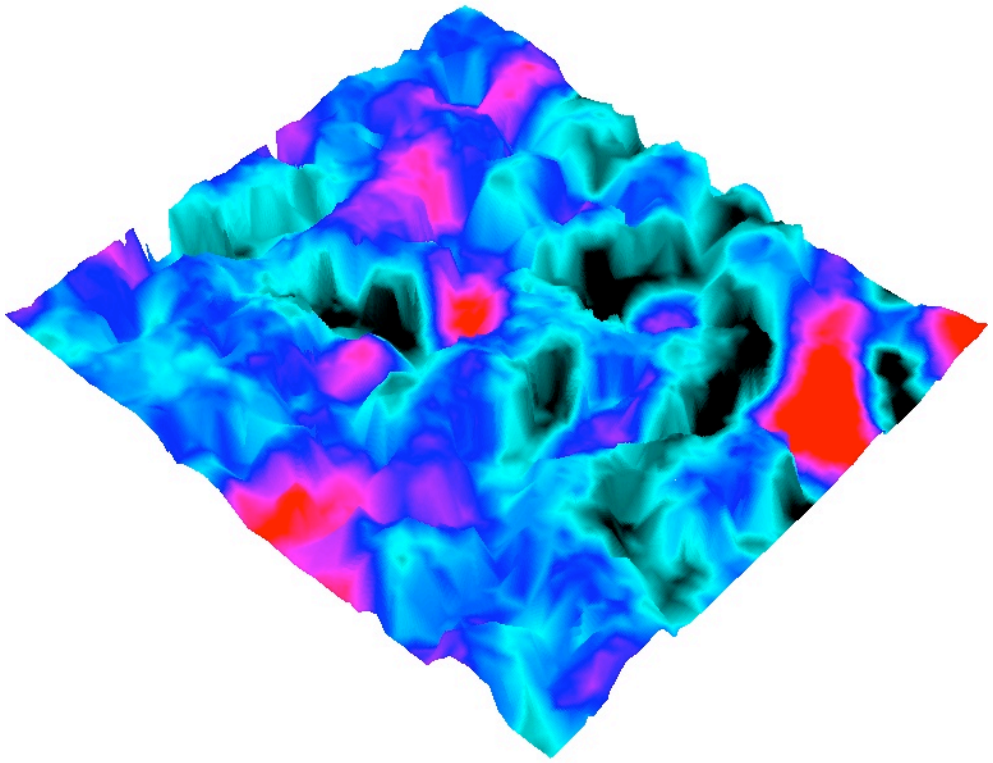
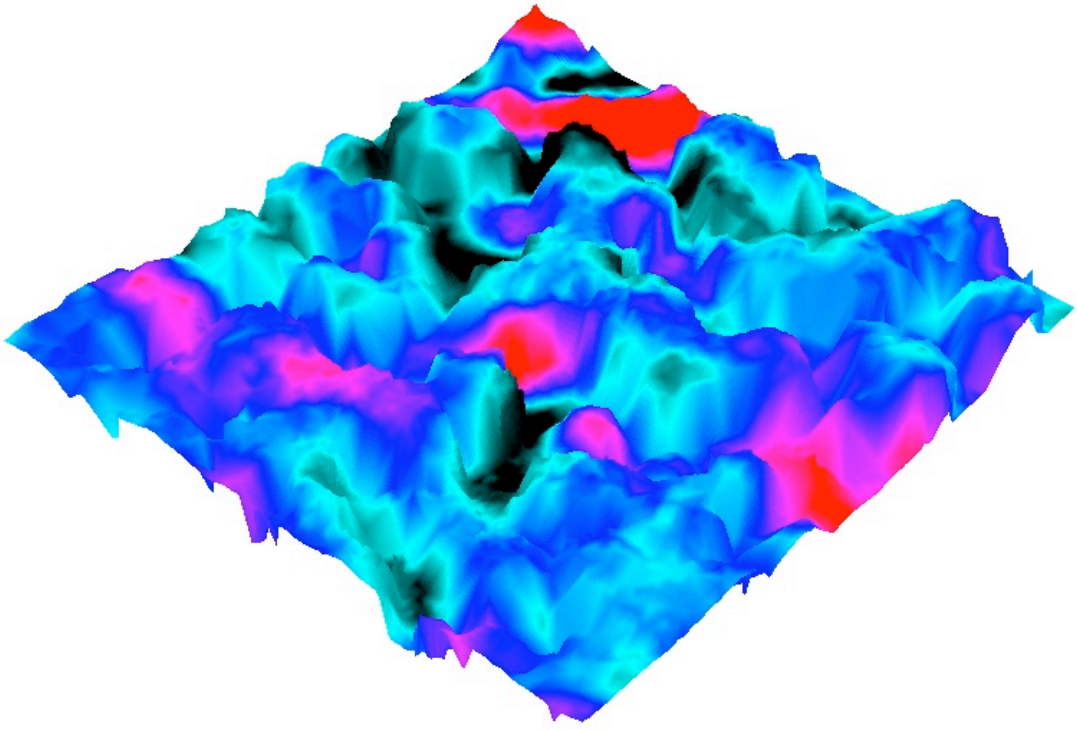


Figure 5-9: G_i^* image with field data overlaid (plot 8). Circles are proportional to mean crown radius. Colours represent tree species: Douglas-fir (white), Western hemlock (yellow), Western redcedar (orange). This image shows the complex nature of the relationship between G_i* and tree-level attributes.

In an effort to better understand the nature of the G_i^* image it was draped over a three-dimensional surface representing canopy height. This visualization allows for interpretation of G_i^* with respect to vertical structures (e.g., crowns and gaps). The highest G_i^* values (red) identify an old logging road northeast of the plot boundary. High G_i^* is expected here because the road is relatively homogenous in height, texture, and composition. Most tree crowns have G_i^* values ranging between negative and positive two. There are a few crowns with G_i^* values exceeding positive two (pink). Only one crown, located near the road, had a G_i^* value below negative two (black). Upon further inspection of the flightline trajectory, terrain slope, and discontinuity in the canopy (i.e., the road), it is probably that the laser points "piled up" on the side of this crown, and reduced its intensity. While crowns were represented by a full range of G_i^* values, gaps are characterized by either low negative (five clusters), or high positive G_i^* (two clusters). Of the five negative clusters, three are within the field plot boundaries, and all of these gaps have no understorey trees, with the exception of one very small Douglas fir. Only one of the positive clusters is within the field plot boundaries, and it has a vigorous Western hemlock understorey. It also appears that positive clusters occur in larger, more open gaps, while negative clusters occur in smaller gaps that are more obscured by overstorey crowns. Smaller, potentially obscured gaps would have low negative G_i^* values because the laser energy is much more likely to be intercepted and absorbed (e.g., by overstorey branches) along the path to gap, and therefore intensity would be reduced. High positive G_i^* values is expected in the open gap because there is a dense young understorey able to reflect laser energy more strongly and coherently than overstorey crowns, leading to an intensity "hot spot" seen in the centre of the plot.



a)



b)

Figure 5-10: G_i^* image draped over canopy height surface (plot 8). These 3D images are shown from a southeast (a) and a southwest (b) viewpoint. The G_i^* colour gradient is the same as in Figure 5-9.

5.4.3.1 Raster sampling approach

In the raster approach, canopy height, intensity and G_i* were rasterized using the plot-specific scale of autocorrelation identified by Moran's *I*. Each pixel in the resulting rasters represents one case in the following plot-level analyses.

Scatterplots of intensity and G_i* vs. canopy height reveal contrasting trends for young and old stands. Scatterplots of the regeneration (Figure 5-11) and immature (Figure 5-12) plots indicate that returns from the uppermost part of the canopy had moderate to high intensity and G_i*, while returns near the ground had a low intensity and G_i*. In old growth plots (Figure 5-13), the situation was reversed. The upper canopy had low intensity and G_i*, while the lower parts had moderate to high intensity and G_i*. To determine the direction, and assess the strength of the association between canopy height and intensity and G_i*, correlation analysis was applied to all plots showing a linear trend. A summary of the results is presented in Table 5-9. As observed in the scatterplots, regenerating and immature plots have weak positive trends, while old growth plots have weak negative trends.

Table 5-9: Direction and strength of association between canopy height, intensity, and G_i*

Plot ID	Age	Intensity		G_i* of intensity		N
		Pearson-*r*	*p*	Pearson-*r*	*p*	
1	I	.010	.916	-.031	.747	110
2	O	-.157	.485	**-.641**	**.001**	19
3	R	.369	.120	.314	.191	22
4	I	.128	.385	.109	.459	48
5	O	-.170	.513	-.159	.541	17
6	R	-.011	.961	.327	.137	22
7	I	.233	.215	**.568**	**.001**	30
8	O	.030	.750	.127	.183	112
9	R	**.622**	**.000**	**.524**	**.000**	53

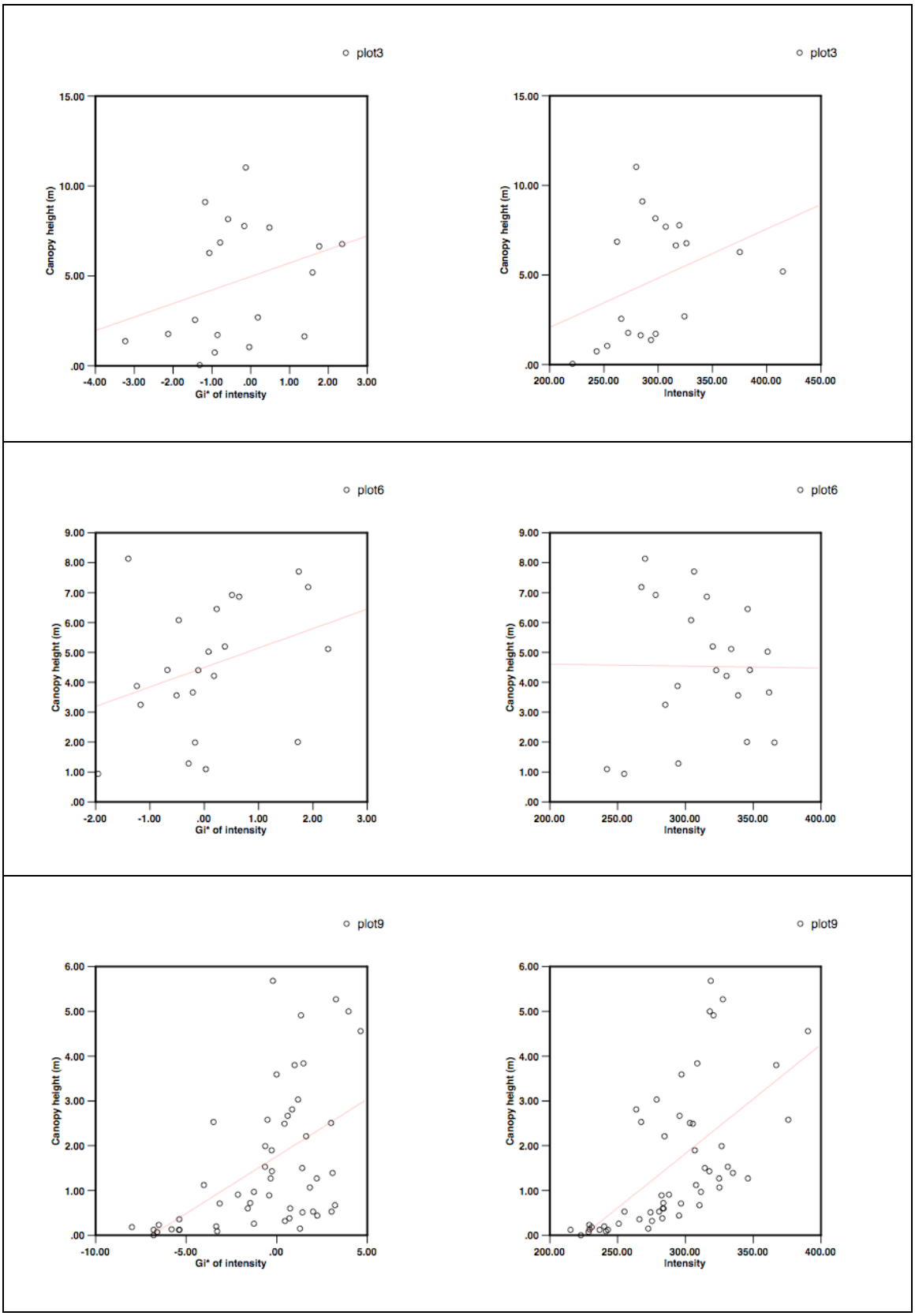


Figure 5-11: Intensity and G_i* vs. canopy height for *regenerating* plots (<=20 years)

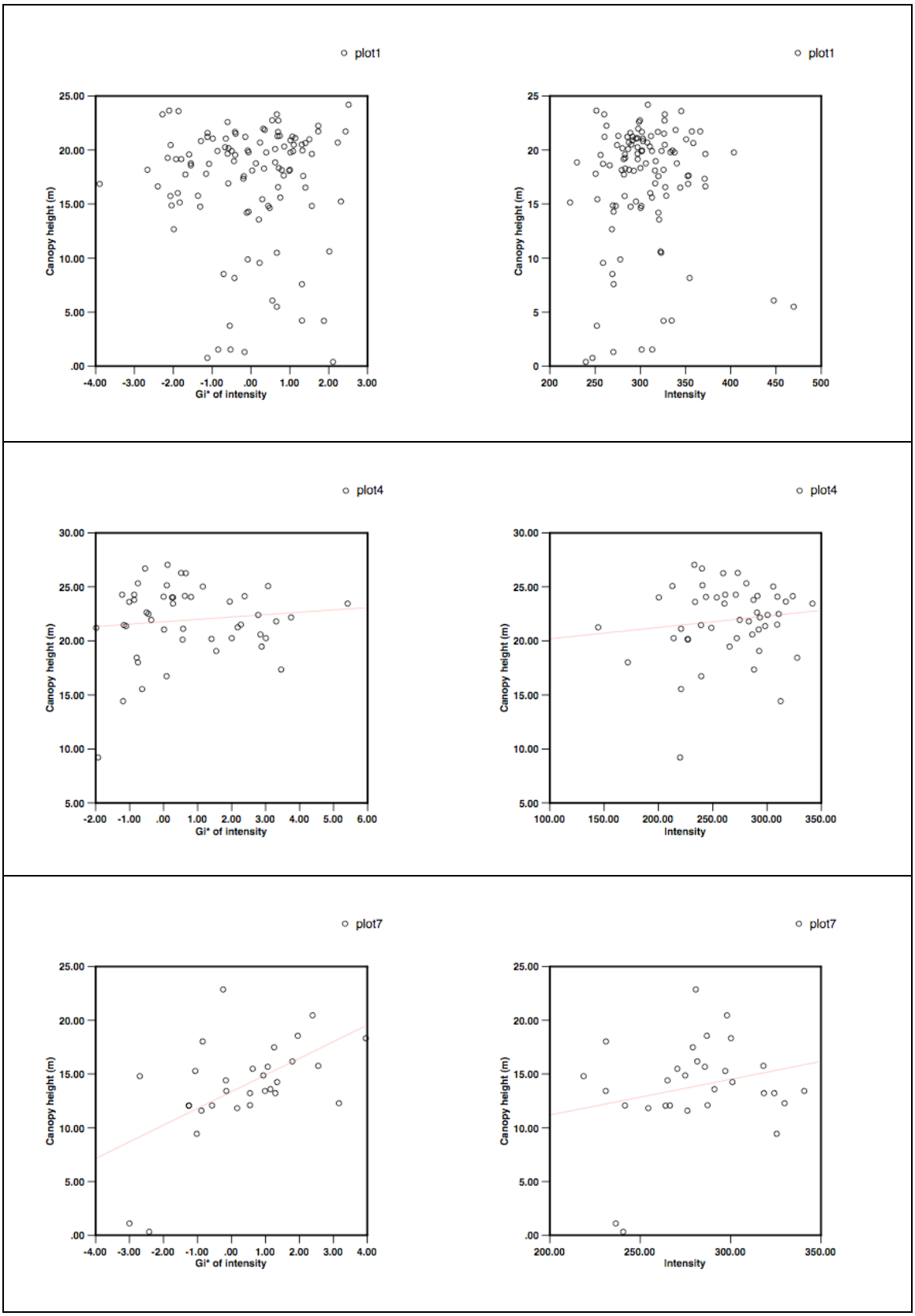


Figure 5-12: Intensity and G_i* vs. canopy height for *immature* plots (20-45 years)

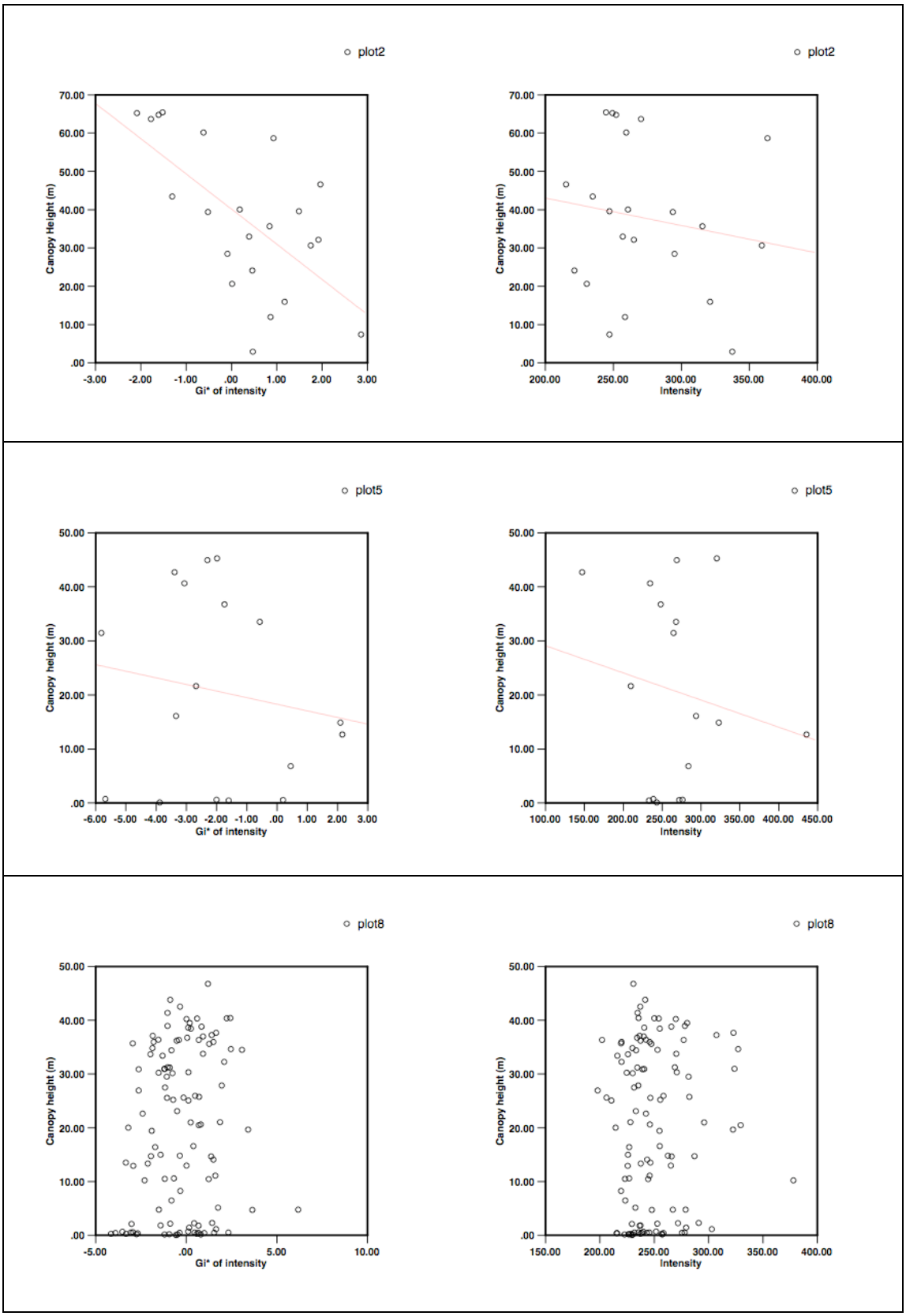


Figure 5-13: Intensity and G_i* vs. canopy height for *old growth* plots (>200 years)

5.4.3.2 Vector (point) sampling approach

By sampling one lidar point for each tree surveyed in the field, tree-level attribute information can be added to the intensity and G_i* vs. canopy height scatterplots (Figure 5-14). Each point on these plots represents a single tree. At first glance these scatterplots reveal that there is substantial class mixing. Many suppressed trees (which should have very low heights) are incorrectly sampled at lidar heights indicative of the overstorey canopy. This sampling problem is likely caused by poor laser penetration through the overstorey, especially in dense plots. Another contributing factor could be poor correspondence of GPS positions between the lidar and field data sets. These problems will be described further in the discussion section.

In an effort to improve sampling, a height threshold was established, such that all samples with a lidar-to-field height difference greater than the threshold value were excluded. To determine the threshold amount, the regression models used to predict height from diameter (dbh) measurements were revisited. The absolute values of the residuals in these models produce a fan-shape when plotted against height, indicating heteroscedascity in the relationship. Heteroscedascity exists in a relationship when the variances are unequal for different levels of the independent variable. In this case, as tree diameter increases, the variance of tree heights also increases, making it harder to accurately predict height. The slope of the outer edge of the residual fan was approximately 0.4 for the Douglas-fir, Western hemlock, and Western redcedar models. Lidar points were excluded from the analysis if the absolute value difference between lidar and field height (the height difference) was more than 40% of field-height (the height threshold). Application of the threshold resulted in 58.7% of the samples being

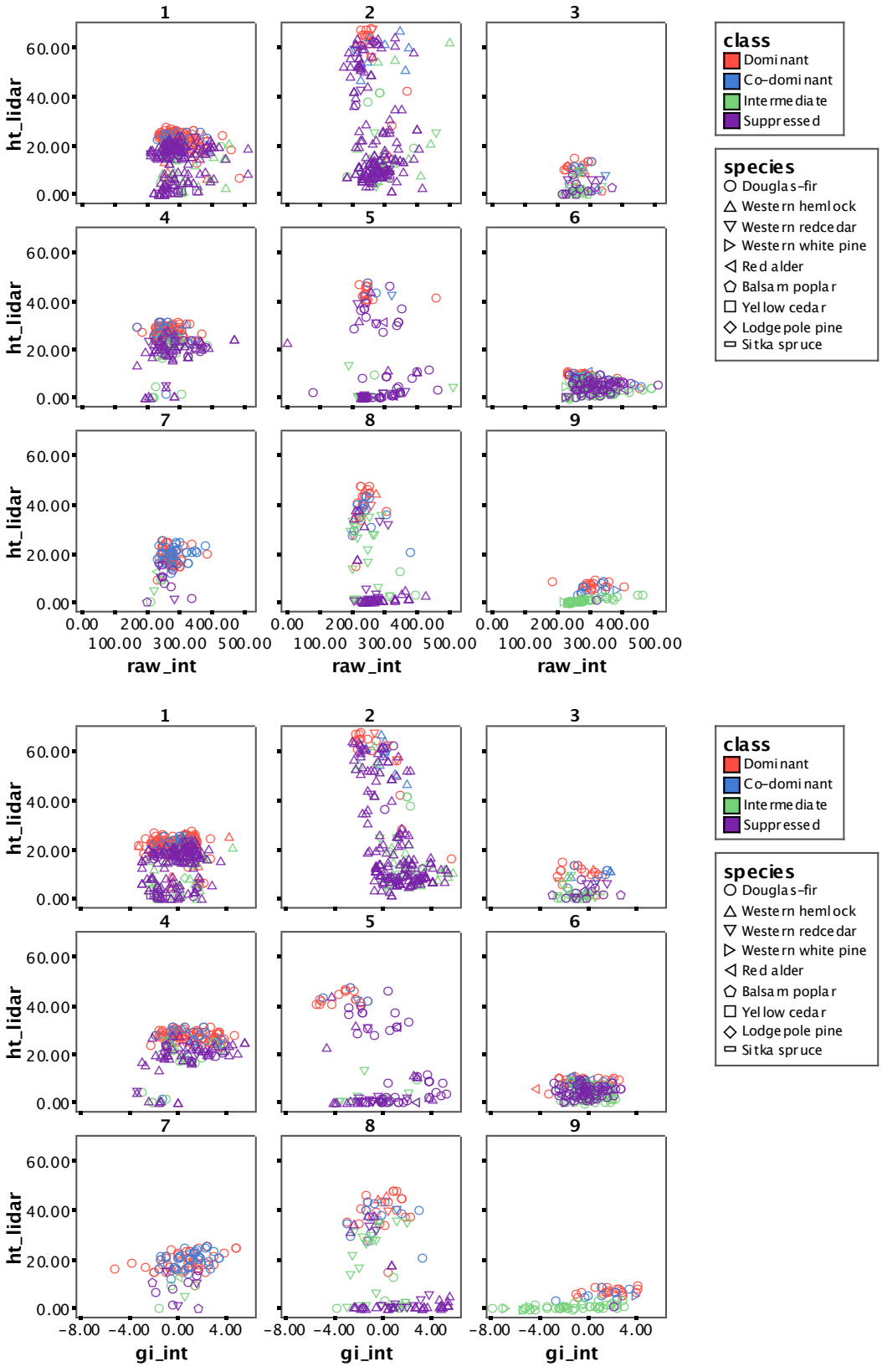


Figure 5-14: *Unfiltered* scatterplots of intensity and G_i* vs. height by crown class and species

excluded. While the exclusion of a large portion of data is not ideal, it is justified to uphold the validity of each sample. Field attributes cannot be applied to lidar points that clearly do not originate from the same tree.

New scatterplots were generated using only points meeting threshold criteria. The most obvious trends exist in old growth plots (Figure 5-15). Douglas-fir trees had the lowest G_i*. Western hemlock trees had the highest G_i*. Western redcedar trees had a range of G_i* values, but tend to be near to zero. For crown class, dominant and co-dominant trees had low G_i*, while intermediate and suppressed had moderate to high G_i*. Similar patterns can be seen in the intensity scatterplots but the species and crown classes are not as separable. The species and crown class information offer some possible explanation for the negative association between intensity and G_i* and height previously shown in old growth plots (Figure 5-13). In old growth plots, the understorey species and crown classes generally have higher intensity and G_i* than the overstorey species. A clustering of high intensity and G_i* values relates to above average near infrared (1089nm) reflectance in the understorey and below average in the overstorey. Scatterplots for the intermediate (Figure 5-16) and regenerating (Figure 5-17) plots show no clear separation by species or crown class in either intensity or G_i*.

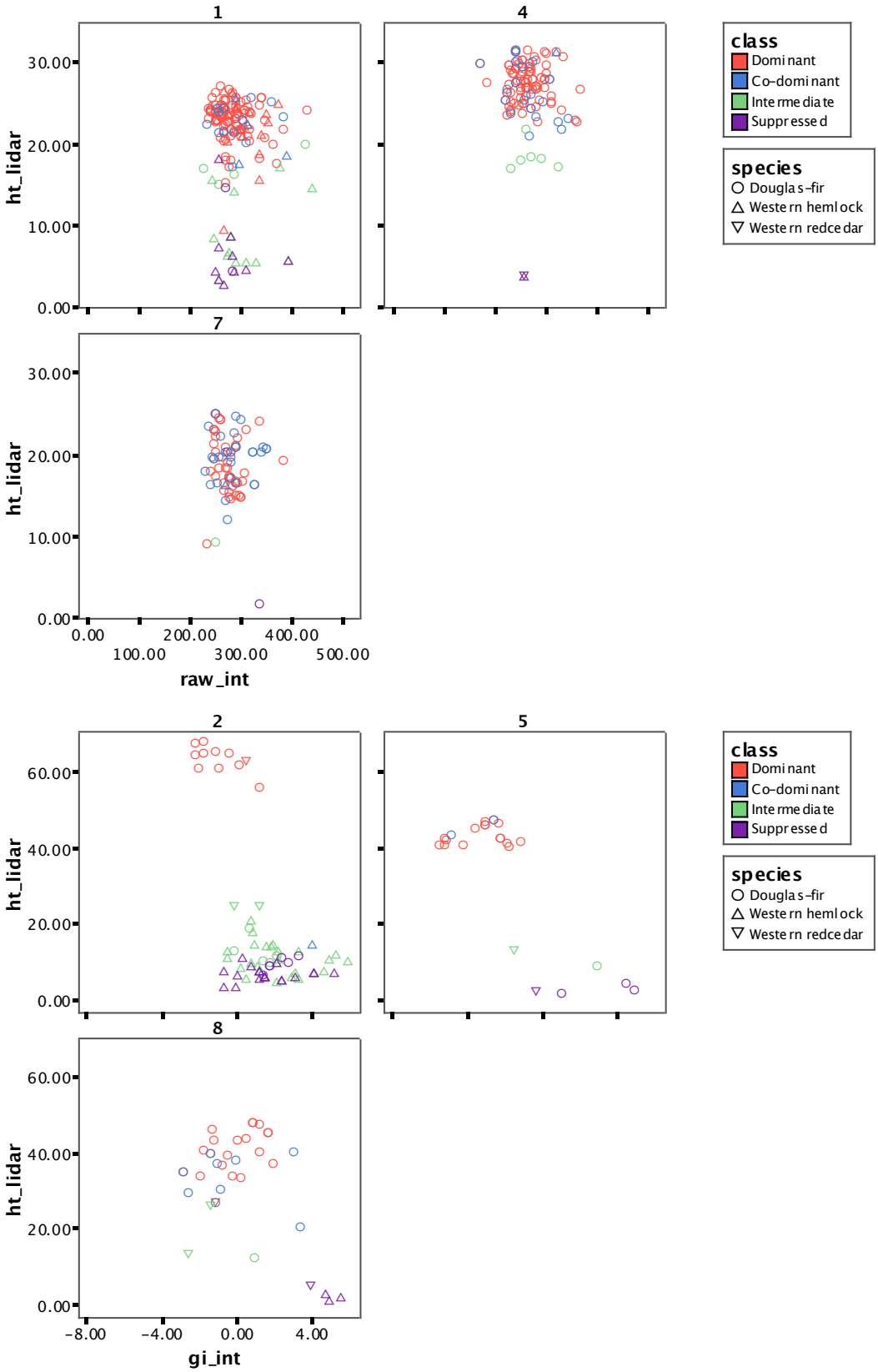


Figure 5-15: Filtered intensity (top) and G_i* (bottom) scatterplots for *old growth* (>200 yrs)

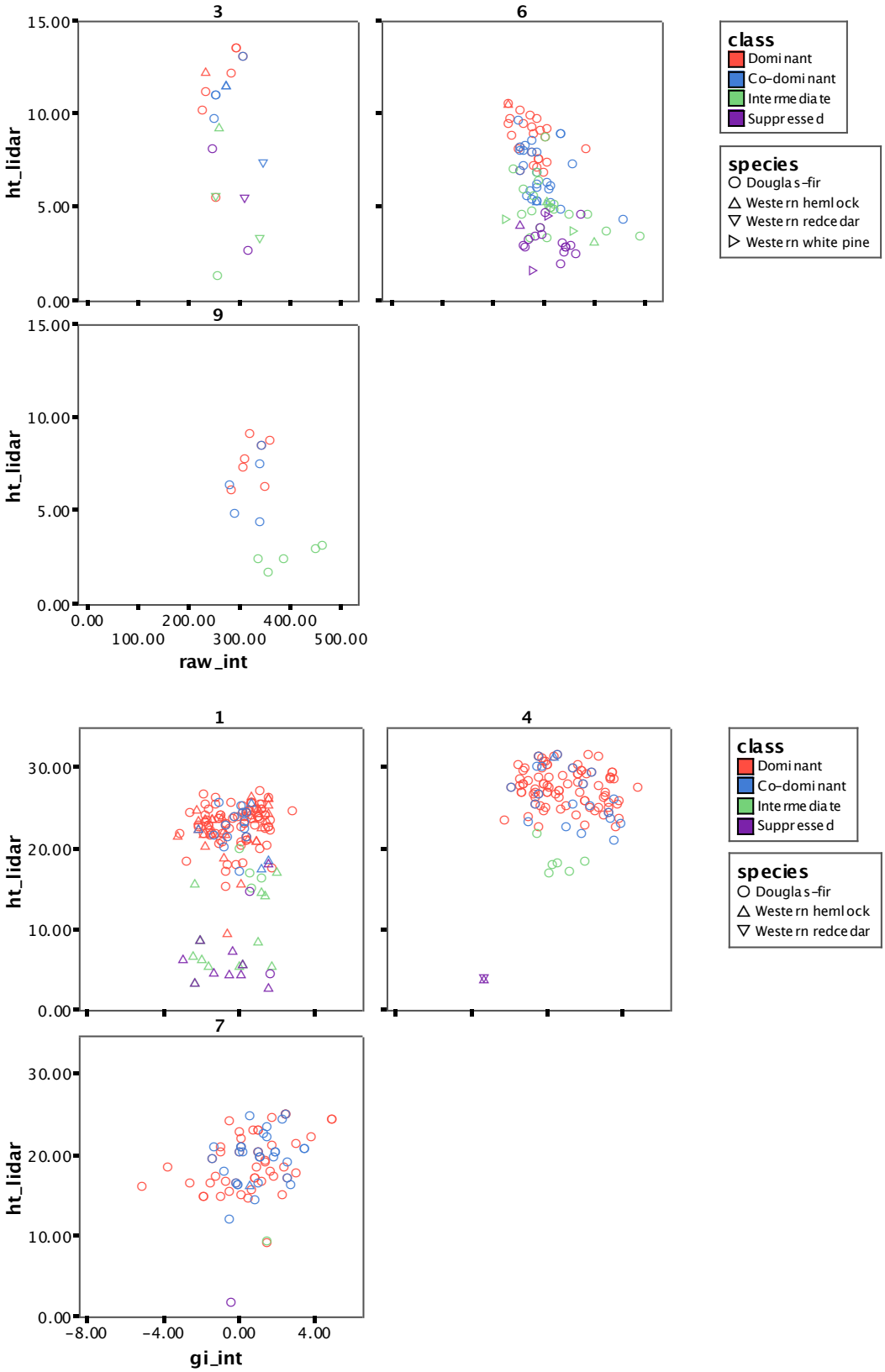


Figure 5-16: Filtered intensity (top) and G_i* (bottom) scatterplots for *immature* (20-45 yrs)

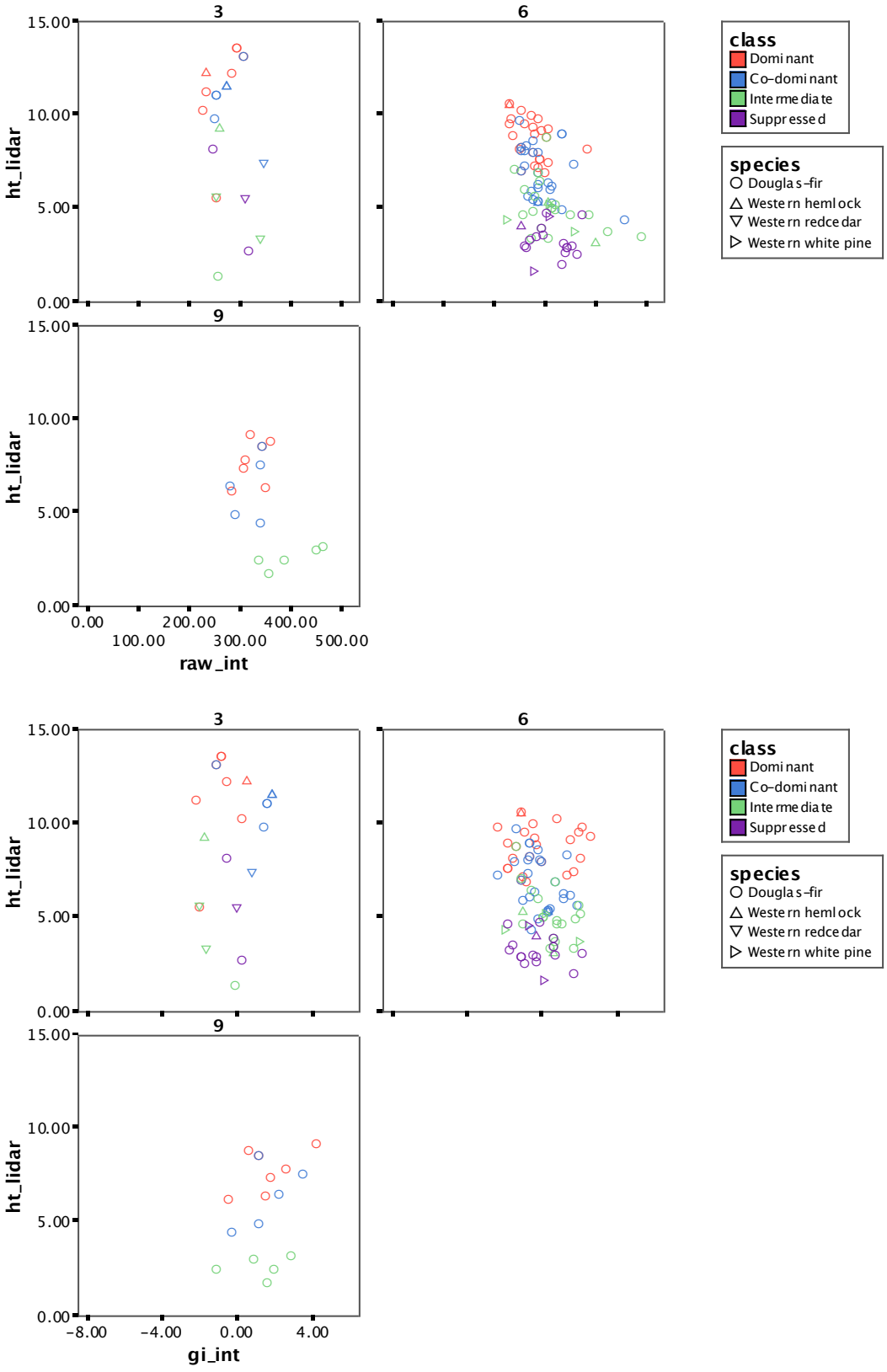


Figure 5-17: Filtered intensity (top) G_i* (bottom) scatterplots for *regenerating* (<=20 yrs)

The canopy height scatterplots illustrate the behaviour of intensity and G_i* through the vertical (height) structure of each plot. It has already been suggested that the weak linear trends in these scatterplots (positive for regenerating/immature; negative for old growth) could be due to compositional (i.e., species) and spatial differences (i.e., gaps) in canopy layers. Three-way ANOVA was used to test for significant between-group variances for three categorical tree attributes: species, crown class and mortality (i.e., living or dead). Species and mortality are compositional attributes. Crown class, the relative vertical position of the tree's crown, is a structural attribute that relates to light availability. Lidar can be considered a proxy for sunlight, and therefore crown class may have an effect on intensity.

To ensure adequate sample and cell size for the three-way ANOVA, plot level data are pooled into the three stand age-class groups (regenerating, immature, and old growth). G_i* will be used exclusively for this investigation because multi-plot analysis of raw intensity values is problematic due to flightline normalization issues. G_i* is considered comparable between plots because it transforms intensity into a standard measure, relative to the mean intensity of each plot. Although the mean intensity may be different, the spatial dependence of values relative to the mean is assumed to be similar. Remember that intensity is essentially a 1089 nm reflectance measurement. G_i* is a transformed intensity value, containing information about the degree of spatial dependence as well as the magnitude of the original intensity value relative to its mean.

Results indicate that species has a significant effect on G_i* for immature and old growth plots. This was also found in previous work using G_i* statistics with optical reflectance data (Wulder & Boots 2001). For old growth plots, there were also significant

interaction effects between species & crown class, and crown class & mortality (Table 5-10). The first interaction suggests that the effect of species on G_i* is modified by crown class. G_i* values for Douglas-fir increase towards the lower (more suppressed) crown classes, while G_i* values for Western hemlock tend to decrease as trees become more suppressed (Figure 5-18). The sample size for Western redcedar (n=9) is too low for a meaningful interpretation of its interaction with crown class. As for the second interaction, crown class & mortality, the sample size of dead trees (n=11) is also too low for meaningful interpretation. Finally, r^2 values indicate that ANOVA models explain 7.4, 13.1 and 54.5 per cent of the variation in G_i* for immature, regenerating and old growth plots respectively (Table 5-10).

Table 5-10: Differences in G_i* for species, crown class and mortality

	Regenerating (n=134)[1]			Immature (n=368)			Old Growth (n=132)		
	F	*df*	*p*	*F*	*df*	*p*	*F*	*df*	*p*
Species	0.341	3	.796	3.245	2	**.040**	5.903	2	**.004**
Class	1.748	3	.161	0.674	3	.569	1.737	3	.163
Mortality	0.004	1	.947	0.021	1	.886	1.904	1	.170
Species*class	1.616	6	.149	0.369	3	.775	4.165	4	**.003**
Species*mortality	n/a[2]	0	-	0.154	1	.695	3.691	1	.057
Class*mortality	.515	2	.599	0.086	1	.769	3.524	2	**.033**
Species*class*mortality	n/a[2]	0	-	0.029	1	.864	n/a[2]	0	-
r^2	.131			.074			.545		

[1] Levene's test reports significantly different group variances (p=0.024) for the *regenerating* sample. This column of ANOVA results should be interpreted with caution.

[2] One or more categories has fewer than two trees

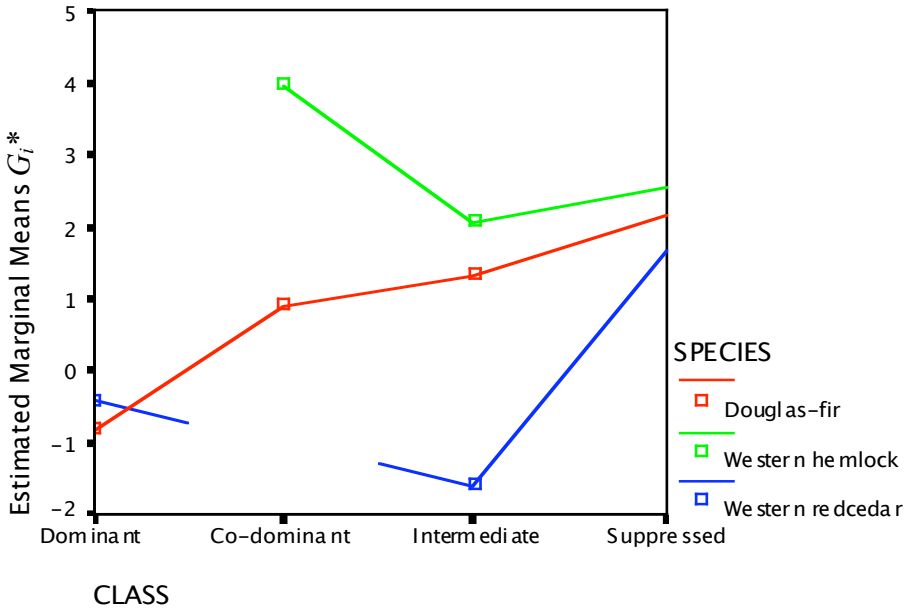


Figure 5-18: Interaction effect between species and crown class (old growth). There were no co-dominant western redcedar and therefore a mean was not plotted.

5.5 Discussion

Significant differences between intensity of points above and below a 2m height threshold, suggest that there are indeed differences between canopy and ground classes. It was further confirmed, by running the analysis using range-standardized intensity values that the relative differences between canopy and ground classes are not due to footprint area, which is larger at the ground. This finding is important because it confirms that intensity is at least partially driven by the composition and structure of the forest canopy.

In order to learn how the spatial structure of the forest canopy drives the intensity pattern, Moran's *I* was used to determine the scale at which spatial autocorrelation is strongest (assuming a stationary pattern for each plot). The results showed that the spatial scale of the intensity pattern increased with stand age. The scale of this pattern appears to relate directly to the mean crown diameter, and mean nearest neighbour distance—measures of crown and gap sizes, respectively. These two attributes increase in both

magnitude and variability as stands progress towards the later stages of forest development.

Transforming raw intensity using the G_i* statistic, reveals local trends that are distinct from the lidar canopy height data. This suggests that, in addition to responding to the canopy and gap structure, intensity may contain information about forest composition. Another finding is that the G_i* is an effective noise-reduction and enhancement technique for lidar intensity data. In addition to highlighting locally autocorrelated spatial clusters, G_i* includes information about the magnitude of the clustered (intensity) values relative to the mean.

When plot-level G_i* of intensity is compared with lidar measurements of canopy height, there are positive associations in regenerating and immature plots, and negative associations in old growth plots. Two possible explanations for the *positive* associations are: (1) intensity increases with height because the footprint area becomes smaller, and (2) intensity decreases as the laser pulse passes deeper into the forest canopy, and is attenuated by absorption or partial reflection. The *negative* associations, particularly bright intensity coming from the lowest parts of the canopy, were interesting. This effect is probably not caused directly by height, but rather different compositional layers that relate to height.

In order to determine if compositional factors are at play, tree-level field data were matched with lidar point data. This procedure was problematic. Numerous understorey trees were incorrectly sampled too high in the lidar point cloud. Two potential sources of error are: (1) the laser did not penetrate deep enough into the canopy to capture all understorey trees, and (2) GPS error (in both lidar and field data sets) contributed to

uncertainty of the tree stem position, which resulted in poor matching, especially for the small trees that populate the understorey. As a consequence, the tree-level dataset was filtered using a height difference threshold. Scatterplots of intensity and G_i* vs. canopy height coded by species and crown class illustrate that the negative association found in old growth plots may be due to compositional and spatial (i.e., gap) differences between two canopy layers (understorey species have higher intensity and G_i* than overstorey species).

To test for significant differences in G_i* of intensity due to tree-level attributes (species, crown class, and mortality) the plot-level data were aggregated into three stand age-class groups. ANOVA reports significant differences for species in the immature and old growth plots. Significant differences are not detected in regenerating plots, probably because these are the most compositionally homogenous. As stands mature, patches and layers of shade tolerant species (e.g., Western hemlock and Western redcedar) develop, and G_i* of intensity appears to be detecting these.

In old growth plots, there was also a significant interaction effect between species and crown class. Although crown class alone does not have a significant effect, it does modify the effect of species on G_i*. This is an interesting finding in itself. For example, dominant classes of Douglas-fir had much lower G_i* than suppressed trees of the same species. Crown class is a proxy measure for tree age. Suppressed trees are the youngest, and dominant trees are the oldest. The reflectivity of Douglas-fir is known to decrease with age (Ogunjemiyo *et al.* 2005). Branches and needles are arranged to capture of a small amount light, and scatter the remainder for use by other areas of the tree. The sharing of light is critical as the tree becomes older and larger, resulting in lower

intensity. Western hemlock and Western redcedar, in contrast, are shade tolerant species designed to capture as much light as possible. As these trees grow, their ability to capture light improves, resulting in higher intensity. Intensity, through G_i*, appears to be detecting these differences in old growth stands where compositional (species) stratification and gap structure are the most developed.

Mortality does not appear to have a significant effect on G_i* of intensity, however, in old growth plots it does have a significant interaction with crown class. This makes sense, because the bark of older snags (dead trees) is different in both texture and composition (e.g., presence of epiphytes) than that of younger snags. Since neither crown class nor mortality have a significant effect on G_i*, intensity does not appear to capture differences between living and dead trees. The sample size of dead trees is low, as they are small in size and weak reflectors in the near infrared wavelength used by the lidar system.

The effects of forest succession are observed by intensity, particularly in old growth plots, as they are more heterogeneous both in composition (demonstrated by G_i* differences) and spatial arrangement (demonstrated by the scale of global autocorrelation identified by Moran's *I*). Reflectivity differences between shade intolerant, and shade tolerant species are also observed in old growth plots through a significant interaction between species and crown class in their effect on G_i* of intensity.

5.6 Conclusion

The goal of this thesis was to better understand intensity, and to find out what it can tell us about forest structure. Despite considerable noise at the individual point level, intensity has a well-defined spatial pattern demonstrated by statistically significant spatial autocorrelation. The distance or scale at which spatial associations are the strongest is

linked to the canopy and gap structure within each plot, as measured by the mean crown radius, and mean nearest neighbour distance. Local variations in spatial autocorrelation are associated with canopy height, and, in some cases, composition. Positive associations, found in regenerating and immature plots, may be due to footprint differences, and/or the absorption of laser energy through upper sections of canopy. Negative associations, found in old growth plots, are linked to canopy gaps that expose different compositional (i.e., species) layers. The potential for intensity to detect species differences is important because lidar height information tells us little about the compositional aspects of forest structure. Further research is needed to include other compositional factors (e.g., moisture content, epiphytes, dead and decadent material, etc). The results from this study demonstrate that intensity data has the potential to be integrated into lidar remote sensing of forest structure, which currently focuses exclusively on height.

Chapter 6 Conclusions

6.1 Thesis summary

This thesis began by stressing that forest stands, especially old growth ones, are heterogeneous in both spatial arrangement and composition. While traditional measures of forest structure (e.g. stand height, stem density, etc.) imply homogeneity, there is increasing interest in capturing within-stand diversity and structural complexity. The ability of lidar measurements to characterize the spatial distribution of canopy material has generated significant interest. However, spatial data alone contain very little information about the compositional aspects of the stand. The purpose of this work was to explore the utility of lidar intensity to aid in the characterization of forest structure.

In a literature review, it was found that, although some potential has been identified, intensity has not been exploited as an information source in the characterization of coniferous forests. There are several possible reasons for this: (1) intensity, a measure of laser backscatter, is not calibrated to known energy units, (2) intensity is not normalized for survey geometry, and (3) the laser's interaction with forest elements is highly complex.

The three dimensional nature of lidar data also present unique processing and visualization challenges. While lidar data can be processed with three spatial dimensions, it is best visualized with two. This thesis has outlined procedures for rasterizing intensity data, and improving visualization through contrast enhancement and adaptive spatial filtering.

A systematic procedure for normalizing intensity data based on laser range was presented; the results were promising, but not definitive. Normalization improved correspondence of intensity means, but significant differences remained in five out of six validation plots. This result had an impact on the comparative analyses that could be performed.

Due to inconclusive normalization results, and high variability in the data generally, spatial pattern analysis was chosen to explore potential links between intensity and forest structure. Global estimates of spatial autocorrelation suggested that intensity is driven by the canopy and gap structure of each stand. The scale of global autocorrelation related to both the mean crown diameter (canopy patch size), and to mean nearest neighbour distance between stems (gap size), and it increased, in magnitude and variability, with stand age. Local measures of spatial autocorrelation appeared to vary with canopy height. Weak positive associations between intensity and height were found in regenerating and immature stands. In contrast, negative associations between intensity and height were found in old growth stands. Positive associations were expected to occur due to: (1) decrease in footprint area (stronger intensity) with height, and (2) attenuation of laser energy through the vertical dimension of the canopy. Negative associations were interesting, because they could not be explained by laser path geometry. The association between intensity and height in old growth stands was linked to a gap structure that exposed two well-defined canopy layers: (1) an overstorey composed of Douglas-fir, and (2) an understorey composed primarily of shade-tolerant Western hemlock and Western redcedar. The understorey layer, composed of vigorous shrubs, herbaceous plants and young trees, had higher intensity than the overstorey. ANOVA results attributed the

difference to tree species. The ANOVA also suggested that the effect of species on intensity is qualified by crown class (a proxy measure for age). This result supports what is already known about the decline in reflectivity for Douglas-fir with age.

6.2 Suggestions for future research

The results of this thesis suggest that lidar intensity has promising potential to improve characterization of forest structure. In order for its potential to be fully realized, intensity needs to be calibrated through laboratory testing, *in situ* measurements, and/or complex modelling. Relationships between intensity and survey geometry factors such as beam divergence, range, scan angle, and angle of incidence need to be better understood. Another investigation is needed to study the reflective behaviour of various materials to laser energy.

The following suggestions are made for future studies of lidar intensity in the forest context:

1) Understand the recording characteristics of the sensor used. Intensity of single returns may not be comparable to that of multiple returns.
2) Be aware that laser range will influence intensity.
3) Consider the wavelength involved when interpreting intensity values.
4) Analyse intensity at the plot-level, rather than at the point-level.
5) Know that intensity is spatial in nature, and depends, to a large degree, on the size, location, and compositional aspects of gaps.

It is unlikely that intensity will ever approach the quality of other forms of remote sensing that specialize in the collection of spectral reflection data (i.e., hyperspectral sensors). Intensity data can, however, complement lidar height data in the

characterization of forest structure. The integration of lidar intensity and height data, as presented in this thesis, shows that there exists an opportunity to examine spatial and compositional structure of forests based on the spatial pattern of lidar intensity.

Bibliography

Arp H. & Griesbach J. C. (1982) Mapping in Tropical Forests: A New Approach Using the Laser APR. *Photogrammetric Engineering & Remote Sensing* 48: 91-100.

Baltsavias E. P. (1999) Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing* 54: 199-214.

Bannari A., Omari K., Teillet P. M. & Fedosejevs G. (2005) Potential of Getis statistics to characterize the radiometric uniformity and stability of test sites used for the calibration of earth observation sensors. *IEEE Transactions on Geoscience and Remote Sensing* 43: 2918-2926.

Bishop M. P., Schroder J. F. & Colby J. D. (2003) Remote sensing and geomorphometry for studying relief production in high mountains. *Geomorphology* 55: 345-361.

Boots B. (2002) Local measures of spatial association. *Ecoscience* 9: 168-176.

Brandtberg T., Warner T. A., Landenberger R. E. & McGraw J. B. (2003) Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment* 85: 290-303.

Brokaw N. V. L. (1982) The definition of treefall gap and its effect on measures of forest dynamics. *Biotropica* 14: 158-160.

Cliff A. & Ord J. K. (1981) *Spatial processes: models & applications*. Pion, London.

CRD (2006) Sooke and Goldstream Watersheds. In: *CRD Water Services* http://www.crd.bc.ca/water/watersupplyarea/watersheds.htm. Capital Regional District, Victoria.

Derksen C., Wulder M., LeDrew E. & Goodison B. (1998) Associations between spatially derived autocorrelation patterns of SSM/I-derived prairie snow cover and atmospheric circulation. *Hydrological Processes* 12: 2307-2316.

Durrand J. M., Gimonet B. J. & Perbos J. R. (1987) Speckle in SAR images: an evaluation of filtering techniques. *Advances in Space Research* 7: 301-304.

Ennos A. E. (1996) Laser speckle experiements for students. *Physics Education* 31: 138-142.

Franklin J. F. & Van Pelt R. (2004) Spatial Aspects of Structural Complexity. *Journal of Forestry* 102: 22-28.

Frazer G. W., Wulder M. A. & Niemann K. O. (2005) Simulation and quantification of the fine-scale spatial pattern and heterogeneity of forest canopy structure: A lacunarity-based method designed for analysis of continuous canopy heights. *Forest Ecology and Management* 214: 65.

Griffith P., Getis A. & Griffin E. (1996) Regional patterns of affirmative action compliance costs. *The Annals of Regional Science* 30: 321-340.

Hofton M. A., Rocchio L. E., Blair J. B. & Dubayah R. (2002) Validation of Vegetation Canopy Lidar sub-canopy topography measurements for a dense tropical forest. *Journal of Geodynamics* 34: 491-502.

Holmgren J. & Persson Å. (2004) Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment* 90: 415-423.

Jensen J. R. (2000) *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice-Hall, Upper Saddle River, NJ.

Jensen J. R. (2005) *Introductory digital image processing: a remote sensing perspective*. Pearson Prentice Hall, Upper Saddle River, NJ.

Jin Y. & Yan F. (2004) Monitoring sandstorms and desertification in northern China using SSM/I data and Getis statistics. *International Journal of Remote Sensing* 25: 2053-2060.

Kaasalainen S., Ahokas E., Hyppa J. & Suomalainen J. (2005) Study of Surface Brightness From Backscattered Laser Intensity: Calibration of Laser Data. *IEEE Geoscience and Remote Sensing Letters* 2: 255-259.

Krabill W. B., Collins J. G., Link L. E., Swift R. N. & Butler M. L. (1984) Airborne laser topographic mapping results. *Photogrammetric Engineering & Remote Sensing* 50: 685-694.

LeDrew E., Holden H., Wulder M., Derksen C. & Newman C. (2004) A spatial statistical operator applied to multidate satellite imagery for identification of coral reef stress. *Remote Sensing of Environment* 91: 271-279.

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.

Lillesand T. M. & Kiefer R. W. (2000) *Remote Sensing and Image Interpretation*. John Wiley & Sons Inc., New York.

Lim K., Treitz P., Baldwin K., Morrison I. & Green J. (2003b) Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing* 29: 658-678.

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

Lovas T., Toth C. K. & A. Barsi (2004) Model-based vehicle detection from LiDAR data. In: *XXth ISPRS Congress*. International Archives of Photogrammerty & Remote Sensing, Istanbul, Turkey.

Lutz E., Geist T. & Stötter J. (2003) Investigations of airborne laser scanning signal intensity on glacial surfaces - utilizing comprehensive laser geometry modelling and orthophoto surface modelling (A case study: Svartisheibreen, Norway). In: *ISPRS Workshop on 3-D reconstruction from airborne laserscanner and INSAR data*, Dresden, Germany.

Luzum B., Starek M. & Slatton K. C. (2004) Normalizing ALSM Intensities pp. 1-8. Geosensing Engineering and Mapping (GEM), Civil and Coastal Engineering Department, University of Florida, Gainesville, FL.

Maclean G. A. & Krabill W. B. (1986) Gross-merchantable timber volume estimation using an airborne lidar system. *Canadian Journal Forest Research* 12: 7-18.

Magnussen S. & Boudewyn P. (1998) Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research* 28: 1016-1031.

Maltamo M., Eerikä inenb K., Pitkä nena J., Hyyppä J. & Vehmasa M. (2004) Estimation of timber volume and stem density based on scanning altimetry and expected tree size distribution functions. *Remote Sensing of Environment* 90: 319-330.

Markwardt C. B. (2005) Markwardt IDL Library.

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.

Moffiet T., Mengersen K., Witte C., King R. & Denham R. (2005) Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing* 59: 289-309.

Moran P. A. P. (1950) Notes on continuous stochastic phenomena. *Biometrika* 37: 17-23.

Naesset E. (1997) Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment* 61: 246-253.

Natural Resources Canada (2006) The State of Canada's Forests, 2005-2006. Canadian Forest Service, Ottawa, Ontario.

Nelson R., Krabill W. & Tonelli J. (1988) Estimating Forest Biomass and Volume Using Airborne Laser Data. *Remote Sensing of Environment* 24: 247-267.

Nigh G. D. & Love B. A. (2004) Predicting crown class in three western conifer species. *Canadian Journal of Forest Research* 34: 592-599.

Ogunjemiyo S., Parker G. & Roberts D. A. (2005) Reflections in bumpy terrain: implications of canopy surface variations for the radiation balance of vegetation. *IEEE Geoscience and Remote Sensing Letters* 2: 90-93.

Ord J. K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24: 189-206.

Parish R. & Thomson S. (1994) *Tree Book: learning to recognize trees of British Columbia*. Ministry of Forests / Canadian Forest Service, Victoria.

Parker G. G. (1995) Structure and microclimate of forest canopies. In: *Forest Canopies* (eds. M. D. Lowman & N. M. Nadkarni) pp. 73-98. Academic Press, Orlando, FL.

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.

Riaño D., Meier E., Allgöwer B., Chuvieco E. & Ustin S. L. (2003) Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment* 86: 177-186.

Ritchie J. C. (1995) Airborne laser altimeter measurements of landscape topography. *Remote Sensing of Environment* 53: 91-96.

RSI (2005) IDL Online Help. In: *IDL Assistant*. Research Systems Inc., Bolder, CO.

Sawada M. (1999) Rookcase: An Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. *Bulletin of the Ecological Society of America* 80: 231-234.

Schreier H., Lougheed J., Tucker C. & Leckie D. (1985) Automated measurements of terrain reflection and height variations using an airborne infrared laser system. *International Journal of Remote Sensing* 6: 101-113.

Soenen S. A., Peddle D. R. & Coburn C. A. (2005) SCS+C: A Modified Sun-Canopy-Sensor Topographic Correction in Forested Terrain. *IEEE Transactions on Geoscience and Remote Sensing* 43: 2148-2159.

Song J.-H., Han S.-H., Yu K. & Kim Y.-I. (2002) Assessing the possibility of land-cover classification using lidar intensity data. In: *Photogrammetric Computer Vision (ISPRS Commission III)* pp. 259-262, Graz, Austria.

Staudhammer C. & LeMay V. (2000) Height prediction equations using diameter and stand density measures. *The Forestry Cronicle* 76: 303-309.

Tan S. & Narayanan R. M. (2004) Design and performance of a multiwavelength airborne polarimetric lidar for vegetation remote sensing. *Applied Optics* 43: 2360-2368.

Trofymow J. A., Porter G. L., Blackwell B. A., Arksey R., Marshall V. & Pollard D. (1997) Chronosequences for research into the effects of converting coastal British Columbia old-growth forest to managed forests: an establishment report. Natural Resources Canada, Pacific Forestry Centre, Information Report BC-X-374, Victoria, Canada.

Wehr A. & Lohr U. (1999) Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 68-82.

Wells L. & Getis A. (1999) The spatial characteristics of stand structure in Pinus torreyana. *Plant Ecology* 143: 153-177.

Wulder M. & Boots B. (1998) Local spatial autocorrelation characteristics of remotely sensesd imagery assessed with the Getis statistic. *International Journal of Remote Sensing* 19: 2223-22231.

Wulder M. & Boots B. (2001) Local spatial autocorrelation characteristics of Landsat TM imagery of managed forest area. *Canadian Journal of Remote Sensing* 27: 67-75.

Yu K., Han S.-H., Chan H. & Ha T.-J. (2002) Potential of reflected intensity of airborne laser scanning systems in roadway features identification. *Geomatica* 56: 363-374.

Zeide B. & Vanderschaaf C. (2002) The effect of density on the height-diameter relationship. In: *Proceedings of the eleventh biennial southern silvicultural research conference* (ed. K. W. Outcalt) pp. 463-466. US Department of Agriculture, Asherville, NC.

Appendix A – Generation of canopy height model

```
PRO GENERATE_CHM
grid_size = 2
; make DTM
ground_file = '/Volumes/C2RAID/jaden/Full_Ground.txt'
OPENR, lun, ground_file, /GET_LUN
nLg = FILE_LINES(ground_file)
ground = DBLARR(10,nLg)
READF, lun, ground
FREE_LUN, lun
x = ground[0,*]
y = ground[1,*]
z = ground[2,*]
ground = 0
TRIANGULATE, x, y, triangles
DTM_grid_size = [grid_size, grid_size]
z_grnd = TRIGRID (x, y, z, triangles, DTM_grid_size, xgrid=x_grnd, ygrid=y_grnd)
; begin main loop
FOR f=1, 28 DO BEGIN
input_file = '/Volumes/C2RAID/jaden/output/file'+string(Format='(i2)',f)+'.txt'
output_file = '/Volumes/C2RAID/jaden/output/subtracted'+string(Format='(i2)',f)+'.txt'
; separate vegetation from ground
OPENR, lun, input_file, /GET_LUN
nL = FILE_LINES(input_file)
data = DBLARR(11,nL)
READF, lun, data
FREE_LUN, lun
class = data[8,*]
veg = WHERE (class EQ 1, count)
nLv = count
```

```
vegArray = data[*,veg]
data = 0
x = vegArray[0,*]
y = vegArray[1,*]
; isolate ground points for this file
xbound = WHERE (x_grnd GE MIN(x) AND x_grnd LT MAX(x))
ybound = WHERE (y_grnd GE MIN(y) AND y_grnd LT MAX(y))
grndArray = z_grnd[MIN(xbound):MAX(xbound),MIN(ybound):MAX(ybound)]
minx = FLOOR(x_grnd[MIN(xbound)])
miny = FLOOR(y_grnd[MIN(ybound)])
maxy = CEIL(y_grnd[MAX(ybound)])
; subtract vegetation from DTM
geo_size_y = y_grnd[MAX(ybound)] - y_grnd[MIN(ybound)]
geo_size_x = x_grnd[MAX(xbound)] - x_grnd[MIN(xbound)]
rows = CEIL(geo_size_y/grid_size)+1
cols = CEIL(geo_size_x/grid_size)+1
cell_space = grid_size / 2.
veg_sub = DBLARR(11,nLv) ; establish array for output file
hit = 0L
FOR i=0, rows-1 DO BEGIN
   lowerBound = (miny - cell_space) + (grid_size * i)
   upperBound = lowerBound + grid_size
   bounds = WHERE ((y GE lowerBound) AND (y LT upperBound))
   IF bounds[0] NE -1 THEN BEGIN
      veg_block = vegArray[*,bounds]
      x = veg_block[0,*]
      FOR j=0, cols-1 DO BEGIN
         leftBound = (minx - cell_space) + (grid_size * j)
         rightBound = leftbound + grid_size
         cell = WHERE ((x GE leftbound) AND (x LT rightbound), count)
         veg_points = count
```

```
            IF cell[0] NE -1 THEN BEGIN
               veg_block[2,cell] = TEMPORARY(veg_block[2,cell] - grndArray[j,i])
               veg_sub[*,hit:hit+(veg_points-1)] = veg_block[*,cell]
               hit = hit + veg_points
            ENDIF
         ENDFOR
      ENDIF
   ENDFOR
   vegArray = 0
   IF (hit LT nLV) THEN veg_sub = TEMPORARY(veg_sub[*,0:hit-1])
   ; write output
   OPENW, temp, output_file, /GET_LUN
   output_format = '(f9.2,f11.2,f7.2,i4,2i2,f8.5,i3,i2,i4,f14.6)'
   PRINTF, temp, veg_sub, FORMAT=output_format
   FREE_LUN, temp
   veg_sub = 0
   ENDFOR
   END
```

Appendix B – Laser range calculation and normalization

```
PRO NORMALIZE
trajectory_file = '~/Desktop/terra/nav/trajectory.csv'
OPENR, lun, trajectory_file, /GET_LUN
trajectory = DBLARR(4,FILE_LINES(trajectory_file))
READF, lun, trajectory
FREE_LUN, lun
FOR file=1,28 DO BEGIN
input_file = '~/Desktop/output/file'+STRING(FORMAT='(i2)',file)+'.txt'
OPENR, lun, input_file, /GET_LUN
nL = FILE_LINES(input_file)
data = DBLARR(11,nL)
READF, lun, data
FREE_LUN, lun
t_aircraft = trajectory[0,*]
t_point = data[10,*]
output_file = '~/Desktop/output/norm/file'+STRING(FORMAT='(i2)',file)+'.txt'
output_format = '(f9.2,f11.2,f7.2,2i4,2i2,f8.2,i3,i2,2i4,f12.4)'
OPENW, lun, output_file, /GET_LUN
FOR i=0L, nL-1 DO BEGIN
   t_range = WHERE(t_aircraft GT t_point[i]-1 AND t_aircraft LT t_point[i]+1)
   t_diff = ABS(t_point[i]-t_aircraft[t_range])
   t_sort = SORT(t_diff)
   t_asc = t_range[t_sort]
   xv1 = trajectory[1,t_asc[0]] ; coordinates of closest trajectory point
   yv1 = trajectory[2,t_asc[0]]
   xv2 = trajectory[1,t_asc[1]] ; coordinates of second closest trajectory point
   yv2 = trajectory[2,t_asc[1]]
   xp = data[0,i] ; coordinates of point
   yp = data[1,i]
```

```
   d = PNT_LINE([xp,yp],[xv1,yv1],[xv2,yv2]) ; perpendicular distance from line
   t1 = t_aircraft[t_asc[0]] ; nearest timestamp
   t2 = t_aircraft[t_asc[1]] ; second nearest timestamp
   tp = data[10,i] ; timestamp of point
   h1 = trajectory[3,t_asc[0]] ; height of aircraft at closest trajectory point
   h2 = trajectory[3,t_asc[1]] ; height of aircraft at second closest trajectory point
   alt = h1+(((h2-h1)/(t2-t1))*(tp-t1)) ; estimated altitude of aircraft (MSL)
   h = alt - data[2,i] ; height of the aircraft above the point (m)
   path_length = SQRT(d^2+h^2) ; path length between aircraft and point
   scan_angle = ATAN(d/h) ; scan angle between aircraft and point (radians)
   angle_deg = scan_angle*(180/!pi) ; scan angle in degrees
   path = 0.1658*(path_length-800)
   int_norm = data[3,i]+path
   IF (int_norm GT 511) THEN int_norm = 511
   PRINTF, lun, data[0,i], data[1,i], data[2,i], data[3,i], int_norm, data[4,i], data[5,i], $
   path_length, data[7,i], data[8,i], data[9,i], angle_deg, data[10,i], $
   FORMAT=output_format
ENDFOR
FREE_LUN, lun
ENDFOR
END
```

Appendix C – Height prediction

Predicting height of Douglas-fir, Western hemlock and Western redcedar using diameter at breast height (dbh) field measurements

Introduction

It is difficult to comprehensively measure individual tree heights within forest stands, especially in mature closed canopies where the tree tops are occluded. It is a common practice in forest research to measure diameter of the outer bark at breast height 1.3m (dbh) as a proxy measure for height. Allometric relationships link dbh to height through regression equations fitted separately for each species.

In this study, dbh was measured on all 1738 individuals and height was measured on a subsample as outlined in Table 1. The allometric model was developed using 4336 measurements from complete plot enumerations conducted in the same study area (Frazer, 2001 unpublished). To build the best model, dead trees and trees with damaged tops were intentionally excluded from the analysis. When the model was parameterized, it is then tested on the 283 height-dbh measurements from the current study. Parameters for Red alder are taken from another study (Staudhammer & LeMay 2000) due to small sample size. Other species are omitted due to low relative population size (1.4%).

Table 1 – Sample details

	Douglas-fir	Western hemlock	Western redcedar	Red alder	Western white pine	Lodgepole pine	Other	Total
N (2005)	878	664	115	30	22	4	25	1738
height-dbh sub-sample (2005)	184 (1)	52	32	4	8	1	2	283
height-dbh (2001)	2073 (4)	923 (1)	1266 (1)	-	43	31	-	4336

The numbers in parenthesis represent outliers that were excluded from the regression

Analysis

Separate analyses proceeded for each species. The scatterplot for Douglas-fir is presented as Figure 1. A curve was fit to these data using Equation 1, which is based on a Weibull distribution (Staudhammer & LeMay 2000). This equation was chosen because it resulted in the highest r^2 and lowest root mean squared error (RMSE) for each species, compared to results from other models such as second order polynomial.

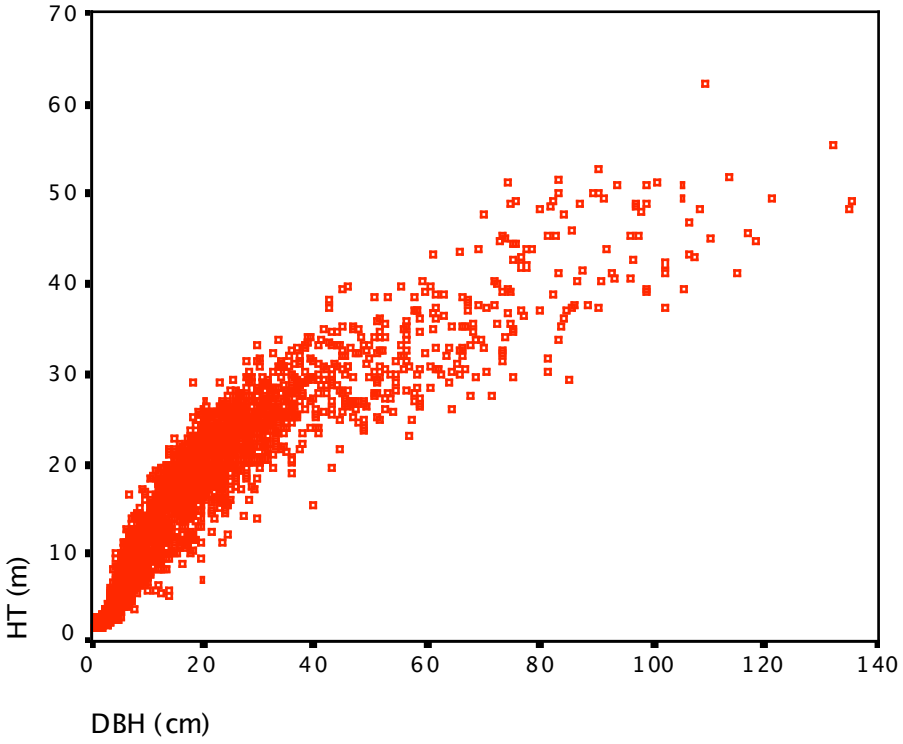


Figure 1 – Height vs. dbh for Douglas-fir

height = 1.3+A(1-exp(B*dbh^C)) + error (1)

Parameter starting values were obtained from published results for this species (Staudhammer & LeMay 2000), and re-estimated in SPSS nonlinear regression using the Levenberg-Marquardt method. The parameterized model is shown in Equation 2 below.

height = 1.3+48.64(1-exp(-.02960*dbh^.9071) + error (2)

R^2 = .89435, RMSE = 3.2789m

A plot of the residuals (Figure 2) indicates that heteroscedacticity may exist. Significance of unequal variance in the height-dbh relationship was confirmed by a Goldfeld-Quandt test. Heteroscedacticity is problematic because it violates a statistical assumption for ordinary least squares (OLS) regression. Because the OLS method minimizes squared differences between predicted and observed values, cases with larger variance will have a greater impact on the estimated coefficients than cases where variance is relatively low. Heteroscedacticity is a statistical problem, however its occurrence is likely due to biological factors that underlie the growth of trees and external factors such as stand density (Zeide & Vanderschaaf 2002).

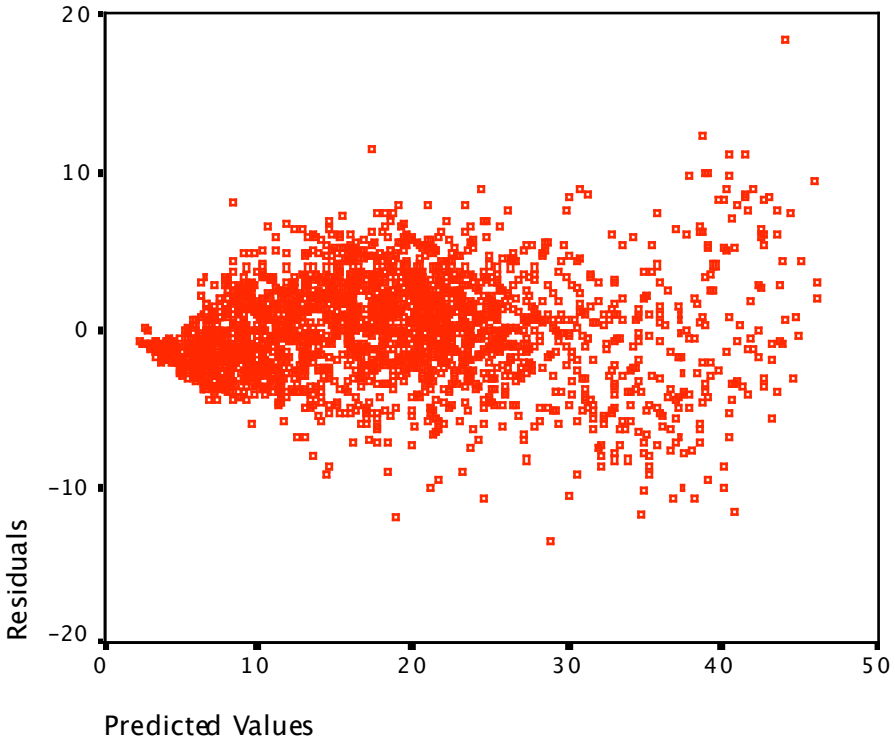


Figure 2 – Heteroscedacticity is evident by looking at the residuals from Equation 1

Dealing with heteroscedacticity

Several strategies were employed to deal with heteroscedacticity including: natural log transformation, multiple regression and weighted least squares (WLS) regression. The first strategy was ineffective because the distinctive curvilinear relationship was not

preserved by the transformation. Partitioning the data was explored at length, but it resulted in discontinuous curves, which are not realistic for tree growth. A multiple linear regression was also attempted, using dbh, density and occupancy variables, however the additional independent variables were found to be marginally significant and the residuals remained heteroscedactic.

WLS regression was found to be the most effective, in terms of the visual curve fit and root mean squared error (RMSE). To determine weighting, a custom IDL script calculated the variance of the dependant variable (range of height) for each level of the independent variable (dbh) and a linear regression was performed to develop a continuous weighting function (see Figure 3). An IDL function called MPFIT (Markwardt 2005) was used to fit the curve, weighted by 1/variance(dbh_i). MPFIT also uses the Levenberg-Marquardt iterative method for optimizing the coefficients.

Results

The model parameter estimates and fit statistics are presented in Table 2, and the graphical plots are shown in Figures 3 through 7. The variance graph, from which the weights are derived, is shown for Douglas-fir as an example. Note that negative bias indicates model over prediction, while positive bias indicates model under prediction.

Table 2 – Model Development

Equation	*height = 1.3+A(1-exp(B*dbh^C) + error*			r^2	Bias (m)	RMSE (m)
Parameter	A	B	C			
Douglas-fir	90.8031	-0.0306115	0.658409	.842	-0.573	3.37
Western hemlock	59.6064	-0.0174121	0.995420	.904	-0.261	2.34
Western redcedar	35.5749	-0.00927767	1.29844	.924	0.089	1.79
Red alder	26.5495*	-0.03079*	1.20438*	.770	-0.083	3.95
Western white pine	49.3006	-0.0156818	1.12385	.936	-0.539	2.19
Lodgepole pine	50.5733	-.0353412	0.794461	.764	-0.088	2.38

*Paramaters from Staudhammer & LeMay, 2000

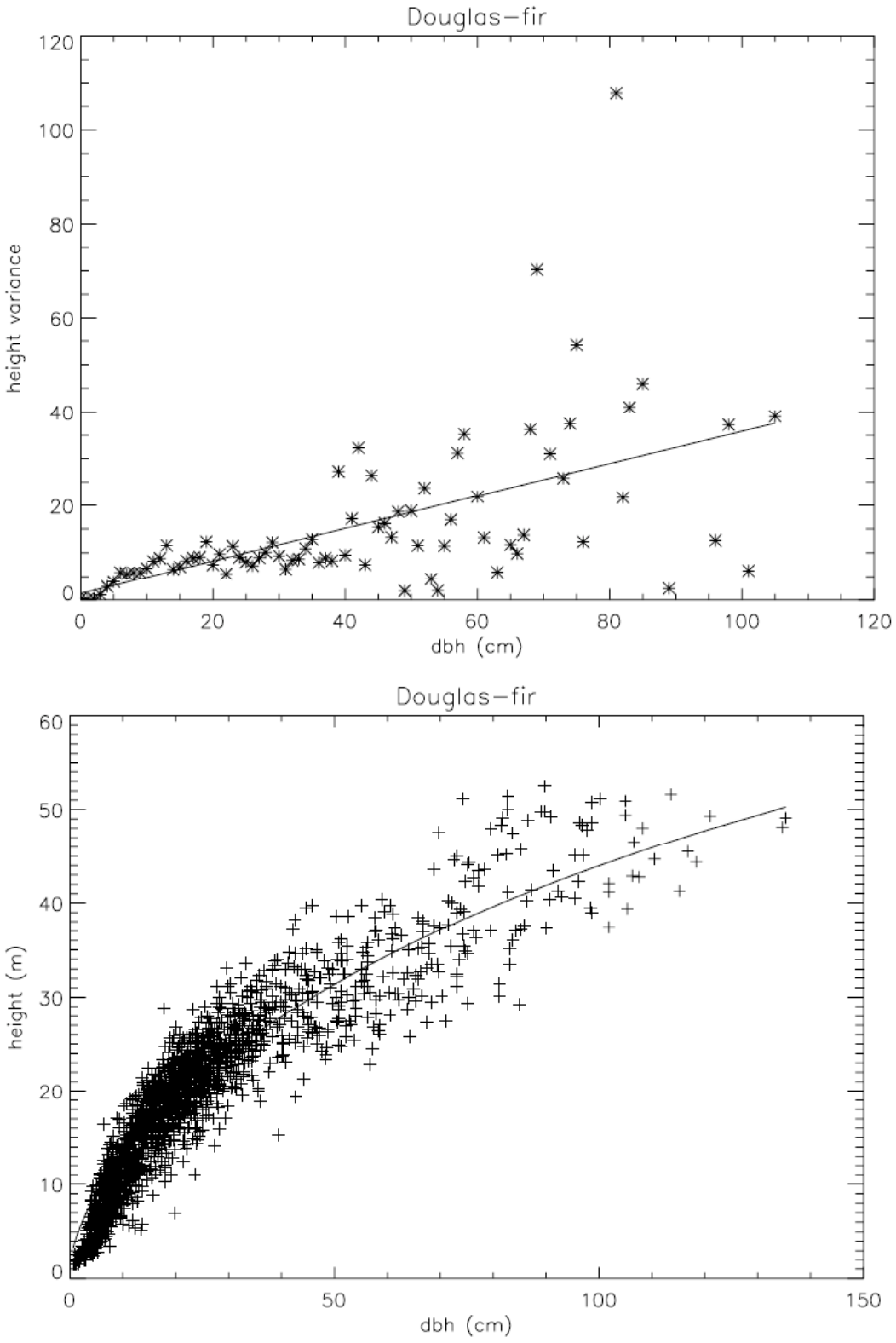


Figure 3 – Douglas-fir dbh vs. height variance (top) and dbh v. height (bottom) (2001 data)

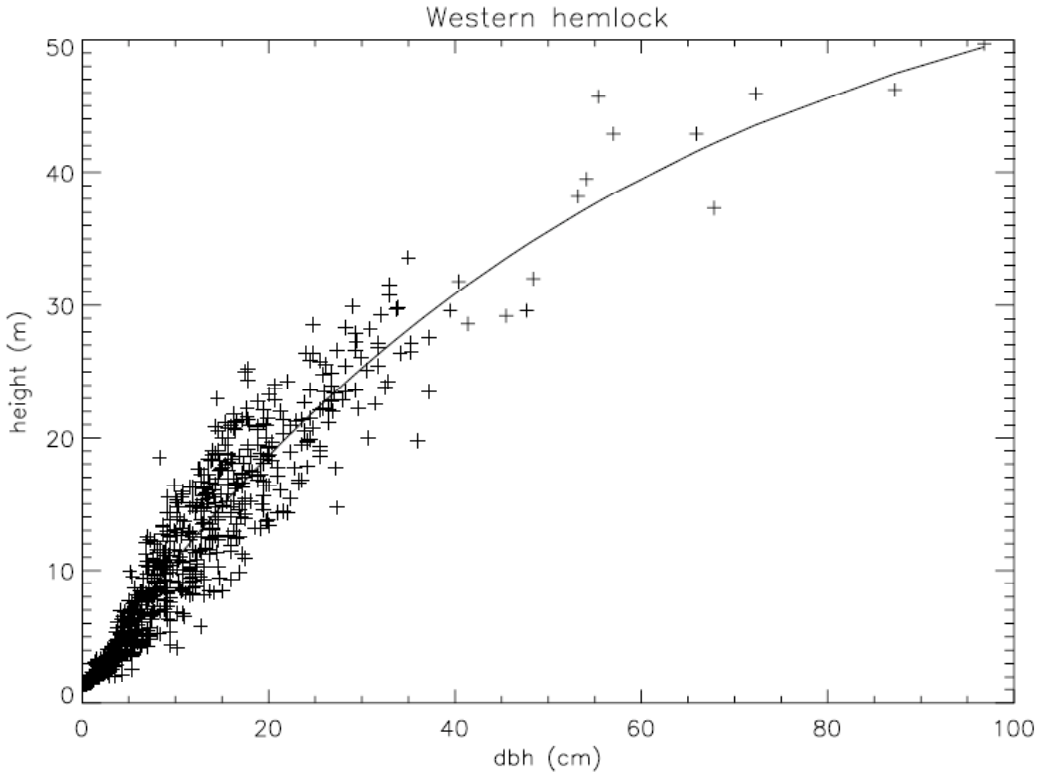


Figure 4 – Western hemlock (2001 data)

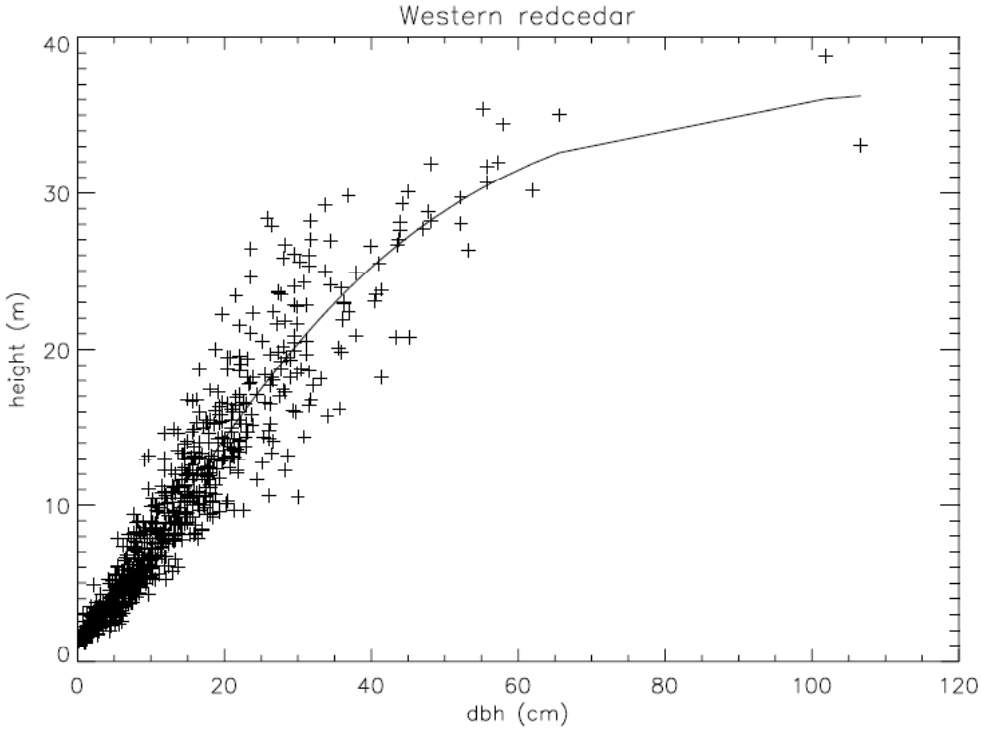


Figure 5 – Western redcedar (2001 data)

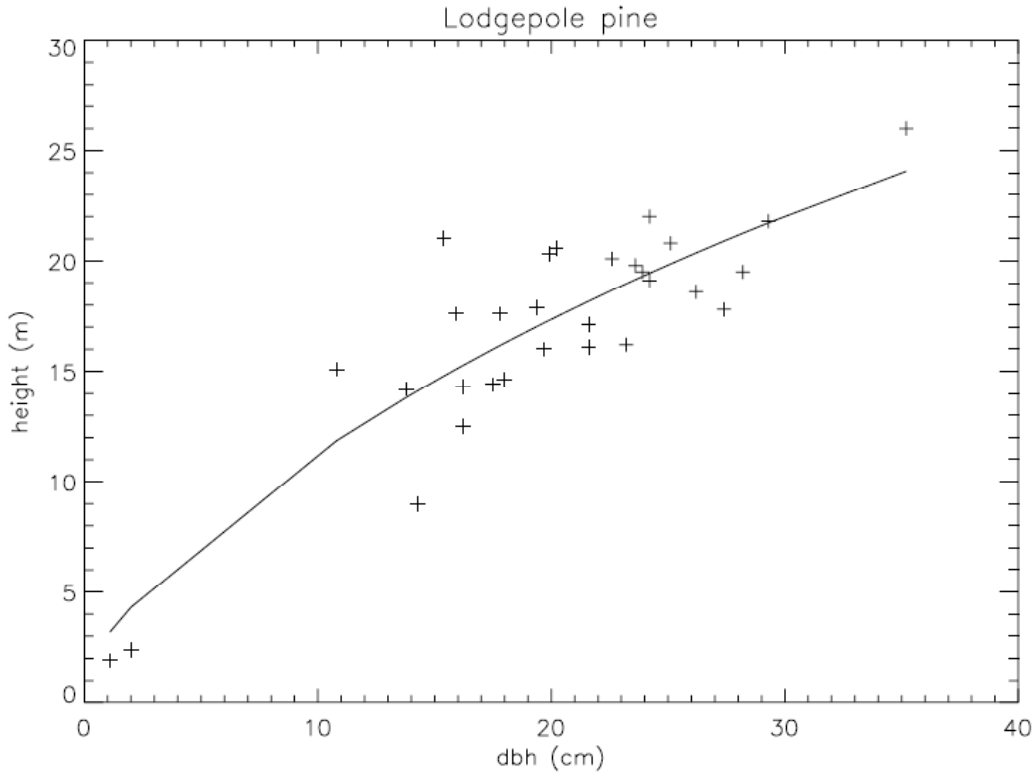


Figure 6 – Lodgepole pine (2001 data)

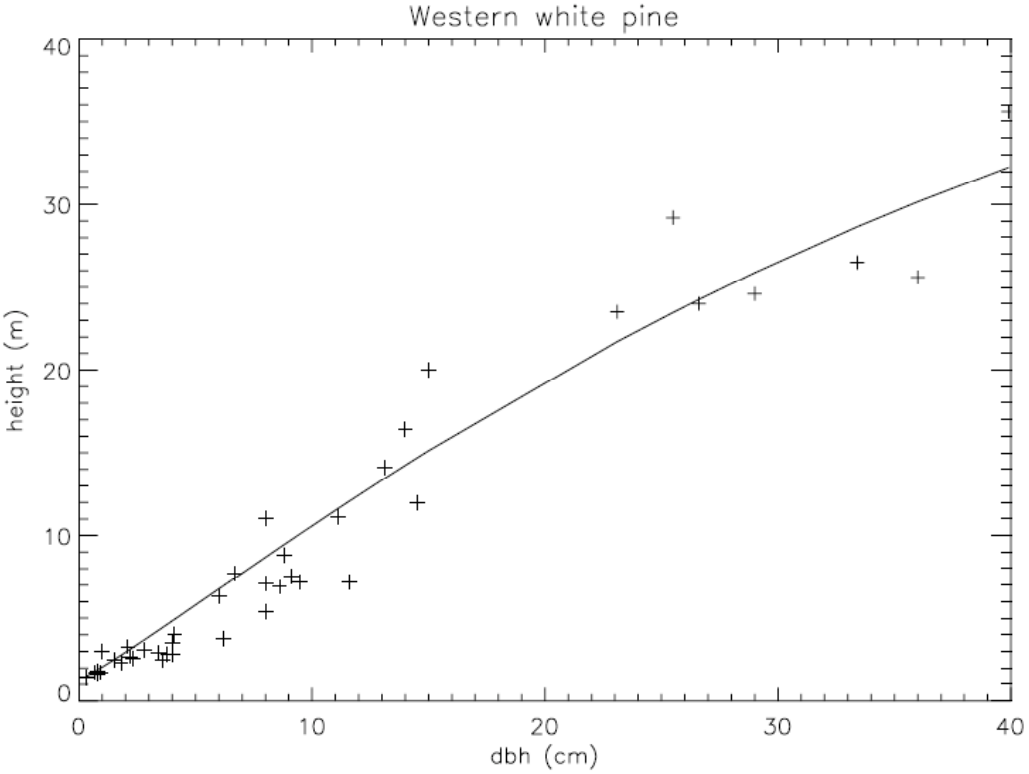


Figure 7 – Western white pine (2001 data)

The curves, developed above using the comprehensive 2001 dataset, were then plotted over the 2005 hight-dbh sub-sample to check for model correspondence. The bias, RMSE, max error and standard error between the model and 2005 data points are shown in Table 3. The prediction curves for each species are shown in Figures 8 through 12. Lodgepole pine was not included in the testing phase, since only one heigh-dbh pair exists in the 2005 data.

Table 3 – Model Testing

Species	n	Bias (m)	RMSE (m)	Max Error (m)	Std Error (m)
Douglas-fir	183	-2.07	4.70	16.5	4.21
Western hemlock	52	-2.01	2.82	6.60	1.96
Western redcedar	32	-1.88	3.12	7.82	2.46
Red alder	4	-0.06	1.00	1.31	1.00
Western white pine	8	-1.66	2.01	2.69	0.948

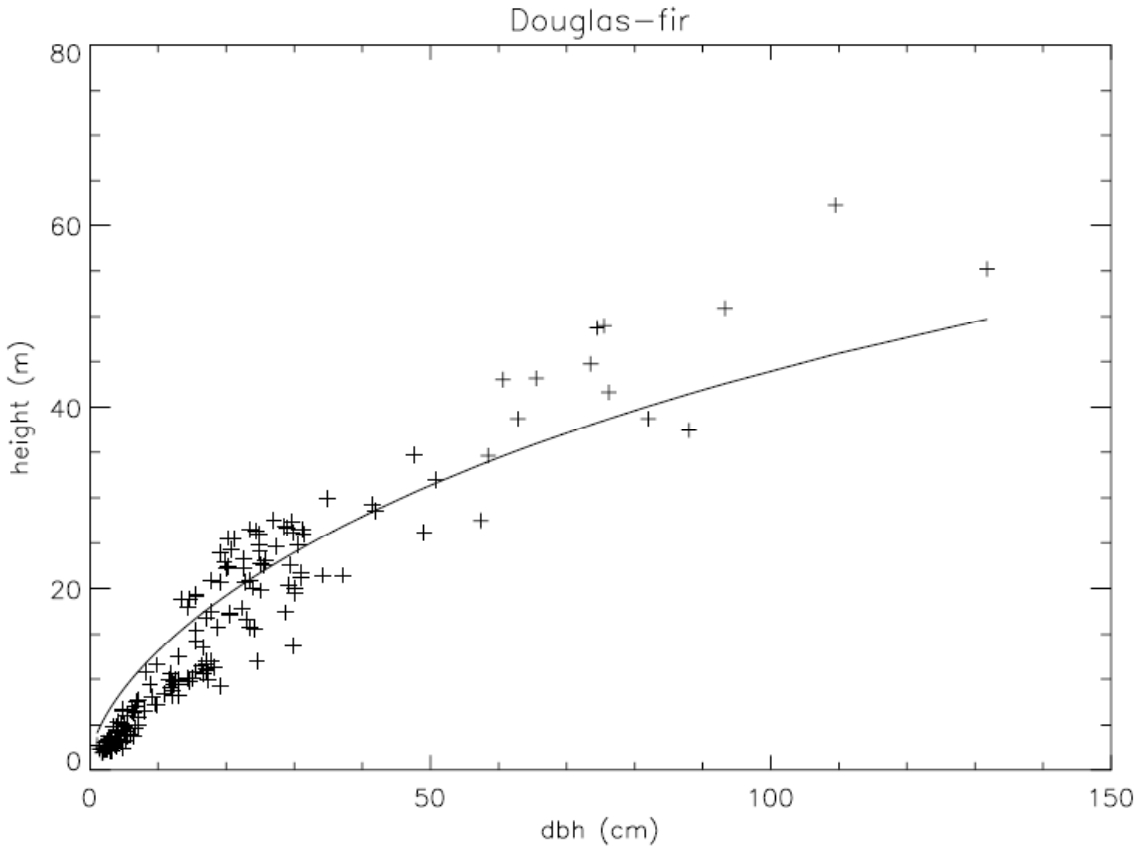


Figure 8 – Douglas fir (2005 data)

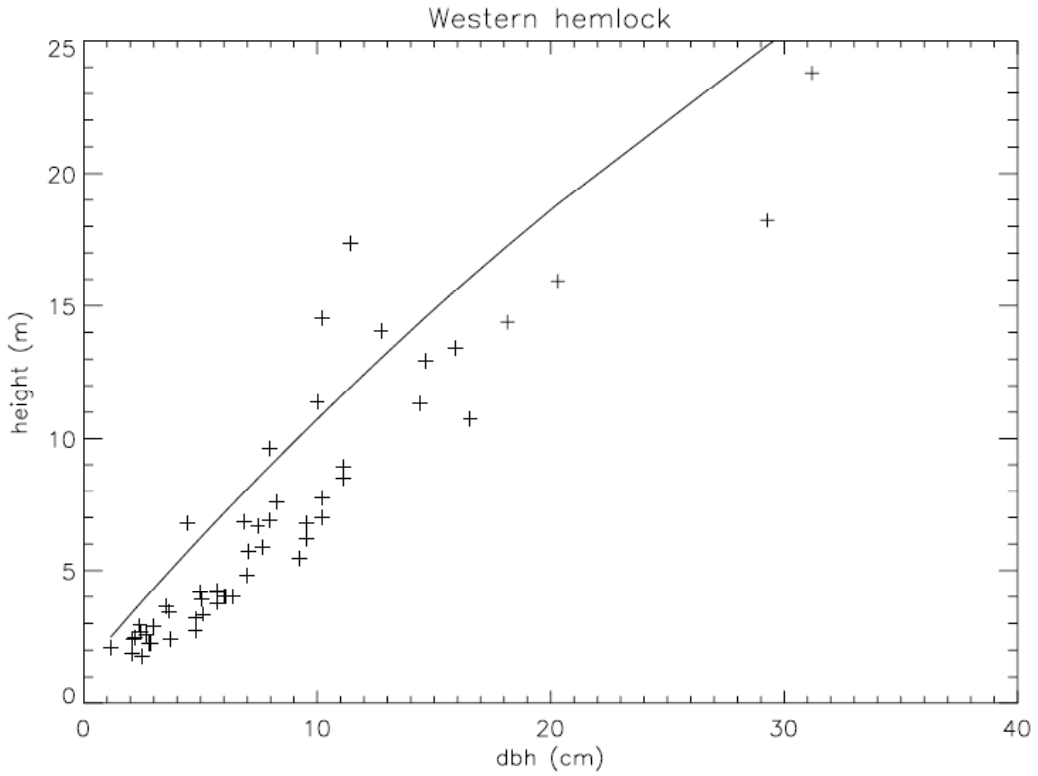


Figure 9 – Western Hemlock (2005 data)

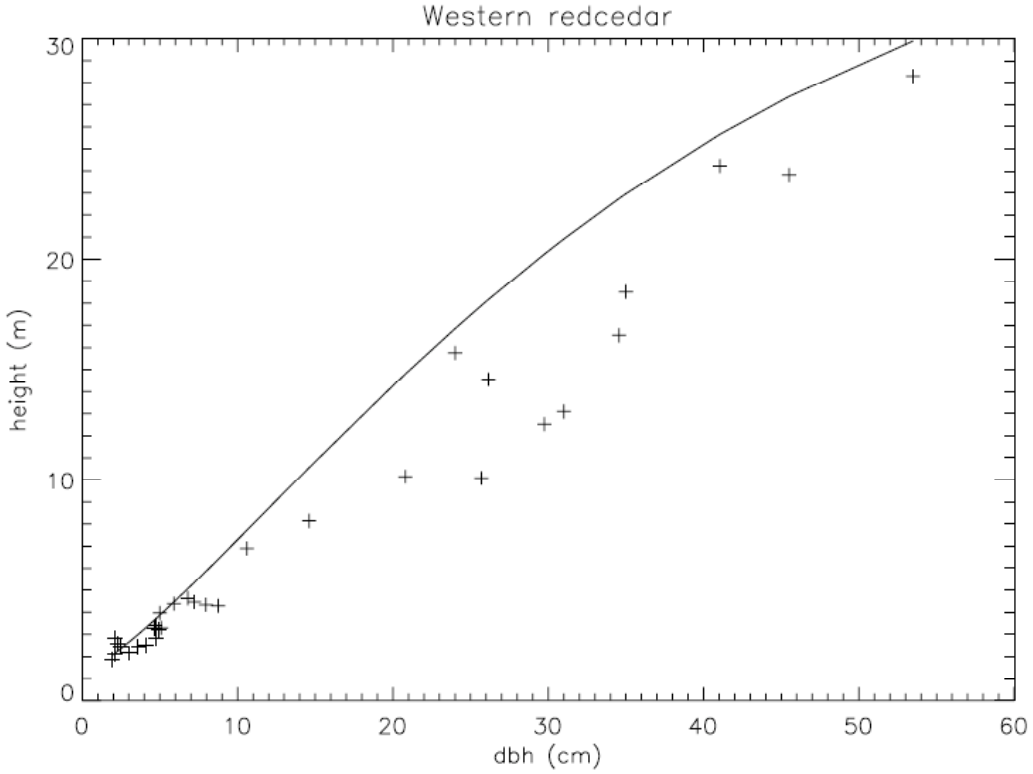


Figure 10 – Western redcedar (2005 data)

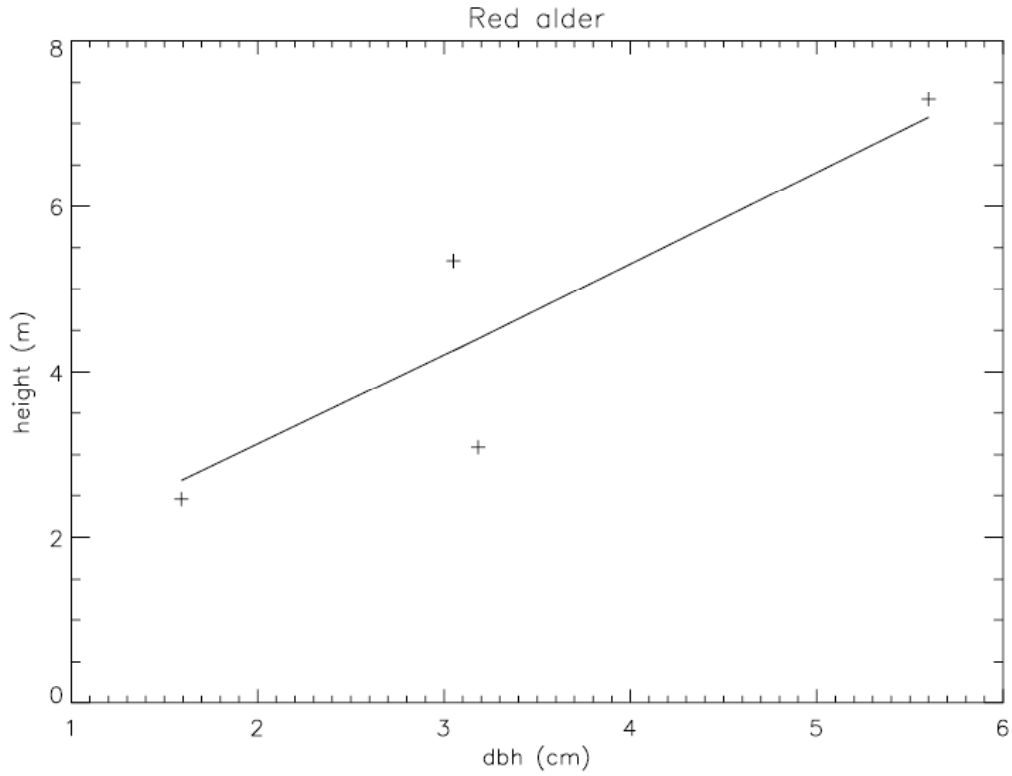


Figure 11 – Red alder (2005 data)

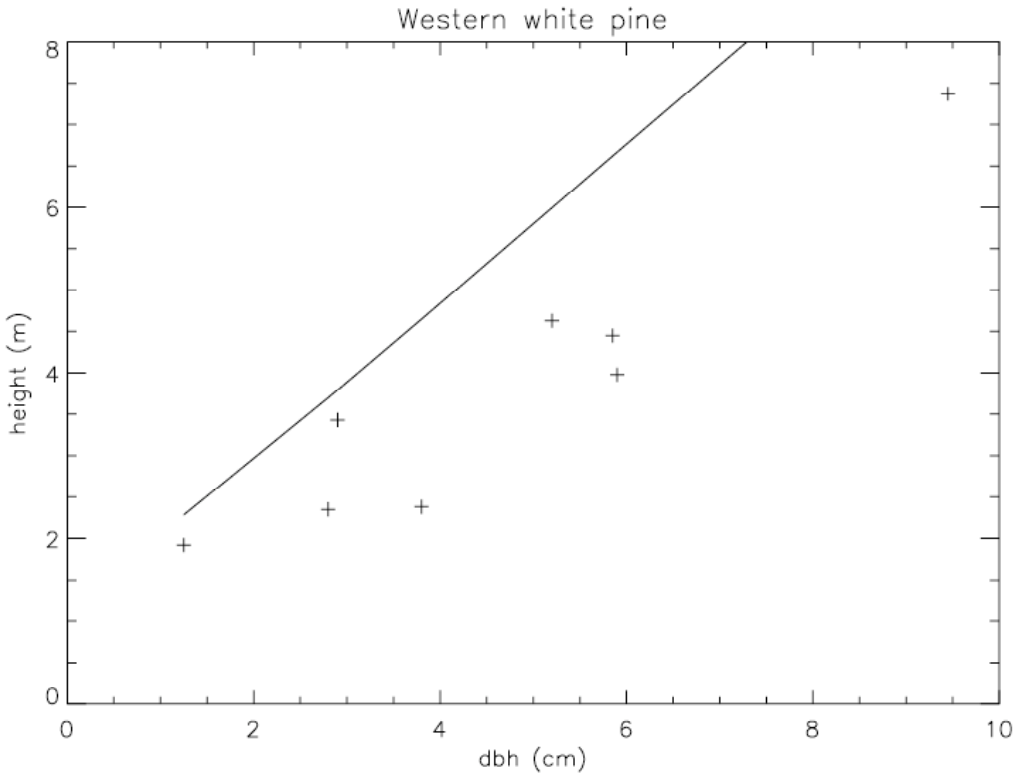


Figure 12 – Western white pine (2005 data)

Appendix D – G_i^* statistic computation

```
FUNCTION GETIS, x, y, values, radius
; initialize output array
num = FLOAT(N_ELEMENTS(values))
output = FLTARR(6,num)
output[0:2,*] = [x, y, values] ; copy original data into output array
values = REFORM(output[2,*]) ; re-set values to ensure they are FLOAT
; calculate mean and variance (for use in standardized Gi* calculations)
mean = MEAN(values)
s2 = TOTAL((values-mean)^2)/num
; begin computing the statistic for each element
missing = 0
FOR i=0, num-1 DO BEGIN
   in = WHERE ( (x-x[i])^2 + (y-y[i])^2 LE radius^2, neighbours) ; find points
   w = FLTARR(num)
   w[in] = 1 ; if within radius, assign weight of one (default is 0)
   output[3,i] = neighbours ; number of neighbours
   output[4,i] = TOTAL(w*values)/TOTAL(values) ; Gi* statistic
   ; standardized Gi* statistic calculations
   Wi = TOTAL(w)
   numerator = TOTAL(w*values) - (wi*mean)
   denomanator =  SQRT(s2) * ( ( Wi * (num-Wi) ) / (num-1) )^0.5
   output[5,i] = TOTAL(numerator)/denomanator ; zGi* statistic
   IF (neighbours LE 8) THEN missing = missing + 1
ENDFOR
IF (missing NE 0) THEN PRINT, missing, ' points with <8 neighbours'
RETURN, output
END
```

Appendix E – Raster sampling routine

```
PRO RASTER_ANALYSIS
plot_file = '~/Field/plot_corners.csv'
OPENR, lun, plot_file, /GET_LUN
corners = DBLARR(3,FILE_LINES(plot_file))
READF, lun, corners
FREE_LUN, lun
output_file = '~/Analysis/raster/table-scale.txt'
OPENW, out, output_file, /GET_LUN
scale = [4,7,5,6,8,5,6,4,3]
FOR p=1,9 DO BEGIN
   grid_size = scale[p-1]
   input_file = '~/Data/plot'+STRING(FORMAT='(i2)',p)+' single nadir.txt'
   OPENR, lun, input_file, /GET_LUN
   data = DBLARR(13,FILE_LINES(input_file))
   READF, lun, data
   FREE_LUN, lun
   x = data[0,*]
   y = data[1,*]
   int = data[3,*]
   getis = GETIS(x,y,int,scale[p-1])
   ht = data[2,*]
   gi = getis[5,*]
   inplot = WHERE(corners[0,*] EQ p)
   px = REFORM(corners[1,inplot])
   py = REFORM(corners[2,inplot])
   inplot = INSIDE_VECTOR(x, y, px, py, /INDEX)
   data = TEMPORARY(data[*,inplot])
   x = data[0,*]
   y = data[1,*]
```

```
   minx = FLOOR(MIN(x)) ; use even integer UTM pixel coords
   maxx = MAX(x)
   miny = FLOOR(MIN(y))
   maxy = MAX(y)
   rows = CEIL( FLOAT(maxy-miny)/FLOAT(grid_size) )
   cols = CEIL( FLOAT(maxx-minx)/FLOAT(grid_size) )
   FOR i=0, rows-1 DO BEGIN
      lowerBound = (miny) + (grid_size * i)
      upperBound = lowerBound + grid_size
      bounds = WHERE ( (y GE lowerBound) AND (y LT upperBound) )
      IF bounds[0] NE -1 THEN BEGIN
         x1 = x[bounds]
         ht = data[2,bounds]
         int = data[3,bounds]
         gi = getis[5,bounds]
         FOR j=0, cols-1 DO BEGIN
            leftBound = (minx) + (grid_size * j)
            rightBound = leftbound + grid_size
            cell = WHERE ( (x1 GE leftbound) AND (x1 LT rightbound) )
            IF cell[0] NE -1 THEN BEGIN
               max_ht = MAX(ht[cell],highest)
               int_c = int[cell]
               gi_c = gi[cell]
               PRINTF, out, p, max_ht, int_c[highest], gi_c[highest]
            ENDIF
         ENDFOR
      ENDIF
   ENDFOR
ENDFOR
FREE_LUN, out
END
```

Appendix F – Vector (point) sampling routine

```
PRO POINT_ANALYSIS
field_file = '~/Field/inventory.csv'
OPENR, lun, field_file, /GET_LUN
inventory = DBLARR(7,FILE_LINES(field_file))
READF, lun, inventory
FREE_LUN, lun
output_file = '~/Desktop/point_analysis_table.txt'
OPENW, out, output_file, /GET_LUN
lag = [4,7,5,6,8,5,6,4,3]
FOR p=1,9 DO BEGIN ; for each plot
   PRINT, 'Plot', p
   lidar_file = '~/Data/Normalized/plot'+STRING(FORMAT='(i2)',p)+' single nadir.txt'
   OPENR, lun, lidar_file, /GET_LUN
   lidar = DBLARR(13,FILE_LINES(lidar_file))
   READF, lun, lidar
   FREE_LUN, lun
   x = lidar[0,*]
   y = lidar[1,*]
   ht = lidar[2,*]
   int = lidar[3,*]
   getis = GETIS(x,y,int,lag[p-1])
   gi = getis[5,*]
   plot_trees = WHERE(inventory[0,*] EQ p, nT)
   tid = inventory[1,plot_trees]
   xt = inventory[2,plot_trees]
   yt = inventory[3,plot_trees]
   tht = inventory[4,plot_trees]
   tcr = inventory[5,plot_trees]
   tcl = inventory[6,plot_trees]
```

```
subscripts = INTARR(nT)
missing = 0
multiple = 0
FOR t=0, nT-1 DO BEGIN ; for each tree in the pth plot
   IF (tcr[t] NE 0) THEN radius=tcr[t] ELSE radius=1
   ; find lidar points within crown
   crown = WHERE(((x-xt[t])^2 + (y-yt[t])^2) LE radius^2)
   IF (crown[0] NE -1) THEN BEGIN
      cx = x[crown]
      cy = y[crown]
      cht = ht[crown]
      cint = int[crown]
      cgi = gi[crown]
      class= tcl[t]
      IF (class GT 0 AND class LE 4) THEN BEGIN
         IF (class LE 2) THEN BEGIN ; dominants and codominants
            sht = MAX(cht,sampled)
            sint = cint[sampled]
            sgi = cgi[sampled]
         ENDIF
         IF (class GT 2) THEN BEGIN ; intermediate and suppressed
            diff = cht-tht[t]
            ht_diff = MIN(diff,sampled)
            sht = cht[sampled]
            sint = cint[sampled]
            sgi = cgi[sampled]
         ENDIF
         sx = cx[sampled]
         sy = cy[sampled]
         sid = WHERE(x EQ sx AND y EQ sy AND ht EQ sht)
         ind = WHERE(subscripts EQ sid)
```

```
            IF ind[0] EQ -1 THEN indie=1 ELSE indie=0
            IF (indie EQ 0) THEN multiple = multiple + 1
            subscripts[t] = sid
            PRINTF, out, p, tid[t], sint, sgi, sht, indie
          ENDIF
        ENDIF ELSE BEGIN
          missing = missing + 1
        ENDELSE
    ENDFOR
    PRINT, missing, ' Missing trees'
    PRINT, multiple, ' Multiple samples'
ENDFOR
FREE_LUN, out
END
```