Swansea University
Prifysgol Abertawe

Cronfa
Setting Research Free

Swansea University E-Theses

Improved estimates of vegetation and terrain parameters from waveform LiDAR.

Mahoney, Craig

How to cite:

Mahoney, Craig (2014) *Improved estimates of vegetation and terrain parameters from waveform LiDAR..* thesis, Swansea University.
http://cronfa.swan.ac.uk/Record/cronfa42490

Use policy:

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence: copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder. Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

Please link to the metadata record in the Swansea University repository, Cronfa (link given in the citation reference above.)

http://www.swansea.ac.uk/library/researchsupport/ris-support/

Swansea University
Prifysgol Abertawe

SUBMITTED TO SWANSEA UNIVERSITY IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Improved Estimates of Vegetation and Terrain Parameters from Waveform LiDAR

Author:
Craig Mahoney (MPhys)

Supervisors:
Dr. Natascha Kljun
Dr. Sietse Los
Prof. Peter North

2014

ProQuest Number: 10801720

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10801720

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Declaration

This work has not been previously accepted in substance for any degree and is not concurrently being submitted in candidature for any other degree.

Signed ....(candidate)

Date .....14/02/2014.....

Statement 1

This thesis is the result of my own independent work, except where otherwise stated. Other sources are acknowledged by in text citations. A bibliography is appended.

Signed ....(candidate)

Date .....14/02/2014.....

Statement 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ....(candidate)

Date .....14/02/2014.....

Summary

Light Detection And Ranging (LiDAR) technologies have evolved rapidly over the last decade, contributing to our knowledge of the Earth's surface evolution from local to global scales. A relatively young form of LiDAR is continuous waveform, which has not yet been fully exploited. The current research investigates and develops new methods, highlighting the potential and possible pitfalls of working with continuous waveform LiDAR.

The first piece of research investigates the effects of shadowing in LiDAR waveforms in physically observed, large footprint LiDAR waveforms, based on previous works noting shadowing effects in radiative transfer models, and in a controlled environment experiment. For this investigation airborne LiDAR derived digital elevation models were employed in conjunction with spatially corresponding waveform returns to identify possible shadowing effects. It was found that shadows occur more frequently over more severely sloped terrain, affecting the accuracy of waveform derived vegetation parameters. The implications of shadows in waveform data are also discussed.

The second piece of research develops and tests two methods, the Slope Screening Model and Independent Slope Model, such to determine ground slope information from LiDAR waveforms. Both methods were validated against discrete return airborne LiDAR data, and British Ordnance Survey data, such to identify which method is most suited to retrieving slope.

The third piece of research utilises the favoured method for slope prediction from the second research topic to correct vegetation height estimates for slope. Two methods (Lee and modified) are investigated and tested, and validated against airborne LiDAR equivalent results at the regional scale, and against normalised difference vegetation index at the near global scale. Both correction methods produced statistically significant differences in mean global vegetation heights with regards to a control dataset.

Contents

Acknowledgements

I am eternally thankful for the help I received from each of my supervisors: Natascha Kljun for her guidance and the opportunity to mold a personal Ph.D though suggesting many possible research paths, Sietse Los for his guidance, continuous support, and invaluable knowledge of working with such data, and Peter North for supporting me through in my turbulent use of his model. Without such academic support this work would not have been possible.

I would like to thank the respective field teams that collected in situ data at each study site used within this project. In particular I wish to thank Laura Chasmer of Wilfrid Laurier University, Ontario, Canada; Eva van Gorsel of CSIRO, Canberra, Australia; and Natascha Kljun, Marianne Hall, and Alex Soudant of Swansea University (at the time of data collection) for their efforts.

ICESat/GLAS data were obtained from the National Snow and Ice Data Center (NSIDC), British Ordnance Survey data were obtained from the EDINA Nataional Data Centre; and the interpolated SRTM-DEM version 4.1 data were obtained from the Consultative Group for International Agriculture Research - Consortium for Spatial Information (CGIAR-CSI). Airborne LiDAR data from the Canadian sites were obtained with support from NERC, from Europe with support from NERC, ARSF, FSF, and GEF; and from Australia with support from NCEO.

I would also like to thank Steve Shaw for his IT knowledge, patience with sometimes persisting problems. I also thank Jamie Williams for persuading me to use the R project for Statistical Computing, the language on which this thesis was built, and taking time to help with issues I may have encountered along the way.

To my office compatriots and friends: Adam, Alex, Eliza, Hazel, Iain, Jamie, Sujung, Will, postgraduates, and staff members at Swansea University for the support offered through laughs, beach BBQs, and all around good times.

I am indebted to my family who have been nothing but proud, supportive, and encouraging from day one. Without this invaluable love and support this work would most certainly not have been possible.

Finally, I am extremely thankful to the Natural Environment Research Council (NERC) for funding this research through the National Centre for Earth Observation (NCEO).

Dedication

I dedicate my Ph.D to my late Grandfather, whom I miss very much. Without his love, guidance, and continuous support of me, no matter the path I chose, I wouldn't be the person I am today.

Acronyms and Abbreviations

AGB	Above-ground Biomass
AGRG	Applied Geomatics Research Group
AMSL	Above Mean Sea Level
AOD	Aerosol Optical Depth
ARA	Airborne Research Australia
ARSF	Airborne Research and Survey Facility
ASPRS	American Society for Photogrammetry and Remote Sensing
ATLAS	Advanced Topographic Laser Altimeter System
BCAL	Boise Center Aerospace Laboratory
BEW	Bare-Earth Waveform
BOREAS	Boreal Ecosystem Atmosphere Study
BRDF	Bidirectional Reflectance Distribution Function
$\mathbf{C_D}$	Crown Diameter
CVM	Canopy Volume Method
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DESDynI	Deformation Ecosystem Structure and Dynamics
DIAL	Differential Absorption LiDAR
DSM	Digital Surface Model
DVGM	Dynamic Global Vegetation Model
EDINA	Edinburgh Data and Information Access
EGM	Earth Gravitational Model
EM	Electromagnetic
EOS	Earth Observing Satellites
ESDS	Earth Science Decadal Survey
ESSP	Earth System Science Pathfinder
$\mathbf{F_{cov}}$	Fractional Cover
FAO	Food and Agriculture Organization
FLIGHT	Forest Light Interaction
fPAR	Fraction of Photosynthetically Active Radiation

FWHM	Full Width at Half Maximum
GCM	Global Circulation Model
GEF	Geophysical Equipment Facility
GFRA	Global Forest Resources Assessment
GLAS	Geoscience Laser Altimeter System
GPS	Global Positioning System
GSFC	Goddard Space Flight Center
ICESat	Ice, Cloud, and Land Elevation Satellite
ICOS	Integrated Carbon Observation System
IMU	Inertial Measurement Unit
InSAR	Interferometric Synthetic Aperture Radar
IPCC	Intergovernmental Panel on Climate Change
ISM	Independent Slope Model
JULES	Joint UK Land Environment Simulator
LAI	Leaf Area Index
LAS	Log-Ascii Standard
Laser	Light Amplification by Stimulated Emission of Radiation
LiDAR	Light Detection And Ranging
LIST	LiDAR Surface Topography
LSM	Land Surface Model
LULUCF	Land Use, Land Use Change, and Forestry
LUT	Look-Up Table
LVIS	LiDAR Vegetation Imaging Sensor
MABEL	Multiple Altimeter Beam Experimental LiDAR
Maser	Microwave Amplification by Stimulated Emission of Radiation
MODIS	Moderate Resolution Imaging Spectroradiometer
MOLA	Mars Observed Laser Altimeter
NAP	National Academic Press
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index
NERC	Natural Environment Research Council
NRC	National Research Council
NSIDC	National Snow and Ice Data Center

OS	Ordnance Survey
RMSE	Root Mean Squared Error
RF	Random Forest
SDT	Science Definition Team
SIMPL	Slope Imaging Multi-Polarization Photon-Counting LiDAR
SLA	Shuttle Laser Altimeter
SLICER	Scanning LiDAR Imager of Canopies by Echo Returns
SNR	Signal to Noise Ratio
SOA	Southern Old Aspen
SOBS	Southern Old Black Spruce
SOJP	Southern Old Jack Pine
SONAR	Sound Navigation and Ranging
SRTM	Shuttle Radar Topography Mission
SSM	Slope Screening Model
SW	Shadow Waveform
TRIFFID	Top-down Representation of Interactive Foliage and Flora Including Dynamics
UNFCCC	United Nations Framework Convention on Climate Change
USFS	United States Forest Service
VCF	Vegetation Cover Fraction
VCL	Vegetation Canopy LiDAR

List of Tables

List of Figures

Chapter 1

Introduction and Research Context

Remote sensing technologies have given the scientific community a great wealth of knowledge on the environment. Particularly in areas pertaining to the storage of carbon in biomass, and the carbon and water cycles. A relatively new development in remote sensing, Light Detection And Ranging (LiDAR) offers a source of information and the potential to measure key parameters, such as vegetation height and biomass, contributing to the understanding of global environmental processes that occur in the carbon and water cycles.
In its short exposure, LiDAR has become well established and still supports potential for methodological developments in order to explore and pioneer methods to extract terrestrial and vegetation parameters. The aims of this thesis are therefore to identify and develop methods such to infer important terrestrial and vegetation parameters. In this chapter key terrestrial and vegetation parameters are discussed with regards to the current research environment and context.

1.1 Research Environment

Vegetation height, timber volume and biomass are commonly used parameters for the assessment of forest productivity and carbon sequestration rates (Allouis et al., 2012). Accurate measurements of these parameters are paramount for defining ef-

fective strategies for sustainable forest management and climate mitigation.
This project addresses these needs through exploring the potential of an innovative means of reducing uncertainties in remotely sensed vegetation products on a regional to global scale, particularly vegetation height estimates, key for inferring above ground biomass (Allouis et al., 2012; Lefsky et al., 2005) and timber volume (Nelson et al., 2009). This work helps address research objectives identified by the UK Natural Environment Research Council (NERC) such as understanding the environmental implications of CO_2 storage, and the response of hydrological and biogeochemical processes to changing climate and land use (NERC, 2013). In addition to NERC, respective nations within the Kyoto protocol have goals to meet regarding carbon emissions, to which managing forest stocks is key (UNFCCC, 1999).
The current chapter (in conjunction with Chapter 2) outlines the requirement of improved vegetation biophysical parameter knowledge in the context of improving environmental process understanding, and the role of LiDAR technologies in understanding these processes. To date contributions of LiDAR in this field are identified, and current limitations discussed.
These points provide the motivation for the identified research questions, and aims and objectives. The section concludes with a brief description of the thesis structure, generated in order to meet these objectives.

1.2 Research Context

The understanding of climate change and the development of strategies for sustainable use of the Earth's natural environmental resources are major scientific and political challenges of the current world. Rising levels of atmospheric CO_2 are believed to contribute to the change of the global climate (IPCC, 2007). This has lead to international negotiations regarding carbon emissions such as the 1997 Kyoto Protocol, pioneered by the United Nations Framework Convention on Climate Change (UNFCCC, 1999), where member states are required to report annually on their unnaturally produced carbon emissions and their efforts towards the removal of atmospheric carbon dioxide (CO_2).
In order to accurately estimate such emissions, prior knowledge of the existing

carbon distribution and sinks are required. This distribution is currently largely unknown, although there is a widespread consensus that forests constitute the largest of the terrestrial carbon sinks located in the Northern hemisphere (Betts, 2000; Chambers et al., 2001; Pacala et al., 2001; Six et al., 2002). Forests are recognised as playing a significant part in mitigating climate change, a statement underpinned by global initiatives and policies put in place, such as: Reducing Emissions from Deforestation and Forest Degradation in Developing countries (REDD), established by multiple agencies of the United Nations (FAO et al., 2008; UNEP, 2009), and reporting on Land Use, Land Use Change and Forestry (LULUCF) establised by the UNFCCC (IPCC, 2003). Forests are significant in this regard due to the role of vegetation in carbon budgets, which is sensitive to disease, wildfires, drought, flooding and/or (perhaps more importantly) anthropogenic actions such as: felling, pollution, land use changes, afforestation or increased CO_2, which in turn can induce climatic changes and impact ecosystem processes.

Remote sensing technology offers a means of quantifying vegetation carbon stocks, however, measuring the Earth's terrestrial landscape is challenging. Its complex characteristics are a result of great variability in surface coverage (ice, water, vegetated/unvegetated), elevation, slope, roughness, reflectance, vegetation height and structure, and varying soil/water cover. The identification, quantification, and characterisation of these surface parameters is essential to reduce uncertainty in important ecological parameters, thus promoting our understanding of carbon and water-based processes.

Such parameters (surface and by extension ecological) are useful as boundary conditions in biosphere models for predictive purposes. Thus such quantifications will allow well informed decisions to be made with regards to managing and preserving our natural environment, whilst still being able to provide natural resources to underpin a growing global population (Peng, 2000).

The contribution of remotely sensed data towards quantifying surface parameters has been invaluable, allowing the acquisition of a multitude of data at large scales. Passive (e.g. aerial imagery) and active (e.g. LiDAR) instruments have been employed by the scientific community to great effect (Hill and Thomson, 2005; Suárez et al., 2005; Rosette et al., 2008). Laser altimetry (or LiDAR) in particular, has already demonstrated its capabilities in this field, providing estimates of numer-

ous biophysical parameters from local to near global extents (Rosette et al., 2008; Nelson et al., 2009; Lefsky, 2010; Simard et al., 2011; Los et al., 2012). Further exploration of this not fully utilised technology will contribute to a more accurate representation of terrestrial parameters.

This study explores some subtleties inherent in LiDAR data that induce uncertainty in derived terrestrial parameters. Particular attention is paid to the possibility of retrieving slope information, correcting vegetation height as a function of slope, and the impact of overlooked subtleties such as shadowing (defined in Section 1.3), in continuous waveform LiDAR. LiDAR is an exciting prospect as at present waveform LiDAR is the only technique offering direct comprehensive global measurements of the Earth's surface in the vertical domain. A more comprehensive overview of this study is given from Section 1.3 to 1.5.

1.2.1 Vegetation Monitoring

Above-ground forest carbon is not directly measurable by remote sensing technologies due to physiological subtleties between vegetation. However, other key parameters, such as vegetation height, can be used to quantify carbon stocks. Remotely sensed data have already shown strengths in indirectly mapping carbon stocks in above ground vegetation (Means et al., 1999; Lefsky et al., 1999b, 2005; Saatchi et al., 2007, 2011). Additionally these technologies offer opportunities for observing vegetation regeneration, and furthering knowledge with regards to carbon flux interactions between the land, atmosphere and oceans on a spatio-temporal basis (Hese et al., 2005). Remote sensing data have been found to be so effective that optical and radar technologies are currently employed with the purpose of mapping the global change in forest extent, as recommended from the Global Forest Resources Assessment (GFRA) 2010 (Ridder, 2007).

For vegetation mapping, both passive systems (Sellers et al., 1996; Los et al., 2000; Hansen et al., 2010) and active systems (Neuenschwander et al., 2008; Nelson et al., 2009) have been employed. Passive sensors harness naturally available energy from the sun, measuring reflected energy from a target. For example, aerial photography requires the visible portion of natural energy to build up a visible image, where spectroradiometers employ visible and non-visible (such as infrared) radia-

tion. The latter of which allows measurements to be made at night due to natural non-visible emission (not reflection). Passive instruments sense only radiation reflected or emitted by the viewed target, hence for vegetation, under layers are inaccessible. In contrast, active sensors illuminate a target with their own energy/radiation, detecting the returned radiation that is reflected or backscattered. For vegetation imaging, LiDAR, an active sensor, allows the direct retrieval of multiple canopy layer structure, not just from the canopy top. This is achieved by reflections and backscatter interactions with intercepted surfaces. The benefit of using LiDAR for such purposes is that the illumination wavelength can be chosen specifically, which is not possible for natural radiation, allowing specific targeted illumination.

The use of LiDAR technology in vegetation imaging has proved useful in retrieving information unobtainable from passive sensors, such as vegetation structure and topographic relief profiles (Rosette et al., 2008, 2010).

1.2.2 Forest Management

Forest management is a key interest in political and scientific decision making processes; as mentioned previously, international efforts have been made in order to quantify carbon stocks through forestry. A focus of this project is to develop new techniques that will indirectly provide more accurate estimates of such stocks and thus aid this decision making process.

In the United Kingdom, the current National Forest Inventory relies on large quantities of field measurements and photographic interpretation, carried out manually to clearly identify forest and non-forested areas (Forestry Commission, 2012a); such processes are costly and time consuming. Such financial and time cost issues may be mitigated with the use of remote sensing data, reducing the requirement for such numerous field campaigns.

Remote sensing, and particularly LiDAR, both airborne surveys (impractical globally), and satellite, provide a wealth of information for vegetation parameter retrieval within forest biomes. However, the current research shows refinement and improvements can be made to associated estimates, thus reducing uncertainty in measurements, aiding refinement in the future of forest management globally.

1.2.3 Modelling Impact

There are numerous computer generated models within the Earth science community, such as: biosphere, hydrological, atmospheric, and general circulation models (GCMs). Vegetation parameters are key inputs in many of these models, particularly those focused towards biogeochemical cycles, even more so those concerning vegetation carbon budgets.

For such models to accurately represent key land surface processes, information regarding plant physiology, surface and soil characteristics, climate characteristics, and radiative interactions are required at varying spatial (and temporal in some cases) resolutions. These parameters provide boundary conditions for models, and more accurate outputs are obtained (provided the internal mechanisms are correct), when less uncertainty is present in inputs. Such variable datasets (over larger extents) are only possible to obtain from remotely sensed data (Dickinson et al., 1981; Dickinson, 1984; Sellers et al., 1986; Myneni et al., 1997).

The popularity of employing models for predictive purposes increased during the mid 1980's with the work of Dickinson et al. (1981) and of Sellers et al. (1986) on global climate modelling through Dynamic Global Vegetation Models (DGVMs) or Land Surface Models (LSMs). Since then, biosphere models have progressed with improved understanding of environmental processes, and rely on accurate input parameters.

Biosphere Models

Biosphere models are a valuable means to improve our understanding of the complex interactions within interlinked Earth systems. By definition the models function as generalisations of reality based on mathematical equations that govern interactions and physical processes that are well understood. Typically, more verbose models are based on a parent-child structure, where component, theme specific models are integrated to incorporate feedback mechanisms, yielding the broader-themed convolution of functions that are present in reality.

The equations upon which such models are built require initial inputs, boundary conditions, and occasionally (process specific) driving information, these are based on physical spatio-temporal data, required to constrain inter-system processes and

govern outputs. Such data requirements are met by a database of in situ field measurements, but remote sensing data are becoming more and more widely utilised for such situations.

Focusing on carbon, DGVMs and LSMs are particularly well suited to predicting change in an ecosystem's carbon balance as a legacy of changing structure and composition as a function of climate change (de Noblet-Ducoudré et al., 1996, 2000; Sitch et al., 2003; Hese et al., 2005). Vegetation class is represented within defined grids cells as generalised functional types, where habitat changes are induced as a function of climate change in order to model/predict successive vegetation states and plant life cycles over specific time frames.

A number of DGVMs and LSMs have been used for predictive purposes (Dickinson et al., 1981; Sellers et al., 1986; Foley et al., 1996; Brovkin et al., 1997; Sitch et al., 2003; Sato et al., 2007; Friend et al., 1997; Woodward et al., 1998). The next generation of Earth system models now include important feedbacks from the biosphere to the atmosphere as a function of changes in vegetation, and thus changes in carbon and hydrological cycles, which are sensitive to climate shifts. For example the Joint UK Land Environment Simulator (JULES) model, employs the Top-down Representation of Interactive Foliage and Flora Including Dynamics (TRIFFID) DGVM to provide vegetation inputs (JULES, 2013).

Remote sensing technologies have provided model input parameters relating to vegetation, and light characteristics by a convenient and non-destructive manner. Vegetation height, fractional cover (F_{cov}), leaf area index (LAI), and normalised difference vegetation index (NDVI) have been derived from such technology (Hopkinson and Chasmer, 2007; Rosette et al., 2008; Hopkinson and Chasmer, 2009; Los et al., 2012). These products are ideal for use in DGVMs and LSMs, for validating model outputs, and as driver data.

Improving the accuracy of such parameters is key in better understanding the role of vegetation in interlinked Earth processes, to which LiDAR is ideally suited. The application of LiDAR for mapping vegetation allows a unique, reliable way to derive important vegetation parameters such as height and structure, which are important for accurate estimates of F_{cov} and LAI. Additionally, the ease with which current LiDAR data can be collected allows DGVMs and LSMs to be executed using up to date input parameter information.

Radiative Transfer Models

Radiative transfer models can be used simulate remotely sensed data allowing a better understanding of the effect of radiation scattering and absorption (especially over vegetated surfaces) on incident radiation. Such models map incident photons from a specific irradiating geometry through a probabilistic scattering path, back to a detector defined at a final geometry; examples of such models are the Forest Light interaction (FLIGHT) model and the Monte-Carlo Ray Tracer developed by North (1996) and Lewis (1999), respectively.

Vegetation structure and associated embedded parameters (LAI, F_{cov}, and FPAR) are important determinants of vegetation productivity. Structural variations within and between canopies can cause multiple scattering events for incident photons, necessitating a three dimensional approach to reflectance simulations. Waveform LiDAR has provided a means of comparing modelled and observed scattering effects, and offered a means of validation of current Bidirectional Reflectance Distribution Functions (BRDF), on which reflectance simulations are based (North et al., 2010).

Conversely, radiative transfer modelling may offer an understanding of incident light interactions with complex vegetated terrain, a current day challenge still not fully understood. Modelling such events can act as "pathfinder" investigations, leading to a better understanding of waveform characteristics by exploring the sensitivity of returned waveforms as a function of vegetation and terrain structural and reflective properties (Ni-Meister et al., 2001; North et al., 2010). At present, inferring vegetation and terrain parameters from physical waveform structures through Look-up Tables (LUTs) based on multiple parameter iterations of FLIGHT runs are being pursued (personal communication, Bye, 2013).

Radiative transfer models offer an idealised situation for radiation scattering (no atmospheric presence) in comparison to observed data. The behaviour of methodological/model outcomes in this somewhat idealised situation allows inferences to be made when such methods are applied to data observations. This is particularly pertinent in the current research where FLIGHT is employed to assess new methods developed for use with LiDAR waveforms.

1.3 Research Questions

Three research questions are addressed throughout this study, each making use of a multitude of collected data from satellite, airborne, and ground based measurements.

In **Chapter 4** it is investigated: **Do shadows occur in large footprint waveform LiDAR returns? If so, what are their cause? And what are the implications of their presence?** Based on the works of Hancock et al. (2012) and Wallace et al. (2012), an investigation into detecting shadows in large footprint ($\geq$ 30 m diameter) LiDAR returns is undertaken. Note, shadows (to which the following is referred to from here on) are defined as partial laser signal attenuation (rather than complete occlusion), suspected to originate from vegetation components obstructing incident laser light. Such an effect is capable of misrepresenting physical features detected by the waveform, potentially leading to spurious derivations of vegetation metrics. The origin and implications of this effect are also investigated, making use of airborne LiDAR data and associated derived data products.

In **Chapter 5** it is investigated: **Is it feasible to accurately retrieve slope angle information directly from LiDAR waveforms?** The development of new techniques are pursued in order to retrieve slope angle information which will allow better understanding of how full waveform LiDAR perceives the ground in complex situations. Understanding complex interactions is important as these have potential to induce spurious derivations of vegetation metrics (Los et al., 2012; Simard et al., 2011; Lefsky, 2010; Chen, 2010b; Duncanson et al., 2010; Nelson et al., 2009; Rosette et al., 2008).

In **Chapter 6** it is investigated: **What are the effects of correcting waveform LiDAR derived vegetation height estimates as a function of slope angle?** This incorporates the findings from the previous research questions (more so from Chapter 5), where the spurious vegetation heights are re-evaluated with corrections applied as a function of slope using the method described by Lee et al. (2011). The sole use of waveform LiDAR derived results, to correct vegetation height metrics will emphasise the potential that LiDAR still offers with regards to vegetation imaging, particularly in mitigating slope as an uncertainty inducing

parameter.
These questions will be asked at the forefront of knowledge, encouraging the gathering and use of high-resolution remote sensing data which will allow more accurate results to be derived for use within the scientific community. Reducing uncertainty in vegetation metrics is important in providing refined quantitative estimates of biomass, timber volume and carbon stocks, in conjunction with the objectives of multiple, global scientific bodies.

1.4 Aims and Objectives

The overall aim of this research is to assess the potential of improving vegetation height estimates, and improving the knowledge of waveform LiDAR interactions with targets, thus allowing refinement of additional parameters derived from this parent parameter (see Chapter 2), from well established waveform LiDAR technology. Uncertainty reductions are examined in alignment with stated research questions, and with respect to local terrain and instrument specifications (where applicable).
The following four objectives have been identified in order to achieve the outlined project aims:

1. Gather field data in conjunction with secondary, local to regional scale datasets for the purpose of validating novel methods and associated results.

2. Hypothesise and develop novel methods, and/or employ established methods (were applicable) in order to refine, and identify, new environmental parameters from waveform LiDAR.

3. Highlight issues associated with retrieving inaccurate vegetation height estimates from waveform LiDAR with respect to reducing uncertainties in metrics derived from such data.

4. Apply, and assess the impact of vegetation height estimate corrections derived from waveform LiDAR.

This research seeks to develop methods that allow refinements to be made to non-specific localities, where possible, with regards to any particular vegetation

size, type, density, and/or terrain relief. These outlined aims and objectives are explored within the following chapters (see Section 1.5).

1.5 Thesis Structure

Available literature is reviewed in **Chapter 2**, highlighting the previous usage of LiDAR remote sensing technologies to quantify parameters important in contributing to our knowledge of interlinked Earth systems. A general overview of LiDAR uses, and gaps in knowledge, is given.

Chapter 3 discusses the remote sensing data used throughout the thesis. The study sites used for research purposes, and associated field work at each site are discussed. The principles and processes of the Geoscience Laser Altimeter System (GLAS) waveform sensor are discussed along with the equivalent for airborne LiDAR. Validation methods and data processing methods, applicable to the airborne, satellite and ground based measurements (throughout the thesis) are discussed. The methods are known as "universal processing methods" in subsequent chapters.

Chapter 4 discusses the quantification of the effect of shadowing in large scale waveform LiDAR returns such as GLAS (over sloped terrain). Airborne LiDAR data are employed to demonstrate the origin of the effect. The consequences of ignoring shadows in waveforms are discussed in the context of vegetation imaging. In addition, the potential of deriving vegetation parameters such as crown diameter is highlighted; previously unobtainable from large footprint waveform LiDAR.

Chapter 5 discusses the development and testing of two novel methods for extracting slope from full waveform LiDAR returns, independently of external data. Initially the hypotheses behind each technique were tested using the FLIGHT radiative transfer model, and later applied to GLAS waveforms which were validated against corresponding airborne LiDAR derived slopes. Slope predictions were made using both developed techniques for GLAS waveforms within the British Principality of Wales and validated against British Ordnance Survey (OS) derived slope; the better of the two developed slope methods is selected for use in Chapter 6.

Chapter 6 employs the findings of Chapter 5 on a regional to near global scale,

allowing the exploration and development of two methods for vegetation height correction, with the goal of obtaining more representative estimates of vegetation height from waveform LiDAR. Vegetation height corrections are applied as a function of slope on a regional and near global scale. The finer points of the slope correction methods are discussed at both scales.

Chapter 7 discusses the findings of each respective research chapter, highlighting positive and negative results. Results are synthesised and the study's contribution to the scientific community is discussed. Furthermore, prospects for each individual result are discussed, along with future research directions that may be pursued to further improve vegetation parameter estimates.

Finally the project conclusions are drawn in **Chapter 8**, highlighting the answers to the overall research questions, aims, and objectives.

Chapter 2

LiDAR Technologies: A Review

LiDAR technologies have provided valuable means for recording 3-dimensional information of the Earth's surface; such as topography, and surface structures, particularly vegetation, for many years to date. This chapter offers a brief review of these technologies, some of only a few ways to record 3-dimensional structure, and how it links with passive, optical systems. A particular interest is paid to discrete multiple return LiDAR and continuous full waveform LiDAR. The differences between systems and their advantages/disadvantages are highlighted. In addition, an overview of the capabilities and applications of different systems is provided.

2.1 History of LiDAR

The oldest known natural version of modern LiDAR evolved millions of years ago, this is the guidance system adopted by bats now understood to be Sound Navigation and Ranging (SONAR). Bats emit short 'bursts' of sound as they travel and receive the echo through their ears, this provides them with a 3-dimensional view of their surroundings, allowing them to navigate around obstacles in darkness. Modern LiDAR systems are based on identical principles to these (discussed in Section 2.2).

The predecessor of LiDAR was Radio Detection and Ranging (RADAR), which is still used today, first developed by German inventor Christian Hülsmeyer in 1904 (Van Loon, 2005). Following this, Albert Einstein laid the foundations of

the laser and maser (acronym's for Light/Microwave Amplification by Stimulated Emission of Radiation) in 1917 (Einstein, 1917). It was not until 1953 that American Physicist Charles Hard Townes developed the first microwave laser (MASER), the details of which he published later in 1958, along with the possibilities of laser exploration in the visible and infra-red spectrum (Schawlow and Townes, 1958). Shortly afterwards, fellow American Theodore Harold Maiman demonstrated the first functioning laser in 1960 employing a ruby crystal as an excitation medium (Maiman, 1960).

Airborne laser ranging was attempted the same year, however, it was not until the 1970s that laser ranging technology became better understood. By the early 1980s LiDAR studies had become so well known, the acronym 'LiDAR' was commonly recognised. During the 1990s the first commercial LiDAR systems became available, making LiDAR data available to industry, where many companies acquire LiDAR campaigns for scientific research and non-scientific applications today. Today LiDAR data are used for a multitude of research and survey purposes, such as forest inventory and flood plain mapping. Such data have been deemed so important that country-wide acquisitions have been pursued by many governments. For example, the Environment Agency's Geomatics Group are currently pursuing this data acquisition for the entire United Kingdom (Environment Agency, 2013).

In the last 20 years, laser altimetry has been applied to planetary sciences, with high impact. Spaceborne laser ranging systems on the jointly supported Clementine mission (1994) by NASA and the Strategic Defense Initiative Organization, produced large-scale topographic images of the Earth's moon between $\pm 60°$ latitude (NASA, 1994). Additionally, the Mars Orbital Laser Altimeter operated in orbit around Mars for 9 years from 1996, as part of the Mars Global Surveyor mission, which produced global topographic maps of the planet's surface. The instrument also noted surface change in some areas, suggested to be caused by the (occasional) flow of water across the Martian surface (NASA, 1996).

Focusing this technology on our home planet, waveform laser ranging has recently become more popular and proved its uses across multiple disciplines. The exploration of spaceborne laser ranging (or LiDAR) data has increased in previous years and shown its ability to enhance our knowledge of the three dimensional distribution of the Earth's surface features and related parameters (Anderson et al., 2006;

Dubayah et al., 2010; Nelson et al., 2009; Los et al., 2012).
The understanding of the methodological processes pioneered in deriving parameters from such systems have also improved, thus yielding queries over the accuracy of some derived parameters (Hancock et al., 2012; Hopkinson, 2013). The exploration and improvement in the understanding of such methods may offer great potential to enhance terrestrial parameter quantifications which, in turn will complement our understanding of the Earth's complex systems. Such investigations of previous or existing data also allow the fine tuning of future missions (e.g. ICESat for ICESat-2; see Section 2.7.3).

2.2 Principles of LiDAR

All LiDAR systems are based on lasers, which are defined in this case as emitting electromagnetic (EM) radiation in the visible and infrared parts of the spectrum; in this case we define this spectral range as 'light'. There are two main classes of laser, known as pulsed and continuous wave (Wehr and Lohr, 1999), the former is of greater interest in modern LiDAR systems. A pulsed laser emits a short finite pulse (or 'burst') of high intensity light towards a target surface. The backscattered (or reflected) photons from said target, whether it be atmospheric or terrestrial, can be used to infer properties of the intercepted surface.
The principles of LiDAR are almost identical to those of radar, which was initially pioneered for range determination (Brenner et al., 2003), although radar by definition makes use of radio waves (from 0.5 m), not 'light' (between 400 nm–2500 nm) as defined above. Feature detection by EM radiation is a function of wavelength (λ), more explicitly, the longitudinal resolution to which EM radiation can detect down to is dependent on its wavelength. Therefore, as a consequence, the far greater wavelengths of radar are incapable of measuring parameters such as atmospheric aerosols (Molero and Jaque, 1999). However, this relatively long wavelength enables surface measurements even in cloudy conditions. This is not possible for the smaller wavelengths of LiDAR, which are capable of measuring down to the hundred-nanometer (nm) scale, including cloud particulates and other atmospheric aerosols.
The use of LiDAR to infer range measurements was founded on 20 years experi-

ence of employing radar for the same purpose (Brenner et al., 2003). As mentioned the principles are almost identical, but due to the differences in the emitted EM radiation of each system, results can vary drastically. For instance, the greater footprint dimensions of a radar system (with respect to a LiDAR system) means that smaller features such as ground roughness are overlooked, and large scale features such as slope dominate. The smaller footprint (typically) produced by LiDAR means that it is the preferred technology over radar in vegetation imaging. This small footprint is a function of shorter wavelengths allowing the retrieval of parameters from higher biomass (denser) canopy structures (Waring et al., 1995; Drake et al., 2003). Consequentially, a smaller footprint allows the identification of terrestrial features with greater precision, hence LiDAR technology would be favored for attempting to account for topographic features. To date attempts to account for topographic subtleties using LiDAR have been documented, but are largely area specific (Lefsky et al., 2007; Lefsky, 2010; Chen, 2010b; Lee et al., 2011).
The characteristics and applications of LiDAR systems are discussed later in this chapter.

2.3 Instrument Types

There are four main types of LiDAR instruments available today, these are: Doppler LiDAR, Bathymetric LiDAR, Differential Absorption LiDAR (DIAL), and Intensity Measurement LiDAR. Each are briefly discussed in their corresponding sections.

Doppler LiDAR

Doppler LiDAR instruments are capable of measuring and mapping atmospheric particle velocities and reflected backscatter with high precision. The velocity of a target is measured by the Doppler effect (or shift) proposed by Christian Doppler in 1842, which states that a change in frequency of an EM wave (or any periodic event) is observable for a target in motion relative to its detector (Toman, 1984). Assuming a stationary instrument (emitter and detector), a return signal's wave-

length is shorter or longer, dependent on whether the target is advancing towards or receding from the instrument. Such systems are typically used to remotely measure wind velocity via a medium of suspended dust and/or aerosol particles in transit.

Bathymetric LiDAR

Bathymetric, or LiDAR for hydrography is used to measure the depth of coastal waters down to approximately 50 m, as well as shoreline topography itself. Such a system fires co-aligned laser pulses, one infrared (IR), the other occupying the blue-green part of the visible EM spectrum. The IR pulse is reflected by the water's surface whilst the visible pulse penetrates to the bed where it is reflected. The time difference between the two returns allows the derivation of the water depth. These systems can also be used for surveying large areas in search of submerged objects of interest.

Differential Absorption LiDAR

Differential Absorption LiDAR, or DIAL, is used in order to make range-resolved concentration measurements of atmospheric constituents, such as: ozone, water vapour, and aerosols. DIAL makes use of two differing wavelengths, usually in adjacent wavebands, selected such that one is absorbed and the other is scattered by the target component. The component concentration is resolved by calculating the intensity difference between the two returned signals (NASA, 2006).

Intensity Measurements

Intensity measurement systems are simple range finding LiDAR, these are the simplest instrument group available on today's markets. This instrument simply measures the distance (or range) between itself and some target surface(s). This instrument is the system of interest for this study. Under this group fall: space-borne, airborne, and terrestrial LiDAR approaches, each successively producing data with increasing spatial resolution, ranging from a footprint encompassing several canopies down to leaf level.

2.4 LiDAR Range Determination

The determination of range from a LiDAR instrument is based on simple physics, namely a time-distance equation (Equation 2.1).

$$R = \frac{c \cdot t}{2} \tag{2.1}$$

where:

R = Range to target [m]
t = Time between pulse emission and detection [s]
c = Speed of light [ms^{-1}] ($\simeq 3 \times 10^8$)

In conjunction with Figure 2.1, consider an emitted laser pulse from an airborne LiDAR instrument, where the emission (Em) and detection (D) positions are located within very close proximity, travelling towards a target (T, e.g. ground or vegetation etc.) at known velocity (c; speed of EM radiation). In the time (t) between pulse emission and (re)detection, the pulse has travelled from Em to T, and reflected back to D; this requires that in finding the range (R) between the altimeter and the target, this two-way travel is accounted for by halving the time of travel (Wehr and Lohr, 1999).

The vertical range resolution (ΔR) of any altimeter system is directly proportional to its time resolution (Δt_r). Mathematically, ΔR is derived using the same form as Equation 2.1, where R and t are replaced by ΔR and Δt_r, respectively (Wehr and Lohr, 1999).

In order to derive range, Δt needs to be accurate; Δt is defined as the difference between the initial pulse emission (t_i) and final detection (t_f) times. A time counter is triggered upon pulse emission, recording t_i, and then re-triggered upon encountering some minimum intensity threshold on the return pulse's leading edge (rising side of the pulse), recording t_f (Baltsavias, 1999). The gradient (or rise time) of the returned pulse is a leading contributor to range calculation accuracy. It is dependent on the combination of numerous factors such as: emitted pulse wavelength, target optical properties at that specific wavelength, laser energy spatial distribution, and atmospheric interference (Baltsavias, 1999; Wehr and Lohr, 1999).

Assuming a static wavelength and laser energy distribution for a given instrument,

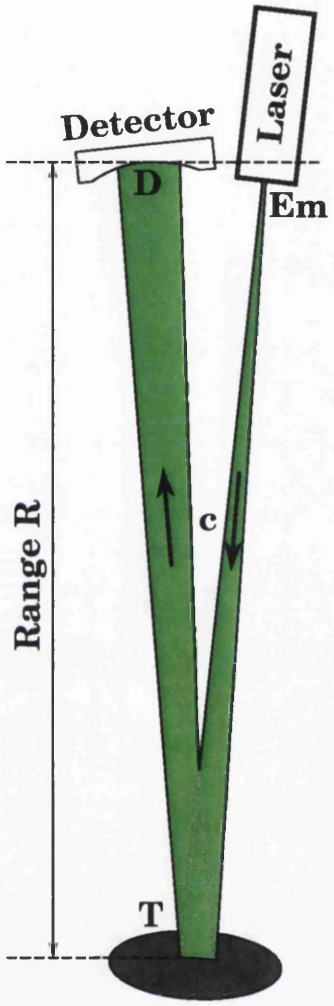


Figure 2.1: Depiction of the determination of range from a LiDAR perspective. Pulse is emitted from the laser (Em) towards targets (T) with a known velocity (c), and reflected towards detector (D). The total distance travelled is twice the range (R), this is the summation of the distance from the laser to the target and from the target to the detector. The angle between beams is typically small, but is dependent on the laser position and orientation when emitting a pulse.

the accuracy of the range is determined by the intercepted target structure and optical properties. For instance, when employed for Earth observation purposes, targeting flat ice sheets with high reflectivity will produce a sharp narrow return with a fast leading edge rise time, whereas a multi-layered vegetated structure over complex terrain will create broader, less consistent returns with a slower leading edge rise time (Harding et al., 1998; Ni-Meister et al., 2001).

The recorded time for an emitted pulse to be (re)detected is typically in the order of nanoseconds (ns), due to the exceptionally fast transit velocity. From Equation 2.1, a single nanosecond of travel equates to an emitted pulse having been displaced by 0.15 m, accounting for two-way distance. Such principles allow the derivation of the range at which objects are intercepted by the laser beam from the altimeter. This allows 3-dimensional information of the target surface to be observed.

Laser pulses are typically fired sequentially in order to avoid multiple returns arriving at the detector simultaneously. Such an occurrence could result in mixed return signals, containing information from multiple targets, and thus yield unusable data.

Laser emissions suffer a beam-width divergence effect whilst in transit, and as a result the beam suffers diminishing energy with distance (Svelto, 2010). However, this also means the target is not illuminated by an infinitesimally small spot, but by a radially measurable footprint. Hence measurements are made of 3-dimensional areas of a target, not singular points. Another consequence of such conditions is that targets closer to the instrument will be more intensely reflected than those at a greater distance.

2.5 Detection Criteria

As with any instrument, there are a set of criteria that must be met for any meaningful data to be recorded. A brief outline of these criteria in the case of laser altimeter systems is discussed in this section.

Detection Restrictions

The minimum detectable object within a laser footprint is dependent, primarily on its reflective properties, and secondarily on its physical size. For example, for a sensor capable of measuring the range to a flat even surface of area A, with a reflectivity $\rho = 0.05$ (5%), then the minimum area of a detectable object where $\rho = 1$ would be equal to A/20 (Rosette, 2009). There are numerous other contributory factors to inaccurate range capture, these can originate from instrument specifications such as: laser power, wavelength, aperture size, detector sensitivity, and background noise level. Additionally, inaccurate range capture can also arise from the target's physical properties and natural surroundings such as: target's optical properties at specific wavelengths, 3-dimensional structure, atmospheric conditions (see below), background irradiation, and terrain (Baltsavias, 1999).

The minimum height at which laser altimetry data can be acquired is typically determined by instrument specifications, safety aspects, and local authority regulations. Typically the most sensitive detectors are available between 800–1000 nm wavelengths, however, at these wavelengths, eye safety is a concern. Therefore if a relatively high laser energy is required, safety aspects need to be addressed, such as exposure time (Rosette, 2009).

With increasing instrument altitude comes increasing footprint dimensions (see Section 2.4). In addition, assuming an unchanged pulse repetition rate, a reduction in the spatial density of footprints occurs (less sampled areas per unit area). This occurs as the travel time between laser pulse emission and interaction with a target increases, thus increasing inter-footprint separation and hence reducing footprint spatial density. Typically, with lower altitudes more samples per unit area are observed when pulse repetition rate remains constant.

Focusing on vegetated surfaces, smaller footprints typically populate a surveyed area with greater density, but can underestimate true canopy height. This is due to a lower probability of encountering the canopy top within a smaller footprint radius (Lim et al., 2003; Zimble et al., 2003; Suárez et al., 2005). However, this problem can be somewhat mitigated in data post-processing by resampling points (by interpolation) at an accurate *tree spaced* resolution; this reduces the chance of misinterpreting the side of a tree canopy as being the canopy top. Larger footprints

tend to yield longer leading edge rise times (see Section 2.4) due to the percentage of the contribution from the canopy top to the total waveform return being small, thus missing the true beginning of signal (Lefsky et al., 2007). This occurs as the cross-sectional area of the canopy top is very small, hence reflections from this part of the canopy are often small enough as not to cause reflected intensities to exceed instrument detection thresholds.

The dimensions of a LiDAR footprint can be optimised to suit its outlined purpose (such as vegetation imaging; Rosette et al., 2013). For instance, large laser footprints ($\geq$ 30 m diameter), with an average circular diameter (F_d), pose difficulties when $F_d > C_w$ (crown width). In this instance, upper canopy variability can be under-estimated (Lefsky et al., 2007), and topographic variation can be misrepresented (Hofton et al., 2002; Chen, 2010a). Medium footprints ($5 - 30$ m diameter) can be advantageous over smaller footprints if their dimensions are similar ($F_d \leq C_w$). In this scenario the majority of the details typical of small footprints ($\leq$ 5 m diameter) are captured due to the increased probability of sampling the canopy top directly. Details are also captured due to the incident laser pulse having a sufficient energy to penetrate the canopy sufficiently to measure the ground surface. For larger footprints, sloped terrain is suggested to be the leading driver of misinterpreting estimates of vegetation height (Chen, 2010a; Simard et al., 2011). The current research explores this claim and the possibility of correcting such large footprint height estimates over sloped terrain.

Acquisition Purpose

The choice of wavelength selected for a specific LiDAR approach is heavily dependent on the purpose of the approach itself. For use in acquiring ice sheet elevations, wavelengths within the visible (VIS) range of the EM spectrum are preferred over those in the near infrared (NIR) range, as ice is known to reflect weakly at approximately 1535 nm (Wehr and Lohr, 1999). In another example, a smaller wavelength is better suited to measure aerosol particle size than a larger wavelength, as the minimum size of a detectable object is a function of the incident light's wavelength (Molero and Jaque, 1999). Therefore larger ($\geq$ IR) wavelengths would not be capable of resolving particles with a size in the order of hundreds of

nm.

Relevance of Atmospheric Conditions

Laser light can be scattered and absorbed dependent on atmospheric conditions (wavelength dependent). Considering all previously mentioned variable conditions to be stable, maximum range capture is proportional to the square root of the reflectivity ($\sqrt{\rho}$) and laser power. The choices of both of which can made in an attempt to account for atmospheric absorption. The best suited conditions are when the atmosphere is cool, dry, and clear of any aerosol/dust particles, when even shorter wavelengths will experience the least number of scattering events whilst in transit. Invariably, laser altimetry cannot always be performed under these conditions, but atmospheric knowledge can allow corrections to be made for acquisition in suboptimal conditions.

Geolocation

The precision of footprint measurement locations is of utmost importance to accurately yield a 3-dimensional view of the environment, especially in the case of large survey areas. An amalgamation of data from Global Positioning Satellite (GPS) receivers and an Inertial Measurement Unit (IMU) on board the altimeter, in addition to laser pointing knowledge upon firing, allow the precise determination of geographical coordinates of every received footprint centroid (Rosette, 2009).

2.6 Laser Altimeter Systems

Laser altimetry, or LiDAR can be divided into two main groups: **Discrete Return**, and **Continuous Waveform**, both of which are used in this research. A description of basic principles, strengths, and limitations of both systems are outlined below.

2.6.1 Discrete Return LiDAR

Discrete return LiDAR records, primarily, discrete points in space (at the point of reflection), with known longitudinal, latitudinal, and elevation positions (x, y, z). In addition, return intensity (I) is often recorded as it offers information regarding target reflectivity, orientation, structure, range, and scanning angle (Boyd and Hill, 2007). There are also numerous other parameters recorded, such as: return number, total number of returns, and return classification etc. The number of different parameters recorded are dependent on the system specifications and its method of acquisition (first and last, all returns; see below).

Survey specifications can be optimised with regards to specific system specifications; such parameters as digital resolution or density of points can be modified to suit the main purpose of the acquired data. For example, a high resolution dataset is required for urban mapping, whereas for the purpose of large area elevation mapping, a coarser resolution will suffice.

A typical acquisition technique for discrete return LiDAR is from on-board an aircraft such as a small plane or helicopter; this allows the acquisition of dense, (sub-metre diameter) footprint data over relatively large areas (typically of the order of 10 km^2 or more) with relative ease and speed (see Figure 2.2). For such a set up, the laser typically scans perpendicularly to the direction of the aircraft's forward motion; the maximum swath angle to which any given instrument can scan is system and application dependent.

The density of footprints for a surveyed area varies as a function of pulse emission frequency, as at present, altimeter systems are unable to perform simultaneous measurements. However, due to the natural speed within systems, it is possible to obtain between 2000 and 100,000 sequential measurements every second, each to a unique location. The footprint density can be defined in two planes, the cross-track, and along-track. The cross-track (defined in the plane of the aircraft's forward motion) spacing between subsequent footprints is dependent on the altitude and angular separation (a function of aircraft roll) at which pulses are emitted (Rosette, 2009). The along-track (defined in the plane of laser scanning) resolution is determined by the pulse emission frequency, the number of cross-track footprints, aircraft ground speed, and pitch attitude (Harding et al., 1998; Balt-

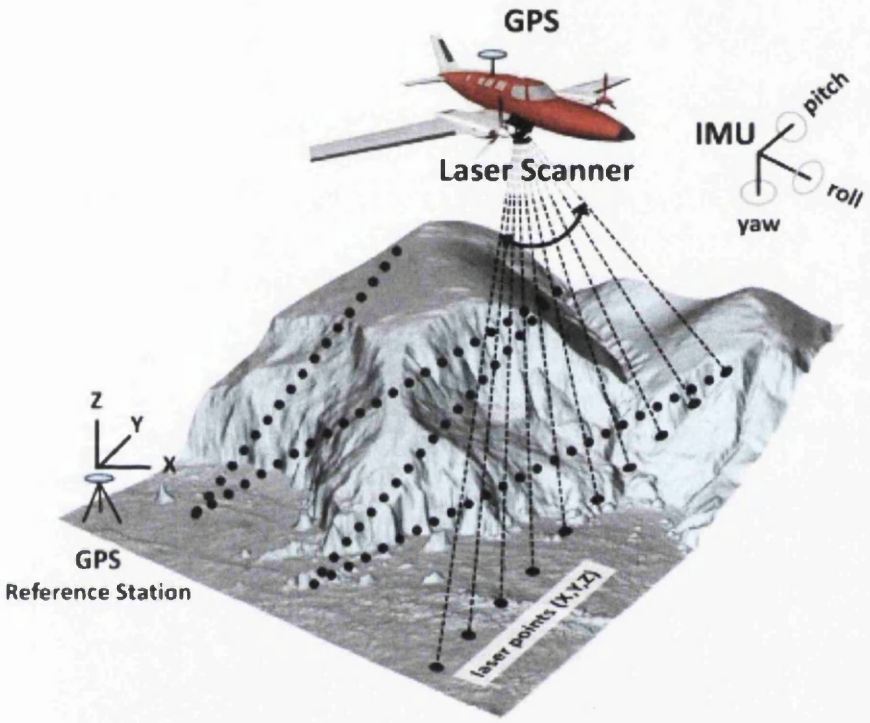


Figure 2.2: Visualisation of aircraft mounted laser altimeter system, scanning perpendicularly to the direction of aircraft travel. The (Global Positioning Satellite (GPS) systems, both on-board the aircraft, and on the ground (reference); in addition to the Inertial Measurement Unit (IMU) recording the aircraft's speed, altitude, pitch, roll, and yaw, allow the accurate determination of footprint geolocations. Modified from Höfle (2010).

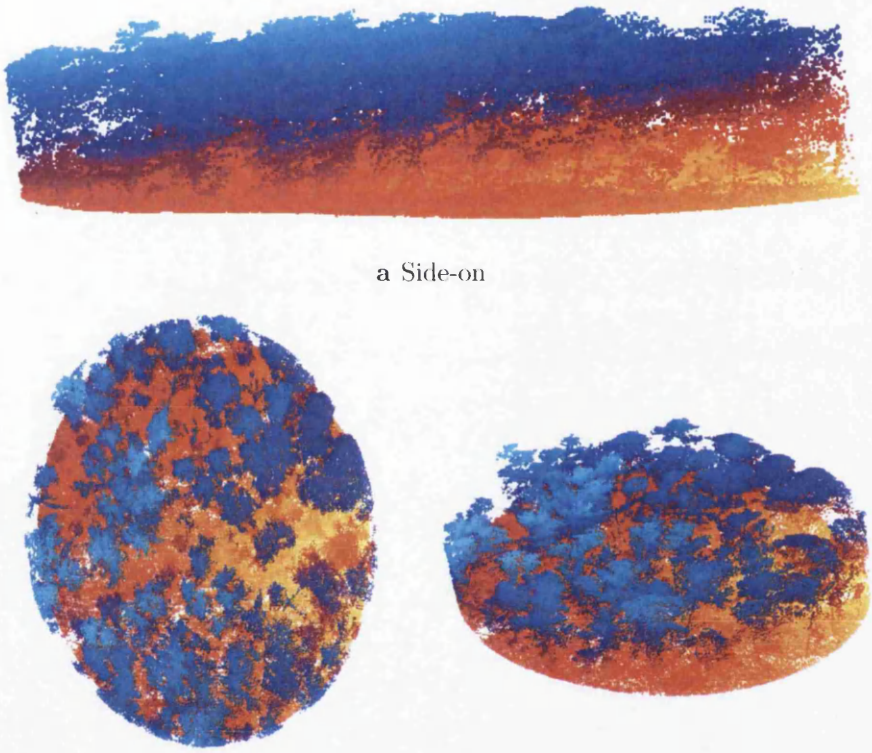
savias, 1999).

During a typical aircraft LiDAR survey, position and orientation data, footprint geolocations, associated range, intensity, and scan angles are recorded for each return, where the number of returns per emitted pulse can vary. The number of returns from an emitted pulse for an aircraft-mounted discrete return system is typically up to a maximum of four, these are known as echoes. Some systems return only the first and last echoes, which theoretically represent the first and last encountered surfaces (e.g. vegetation canopy top and ground, respectively). However, this is not always the case due to specific environmental characteristics, such as canopy density, and/or understorey presence etc. More recently the return of intermediate echoes have been incorporated into most systems, providing further information on structural properties of targets between first and last encounters (e.g. vegetation structures: branches or leaf distribution).

Consideration of all discrete returns from all emitted laser pulses allows the generation of a geographically registered 'point cloud', which is a 3-dimensional, mapped representation of the surveyed area. The distribution of the points within this 3-dimensional space is a function of the instrument's scanning pattern, altitude, ground speed, field of view, pulse emission frequency, and target complexity (Axelsson, 1999; Wehr and Lohr, 1999). Consequentially, the LiDAR point cloud constituents are semi-randomly, irregularly spaced in the spatial and temporal domain. Maximum point density tends to increase with proximity to the maximum allowed swath scan angle (a_{max}). This is due to the beam guiding optic slowing, stopping, and retracting in direction at a position in correspondence to a_{max} (Cobby et al., 2001; Rosette, 2009).

During post-processing, data are filtered to remove noise, anomalous points, measurement errors, and unassigned points (according to post-processing standards; see Section 3.8.1). Also, each individual point is assigned a classification (ground, water etc.), for ease of use. The standards for such classification outlines are made for general application to all forms of discrete LiDAR data. Discrete return LiDAR classification standards for the current research's data are discussed in Section 3.8.1.

Point information can be processed further to yield a number of different products; favoured products include terrain and surface elevation models (Section 2.8



a Side-on

b Top-down

c Angled

Figure 2.3: Example of multiple discrete return airborne LiDAR from the Tumbarumba site (described in Section 3.3) from (a) side-on, (b) top-down, and (c) angled views. The current example is restricted to an elliptical x, y boundary located in a vegetated area that exhibits sloped terrain. Note: point cloud is coloured by elevation.

highlights some uses of LiDAR data). For these products to be spot-continuous (spatially regular) some form of interpolation is required. However, interpolation gives rise to uncertainty, especially where the density of points is low, as is typically the case with low scan angles (Rosette, 2009). For example, poor results are expected when a 1 m resolution spatial product is required from a point cloud where the mean point density is ≤ 1 m^{-3}.

An example of a point cloud generated for a vegetated area from a four return (first, last and intermediate returns), discrete return LiDAR system is shown in Figure 2.3. Note the intermediate return LiDAR gives a relatively good visual representation of the physical environment that it has recorded. The vegetation component is clearly visible, and separate from that of the ground, where individual trees can be identified. Ground features/short vegetation are also visible, as is the elevation difference at ground level throughout the ellipse.

Another discrete return LiDAR acquisition technique is from the ground, known as terrestrial LiDAR. This is a more recent development in LiDAR technology, where similar processing techniques to those mentioned above are followed to obtain results. However, typically a single, fixed acquisition geometry is considered for terrestrial systems, whereas airborne LiDAR requires the consideration of continually changing acquisition geometries with aircraft motion. However, crop characterising measurements have been made from on-board moving vehicles (AndúJar et al., 2013; Sanz et al., 2013).

Terrestrial LiDAR are typically employed at the vegetation structure scale, processed to a registered point cloud (as above). The post-processing of which allows the inference of mean diameter at breast height (DBH), stem density, basal area, leaf area distribution, and above-ground biomass (Yao et al., 2011). Further principles, uses, and limitations of terrestrial LiDAR are discussed in Dassot et al. (2011). Terrestrial LiDAR systems are also available as continuous waveform systems, not discussed below (Section 2.6.2), but also discussed in Dassot et al. (2011).

2.6.2 Continuous Waveform LiDAR

Continuous (or full) waveform LiDAR is based on the same optical principles as those of discrete return systems, however, such systems continuously record the

intensity feedback from a target reflected pulse. Such recordings are possible due to far greater sampling frequencies (in the order of 1 GHz; 1×10^9 per second) than those found in discrete return systems, which allow the entire echoed waveform to be stored for analysis. This waveform is then able to be broken down into sub-components for analysis of the recorded target's structure with highly resolved vertical accuracy.

The returned echo waveform from each emitted pulse is a function of the reflectance characteristics of surfaces intercepted whilst in transit; other physical surface characteristics such as topographic relief, slope, and vegetation structures also influence the shape of the return (North et al., 2010; Rosette et al., 2010; Chen, 2010b). These physical characteristics can be retrieved from waveforms, in addition to inferred reflectance characteristics (Rosette, 2009). However, retrievals can be complicated by issues originating from surface complexities, and atmospheric signal attenuation.

The power of the emitted pulse suffers a natural decay in energy (see Section 2.4), specific to vegetation, it also decays as a function of canopy depth penetration, due to encountered foliage surfaces scattering (single or multiple), and/or absorbing incident photons. Even in cases where the footprint diameter is greater than a given crown width, and where little slope is encountered, sufficient power is available such that ground reflected returns can be expected from almost *all* footprints. However this becomes more problematic with very small footprints as the likelihood of physically encountering the ground decreases (Parker et al., 2001), dependent on vegetation characteristics such as canopy density. Some studies (Fukushima et al., 1998; Lefsky et al., 1999a; Peterson et al., 2005) have attempted to compensate for this power decay, such to minimise those footprints where no ground return is produced, by applying the MacArthur and Horn (1969) algorithm. This algorithm weights the normalised cumulative distribution of waveform energy by Equation 2.2, in order to yield a cumulative distribution of canopy area (Harding et al., 2001).

$$W = -\ln(1 - f_s) \tag{2.2}$$

where:

W = Weights

f_s = The fraction of sky obscured by foliage (Lefsky et al., 1999a)

Until recently, full waveform LiDAR systems (airborne and spaceborne) generally produced coarser resolution footprints, in the order of 10–100 m in diameter, compared with the finer resolution footprints of discrete return approaches. Consequentially, complex surface interactions occur with greater probability as footprint size increases. These interactions are due to the larger area encompassed by the footprint, hence increasing the within footprint surface variability. This produces a vast number of iterations of spatial variation, thus producing a challenging environment from which to extract generalisations of ground and vegetation subcomponent features (Rosette, 2009).

Until approximately 15 years ago, full waveform systems were only used specifically within the research community, this restriction was generally due to the cost of acquisition, and the data storage requirements of such options (Means, 2000; Todd et al., 2003). However, with recent advances in hardware, data storage issues have been abolished, and acquisition costs have been cut, meaning full waveform LiDAR systems have become more commonly applied, not just in research, but in industry also.

Waveform LiDAR can be used on-board an aircraft (for example SLICER; Section 2.7) as noted with discrete return systems (2-dimensional footprint coverage area by swath perpendicularly to aircraft motion). However in this situation, in-depth swath angle information is required. Such information is pertinent as full waveform returns provide a full vertical profile of intercepted surfaces, surface interception at an angle can lead to misrepresented vertical profiles.

Alternatively, such as the case with the spaceborne, GLAS LiDAR profiling system (see Section 3.4), sequential along-track observations can be made, where no deviation from the direction of motion track is made (i.e. no swath). This negates the need to account for scan angle during data processing, as it is expected to be minimal (unless otherwise stated).

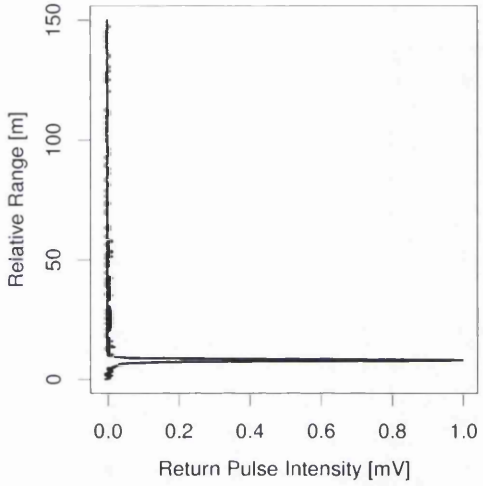
Both (swath and profiling) systems produce a vast amount of vertical profiling data for each individual footprint, allowing the determination of numerous different parameters. Returned information is depicted differently than that from discrete return systems, where an intensity profile is recorded as a function of rel-

ative time, which can be converted to range (noted in Figure 2.4).

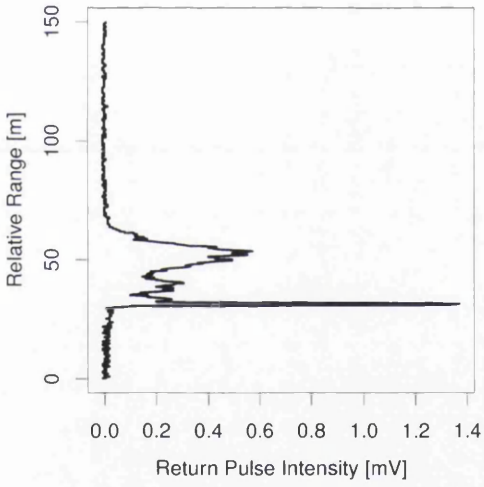
Recent advances have seen the use of small footprint waveform LiDAR for a multitude of applications, from vegetation species classification and vertical structure to urban classification (Reitberger et al., 2006; Mallet et al., 2008; Wagner et al., 2008). The benefits of small footprints in a continuous waveform LiDAR system are in the amalgamation of the high spatial density of footprints, and the continuous vertical profile element.

Further advancements have seen investigations suggesting and employing dual or multiple wavelength continuous LiDAR systems (Morsdorf et al., 2008; Hancock et al., 2012) to alleviate such issues as the effects of topographic complexities etc. However, these are new technologies, still not fully understood. But potential has been shown for their use in vegetation imaging.

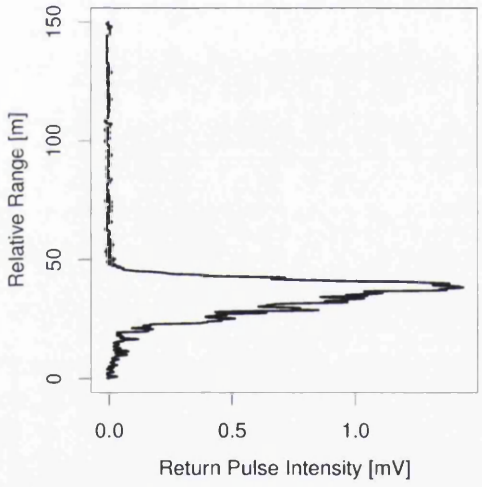
The current study (in part) investigates the impact of terrain on the accuracy of vegetation height retrieval from full waveform systems, namely from GLAS. Figures 2.4a to 2.4c give an example of typical returns from flat-unvegetated, flat-vegetated, and sloped-vegetated surfaces, respectively. The variation between these example returns highlights the role that natural structural properties of targeted surfaces play in shaping waveform LiDAR vertical profiles, and hence encourages their investigation.



a Flat and unvegetated



b Flat and vegetated



c Slope and vegetated

Figure 2.4: Example of GLAS waveform returns, from over the Netherlands for (a) unvegetated, flat surface, (b) vegetated, flat surface, and (c) vegetated area, where slope is present.

2.7 Waveform Instruments

Although GLAS was the first spaceborne LiDAR mission to be operated from a satellite, a number of waveform instruments preceded it, and served as a proof of concept for biophysical parameter retrieval (amongst other purposes), demonstrating consistent results regardless of differing instrument specifications; some such predecessors, both airborne and spaceborne, are discussed in the following. Note ICESat/GLAS is not discussed here as it was employed for analysis throughout this thesis. ICESat/GLAS is discussed in Section 3.4.
Airborne missions are typically useful as pathfinder missions as controls can be imposed more easily. In addition, these missions typically offer the best possible results due to the higher spatial and temporal resolution of data collection offered. This makes them ideally suited for assessing their results for Earth observation purposes. In the case of spaceborne missions, airborne principles can be applied, however far greater areas can be observed in a relatively small time, however, the spatial and temporal resolution of data collection is usually less consistent.

2.7.1 Developed Systems

Of the outlined missions that follow all were developed for their potential importance in vegetation imaging, however, they were not implemented to the stage of deployment due to financial restrictions. The research and development dedicated to each system provided invaluable knowledge for superseding systems (such as GLAS, and future missions), and underpins the recognised importance of quantifying vegetation characteristics.

VCL

The spaceborne Vegetation Canopy LiDAR (VCL) was selected in March, 1997 as the first Earth System Science Pathfinder (ESSP) spaceflight mission (Dubayah et al., 1997; NASA, 2012). It was scheduled for launch in January, 2000, however budget constraints resulted in program closure, at present (NASA, 2012). VCL was part of the proposed Carbon-3D space mission, where it's primary mission objective was, for the first time, to accurately estimate above-ground biomass on

a global scale from a combined, synergistic (Bidirectional Reflectance Distribution Function) BRDF-LiDAR dataset (Hese et al., 2005; Rosette, 2009).
As of yet the only comparable mission proposed since the halt of VCL was the Deformation Ecosystem Structure and Dynamics of Ice (DESDynI) mission (see below), however, this programme was terminated before deployment also. The scientific community continues to promote the use of such a system by means of simulations and tunable LiDAR systems such as (LiDAR Vegetation Imaging Sensor) LVIS, in the hope that development may continue (Rosette et al., 2013).

DESDynI

During operation, the ICESat mission highlighted the importance of vegetation observations for carbon source and sink quantification, and thus prompted the U. S. ESDS to dedicate a mission to this specific area of research: the Deformation Ecosystem Structure and Dynamics of Ice (DESDynI) mission (Donnellan et al., 2008; Abdalati et al., 2010).
DESDynI was proposed for launch in 2021 (NASA, 2013c), with a conceptual design based on the combination of a multi-beam LiDAR sensor for along-track observations, and Interferometric Synthetic Aperture Radar (InSAR) sensors for wall-to-wall coverage (Freeman et al., 2009; Hawbaker et al., 2009). The programme is no longer being fully developed, only the InSAR portion of the mission is currently being pursued. However the scientific community still support its redevelopment as a useful tool in vegetation imaging.
The conceived LiDAR system for DESDynI was proposed to be a multi-beam system (proposed 3–5 beams), operating at a wavelength of 1064 nm producing a footprint diameter of approximately 25 m (Freeman et al., 2009) where footprint separation was expected to be 25–30 m, ideal for vegetation imaging purposes (Rosette, 2009; Nelson, 2010). This system was expected to reach unprecedented accuracy levels, with a ± 1 m vertical resolution (NASA, 2008; Freeman et al., 2009).
DESDynI's InSAR system was designed to address the shortcomings of existing InSAR capable satellites (Freeman et al., 2009). The system was planned as an L-band (1.2 GHz) radar with a 24 cm wavelength capable of operating in several

modes (Freeman et al., 2009). The radar applications were primarily focused on surface deformation originating from volcanic, tectonic, or earthquake sources, in addition to ice-sheet deformation (NASA, 2008; Freeman et al., 2009).
DESDynI would have been the first dedicated satellite focused primarily on carbon budget quantification, promising a challenging data rich environment to work in. Although terminated, knowledge gained throughout the development of this programme can be transferred to the development and implementation of proposed future missions.

2.7.2 GLAS Predecessors

Some of the following missions were deployed as airborne and spaceborne missions prior to the deployment of ICESat/GLAS. Each contributed to realising the importance of LiDAR missions in remote sensing and Earth observation, and vegetation imaging in particular. The knowledge gained from each of these previous missions aided in the design, development, and implementation of similar subsequent missions.

SLICER

The Scanning LiDAR Imager of Canopies by Echo Returns (SLICER) is an airborne laser altimeter system developed at NASA's Goddard Space Flight Center. The rationale behind the design and implementation of SLICER was to perform measurements in support of NASA's Topography and Surface Change Program, Terrestrial and Ecology Program, and the Boreal Ecosystems Atmosphere Study (BOREAS). Results, and the experience gained of full waveform measurements from SLICER were incorporated into the superseding system: the LiDAR Vegetation Imaging Sensor (LVIS; see below).
SLICER was one of the first waveform systems, and hence some instrument specifications were refined as instrumentation evolved. For instance, SLICER's laser pulse shape was based on a Rayleigh distribution, rather than the later selected Gaussian distribution commonly used on waveform instruments today. Such a pulse shape shifted the energy distribution across the laser footprint in a non-symmetrical manner in comparison to what is expected from Gaussian distribu-

tions. thus effectively decreasing the usable footprint diameter.
SLICER also demonstrated that high range accuracy to multiple targets within a recorded footprint was possible. This was achieved by utilising a fast leading edge rise time and high peak power specification optic (Rosette, 2009). SLICER had a swath of five 10 m diameter footprints perpendicular to it's direction of motion. Synthesising this information from multiple targets arose from the designed detector being larger than that of individual laser footprints. SLICER produced vertical and horizontal positioning accuracies of 11 cm (equivalent of vertical sampling every 0.742 ns), and between 5–10 m, respectively (Means et al., 1999).

LVIS

The LiDAR Vegetation Imaging Sensor (LVIS) is a medium altitude airborne sensor developed at NASA's Goddard Space Flight Center in the late 1990s (Dubayah et al., 2010), superseding the previously mentioned SLICER. The instrument was implemented as a trail for conceptual planning for instrumentation for the planned spaceborne Vegetation Canopy LiDAR mission (see above) (Rosette, 2009).
LVIS is capable of operating at 10 km altitude with a swath width in the order of 1 km. This results in footprints with a typical (medium resolution) diameter of 25 m (Dubayah et al., 2010; Rosette et al., 2013), although footprints with diameters between 1 and 80 m are possible (Blair et al., 1999). A 1064 nm laser pulse is emitted with a width of approximately 10 ns at Full Width at Half Maximum (FWHM; defined as a measure of resolution equal to the width of an image line sourced at points where the intensity is reduced to half it's maximum) (Dubayah et al., 2010; Rosette et al., 2013); pulse repetition is possible between 100–500 Hz (Blair et al., 1999).
Unlike most sensors, which calculate range using two separate waveform digitisers and time interval units, LVIS utilises a single detector and oscillator for all roles. Such a set up eliminates any inconsistencies that may exist between instruments, and allows the return pulse to be located precisely (Rosette, 2009). For future development potential, the receiver is also capable of housing two additional sensors. This allows the possibility of conducting investigations into set-ups with dual wavelengths (Rosette, 2009), potential for which has also been investigated

by Hancock et al. (2012).

The LVIS system was designed to produce a Gaussian energy distribution and allow the emitted pulse to pass through filters, thus allowing the modification of energy to optimise the strength of the return signal. This instrumental feature becomes especially useful when considering large footprints, as the energy distribution from such footprints is particularly sensitive to surface spatial variability that can be present. Investigations pertaining to energy distribution sensitivity have provided optimal set-ups for differing uses of waveform systems, and have aided in the elimination of weaknesses identified in previous instruments, and were due to be incorporated into the spaceborne VCL (Blair et al., 1999).

SLA

The Shuttle Laser Altimeter was one of the first spaceborne LiDAR missions to be realised, based on the design and development experience of the Mars Observed Laser Altimeter (MOLA-1) sensor (Garvin et al., 1998).

Two SLA missions were realised, SLA-I operated as part of the STS-72 mission on-board Space Shuttle Endeavour between $11^{th} - 20^{th}$ January 1997. Whereas SLA-II operated as part of the STS-85 mission, from on-board Space Shuttle Discovery, launched later the same year, between $7^{th} - 18^{th}$ August (Garvin et al., 1998; Carabajal et al., 1999). These missions were the first of their kind that NASA trailed in a space environment.

Even though both SLA missions were of short durations they succeeded in obtaining data between $\pm 57°$ latitude (Rosette, 2009); data were collected as 100 m diameter footprints with along-track separation of approximately 700 m (Bufton et al., 1999).

The underlying objectives of the SLA missions were to serve as an engineering pathfinder for high resolution orbital laser altimeter observations of terrestrial surfaces, and provide datasets for global Earth System science problems, particularly those related to land cover dynamics in arid, coastal, and vegetated regions (Garvin et al., 1998).

These objectives were achieved with success as the SLA missions produced approximately 3 million observations each (Garvin et al., 1998; Harding et al., 1999;

Bufton et al., 1999), where over 16.5% ($\sim 475,000$) of valid observations were of land surfaces, whilst approximately 38% (~ 1.1 million) were of the ocean (Garvin et al., 1998; Harding et al., 1999).
Despite the large footprint diameters produced, the mission provided an insight to waveforms produced from differing surface types. The use of SLA footprints for derivations of vegetation metrics was limited by the coarse spatial and temporal resolutions. However, the mission did make significant contributions towards the improvement of geolocating footprints in preparation for the launch of GLAS (Luthcke et al., 2002).

2.7.3 GLAS Proposed Successors

The success of the first ICESat mission, which operated from Feburary, 2003 to October, 2009 (NASA, 2007; Abdalati et al., 2010; NSIDC, 2012a), coupled with recent observations of dramatic changes in polar ice has led the U. S. National Research Council's (NRC) Earth Science Decadal Survey (ESDS) to call for an ICESat follow-on mission (NRC, 2007). The mission is also expected to be suitable for vegetation imaging; this is also true for the LiDAR Surface Topography (LIST) mission. Both missions are discussed below, in addition to two airborne pathfinder missions implemented in order to evaluate possible issues that may arise during ICESat-2 development.

SIMPL

The Slope Imaging Multi-polarization Photon-counting Lidar (SIMPL) is an airborne multi-beam micro-pulse photon counting instrument developed by Sigma Space Corporation (Sigma Space, 2013).
SIMPL operates at 532 nm and 1064 nm using a micropulse laser set up with a 100 picosecond (ps; 100×10^{-12} s) resolution timing unit (Sigma Space, 2013). The Nd:YAG laser can fire up to rates of 22 kHz. The beam is broken into 100 beamlets of a 10 by 10 configuration that are scanned over a scene with a spatial resolution of 15 cm, where individual photons can achieve a ranging precision of 8 cm (Harding et al., 2010; Sigma Space, 2013). The pattern and frequency of the dual wedge scanner, synchronised to the laser pulse rate, are such that contiguous

coverage of a scene is obtained with a single flight pass (Sigma Space, 2013).
The system was employed by NASA in 2009 as a pathfinder for the future space-borne mission ICESat-2, which is set to use photon counting technology. The investigations pursued were primarily focussed on behaviour over ice and water, and used both the 532 nm and 1064 nm wavelength laser set-ups (Harding et al., 2010; Brunt et al., 2011). Results were successful in locating the ice-sheet surface elevation and differentiating between open water and differing types of ice that exhibit variable optical properties (Harding et al., 2010).

MABEL

The Multiple Altimeter Beam Experimental Lidar (MABEL), is a prototype airborne multi-beam micro-pulse photon counting instrument, developed by NASA at the Goddard Space Flight Center, and like SIMPL, was employed as a pathfinder for ICESat-2 (Brunt et al., 2011; NASA, 2011). At this point in ICESat-2 development, data from MABEL are critical for model validation and algorithm development (NASA/GSFC, 2013).
MABEL can operate both 532 nm and 1064 nm wavelength lasers at a pulse repetition rate of up to 25 kHz, providing footprints approximately 2 m in diameter when run at high altitude. High-altitude operation (up to $\sim$ 20 km) is a key aspect of this pathfinder mission, as conditions in the high atmosphere are important in the verification process of this revolutionary laser system, a variation of which is to be used on ICESat-2 (NASA/GSFC, 2013).
MABEL fires a single laser beam which is split into 105 beamlets/channels, which provides a maximum swath width of ± 1.05 km. The total number of channels that can be selected for use at any given time is 24; up to 8 channels from the 1064 nm laser, and up to 16 channels for the 532 nm regime (NASA/GSFC, 2013).
MABEL first flew from NASA's Dryden Flight Research Center in Palmdale, California in December 2010. Data were acquired over a period of days over a multitude of surfaces including: vegetated areas, ocean and freshwater, snow, salt flats, and steep terrain (Brunt et al., 2011; NASA/GSFC, 2013). Such a broad application aided in better understanding this pioneering technology, and it's behaviour over different surfaces, in addition to understanding how ICESat-2 will likely measure

different landscapes.

This test period was deemed a success, proving that MABEL could successfully locate the ground surface for numerous landscapes, however system noise was greater than expected; an example of MABEL data is shown in Figure 2.5. During these test flights some minor operational issues were also noted, these were subsequently investigated and addressed where possible.

At present no definitive literature is available with regards to specific performance qualities, but this will almost certainly be available before the launch of ICESat-2, outlining its expected on orbit performance.

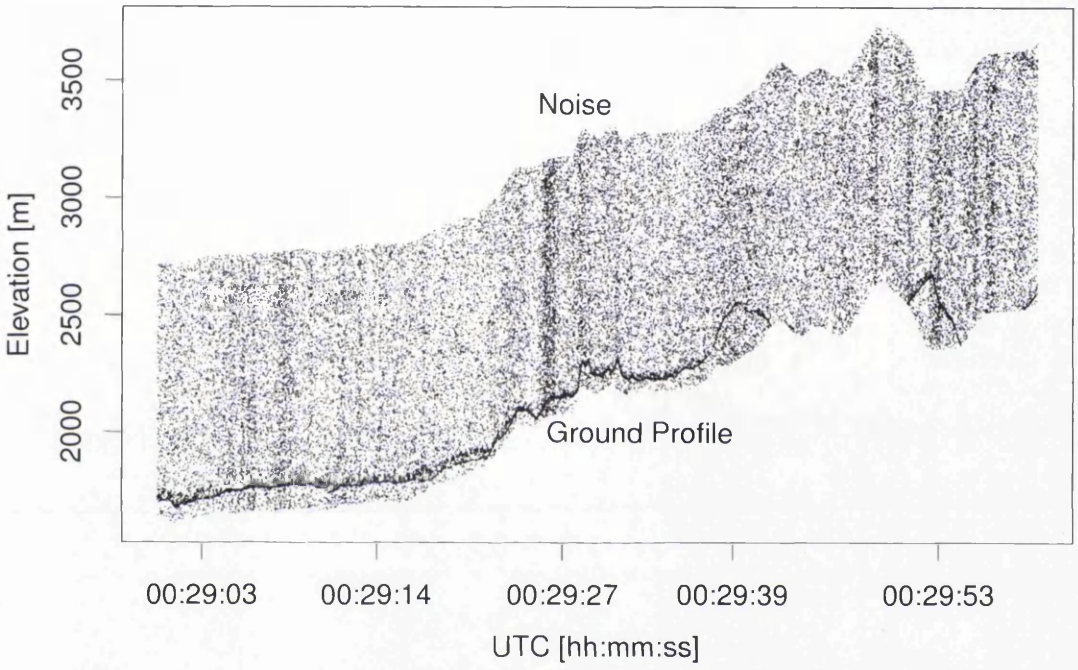


Figure 2.5: Example of photon counting data from the Multiple Altimeter Beam Experimental LiDAR (MABEL), flown over the Sierra Nevada mountains in the USA, during the December of 2010. Note the ground is clearly indicated, where points are most dense, however, heavy noise is also present in the system.

ICESat-2

ICESat-2 will be the direct successor to ICESat, and is currently under development. The mission was originally planned for launch in 2015 (Rosette, 2009; Abdalati et al., 2010), but is now rescheduled for a 2016 launch (NASA, 2013a). The instrument is expected to be the first satellite based micro-pulse photon counting device capable of providing global coverage of the Earth's surface.

ICESat-2 will operate the Advanced Topographic Laser Altimeter System (ATLAS), whose initial instrumental design was to closely mimic that of its predecessor, GLAS, on board ICESat (Yu et al., 2010b; Nelson, 2010). However, in mid-2009, updated science requirements outlined during the ICESat-2 workshop (NASA, 2007) precipitated an instrument redesign (Smith et al., 2001; Yu et al., 2010b). The redesign of ATLAS called for a micro-pulse, multi-beam laser architecture with a high repetition rate (10's kHz), low energy (100's mJ), and a significantly shorter pulse width (FWHM $\sim$ 1 ns) than was operated from on board ICESat (Yu et al., 2010b).

Up to 2010, the ATLAS design consisted of a single 532 nm (Abdalati et al., 2010; Rosette et al., 2011) laser split into a 9-beam pattern, arranged in a 3×3 configuration (Figure 2.6). A 10 m footprint diameter is expected, separated (along track) by 0.7 m (Rosette et al., 2011), where specific footprint energy is dependent on its grid location with regards to the central nadir footprint (central point in Figure 2.6), which exhibits the greatest energy. The energy of footprints in direct lateral or longitudinal alignment with this central footprint exhibit half its energy, whereas the outer footprints, on the grid corners, exhibit a quarter of its energy.

The ATLAS laser configuration design is still under development, and hence is subject to change. It has been reported to have changed from a 9 beamlet set up to a 6 beamlet set up (Rosette et al., 2011), however, specific updates from NASA have not been disclosed.

Like its predecessor, ICESat-2 is expected to support multidisciplinary applications. Following the Decadal Survey's recommendation, the main science goals are expected to be measuring ice-sheet changes, sea ice thickness, and vegetation biomass (Abdalati et al., 2010). The ICESat-2 Science Definition Team (SDT) has defined specific objectives for the mission in each of these key areas.

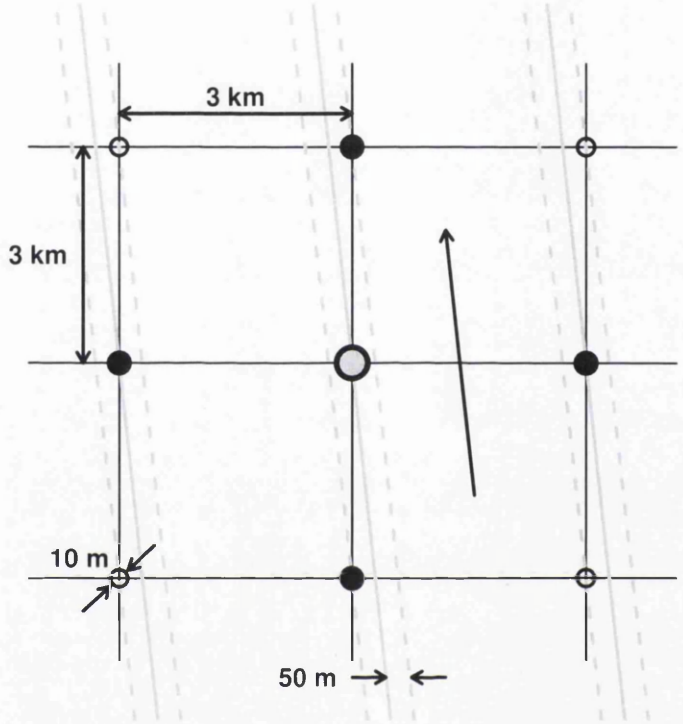


Figure 2.6: 3×3, 6 km grid pattern, configuration currently conceived for ATLAS whilst on board ICESat-2. Three ground tracks are observed in a single pass by 10 m diameter footprints, where lateral separation between within track footprints is 50 m; ground tracks are separated by 3 km each from the central track. Footprint energy varies by location throughout the gridded configuration. Figure based on Yu et al. (2010b).

In cryospheric applications, the primary objective of the mission is to quantify polar ice-sheet contributions to sea-level rise and the linkages to climate change (Abdalati et al., 2010). A secondary objective focusses on sea ice, with the view to estimate sea ice thickness, by means of measuring freeboard height (ice height above open water), and accounting for density differences between the floating ice, snow and water (Abdalati et al., 2010).

For vegetation purposes, the primary objective is to measure vegetation height on a global scale, thus allowing an estimation of global above-ground biomass with high spatial resolution (Abdalati et al., 2010; NASA, 2013a); in addition, ICESat-2 will enhance the utility of other Earth Observation systems through supporting measurements (NASA, 2013a).

Results from the ICESat mission (Lefsky et al., 2007) suggest that extending the ICESat capacity to a 91-day (from 33-day) continuous measurement regime, could make ICESat-2 capable of producing a global vegetation height surface with 3 m vertical accuracy at a 1 km spatial resolution. This assumes that off-nadir observations can be employed to increase the spatial distribution of measurements over terrestrial surfaces (Abdalati et al., 2010).

A combined data pool of ICESat, and the anticipated 5 year acquisition of ICESat-2 measurements will provide a sufficiently long period of surface observations. This will allow the quantification of more than 15 years of change (Abdalati et al., 2010), particularly for the cryosphere (due to NASA's Operation IceBridge programme; Koenig et al., 2010; Kurtz and Farrell, 2011). Unfortunately no coincident program has been launched for vegetation purposes, hence inter-mission data gaps are likely to arise.

In the context of forestry applications, a combination of both ICESat and ICESat-2 measurements are expected to provide a new set of global ecosystem applications. These include mapping forest productivity, observations of tree phenology, forest disease, pest outbreaks, and vegetation height (Abdalati et al., 2010).

LIST

The LiDAR Surface Topography (LIST) mission was originally expected for launch between 2016–2020 (NAP, 2008), but is currently more likely to be launched in

2025 (Yu et al., 2010a). LIST aspires to carry out the most precise global topographic survey to date, with a resolution of 5 m and vertical precision in the order of 10's m, imaging the Earth's surface and it's overlying covers of vegetation, water, snow, ice, and man-made structures (NAP, 2008; Yu et al., 2010a).
The instrument currently proposed is expected to house 1000 parallel micro pulse, photon counting profiling channels, totalling a 5 km swath width. Each profiling channel will measure approximately 5 m in diameter, pulsing with a repetition rate of 1.4 KHz (NAP, 2008; Yu et al., 2010a).
This spaceborne instrument is in its early development stage, necessitating the need for airborne instrument simulations to provide a valuable insight to the possible shortcomings of the proposed mission, allowing improvements to be made if required.
LIST promises to be an exciting leap in accurately characterising the Earth's surfaces at an extremely high spatial resolution.

2.8 General LiDAR Applications

Since the relatively recent establishment of LiDAR technologies, their applications have been widely demonstrated, providing two and 3-dimensional representations of features across a multitude of disciplines; these include atmospheric, oceanographic, cryospheric, and terrestrial applications. Both discrete return and continuous waveform LiDAR approaches have been applied across these disciplines.
Atmosphere LiDAR applications have been used to measure: aerosol optical depth (AOD), cloud top heights and optical depth, and surface properties and constituents (Spinhirne et al., 1982; Sassen and Benson, 2001; Comstock et al., 2002; Weitkamp, 2005).
Applications in the **Ocean** have also been broad (Dai et al., 2004). For example surface elevation, temperature, sound velocity, bulk viscosity have been measured (Popescu et al., 2004; Lancaster et al., 2005; Shi et al., 2008). In addition, ocean parameters such as depth, and sea-floor reflectance have been obtained, as well as the locations of submersed objects being recorded in real time (Muirhead and Cracknell, 1986; Mullen et al., 1996; Irish and Lillycrop, 1999; Qiyang et al., 2002).
Cryosphere LiDAR applications have been focussed on measuring the properties

and changing parameters of land-locked ice sheets, and sea ice in the polar regions (Krabill et al., 1995; Zwally et al., 2002; Kwok et al., 2009).

The application of LiDAR technologies **terrestrially** has yielded quantifications of surface properties such as: elevation and topographic detail, in addition to land-use and land-use change, and above ground biomass, and vegetation structures and their constituents (Dubayah and Drake, 2000; Hofton et al., 2002; Lefsky et al., 2002a; Hese et al., 2005; Antonarakis et al., 2008; Rosette et al., 2008; Hopkinson and Chasmer, 2009; Chen, 2010a).

LiDAR applications for forestry are discussed separately in Section 2.9 as they warrant a more in-depth description, considering the context of the current research project.

2.9 Forest Applications of LiDAR

LiDAR has proven itself as an extremely useful tool fit for many purposes, particularly in measuring forest structure related parameters, which are important for quantifying carbon stocks, and creating sustainable forest management strategies, in conjunction with today's political challenges. Some such parameters as vegetation height, structure, biomass, and timber volume, and forest floor topography effects, are discussed throughout this section.

The majority of these observations have been made with airborne LiDAR, but in recent years, observations have also been made by terrestrial and spaceborne systems (Rosette, 2009).

2.9.1 Forest Floor Topography

Forest floor retrieval has been demonstrated at local scales for small footprint airborne LiDAR by many (Blair et al., 1999; Jansma et al., 2001; Hofton and Blair, 2002). Employing LiDAR for such retrievals has become favoured over traditional surveying techniques in some cases, due to the non-destructive nature and cost efficiency per sampled unit area; additionally, LiDAR surveys alleviate issues with access to remote locations.

Elevation derivations from LiDAR have also demonstrated an improved accuracy

over other remote sensing techniques such as photogrammetric methods due to it's ability to directly detect the ground in most cases (Lefsky et al., 2002b). This capability lends itself to many disciplines. For instance, in archaeology and oceanographic scanning, LiDAR has enabled the location(s) of historic remains and/or submerged objects to be detected in otherwise inaccessible locations (Qiyang et al., 2002; Devereux et al., 2005; Chase et al., 2011).

The process of creating an elevation model over a forested area is reliant on the ability to classify the ground return(s) or component, for discrete return and full waveform systems respectively. Typically for full waveform returns, post-processing is performed and discrete returns extracted from the waveform profile to produce a three dimensional location (x,y,z) (Mallet and Bretar, 2009). This method of waveform decomposition can provide up to 60% more pulses than a real time discrete return system (Reitberger et al., 2008), however, this is dependent on the local environment such as topographic complexity and surface structures.

Considering ground classified points only, interpolation techniques are typically applied to the three dimensional data to yield a gridded (in x, and y) representation of the surface elevation. It has been noted that the interpolation method chosen for each unique dataset can affect the uncertainty in results. This is typically dependent on the density of data available for processing and on the terrain complexity (Hodgson and Bresnahan, 2004); in addition vegetation density can also effect uncertainty (Clark et al., 2004).

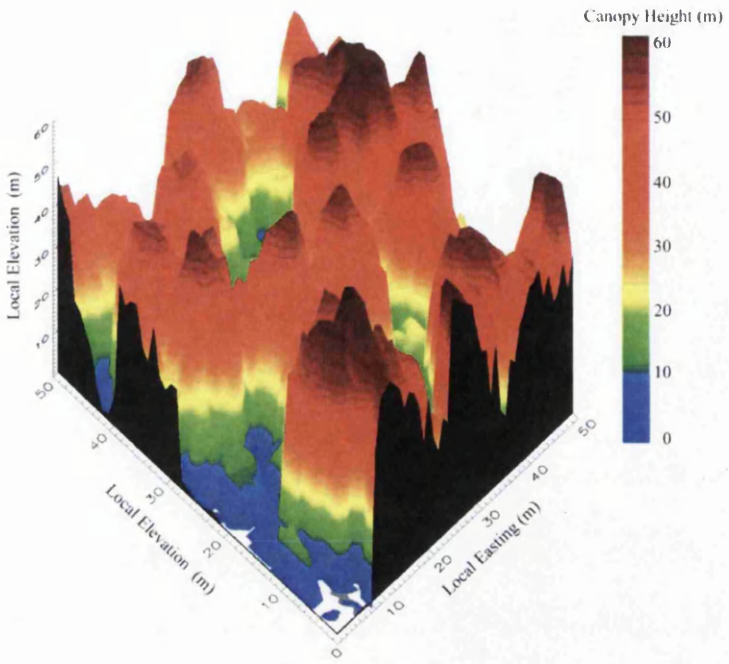
In general, with adequate quality control, elevation models from LiDAR sources are highly accurate, where a root mean square error (RMSE) of 0.5 m can be achieved for high density airborne measurements (Lefsky et al., 2002b). However, it is also typical that accuracy degrades with ground relief. For example, Hodgson and Bresnahan (2004) noted a large increase in the uncertainty of modelled elevation results for a multitude of surface types when retrieved over a 25° slope, when compared to a slope of 4°.

Forest floor topography information is of vital importance in deriving accurate parameters. It is often the cause of spurious parameter derivations (Simard et al., 2011; Rosette et al., 2013). At present the uncertainties induced by complex topography are not fully understood with regards to retrieving vegetation parameters. Quantifying the magnitude of its affects on such parameters will enable steps to

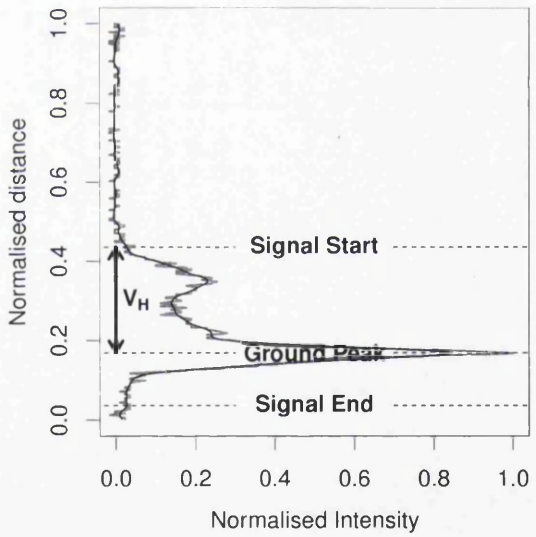
be taken in correcting for its presence.

2.9.2 Vegetation Height

A key parameter measurable by LiDAR is vegetation height, which allows the derivation of further parameters such as timber volume, and above ground biomass (see Section 2.9.3), which in turn can be used to infer vegetation carbon stocks. Laser altimetry offers a remote, non-destructive way of estimating vegetation heights in closed canopies, however, there are some exceptions in cases of extremely dense canopies ($F_{cov} \geq 0.9$, and LAI $\geq 6 - 7$) (Rosette et al., 2013).
Vegetation height estimates, from small to large (near global) scales have been measured by many (Lefsky et al., 2002a; Boudreau et al., 2008; Los et al., 2012). Discrete return (Patenaude et al., 2004; Popescu, 2007), and small and large footprint waveform LiDAR have been employed for such purposes. Waveform observations have been made from LVIS (Drake et al., 2002, 2003; Dubayah et al., 2010), SLICER (Lefsky et al., 1999a; Harding et al., 2001; Lefsky et al., 2002a), and GLAS (Nelson et al., 2009; Simard et al., 2011; Los et al., 2012), where relatively consistent results have been obtained.
Obtaining vegetation height estimates from discrete return LiDAR is a matter of relatively simple raster subtraction, as described by Lefsky et al. (2002b). The process utilises two elevation models, one based only on ground returns (DEM; as in Section 2.9.1), and another based on all returns (Digital Surface Model, DSM), including those from vegetation. The subtraction of the DEM from the DSM yields the canopy height (Figure 2.7a) above the ground surface (Lefsky et al., 2002b; Popescu, 2007).
This same process can be applied for waveform LiDAR returns, provided that these returns are broken down into constituent discrete components (Hofton et al., 2000; Reitberger et al., 2008; Rutzinger et al., 2008). This yields a point cloud representation, similar to that obtained from discrete return LiDAR systems, which in turn enables the creation of the required DEMs and DSMs.
Alternatively, a direct analysis of the waveform itself can be performed, which yields a direct, individual footprint estimate of vegetation height. Numerous methods exist for such estimates from waveform returns (Lefsky et al., 2005; Duong



a Discrete Return



b Continuous Waveform Return

Figure 2.7: Example of (a) 3-dimensional vegetation (denoted canopy here) height representation derived from the subtraction of Digital Surface and Elevation Models (Source: Lefsky et al., 2002b), and (b) a direct measurement of vegetation height (V_H) from a waveform return.

et al., 2006; Lefsky et al., 2007; Rosette et al., 2008; Sun et al., 2008). The principles of each of these methods are very similar, which is to correctly locate the ground and canopy top elevations. Successful location of the ground anchors the lower limit of the vegetation, where the canopy top is generally defined at the elevation where the waveform energy/intensity first exceeds a defined noise threshold. The difference between these two limits represents the vegetation height (V_H; Figure 2.7b).

Regardless of method, vegetation height estimates suffer inaccuracy when derived over complex terrain and dense vegetation (Chen, 2010b; Simard et al., 2011; Rosette et al., 2013). Dense vegetation prevents the laser penetrating the canopy and reaching the ground. The ground (when sloped) covers a broader range of elevations, hence vegetation returns are mixed with those from the ground leading to exaggerated V_H (Rosette et al., 2013). These issues have been noted most commonly in larger footprint systems, due to the increase surface variability encompassed by the footprint area.

Another issue of note is that the highest portion of vegetation canopies are often not detected when they exhibit a small cross-sectional area from a top-down view, thus leading to under-estimation of the canopy top. This is most common in systems were the energies within the footprint energy distribution are small. This dictates that reflected returns are not of a high enough intensity to be distinguished from background noise. Top-down cross-sectional areas are not always problematic as they are a function of vegetation species, which dictates (in part) its structure (e.g. conical or spherical crown shape), and canopy density (i.e. less reflection from smaller area). Whereas instrument specifications are fixed, they are often the limiting factor in the accurate retrieval of V_H; specifically, signal strength and wavelength (affects return signal strength), noise thresholds, and signal to noise sensitivity offer the most restriction.

Additionally, as suggested by Hancock et al. (2012) and Wallace et al. (2012), signal attenuation caused by vegetation subcomponents can effect waveform geometries and hence cause spurious derivations of vegetation height and subsequently vegetation structure, biomass and volume estimates. This signal attenuation by vegetation is referred to as a "shadow", as defined in Section 1.3.

Shadows are suggested to be capable of modifying waveform geometries sufficiently

such that "phantom peaks" exist, which can be misinterpreted to represent physical features (such as ground elevation) incorrectly. This effect has not been noted in large footprint waveform systems, but only under controlled conditions (Wallace et al., 2012), and in model simulations (Hancock et al., 2012).
This study takes steps towards eliminating uncertainties in vegetation height estimations over sloped terrain. Accurate quantification of vegetation heights are important as numerous parameters can be derived from this principle forest parameter.

2.9.3 Biomass and Volume

Timber volume and biomass are important parameters in estimating vegetation carbon content. Numerous approaches have been investigated in order to retrieve such parameters, employing both discrete return (Nelson et al., 1988; Næsset, 1997; Nelson et al., 1997, 2004) and continuous waveform (Sun et al., 2008; Nelson et al., 2009) LiDAR systems. The techniques of such investigations are briefly discussed in this section.
Prediction of forest stand structure using discrete return LiDAR was based on the theory of Maclean and Krabill (1986), who adapted a photogrammetric technique to the interpretation of LiDAR data (Lefsky et al., 2002b). This enabled important parameters, such as vegetation species and height, used for retrieving volume and biomass estimates to be retrieved over large areas. In conjunction with these LiDAR derived parameters, (tree) species specific carbon allocation equations are often used in volume and biomass calculations.
Nelson et al. (1988) successfully predicted the volume and above-ground biomass (AGB) of southern pine forests in southwestern Georgia, USA, employing six laser canopy measurements in conjunction with two logarithmic equations to find the optimal model for such predictions. The best models explained between 53% and 65% of the variance noted in ground measurements of forest biomass and volume (Nelson et al., 1988). Mean total LiDAR estimates of biomass and volume for 38 test sites fell within 2% and 2.6% of ground based measurements respectively; in this particular case species stratification did not consistently improve results for such forests.

Later work by Nelson et al. (1997) in tropical wet forests at La Selva Biological station obtained similar results for basal area, volume and AGB. Here a similar method was used as noted in Nelson et al. (1988), however, it was developed such that optimal spatial configuration of field sampling for LiDAR data comparison was employed (Lefsky et al., 2002b).

Næsset (1997) estimated the volume of 36 stands of Norway spruce (Picea abies) and Scots pine (Pinus sylvestris) in southern Norway. Stand volume was derived from various vegetation canopy height metrics and canopy cover density measurements. The regression techniques employed explained between approximately 45%–89% of the variance in stand volumes, between LiDAR derived results and field data; here species stratification was not employed.

Nelson et al. (2004) showed the potential for state-wide documentation of vegetation volume and AGB based on estimates for Delaware, USA, employing first return LiDAR profiling only. Best results were found when broad vegetation stratification was applied over four groups, i.e. mixed wood, hardwood, conifers, and wetlands. A simple regression model was employed for a combination of mean vegetation height metrics (and canopy cover in the case of hardwoods) estimating merchantable volumes within 1% of US Forest Service (USFS) estimates at the state-wide level; AGB estimates fell within 16% of USFS estimates at this scale (Nelson et al., 2004).

Waveform systems have been employed for such parameter retrievals also. LVIS waveform energy quantiles have been found to highly correlate with mean diameter at breast height (DBH), basal area, and AGB estimates (Drake et al., 2003). These metrics were found to be applicable to such estimates, as they are dependent on both vertical canopy distribution and canopy cover, as less dense canopies allow greater laser penetration, thus reducing the height at which these quantiles are recorded (Rosette, 2009).

Nelson et al. (2009) employed GLAS waveforms, following the suggestion of suitability for timber volume and biomass retrievals from Sun et al. (2008), in Siberia to estimate timber volume. Used in conjunction with the Moderate Resolution Imaging Spectrometer (MODIS) data for land classification, non-linear models were developed to infer timber volume, yielding a regional agreement between derived and field results of 1.1%. The study demonstrated the capabilities of data

synthesis between GLAS and MODIS. However, a limitation of GLAS derived estimates was noted, namely where slopes $> 10°$, these data were excluded from the study due to spurious results being induced.

Canopy Volume Method

The techniques described for obtaining volume and canopy structure have been based primarily on statistical processes. The capability of LiDAR to rapidly measure the 3-dimensional structure of canopies stimulated the development of new methods that exploited this capability. One such method is the Canopy Volume Method (CVM; Figure 2.8) described by Lefsky et al. (1999a) which is capable of representing canopy cover and volume as a function of depth within the vegetation canopy, from LiDAR waveforms.

Waveforms are converted to vertical height profiles with established thresholds (Figure 2.8 left panels) in order to allow classification of the vegetation canopy into four different zones (Figure 2.8 right panel): Empty space consisting of no canopy/ground returns, Euphotic zone comprising the upper 65% canopy closure elements, Oligophotic zone forming the remaining 35% of elements, and the Closed and Open Gap Space elements. The Open Gap Space is derived from the difference between the waveform extent (as defined by Lefsky et al., 2005) and the vegetation height on an individual footprint basis. This allows the proportions of each class to be calculated (represented in Figure 2.9a), which in turn affords information regarding volume, canopy cover, and light regimes (Lefsky et al., 1999a; Lim et al., 2003; Rosette, 2009).

This method is particularly relevant for use with waveform LiDAR, as these systems are capable of revealing distinct characteristics relating to canopy structure, volume, and canopy cover distribution as a function of canopy depth. This is emphasised, particularly as waveform LiDAR is capable of distinguishing individual vegetation elements such as branches and stems in addition to foliage (Chen et al., 2004; Hancock et al., 2012). Such characteristic information can allow probabilistic computations of shade tolerance and light competition (Parker et al., 2001), indicate on stand microclimate, and allow stand development to be observed (Lefsky et al., 1999a; Harding et al., 2001) as a function of vegetation volume (Figure

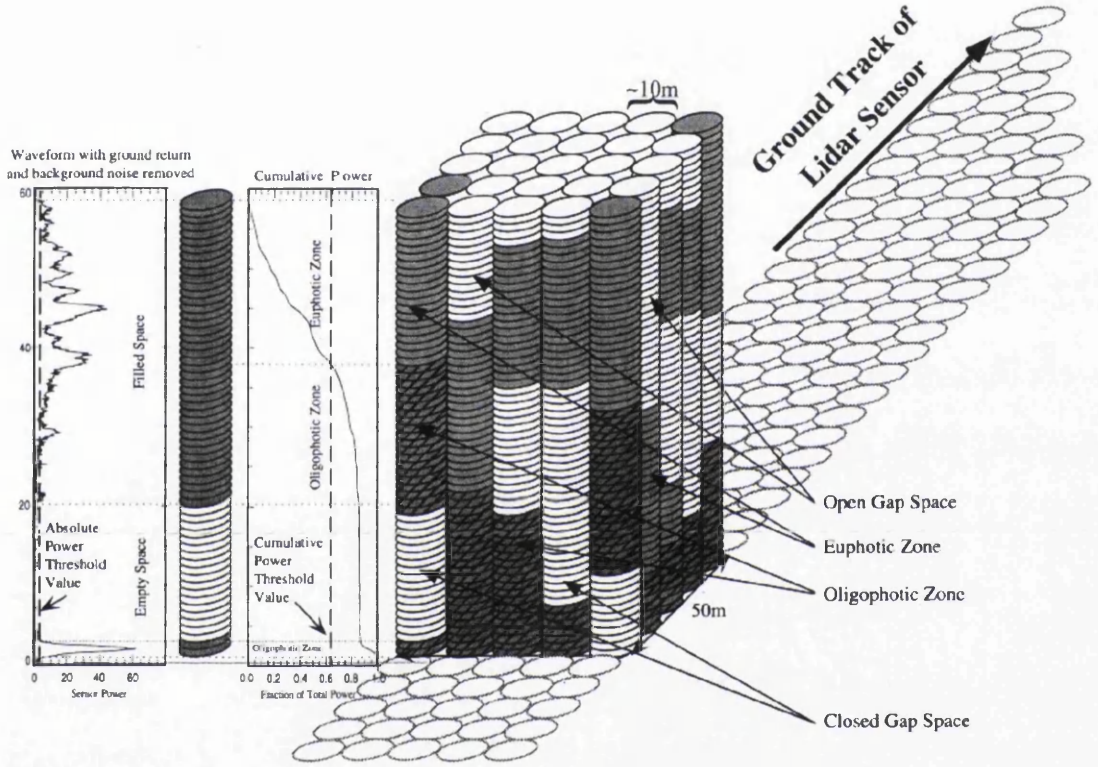
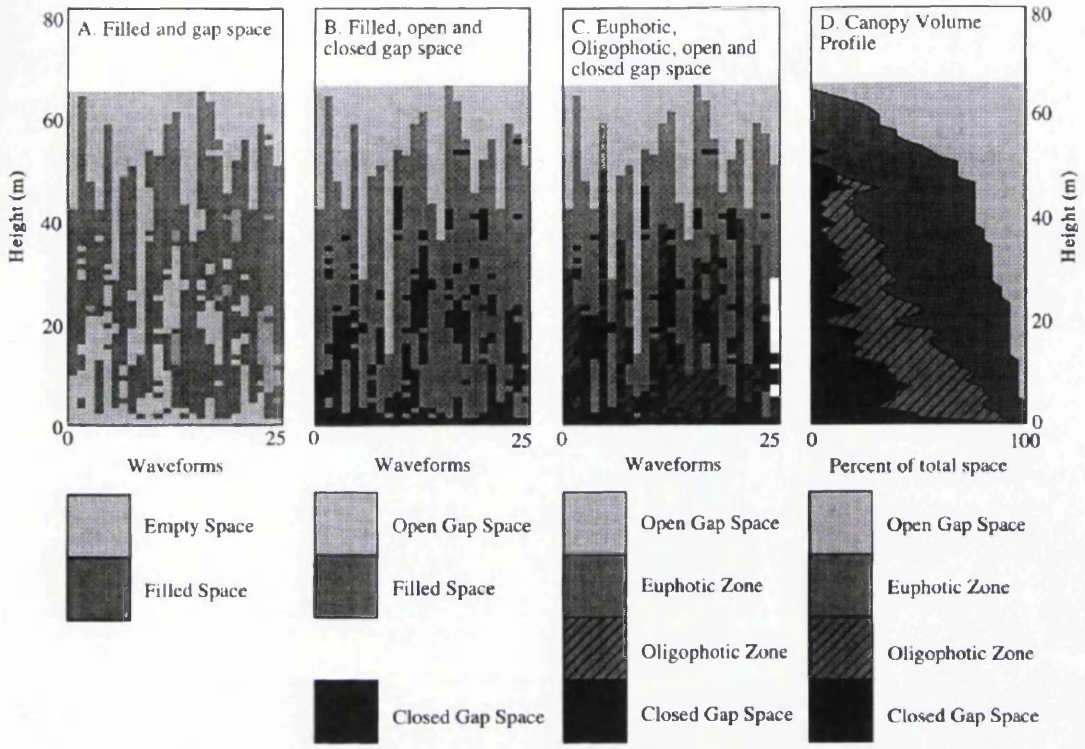
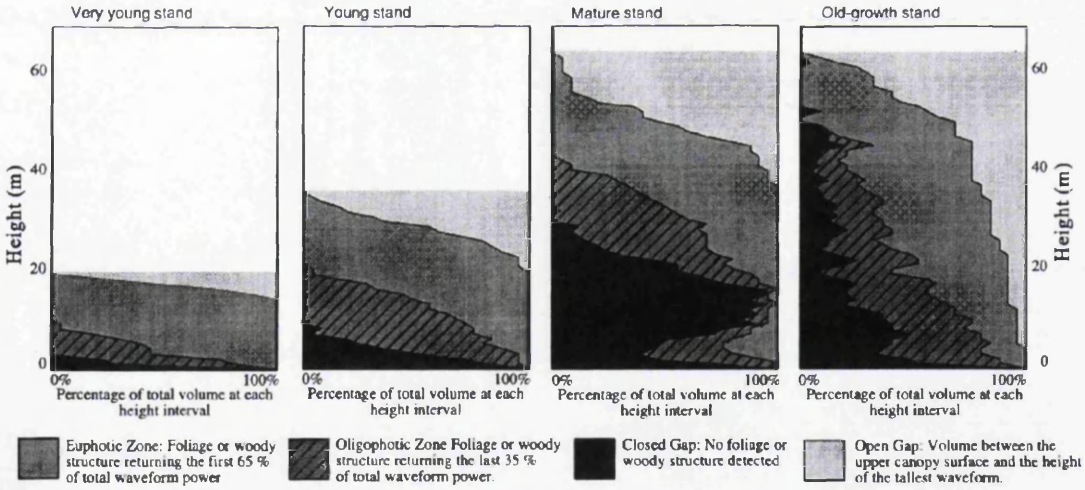


Figure 2.8: Conceptual basis of voxel based Canopy Volume Method (CVM). Source: Lefsky et al. (1999a).



a Canopy Volume Profile Concept



b Canopy Volume Profile

Figure 2.9: Example of (a) conceptual basis of Canopy Volume Method (CVM) profile construction, and (b) example of CVM profiles for very young, young, mature, and old-growth stands. Source: Lefsky et al. (1999a).

2.9b).

Figure 2.9b suggests that older canopies (mature and old-growth stands) tend to exhibit larger gaps and clumping of vegetation elements more so than younger stands. This can yield information pertaining to the probability of discrete return LiDAR systems interpreting returns from the top or side of vegetation (Rosette, 2009).

2.9.4 Vegetation Structure

Forest biophysical parameter retrieval from LiDAR data at stand level has been demonstrated by many at local scales (Hyyppä et al., 2001; Næsset, 2002; Maltamo et al., 2004; Patenaude et al., 2004; Hopkinson and Chasmer, 2007, 2009). The benefits of data fusion with spectral and/or aerial photography have been noted for surface classification (Hill and Thomson, 2005; Suárez et al., 2005), and the isolation and identification of individual trees (Koukoulas and Blackburn, 2005; Suárez et al., 2005; Holmgren et al., 2008); which act as inputs for modelling stand conditions (Morsdorf et al., 2004; Andersen et al., 2005). Furthermore, in recent years, satellite data have allowed the spatial extent of such estimations to extend from regional up to global scales (Gibbs et al., 2007; Lefsky, 2010; Simard et al., 2011; Los et al., 2012).

LiDAR has proved its capability for recording forest understorey vegetation layers (Goodwin et al., 2007; Hill, 2007), and hence offers prospects for monitoring regeneration processes. However, this has only been demonstrated with any success for small footprint LiDAR to date. A more complete characterisation of forest structure lends itself to habitat condition analysis, which in turn allows the inference of bird and animal distributions (Hinsley et al., 2006; Vierling et al., 2008).

At present only the ICESat mission has contributed directly to a near global measurement of vegetation structure parameters. Its inability to quantify such parameters with the same precision as small-scale acquired LiDAR is directly related to its (relatively) much larger footprint diameter (see Section 2.5). Future global missions aiming to quantify such parameters have notably smaller footprint diameters (some contiguous), which should help mitigate problems that come with the sizable footprints recorded by GLAS.

As LiDAR waveforms are capable of retrieving information regarding individual within canopy elements, this allows estimates of plant area index or canopy cover (as a fraction of the area under the entire waveform) to be retrieved, as suggested by North et al. (2010). However, estimation of transmittance profiles from waveform return signals does not directly account for laser light absorption by foliage (and other elements) within the canopy itself. It has been noted that for 1064 nm incident laser light, absorption by both needle and broadleaf foliage is minimal (Parker et al., 2001). Providing atmospheric attenuation can be accounted for, waveforms may contain information pertaining to direct estimates of canopy cover and LAI (Parker et al., 2001; Todd et al., 2003).

2.10 Discussion

Based on previous studies (Sun et al., 2008; Lefsky et al., 2007; Chen, 2010b; Simard et al., 2011; Rosette et al., 2013) this chapter has highlighted that waveform subtleties are suggested as one of the main causes for spurious derivations obtained from waveform LiDAR information. Such subtleties arise from numerous sources, however, terrain complexity appears as the most problematic, capable of inducing numerous sources of uncertainty in waveform derivations, as suggested by Nelson et al. (2009); Hancock et al. (2012), and Rosette et al. (2013).

Terrain complexities (or slopes) are capable of causing signal mixing within returned LiDAR waveforms as this often means that ground elements and vegetation elements occur at similar elevations. This often makes the waveform more difficult to interpret and thus increases the probability of deriving spurious parameters. As these difficulties occur in the waveform itself, *all* products derived from any difficult-to-interpret waveform are subject to great uncertainty.

As a precautionary measure, waveforms that suffer slope afflictions are often removed from data analysis because of the potential uncertainties they carry. However, whilst waveforms cannot (at present) be corrected to account for the influence of sloped terrain, if the correction of waveform derived parameters for slope is possible, waveform datasets will no longer need to be filtered for slope. The quantification of slope angle will allow this to be tested, making the first step towards correcting for (or better filtering) severely slope affected waveforms. Thus miti-

gating (or removing) waveforms that are subject to great uncertainty, increasing the accuracy of vegetation parameters derived from such data.

Additionally, as an expected function of slope, shadows found in LiDAR waveforms by Hancock et al. (2012) and Wallace et al. (2012) present an exploratory opportunity to assess and quantify the implications of such effects on metrics derived from such waveforms. At present such an effect has been noted in simulations (Hancock et al., 2012) and under (small footprint) controlled conditions (Wallace et al., 2012). The possibility of such effects occurring in large footprint LiDAR waveforms also, has the potential to affect the degree of uncertainty of vegetation knowledge at the large scale.

The reduction of uncertainty in LiDAR derived products is paramount with regards to accurately accounting for vegetation structure, timber volumes, biomass, and carbon stocks. All of which are important in establishing sustainable forest management strategies and climate mitigation.

The particular focus of this study has been identified as spaceborne waveform LiDAR, as waveform LiDAR is (at present) the only data retrieval technique offering in-depth 3-dimensional measurements of the Earth's surface at the (near) global scale. In addition, this technology was deemed ground-breaking as it offers real time direct physical measurements of surface vertical structure, unlike passive systems (which are incapable of direct vertical structure measurements), and evolutions of this technology are expected for implementation on future Earth observation missions.

2.11 Summary

This chapter has summarised the principles of two differing specifications of LiDAR systems, discrete return and continuous waveform. Applied studies of both instrument types across a multitude of disciplines have been identified. Comprehensive studies of biophysical parameter retrieval from LiDAR data have been highlighted, within the context of climate model inputs or forest management.

Topographic complexity has been identified as a capacious issue within the LiDAR community, particularly as this has the potential to affect vegetation height estimates, which has been identified as a key parameter from which many other

biophysical parameters are derived.

In this context, this project will focus on reducing uncertainty in derived products from waveform LiDAR returns observed over sloped terrain. This will be addressed as a function of slope itself, and its ability to induce other uncertainty increasing affects such as enabling shadowing presence. In addition, the first steps of correcting vegetation height as a function of these issues will also be explored.

Chapter 3

Data Collection and Processing

This chapter introduces the data collected and/or employed throughout this project. Study sites, collected field data, and remotely sensed airborne and space borne LiDAR and spectral datasets are introduced. In addition, Ordnance Survey Digital Elevation Model (DEM) data are introduced as a means of validating findings. Post-processing methods of both the airborne and spaceborne LiDAR systems are discussed, highlighting the processes undertaken to obtain outputs from raw data. In addition, slope derivations for applicable datasets are discussed.

Data Collection and Description

The first subdivision of this chapter focuses on the collection and description of the data used in this project.

A primary focus of the data collection is the need for in situ field data from different geographic locations; study sites are introduced in Section 3.1. Field data are import such to validate remotely sensed data collected from other sources, in the case of this study, airborne and spaceborne LiDAR data (with regards to vegetation height derivations). Using this knowledge, collected field measurements focused on vegetation information, particularly vegetation height (see Section 3.2).

3.1 Study Sites

A total of seven Fluxnet (2013) study sites were employed throughout this study (Figure 3.1): four boreal forest sites and three temperate or lower latitude forest sites.

Three of the boreal sites are located in Central Saskatchewan, Canada (CA), approximately 170 km north of Saskatoon, with the fourth at Norunda, approximately 20 km north of Uppsala (near Björklinge), Sweden (S).

Two temperate forest sites are located in mainland Europe; one at Loobos, approximately 64 km southwest of Amsterdam, the Netherlands (NL), and one at Tharandt, approximately 25 km southwest of Dresden, Saxony, Germany (D). The seventh site is at Tumbarumba, approximately 180 km southwest of Canberra, New South Wales, Australia (AUS).

Figure 3.1 shows the location of each site; Table 3.1 provides information and references for further reading regarding site-specific characteristics.

Canada

The three Canadian Fluxnet sites are in the Prince Albert National Park in the west, and the Narrows Hills Provincial Park in the east, Saskatchewan. These sites were established in 1993, at the start of the Boreal Ecosystem-Atmosphere Study (BOREAS; Sellers et al., 1997).

These sites are referred to according to location and the stand's dominant species; for instance the Old Aspen site, is referred to as Southern Old Aspen (SOA). The same terminology applies for the other two sites in the region: Southern Old Black Spruce (SOBS), and Southern Old Jack Pine (SOJP).

SOA is a relatively homogeneous forest with flat topography within the Prince Albert National Park (53.63° N, 106.20° W). The dominant species is Trembling Aspen (*Populus tremuloides*; see Table 3.1) up to 94 years of age (Kljun et al., 2007), with a mean vegetation height reported as 21 m as of 2001 (McCaughey et al., 2002). There is a relatively dense Hazelnut (*Corylus cornuta*) understorey present, contributing to a relatively dense canopy structure (LAI ~ 5.6 m^2m^{-2}; Kljun et al., 2007).

SOBS is also a relatively homogeneous forest with flat topography, also within the Prince Albert National Park. The site is dominated by black spruce (*Picea mariana*), with scattered Tamarack (*Larix laricina*; McCaughey et al., 2002), located north east of SOA (53.99° N, 105.12° W). The stem density here is the greatest of all the Canadian study sites, contributing to a relatively dense canopy (LAI $\sim$ 4.2 m^2m^{-2}; Kljun et al., 2007). In addition, SOBS is very slow growing and the oldest of the Canadian study sites, at 134 years old, where vegetation height was reported as 11 m for Black Spruce, and ranging between 10 m and 16 m for Tamarack as of 2001 (McCaughey et al., 2002).

SOJP site is also a relatively homogeneous, flat topography forest, located east of SOBS, in the Narrow Hills Provincial Park (53.92° N, 104.69° W). The site is approximately 84 years of age, dominated by Jack Pine (*Pinus banksiana*), with a mean vegetation height of 13 m as of 2001 (McCaughey et al., 2002; Kljun et al., 2007). The canopy at SOJP is the most open of the Canadian study sites, resulting in a low density canopy structure (LAI $\sim$ 2.4 m^2m^{-2}) (Kljun et al., 2007).

Europe

The three European sites have been part of Euroflux, CarboEurope and are now part of the recently established Integrated Carbon Observation System (ICOS) network (Paris et al., 2012). Of the three European sites, two are temperate/mid-latitude and one ia a boreal forest site. The three sites are Loobos in the Netherlands, Tharandt in Germany, and Norunda in Sweden.

Loobos, the Netherlands, is a homogeneous, temperate forest site located just north of the Hoge Veluwe National Park (52.10°N, 5.44°E), with relatively flat topography, with small-scale (minor) undulations. The 110 year old stand is based on a sandy soil and dominated by Scots Pine (*Pinus sylvestris*) with a mean vegetation height of $\sim$ 15 m as of 2001 (Dolman et al., 2002). The canopy is relatively open with some large gaps between vegetation, resulting in a low density canopy structure (LAI $\sim$ 1.9 m^2m^{-2}; Dolman et al., 2002).

Tharandt is a managed, temperate, heterogeneous forest with undulating terrain located in Saxony, Germany (50.57°N,13.34°E). The site is a mixed coniferous (87%) and deciduous (13%) forest dominated by Norway Spruce (*Picea abies*;

Grünwald and Bernhofer, 2007). The stand is $\sim$ 120 years old, with a mean canopy height reported at 26 m in 1999 (Grünwald and Bernhofer, 2007). There is a relatively dense understorey present contributing to a dense canopy structure (LAI $\sim$ 7.6 m^2m^{-2}; Grünwald and Bernhofer, 2007).

Norunda is a relatively homogeneous coniferous forest site with flat, but rough topography situated in the Swedish boreal zone (60.50°N, 17.29°E). The forest is dominated by Norway Spruce (*Picea abies*; 33%) and Scots Pine (*Pinus sylvestris*; 65%), with a small deciduous (2%) element (Lindroth et al., 1998; Feigenwinter et al., 2010). The 100 year old forest had reported vegetation heights ranging between 24 m and 28 m as of 2008 (Feigenwinter et al., 2010). The canopy is relatively open for the majority of the forest, but is dense in areas, leading to a relatively dense canopy structure (LAI $\sim$ 4.5 m^2m^{-2}; Lindroth et al., 1998; Feigenwinter et al., 2010).

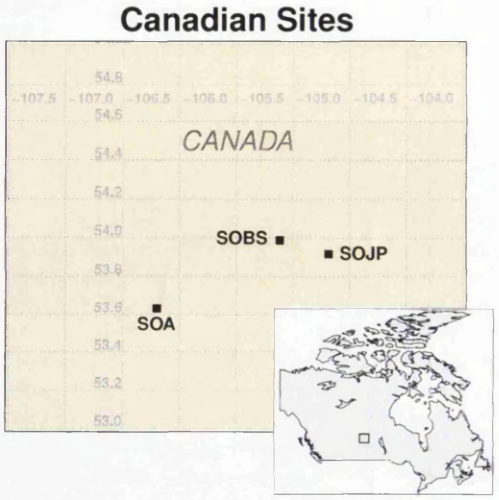
Australia

Tumbarumba is an OzFlux site located in Bago State Forest, New South Wales (35.39° S, 148.09° E). However, the data employed in this project were collected approximately 7 km to the east of this location with the primary focus of coinciding with over 600 measurements made by GLAS, which of utmost importance in this study. Whilst the following site characteristics are sourced from literature describing the OzFlux site itself, forest characteristics at both locations are very similar in all regards.

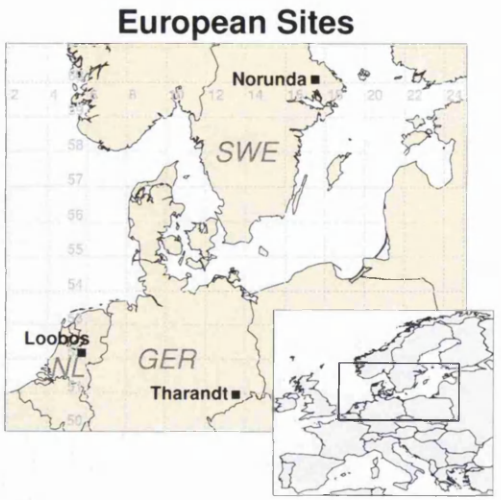
The forest is a heterogeneous temperate/mid-latitude forest dominated by two species of Eucalyptus (*Eucalyptus delegatensis* and *Eucalyptus dalrympleana*) with a mean height of 40 m (see Table 3.1; van Gorsel et al., 2008). The stand is of mixed age, with the maximum age thought to be approximately 90 years old (van Gorsel et al., 2008). The forest is relatively open, with a very open understorey present, contributing to a low density canopy structure (LAI $\sim$ 1.4 m^2m^{-2}; Leuning et al., 2005). The vegetation heterogeneity at the site is largely attributed to terrain complexities (Leuning et al., 2005; van Gorsel et al., 2008).

The complex terrain at the site is a consequence of its location in a mountainous region with a mean elevation of approximately 1200 m above mean sea level (amsl).

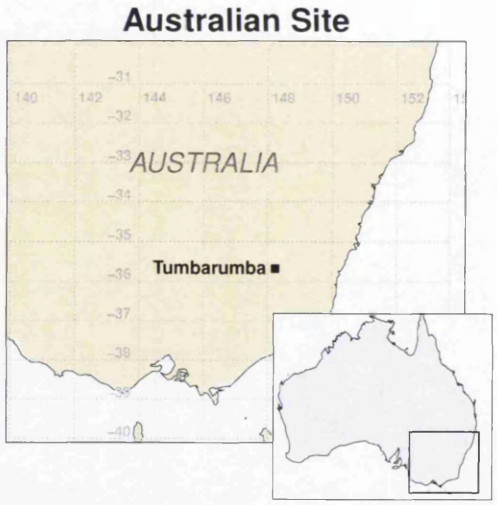
Slopes in the region are up to (and exceed) $\sim 30°$.



a Canada



b Europe



c Australia

Figure 3.1: Locations of each of the seven sites around the globe.

Table 3.1: Brief summary of site characteristics, canopy and understorey vegetation with additional references for further information. Note: E = elevation, and $\bar{V}_H$ = Mean vegetation height.

Site	E[m]	$\bar{V}_H$[m]	Canopy species	Understorey species	Reference(s)
SOA (CA)	601	21	Trembling Aspen *(Populus tremuloides)*	Hazelnut *(Corylus cornuta)*	McCaughey et al. (2002) Chen et al. (2006) Gaumont-Guay et al. (2009)
SOBS (CA)	629	11–16	Black spruce, Tamarack *(Picea mariana, Larix laricina)*	Sphagnum, Feathermoss *(Sphagnum fuscum, Pleurozium)*	McCaughey et al. (2002) Chen et al. (2006) Gaumont-Guay et al. (2009)
SOJP (CA)	579	13	Jack Pine *(Pinus banksiana)*	Reindeer lichen, Green alder *(Cladina mitis, Alnuscripspa)*	McCaughey et al. (2002) Chen et al. (2006) Gaumont-Guay et al. (2009)
Loobos (NL)	25	17	Scots pine *(Pinus sylvestris)*	Wavy hair grass *(Deschampsia flexuosa)*	Dolman et al. (2002)
Tharandt (D)	380	26	Norway spruce *(Picea abies)*	Wavy hair grass *(Deschampsia flexuosa)*	Grünwald and Bernhofer (2007)
Norunda (S)	45	24–28	Norway spruce, Scots pine *(Picea abies, Pinus sylvestris)*	None	Lindroth et al. (1998) Lagergren et al. (2004) Feigenwinter et al. (2010)
Tumbarumba (AUS)	1200	40	Alpine ash, White mountain gum *(Eucalyptus delegatensis, Eucalyptus dalrympleana)*	None	Leuning et al. (2005) van Gorsel et al. (2008)

3.2 Field Work

In situ field data were collected at all flux sites by local field teams for the Canadian and Australian sites respectively. Measurements were collected for the European sites by a field team as part of this study.

Field data were collected in order to validate derived products from remotely sensed data, namely airborne LiDAR vegetation metrics. Considering this, maximum vegetation height, tree density, Leaf Area Index (LAI), and fractional cover (F_{cov}) were of particular interest for this study.

A standard method of data collection was adopted across all study sites regarding plot positions with respect to the central flux-tower location. Field plots were established 100 m from the flux tower in the cardinal directions, and 300 m from the flux tower in the ordinal directions, as illustrated in Figure 3.2.

A field plot radius (r) of either 11.3 m or 13.8 m was established (dependent on site and tree density), sampling a subset of the forest. The exact location and associated radius of each field plot was derived by means of a differential global positioning system (GPS), loaned from the Natural Environment Research Council (NERC) funded Geophysical Equipment Facility (GEF) based at Edinburgh University.

Field data were collected for the majority of field plots, with few exceptions due to inaccessibility. These data are compared with airborne LiDAR equivalent results, namely maximum vegetation height. For this purpose, airborne LiDAR derived 1 m spatial resolution, spatially concurrent maximum vegetation height information was compared with collected field data. These results are shown in Appendix 9.1, indicating that the measures produced from airborne LiDAR data accurately represent the forest sites.

The field data collected at each site are summarised in Table 3.3.

Canada

Field measurements for all BOREAS sites were recorded in the summer of 2008, within a few days of the airborne LiDAR campaign, thus minimising the physical difference between data recordings. A total of 8 field plots were established for

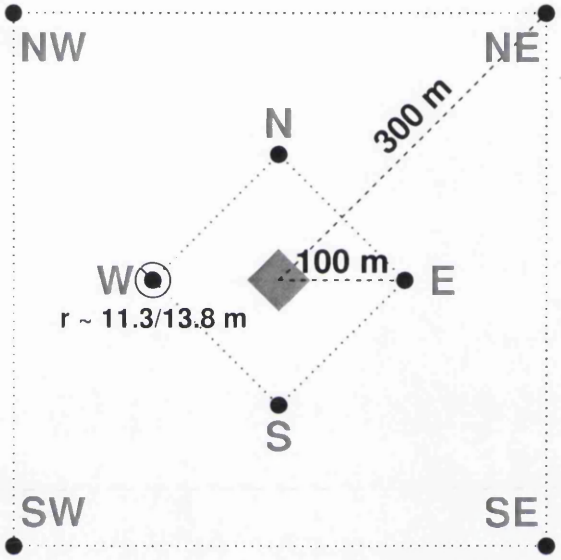


Figure 3.2: Map overview of the positions of established field plots at each study site with respect to the central flux tower (grey diamond). Eight plots were established, the centres of each represented by black dots, which have a radius (r) of either 11.3 m or 13.8 m (see W symbol). Plots along the cardinal directions were set 100 m away from the flux tower, whereas 300 m was chosen for those plots set along the ordinal directions.

SOJP, under the outlined field collection technique, whereas only 3, and 4 field plots were set for SOA and SOBS, respectively, due to lack of personnel during the field data collection period. For each established plot a radius of 11.3 m was used.

Europe

Field measurements were obtained at Loobos and Tharandt during the summer of 2010, where calibration measurements were taken coincidentally to the acquisition of the airborne LiDAR data. Vegetation measurements were acquired within a few days of the airborne LiDAR campaign.

The majority of field measurements at Norunda were made during the summer of 2010 also, however, due to poor weather conditions the airborne LiDAR campaign was abandoned at that time and flown almost a year later during the summer of 2011, when the outstanding field measurements were obtained. This temporal difference, although not ideal, can be neglected as the forest at Norunda is known to be very slow growing, hence little or no difference between airborne LiDAR derived vegetation height estimates and the previous year's field measurement acquisitions is expected. Field plot measurements repeated in 2011 at some of the same locations as in 2010 confirm this.

A total of 8 field plots were established for all sites in Europe, where a radius of 13.8 m was established for most plots.

Australia

Field measurements for Australia were recorded at the OzFlux site, not at the location employed within this study, in the (southern hemisphere) summer of 2009. Airborne LiDAR data were acquired at the OzFlux site as well as at the location 7 km east during the same flight. Field data were collected within days of the airborne LiDAR campaign. A total of 8 field plots were established with a radius of 13.8 m, following the outlined field collection method.

These field observations, whilst not applicable to the specific location of interest, 7 km east of the OzFlux site, act as a validation of the acquired airborne LiDAR data accuracy.

3.3 Airborne LiDAR

A total of seven different airborne LiDAR campaigns were employed in this study, at four boreal forest sites and three temperate or lower latitude sites. Data at all sites within the same continent were collected by the same equipment, but different equipment was used on each continent. Information on each airborne LiDAR campaign is summarised in Table 3.3.

All airborne LiDAR campaigns took place during clear summer days of their respective years in order to minimise the low-lying cloud cover which is measurable by such systems (Spinhirne et al., 1982; Sassen and Benson, 2001). Additionally, restricting data acquisition periods to the summer only, will ensure that all sites are collected in leaf-on conditions, this is especially important for those forests with a substantial broad-leaf component and understorey.

The mean density of airborne LiDAR (all) returns for each site is summarised under the appropriate headings below. The proportion of points that are 1^{st}, 2^{nd} etc. returns are given in Table 3.2; statistics are based on individual Log ASCII Standard (LAS) files, representing individual flightlines. Statistics are in keeping with site characteristics, particularly noticable at SOA which exhibits the greatest percentage of 3^{rd} returns of all sites, suggesting understorey presence.

Table 3.2: Summary of Airborne LiDAR return statistics by site. Note: N = Mean number of returns per LAS file (flightline), Px = Mean percentage of N that are return number x within a given pulse.

Site	N	P1 [%]	P2 [%]	P3 [%]	P4 [%]	P5 [%]
Loobos	18759642	84.54	14.26	1.17	0.03	0
Norunda	24310869	78.82	19.53	1.6	0.04	0
SOA	12249316	56.21	33.8	9.35	0.65	0
SOBS	7013355	70.87	26.08	2.94	0.11	0
SOJP	16868062	67.15	28.9	3.81	0.14	0
Tharandt	22786827	76.92	17.38	4.52	1.18	0
Tumbarumba	3927948	62.76	29.24	6.97	0.93	0.08

Canada

The Canadian airborne LiDAR data were acquired in August 2008, by the Applied Geomatics Research Group (AGRG), using an Optech Incorporated ALTM 3100. This is a discrete multiple return system capable of recording up to four returns per emitted laser pulse; for further details see Optech (2013).

Data were acquired at a mean altitude of $\sim$ 950 m over the three sites, with a common laser pulse emission rate of 70 KHz, collected over multiple flight lines. Each flight line constitutes data collected across the full swath width range of the aircraft mounted laser instrument (where swath width is defined as in section 2.6.1). Flight lines were flown with a 50% overlap for high-resolution data, which resulted in an approximate point density of 4 returns per square metre (pm^{-2}).

Europe

The European airborne LiDAR data were acquired in June 2010, for both Loobos and Tharandt, and a year later for Norunda in June 2011, due to unsuitable weather conditions during the original flight window in 2010. Data were collected by the Natural Environment Research Council Airborne Research and Survey Facility (NERC ARSF), using a Leica ALS50-II LiDAR system, a multiple, discrete return recording instrument with a waveform upgrade as of 2011; for details see ARSF (2013).

Data were acquired at a mean above ground altitude of 1900 m, 2250 m, and 1550 m for Loobos, Tharandt and Norunda respectively, with respective laser pulse repetition rates of 127 KHz, 124 KHz, and 150 KHz. Flight lines were flown with a 50% overlap to increase return resolution, which was reported at a mean value of $\sim$ 5–6 pm^{-2} for all three sites.

Australia

Airborne LiDAR data were acquired at Tumbarumba in November 2009 by Airborne Research Australia (ARA), using a Riegl LMS-Q560 continuous waveform airborne laser scanner, for details see Riegl (2013). Waveform data here were post-processed into discrete point cloud information by ARA. This technique,

transforming waveform information to point cloud, has been suggested to offer greater detail of vegetation structure than is retrievable from standard discrete return systems (Rutzinger et al., 2008; Wagner et al., 2008).

Data were collected at a repetition rate of ~100 KHz, which resulted in a point return density of approximately of 4–5 pm^{-2}. Data were collected with a 50% flightline overlap for high density coverage. An absolute above mean sea level (amsl) altitude of 1500 m was maintained by the aircraft over the entire region, resulting in an average above-ground altitude of $\sim$ 900 m. The minimum and maximum above-ground altitudes were $\sim$ 350 m and $\sim$ 1500 m respectively, resulting from the terrain complexity.

Table 3.3: Summary of in-situ field data, airborne LiDAR equipment/data, and weather conditions for each study site. Note: FW_{Date} = date of field work collection, AB_{Date} = date of airborne LiDAR collection, P_D = point density, Software = data processing software.

Site	FW_{Date}	Data Collected	AB_{Date}	Instrument	P_D	Conditions	Software
SOA (CA)	August **2008**	V_H, LAI	August **2008**	Optech ALTM 3100	4 pm^{-2}	Scattered cloud	BCAL
SOBS (CA)	August **2008**		August **2008**				
SOJP (CA)	August **2008**		August **2008**				
Loobos (NL)	June **2010**	V_H, LAI, V_D, DBH	June **2010**	Leica ALS50-II	5 pm^{-2}	Cirrus clouds	BCAL
Tharandt (D)	June **2010**		June **2010**		5 pm^{-2}	Scattered cloud	
Norunda (S)	September **2010**/June, **2011**		June **2010**		6 pm^{-2}	Clear	
Tumbarumba (AUS)	November **2009**	V_H, LAI	November **2009**	Riegl LMS-Q560	5 pm^{-2}	Clear	BCAL

Note: V_H = vegetation height, V_D = field plot vegetation density, LAI = leaf area index, DBH = tree diameter at breast height, pm^{-2} = points per square metre

3.4 ICESat/GLAS

The Ice, Cloud, and land Elevation Satellite (ICESat) housed the Geoscience Laser Altimeter System (GLAS), which operated with the primary objective of determining inter-annual and long-term changes to ice-sheet volume (Zwally et al., 2002). However the multi-disciplinary application of GLAS was noted by many and thus applied for research on: surface topography, hydrology, vegetation canopy metrics, cloud heights, and aerosol distributions (Zwally et al., 2002; Schutz et al., 2005; Chen, 2010b; Los et al., 2012). At present, ICESat has been the only satellite mission of its kind, offering high vertical accuracy, typical of LiDAR systems, at an almost global scale.

3.4.1 Operation and Data Storage

GLAS was operated in a 600 km orbit, aboard ICESat between 2003 and 2009, where almost one billion full waveform measurements of the Earth's surface were collected by 2005 (Abshire et al., 2005). GLAS actively sampled the Earth's surface between ±86° latitude by means of three different lasers, each fired at strategic times (Schutz et al., 2005; Rosette et al., 2008). Each laser recorded in the 532 and 1064 nm wavebands at a repetition rate of 40 Hz (Zwally et al., 2002).

As a consequence of the high orbit, large geolocated footprints were produced (laser campaign dependent, see below). Considering all lasers, a mean footprint diameter of 64 m was produced, each separated by an almost constant 170 m (Zwally et al., 2002; Rosette et al., 2008).

Strategic laser firing was introduced for GLAS after its launch, when it was noted that the pumping energy required for firing each of the lasers diminished rapidly (Schutz et al., 2005). As a result, GLAS was fired for shorter continual (typically month long) periods, known as laser campaigns, typically 2–3 times per year (Zwally et al., 2002; Harding and Carabajal, 2005). This was executed with the aim to repeat orbital tracks over many campaigns with the goal to monitor change with time.

Each laser campaign was abbreviated in the form LXY; where X refers to the laser number (1, 2, or 3), and Y is a letter denoting sequential data acquisitions for

individual lasers. For example, laser 3 acquired data for the fourth time between 21st October and 24th November, 2005, and is hence abbreviated to L3D.
Collected data are stored and post-processed at the National Snow and Ice Data Center (NSIDC), where they are freely available for download. Raw data are processed into data releases, which are essentially post-processed versions of raw data (release 33 is used in the current study, see Zwally et al., 2011). Each release consists of unique data products, refined for complementary data (abbreviated as GLAXX, where XX ranges from 01-15), see Table 3.4 for data product descriptions. Each subsequent data release is a revised version of its predecessor, hence eliminating known issues with certain data and revising uncertainty in certain measurements; further information regarding data products see http://nsidc.org/data/icesat/data.html.

Table 3.4: Summary of the 15 differing data products available from the NSIDC. Note: L1A, L1B, and L2 description heading refer to data processing levels, not laser campaign identifiers.

Product	**Description**
GLA01	L1A Global Altimetry Data
GLA02	L1A Global Atmosphere Data
GLA03	L1A Global Engineering Data
GLA04	L1A Global Laser Pointing Data
GLA05	L1B Global Waveform-based Range Corrections Data
GLA06	L1B Global Elevation Data
GLA07	L1B Global Backscatter Data
GLA08	L2 Global Planetary Boundary Layer and Elevated Aerosol Layer Heights
GLA09	L2 Global Cloud Heights for Multi-layer Clouds
GLA10	L2 Global Aerosol Vertical Structure Data
GLA11	L2 Global Thin Cloud/Aerosol Optical Depths Data
GLA12	L2 Antarctic and Greenland Ice Sheet Altimetry Data
GLA13	L2 Sea Ice Altimetry Data
GLA14	L2 Global Land Surface Altimetry Data
GLA15	L2 Ocean Altimetry Data

source: NSIDC (2012a)

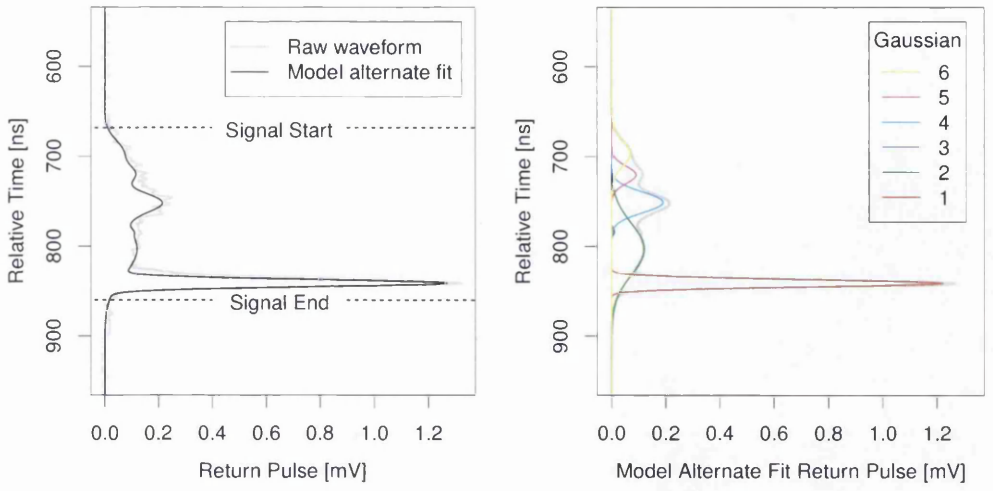
3.4.2 Pulse Emission and Detection

GLAS was capable of emitting a Gaussian pulse waveform in the 532 and 1064 nm wavebands simultaneously, the width of which was approximately 1 m at its full width at half maximum (equivalent to a ~5–6 ns emission duration; Rosette et al., 2008; Los et al., 2012). Whilst in transit, between emission and redetection, the pulse width remains constant. Upon arrival at the detector, the returned pulse is measured for a variable duration equivalent length of either 30. 82, or 150 m (NSIDC, 2012a), dependent on laser campaign, at 0.15 m (0.5 ns) intervals (Los et al., 2012; NSIDC, 2012a). At this time the returned waveform contains information pertaining to individual intersected target reflective properties, and scattering events that may have occurred at the target surface.

The emitted laser intensity (energy) distribution is also Gaussian shaped, hence the greatest intensity occurs at the footprint centre and decreases radially outwards. The area illuminated by the laser at the target surface is referred to as a footprint, which varies in shape and size (see Section 3.4.3) according to laser campaign. The footprint region is restricted up to the points at which the intensity distribution falls below a $1/e^2$ threshold (where e = Euler's constant $\simeq 2.718$), equivalent to 13.5% of its maximum (Harding and Carabajal, 2005).

As a consequence of a Gaussian like, radially decreasing energy distribution, each footprint is most representative of interactions that occur nearest its centre (Rosette et al., 2008). This was noted by Chen (2010b), where a considerable improvement in correlation was found between GLAS and airborne derived vegetation heights with increasingly restricted footprint diameter. A mean correlation of R=0.89 increased to R=0.98 as the compared GLAS and airborne LiDAR data areas were restricted from a mean diameter of 40 m down to 10 m.

For vegetated surfaces, the returned waveform typically contains information on two major features, vegetation and ground. The vegetation component contains information on scattering events and interactions dependent on vegetation structure. The ground component is generally only a singular peak from an unbroken surface, but local topographic complexities can induce peak distortion. Distortions in the ground component can occur due to slope, terrain roughness or highly reflective surfaces such as water, where signal saturation can occur.



a Model Fit

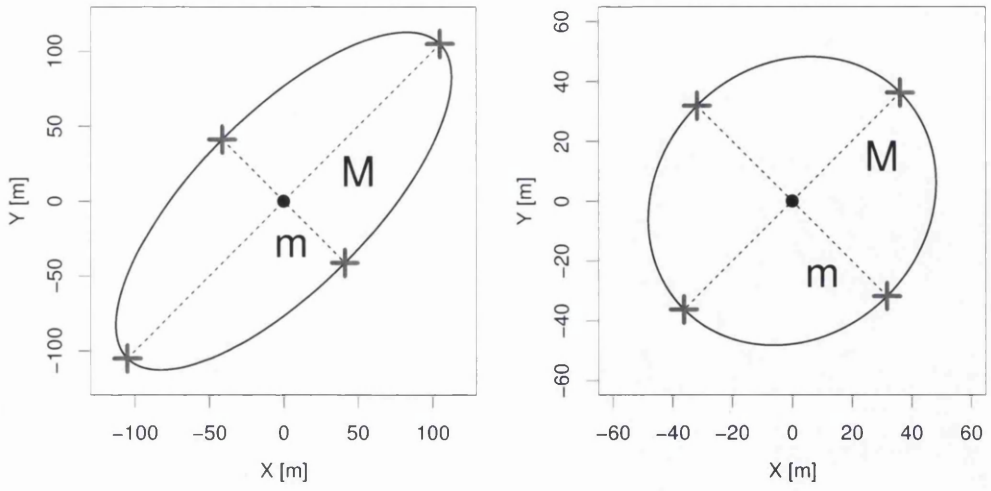
b Gaussian Decomposition

Figure 3.3: Example of a GLAS waveform return over a vegetated area in Wales with (a) modelled waveform fit from (b) six fitted Gaussians to the raw waveform.

The returned waveform is analysed and fitted with up to 6 Gaussians (Brenner et al., 2003; Figure 3.3b) according to the fitting algorithm described by Duong et al. (2006). This allows a model alternate fit return pulse (standard terminology) to be established by summing the intensities of each fitted Gaussian as a function of relative time (which can be converted to range), and the signal start and end locations to be identified (Figure 3.3a). Gaussian decomposition of returned waveforms has proven popular in the derivation of maximum canopy height from GLAS waveforms by many (Duong et al., 2008; Rosette et al., 2008; Sun et al., 2008; Duong et al., 2009; Chen, 2010a; Los et al., 2012).

3.4.3 Footprints

Footprint shape and dimensions are almost unique for each laser campaign. Some footprints were almost circular, whereas others were highly eccentric ellipses with varying semi-major and semi-minor axes (Figure 3.4). The NSIDC (2012a) published semi-major axis values ranging from 148.6 ± 9.8 m to 51.2 ± 1.7 m. Only semi-major axis information are available from the NSIDC; however, the semi-



a L1A b L3F

Figure 3.4: Nadir view of GLAS footprints from campaigns L1A and L3F, demonstrating the differences in shape and size of each footprint's coverage region dependent on campaign. Note: M = semi-major axis length, and m = semi-minor axis length.

minor axis can be calculated by Equation 3.1, using the semi-major axis and associated footprint eccentricity. A full listing of GLAS footprint semi-major axes and eccentricities are given in Table 3.5.

$$S_{minor} = S_{major}\sqrt{1 - e^2} \tag{3.1}$$

The pointing knowledge accuracy is unique to each laser campaign also, this pertains to the laser orientation when fired from the satellite. The smallest misalignment in laser direction translates to large horizontal and vertical geolocation uncertainties at the target surface. For example, data release 23 exhibited a horizontal geolocation accuracy as poor as 37.7 ± 53.4 m in some extreme cases. However, data release 33, used in this study, has evolved allowing for much greater accuracy associated with horizontal and vertical geolocation, up to 4.6 ± 15.6 m and 3.2 ± 10.9 m, respectively (NSIDC, 2012a). Campaign associated pointing knowledge is given in Table 3.5.

Table 3.5: Summary of the change of physical parameters from data release 33 (with associated uncertainties) for differing GLAS acquisition periods (laser campaigns) affecting footprint size and shape. Note: e = eccentricity, S_M = semi-major axis, Δ_V = vertical pointing knowledge accuracy, and Δ_H = horizontal pointing knowledge accuracy.

Campaign	**Acquisition Period**	**e**	**S_M[m]**	**Δ_V[m]**	**Δ_H[m]**
L1A	20/02/2003–21/03/2003	0.920 ± 0.017	148.6 ± 9.8	0.60 ± 3.30	0.79 ± 4.77
L1B	21/03/2003–29/03/2003	NA	NA	NA	NA
L2A 8-day	25/09/2003–04/10/2003	0.874 ± 0.001	86.7 ± 0.7	0.20 ± 2.94	0.33 ± 4.21
L2A Pre 91-day	04/10/2003–13/10/2003	0.877 ± 0.003	91.4 ± 3.6	0.20 ± 2.94	0.33 ± 4.21
L2A Post 91-day	13/10/2003–19/11/2003	0.884 ± 0.012	105.3 ± 4.3	0.20 ± 2.94	0.33 ± 4.21
L2B	17/02/2004–21/03/2004	0.820 ± 0.050	89.8 ± 5.1	0.70 ± 2.60	0.93 ± 3.73
L2C	18/05/2004–21/06/2004	0.892 ± 0.044	88.4 ± 19.1	0.30 ± 7.20	0.37 ± 10.30
L3A	03/10/2004–08/11/2004	0.570 ± 0.040	55.8 ± 0.4	0.40 ± 2.30	0.56 ± 3.29
L3B	17/02/2005–24/03/2005	0.750 ± 0.050	79.3 ± 11.6	0.10 ± 2.93	0.07 ± 4.20
L3C	20/05/2005–23/06/2005	0.630 ± 0.030	55.4 ± 1.8	0.20 ± 2.04	0.29 ± 2.92
L3D	21/10/2005–24/11/2005	0.520 ± 0.010	52.0 ± 1.1	0.10 ± 2.00	0.07 ± 2.86
L3E	22/02/2006–28/03/2006	0.483 ± 0.040	52.3 ± 1.6	0.00 ± 2.38	0.00 ± 3.41
L3F	24/05/2006–26/06/2006	0.479 ± 0.025	51.2 ± 1.7	1.00 ± 3.08	1.35 ± 4.42
L3G	25/10/2006–27/11/2006	0.510 ± 0.040	53.4 ± 1.5	0.10 ± 2.35	0.07 ± 3.37
L3H	12/03/2007–14/04/2007	0.520 ± 0.020	55.6 ± 0.5	0.00 ± 3.00	0.00 ± 4.29
L3I	02/10/2007–05/11/2007	0.590 ± 0.010	57.3 ± 0.6	0.50 ± 1.55	0.65 ± 2.22
L3J	17/02/2008–21/03/2008	0.520 ± 0.040	58.7 ± 0.6	0.40 ± 3.25	0.62 ± 4.66
L3K	04/10/2008–19/10/2008	0.520 ± 0.040	58.7 ± 0.6	0.10 ± 2.61	0.15 ± 3.74
L2D	25/11/2008–17/12/2008	0.520 ± 0.040	58.7 ± 0.6	1.20 ± 5.14	1.72 ± 7.36
L2E	09/03/2009–11/04/2009	0.520 ± 0.040	58.7 ± 0.6	1.60 ± 5.59	2.32 ± 8.01
L2F	30/09/2009–11/10/2009	0.520 ± 0.040	58.7 ± 0.6	3.20 ± 10.92	4.63 ± 15.64

source: NSIDC (2012a)

3.5 MODIS Spectral Data

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard NASA's Aqua and Terra Earth Observing Satellites (EOS), capable of spectrally imaging the entire Earth's surface every 1–2 days in 36 spectral bands (NASA, 2013b).

For the purposes of this project, related vegetation products from MODIS were employed. These were the 0.5 km $\times$ 0.5 km spatial resolution vegetation continuous fields (VCF) (Hansen et al., 2003; Townshend et al., 2011), and Normalised Difference Vegetation Index (NDVI; Huete et al., 2002, 2011) products.

The VCF product is a measure of the percentage of vegetation coverage at a specific location, where NDVI, by its definition, can be viewed as a measure of how green vegetation is. NDVI is defined as:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{VIS}}}{\rho_{\text{NIR}} + \rho_{\text{VIS}}} \tag{3.2}$$

Where ρ_{NIR} and ρ_{VIS} are spectral reflectance measurements from the near-infrared (approximately 700–1100 nm wavelengths) and visible (approximately 400–700 nm wavelengths) regions of the electromagnetic spectrum.

The pigments in live green plant leaves, chlorophyll, strongly absorb in the visible part of the spectrum, where this radiation is used in photosynthesis. At the same time the cell structure of the leaves strongly reflects near-infrared radiation. By this definition (Equation 3.2), the greater the value of NDVI, the more live green vegetation is present in the observation window.

As ρ_{NIR} and ρ_{VIS} are ratios of reflected and incident radiation for individual spectral bands, they are restricted to values between 0–1. This means that NDVI itself varies between -1 and 1.

3.6 Ordnance Survey Elevation Data

The Ordnance Survey (OS) is the national mapping agency for Great Britain, established in 1791. The OS provides a wealth of digital and paper-based products ranging from business and educational mapping to leisure maps covering 100% of

the UK mainland (excluding Northern Ireland).

Ordnance Survey data were used in this study for the purpose of independently testing respective methods for inferring slope values from waveform LiDAR (as discussed in Chapter 5). OS data were acquired for the entire area of Wales, a region chosen for its topographic variability (see Figure 3.5). Note, Wales is not employed as a *study site*, but as a means of independently testing methodological principles under highly variable topographic conditions. See Chapter 5 for further details on the use of OS data.

The Land-Form PROFILE Digital Elevation Model (DEM) data were employed. This dataset is based on the Land-Form PROFILE Contours dataset, which maps terrain elevation at a contour interval of 5 m for the majority of the country, where a 10 m interval is used in some mountain and mooreland areas (EDINA, 2013). Digital contour accuracies are typically better than half of the contour interval, i.e. ± 2.5 m for 5 m, and ± 5 m for 10 m vertical contour intervals. The DEM itself is a mathematically derived product consisting of a 10 m resolution grid of elevation values interpolated from the contour dataset. Accuracy varies according to the complexity of terrain, but in general is the same as the contour dataset from which the product is derived. The uncertainty is OS data is deemed acceptable due to model assumptions when inferring slope values (see Chapter 5).

Data are provided by http://edina.ac.uk/digimap projected in British National Grid format, where the most up-to-date (2013) data were used in this study.

3.7 Shuttle Radar Topography Mission Data

The Shuttle Radar Topography Mission (SRTM) was a NASA mission during 2011, where global elevation data were collected between $\pm 60°$ latitude. SRTM elevation data is available at a 90 m resolution (at the equator), available in $5° \times 5°$ seamless tiles distributed by the Consultative Group for International Agriculture Research - Consortium for Spatial Information (CGIAR - CSI). Data are reported with a vertical rmse of up to 8 m, constituting the best global elevation dataset ever realised.

For this study the interpolated version of the SRTM DEM 4.1 data is used, in which any "no-data" holes (caused by water or heavy shadowing) have been addressed

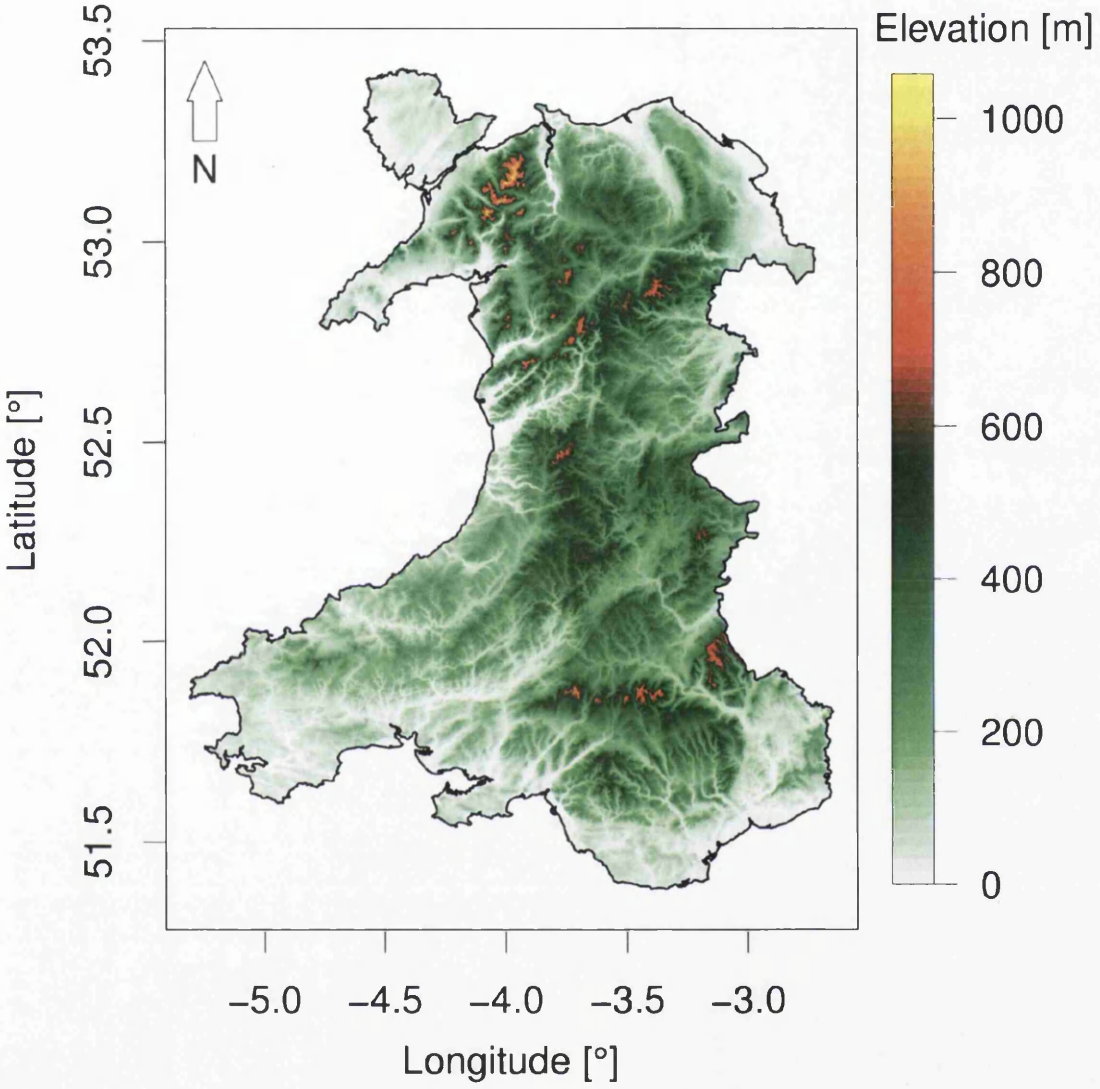


Figure 3.5: Elevation map of Wales, illustrating topographic variability across the region.

by the method described by Rodriguez et al. (2005) and Jarvis et al. (2008).

Data Processing

The second subdivision of this chapter focuses on the processing of collected data. The processing methods for the airborne and spaceborne LiDAR are outlined, in addition to how slope values were derived for each dataset, and from the Ordnance Survey DEM dataset. All data manipulation and/or processing was executed using the R-project for statistical computing (R Core Team, 2013) unless otherwise stated.

3.8 Airborne LiDAR Processing

Airborne LiDAR point cloud data were provided in Log ASCII Standard (LAS) files and were analysed using Idaho State University's Boise Center Aerospace Laboratory LiDAR Tools (BCAL, 2011). BCAL performs analysis on the LAS point cloud data directly, filtering, classifying, and resolving outputs. BCAL processes are reliant on each other, hence for vegetation metric retrieval the recommended work flow is:

- Tile data (if LAS files > 200 MB)
- Buffer data (overlapping data at tile edges to give a seamless transition between adjacent tiles)
- Height filter data to classify ground/non-ground returns according to the algorithms of Streutker and Glenn (2006)
- Create outputs (rasters of height quantiles, DEM etc.)

The height filtering process for vegetated areas requires some forest information, physical thresholds, and methodological details. Forest information such as the the canopy spacing (CS) and vegetation height limit (H_{max}) are required for use as a measure of how open the canopy is, and a maximum limit to which vegetation heights are expected to reach respectively.

The threshold for distinguishing ground features (T_G) is set in accordance with the level of detail requested from ground returns, such as distinguishing ridges or rock outcrops; the greater the threshold, the more detail is preserved. In order to generate more ground returns a higher value of T_G is used, further details are available from Streutker and Glenn (2006). The maximum number of iterations (I_{max}) is set as the maximum number of iterative increments. This becomes more important as more detail is requested from the resulting outputs, such as the ground.
Methodological inputs such as the return type, and interpolation method define whether the use of all LiDAR returns or specific returns (ground only, non-ground only) should be used, and the method to be used for the interpolation process. Further details available from BCAL (2011).
For this study, CS, H_{max}, and I_{max} varied between sites due to differing vegetation and topographic characteristics. Specifically, CS and H_{max} were chosen according to each site's vegetation characteristics, and I_{max} according to local terrain complexity. Exact values for these variable parameters are given in Table 3.6 on a site specific basis.
T_G, return type, and interpolation method were held constant. All return types (ground and non-ground) were considered for "natural neighbour" interpolation, where the ground threshold set as 0 (as *detailed* ground roughness features were of little importance, and irretrievable from GLAS waveforms). The natural neighbour method of interpolation was chosen for its simplicity and speed.
Products derived from BCAL were verified against equivalent products derived by AGRG using the Terrascan software package (Terrasolid, 2011). Minor irregular peaks were found in BCAL derived products, but these were deemed to be of negligible effect at the scales at which data were analysed, i.e. correlation of vegetation height with spaceborne data was not greater for either of the products derived from BCAL or Terrascan.

3.8.1 Post-Processing

All airborne LiDAR point clouds were classified according to the American Society for Photogrammetry and Remote Sensing (ASPRS, 2010) standard, outlined in Table 3.7. LiDAR point cloud classification was performed by the data gathering

Table 3.6: Summary of values used in processing LAS point cloud data at each site using BCAL LiDAR tools software. Note: CS = canopy spacing, H_{max} = vegetation height limit, and I_{max} = maximum number of iterations.

Site	CS [m]	H_{max} [m]	I_{max}
Loobos	3.5	40	15
Norunda	5	40	15
Tharandt	3	50	20
SOA	5	40	15
SOBS	3	40	15
SOJP	5	40	15
Tumbarumba	5	100	20

team in the data post-processing stage.

Not classified and unassigned points were discarded from the analysis, as these are unreliable and tend to yield outliers. Data points classified as clouds were also removed, as these can also produce spurious elevations with regards to vegetation height measurements. Only a few building classified points were obtained in the datasets, due to the forest location of the data acquisition; however, any scattered structures that were identified were also removed so as not to misrepresent the forest data. Water points were not present at the majority of sites with the exception of SOA and Norunda, where small pools of water were present. Wherever identified, these points were also removed, leaving only ground and the three (low, medium, and high) vegetation classified points for a more strict forest representation.

Two output products of interest were extracted from the LiDAR point cloud data for all sites. These were extracted in raster format at a 1 m spatial resolution, and were the 95% values of vegetation height, and a DEM representing the ground.

3.8.2 Airborne LiDAR Footprint Extraction

Employing the raster products created by BCAL, exact areas corresponding to GLAS footprints can be extracted. Initially the full extent of the LiDAR raster dataset was considered, over which the corresponding GLAS footprints were overlayed.

Table 3.7: ASPRS classification codes assigned to LiDAR point cloud and their respective meanings.

Code	Meaning
0	Never classified
1	Unassigned
2	Ground
3	Low vegetation
4	Medium vegetation
5	High vegetation
6	Building
7	Noise
8	Model key
9	Water
10–11	Reserved for ASPRS definition
12	Overlap
13–31	Reserved for ASPRS definition

The footprint centre locations (i_lat, i_lon), associated dimensions (S_{major}, S_{minor}), and orientations (i_Azimuth) from the corresponding GLA14 data products (see Table 3.8; unique to laser campaign) were employed for defining the footprint perimeter. The centres of grid cells from the BCAL raster product that fell within these established GLAS footprint perimeters were exported for individual use. This yielded a 1 m spatial resolution, airborne LiDAR representation of individual GLAS footprints.

GLAS footprint locations, and orientations are extracted from corresponding GLA14 data products. However, the dimensions of each footprint are functions of their respective parent laser campaign, the details of which can be found in NSIDC (2012a).

3.9 GLAS Processing

GLAS waveforms are accompanied by a wealth of ancillary data spread across 15 different data products (Section 3.4). A number of these products were employed for various uses throughout this study, see Table 3.8 for reference purposes.

GLAS waveforms are used for a variety of purposes in this study, hence specific methodological processes for GLAS waveform analyses are outlined in respective chapters.

3.9.1 Footprint Elevation

In accordance with Los et al. (2012) the GLAS waveform reference elevation (Equation 3.3) was compared with corresponding SRTM 4.1 elevations (i_DEM_elv), where it is assumed that large elevation differences ($\geq$ 8 m, in conjunction with SRTM) between the two indicate problems in either dataset (Los et al., 2012). The calculation of this parameter (Equation 3.3) is essential for filtering spurious elevation data.

$$h = h_e + \Delta h_e - \Delta h_g + \Delta h_l \tag{3.3}$$

where:

$h =$ GLAS waveform reference elevation
$h_e =$ GLAS elevation (i_elev)
$\Delta h_e =$ Saturation elevation correction (i_satElevCorr)
$\Delta h_g =$ Height of EGM2008 geoid above TOPEX/Poseidon ellipsoid (i_gdHt)
$\Delta h_l =$ Difference WGS84 and TOPEX/Poseidon ellipsoid
$= \Delta r_a (\cos\phi)^2 + \Delta r_b (\sin\phi)^2$

where:

$\Delta r_a =$ Equatorial radial difference between WGS84 and TOPEX/Poseidon ellipsoids (0.7 m)
$\Delta r_b =$ Same as Δr_a for meridian (0.713682 m)
$\phi =$ Latitude

3.9.2 Estimating Vegetation Height

The accuracy of vegetation height estimates from GLAS waveforms is highly dependent on the ability of detecting the canopy top and ground elevations (Rosette et al., 2010). A number of different methods exist for vegetation height (V_H) retrieval from GLAS data. Here we consider the method of Rosette et al. (2008):

Table 3.8: Summary of GLAS ancillary data products used in this study. Note: † indicates a self assigned code as no GLAS assigned code was present within the respective data product itself. "NSIDC" in Table 3.8 is in reference to GLAS campaign information from the NSIDC (2012a), not from GLAS ancillary data products themselves.

Product	Code	Description
GLA01	Raw†	Raw GLAS return wave
	Nbin†	Number of 0.5 ns bins present in current waveform
GLA05	i_wfnoiseObl	Mean background noise of waveform
	i_sDevNsObl	Standard Deviation of background noise of waveform
	i_parmTr	Parameters of Gaussian fitted to transmitted pulse
GLA14	i_rec_ndx	Unique waveform batch number
	i_UTCTime	Time code referenced with respect to 01/01/2003 00:00:00 UTC
	i_lat	Latitude of waveform centroid
	i_lon	Longitude of waveform centroid
	i_elev	Waveform reference elevation (often located at waveform centroid)
	i_gdHt	Geoid height (EGM2008 geoid)
	i_DEM_elv	SRTM 4.1 DEM elevation
	i_SigBegOff	Signal begin range increment
	i_ldRngOff	Land range offset
	i_SigEndOff	Signal end range offset
	i_gpCntRngOff	Centroid range increment for up to six peaks
	i_Gamp	Maximum amplitude of up to six Gaussians
	i_satElevCorr	Saturation elevation correction
	i_satCorrFlg	Saturation correction flag
	i_FRir_qaFlag	Cloud detection flag
	i_Azimuth	Azimuth angle of footprint with respect to North
	Sn†	Shot number within subset of "i_rec_ndx" (1 to 40)
	ID†	Footprint ID (combination of "i_rec_ndx" and Sn, separated by "_")
NSIDC	$S^{\dagger}_{major}$	Semi-major axis of GLAS footprint
	$S^{\dagger}_{minor}$	Semi-minor axis of GLAS footprint

$$V_H = 1.06(r_1 - G_{A_{1,2}}) \quad (3.4)$$

where:

V_H = Maximum vegetation height

r_1 = Elevation of canopy top (i_SigbegOff)

$G_{A_{1,2}}$ = Centroid range increment (i_gpCntRngOff) for greatest amplitude (i_Gamp) of either Gaussian 1 or 2

Equation 3.4 was derived for the Forest of Dean in the UK, an area of complex terrain, populated by mixed broadleaf and needleleaf trees. The choice of the maximum amplitude Gaussian of the least elevated of the first two fitted Gaussians to represent the GLAS waveform ground elevation reduces the effect of slope over areas of low to moderate topography (Rosette et al., 2008).

3.10 Slope Calculations

Slope calculations are employed throughout this study, applied to different platforms and elevation datasets. This section introduces the general form of the equation employed for these calculations and briefly mentions the modifications applied to suit the equation's application.

3.10.1 Airborne LiDAR Slope

High-spatial resolution (1 m) bare Earth DEMs produced from BCAL raster datasets (for each site) allowed the derivation of high-resolution slope information within each GLAS footprint. This slope information can be related to related to GLAS waveform ground component features.

Using the footprint extraction method outlined in Section 3.8.2, GLAS footprint areas comprised of airborne LiDAR derived DEM data were analysed on an individual basis for continuous planar (not modal/most common) slope. This technique assumes zero ground roughness between the 5% (E_{05}) and 95% (E_{95}) quantile elevations within each footprint; these values were chosen to minimise the impact of vertically outlying elevations. An elevation gradient was formed between

this within-footprint quantile elevation difference over a distance equivalent to the mean of the dimensions (S_{major}, S_{minor}) of the footprint ($\bar{X}_{fp}$) and converted to slope angle (θ) by simple trigonometry (see Equation 3.5).

Note, the mean diameter of a given footprint is employed to calculate slope rather than a weighted mean footprint diameter (according to proportional fractional values of the semi-major and semi-minor axes). For a weighted mean diameter, directional bias is introduced, i.e. it is assumed that the calculated slope is more aligned in the direction of the semi-major axis. By employing the (unweighted) mean footprint diameter, directional bias is minimised.

$$\theta = \arctan\left(\frac{E_{95} - E_{05}}{\bar{X}_{fp}}\right) \tag{3.5}$$

Equation 3.5 is often used throughout this document when discussing slope derivations from multiple discrete return airborne LiDAR, continuous waveform spaceborne LiDAR, and ground-based elevation datasets. The form of Equation 3.5 remains constant, however the meanings of some parameters may change; these changes are explained in corresponding discussions.

3.10.2 Ordnance Survey Slope

Ordnance Survey (OS) slope derivations were made for validation purposes. OS data was provided by EDINA Digimap Ordnance Survey Service (EDINA, 2013), as a 10 m resolution DEM.

Data for the entire of the principality of Wales were employed, however, slopes were only calculated for GLAS footprint areas. In order to get an accurate OS representation of each GLAS footprint, only DEM grid cells with 90% of their area within corresponding GLAS footprint perimeters, when overlayed in British National Grid, were considered for slope calculations (see Figure 3.6).

The OS representation of slope for each footprint was calculated in much the same way as in Section 3.10.1, where E_{95} and E_{05} in Equation 3.5 represent the maximum and minimum within footprint elevations respectively. However, pixel size is greater in the case of OS data, which can lead to some elevation features being overlooked, thus inducing uncertainty in measurements. $\bar{X}_{fp}$ remains the

same as in Equation 3.5, the mean of the footprint's semi-major and minor axes.

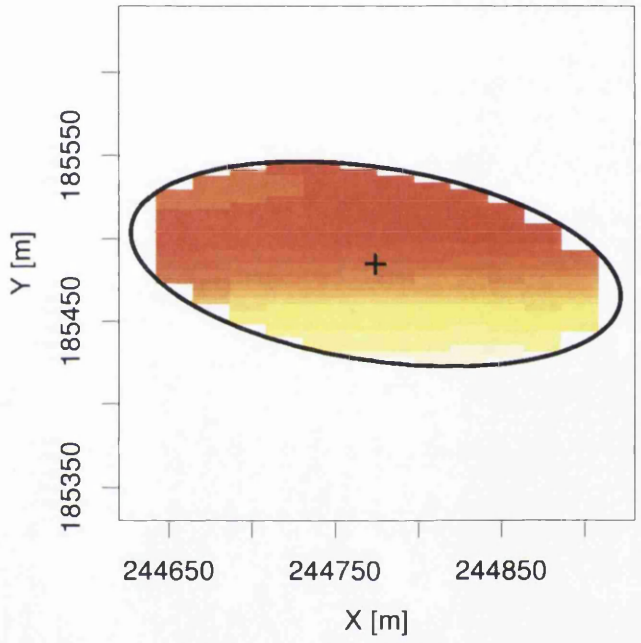


Figure 3.6: OS elevation tiles within a typical GLAS footprint (overlayed) in Universal Transverse Mercator (UTM) coordinate system (Zone 31U). Slope was derived as described in Section 3.10.1.

3.11 Summary

This chapter has introduced the study sites, and associated field data used in this study. Additionally, the variety of data sources employed throughout this study, specifically airborne and GLAS spaceborne LiDAR, have been discussed. Associated methods of data post-processing have been outlined for both of these approaches. In addition, data from MODIS, the Ordnance Survey, and SRTM have been introduced, employed for use as validation tools in specific research chapters that follow.

The following 3 research chapters make use of the introduced data, with the view of answering the research questions established in Chapter 2. Specific processing methods designed/developed as a means of answering each research question are described in the corresponding chapter.

Chapter 4

Shadows in Full Waveform LiDAR

This chapter seeks to confirm the existence of shadows (as defined in Section 1.3) in large footprint waveform LiDAR as suggested by Hancock et al. (2012), where this phenomena was noted in simulated waveforms. The presence of shadows in real waveforms can have a detrimental effect on results derived from such waveforms due to misinterpretation.
A method was developed and tested, such to obtain a 2-dimensional waveform derived from 3-dimensional airborne LiDAR data (spatially concurrent with GLAS footprints), indicating within footprint ground characteristics. This waveform was compared with its GLAS observed equivalent, and investigated for shadow presence. The frequency of shadow occurrences, and the conditions under which they occur were investigated, in addition to the implications of finding shadows in waveforms with regards to physical parameter observations from waveform LiDAR.

4.1 Method - Waveforms from Airborne LiDAR

In order to identify shadows in individual waveforms, the construction of a bare-Earth waveform (BEW) is required for comparison with a corresponding, observed, LiDAR waveform (GLAS in this case). The BEW was inferred from an airborne LiDAR derived bare-Earth DEM at the GLAS footprint scale, such that only information within the GLAS footprint (see Section 3.8.2) was processed (according to Section 3.8). The individual processes implemented to identify shadows within

(GLAS) waveforms are described in Sections 4.1.1 to 4.1.5, where the work flow takes the following structure:

- Perform Airborne LiDAR data quality check
- Resolve and populate vertical waveform elevation bins by considering emitted laser pulse energy distribution and intensity, and ground elevation (from the bare-Earth DEM)
- Convolve resulting waveform over a defined pulse width
- Align GLAS and bare-Earth waveform ground components, quantifying the differences between them

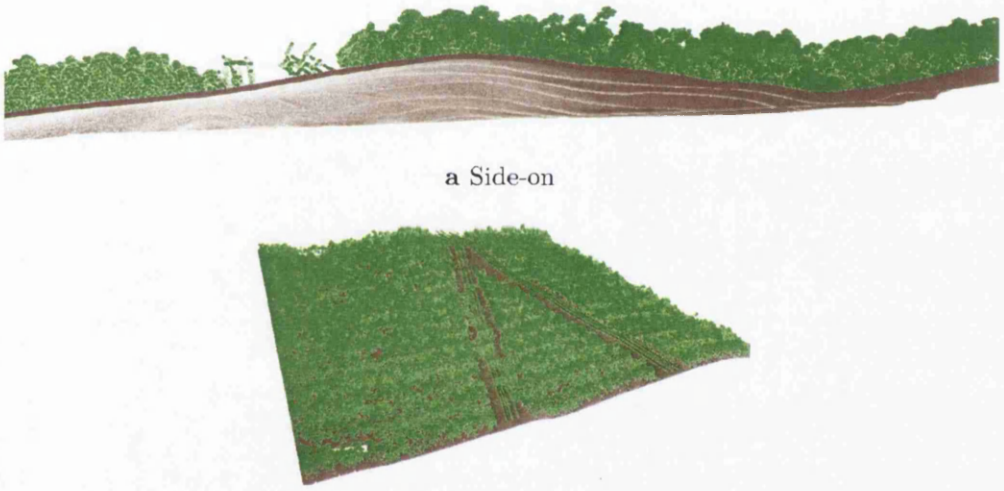
As GLAS waveforms represent the observations to which the BEW is compared, Gaussian pulse parameters, as used in the bare-Earth waveform construction, are required to be explicitly known. These are available from the GLAS GLA05 product (i_parmTr), where the transmitted Gaussian laser pulse amplitude and width are recorded. For other LiDAR systems this information will need to be obtained by different means.
Sloped terrain is suspected to be a contributory factor to shadowing effects (Hancock et al., 2012). Hence, for this study, only data from the Tumbarumba site were employed as this is the only site that exhibits suitable topographic conditions, i.e. complex terrain (see Section 3.1).

4.1.1 Airborne LiDAR Quality

During airborne LiDAR processing, it is vital to ensure that produced data products are representative of the area over which the original data was collected. This is paramount for the bare-Earth DEM as it is used as a reference against which other information (e.g. vegetation heights) are often compared. In situ field data collected from the Tumbarumba site suggests that airborne LiDAR quality collected is of a high standard. However, this does not prove the quality of the derived elevation and canopy height models from BCAL LiDAR tools (Section 3.8). Problems in shadow identification analysis will arise if the derived bare-Earth DEM exhibits data that are sourced from mixed LiDAR points belonging to both ground

and vegetation, this will give a false reading of the ground location. Occurrences of this are simple to identify visually, they appear as a sharp peak in the relatively smooth ground contour that is otherwise present.

To test the quality of the data, five (of a possible 12) relatively large (2 km × 2 km), random tiles were chosen from the processed Tumbarumba airborne LiDAR point cloud, which includes information regarding terrain and vegetation. Figure 4.1 shows one of the five tiles from side-on and angled views as an example. No peaks were found in the ground contour of any of the selected tiles, hence airborne LiDAR processed products were deemed of adequate quality for the purpose of this investigation. As terrain at this site is consistently rough throughout, it is assumed that the full point cloud, and thus subsequently processed products are free of quality issues also.



a Side-on

b Angled

Figure 4.1: (a, b) Example of terrain (brown points) and vegetation (green points) components derived from the Tumbarumba site airborne LiDAR point cloud viewed from a side-on and angled geometry respectively. The ground contour is visually checked for peaks (such not to confuse real surface variability with quality issues), which would be an indicator of poor quality.

4.1.2 Bare-Earth Waveform

The post-quality checked airborne LiDAR point cloud was used to derive a 1 m resolution DEM, GLAS footprint representations of which were inferred according to Section 3.8.2. Only GLAS footprints that passed the elevation filter described in Section 5.4.3 were analysed to create a bare-Earth waveform (BEW).

The 1 m resolution airborne LiDAR derived DEM representation of the GLAS footprint (Figure 4.2a) was binned with regards to elevation, the resolution of each bin was selected at 0.15 m (or 0.5 ns), equivalent to that used in GLAS processing (Section 3.4.2).

A 2-dimensional Gaussian (Figure 4.2b; of the form given in Equation 4.1; UQU, 2013) is selected to represent the radial energy/intensity distribution across each DEM footprint. This was chosen such to match the radial energy distribution that is known to exist for GLAS footprints (Abshire et al., 2005; Harding and Carabajal, 2005). The 2-dimensional Gaussian was created, unique to each footprint, where size and orientation parameters for each are sourced from specific GLAS campaign information from NSIDC (2012a) and/or the GLAS GLA14 data product. The Gaussian extent, in x and y, is defined at a radius where its returned energy (z) drops below a $1/e^2$ ($\sim 13.5\%$) threshold, such to correspond to GLAS footprint dimensions (Section 3.4.2).

$$z = A \cdot \exp\left(-a(x - x_0)^2 + 2b(x - x_0)(y - y_0) + c(y - y_0)^2\right) \tag{4.1}$$

where:

$$a = \frac{\mathrm{Cos}^2(\theta)}{2\sigma_x^2} + \frac{\mathrm{Sin}^2(\theta)}{2\sigma_y^2} \tag{4.2}$$

$$b = \frac{-\mathrm{Sin}(2\theta)}{4\sigma_x^2} + \frac{\mathrm{Sin}(2\theta)}{4\sigma_y^2} \tag{4.3}$$

$$c = \frac{\mathrm{Sin}^2(\theta)}{2\sigma_x^2} + \frac{\mathrm{Cos}^2(\theta)}{2\sigma_y^2} \tag{4.4}$$

where:

$z =$ Gaussian energy/intensity at positions of x and y

$A =$ Gaussian maximum amplitude ($= 1$)

$\theta =$ Angle of rotation of footprint (i_Azimuth)

$x, y =$ Location along x-axis and y-axis respectively

$x_0, y_0 =$ Coordinates at which A occurs (i_lon, i_lat)

$\sigma_x, \sigma_y =$ Gaussian width in x and y respectively (S_{major}, S_{minor})

The DEM and Gaussian as in Figure 4.2 occupy the same 2-dimensional space by definition, however, some minor discrepancies occur at their edges. This is attributed to the coarser resolution of the DEM (1 m), with respect to the definition of the 2D Gaussian distribution. However, investigations into resolution subtleties have shown that these minor discrepancies have no significant effect on the resulting BEW.

Considering the DEM and 2D Gaussian only, these spatial data are binned/gridded by Easting and Northing at a 1 m resolution, a limitation imposed by the DEM resolution. First considering the DEM only, the maximum number of bins required to sufficiently cover the elevation range present within the DEM is derived according to Equation 4.5. Note: $\lfloor\ \rfloor$ represents the floor function (e.g. $\lfloor 1.68 \rfloor = 1$).

$$B_{max} = \left\lfloor \frac{Z_{max} - Z_{min}}{R_{bin}} \right\rfloor \tag{4.5}$$

where:

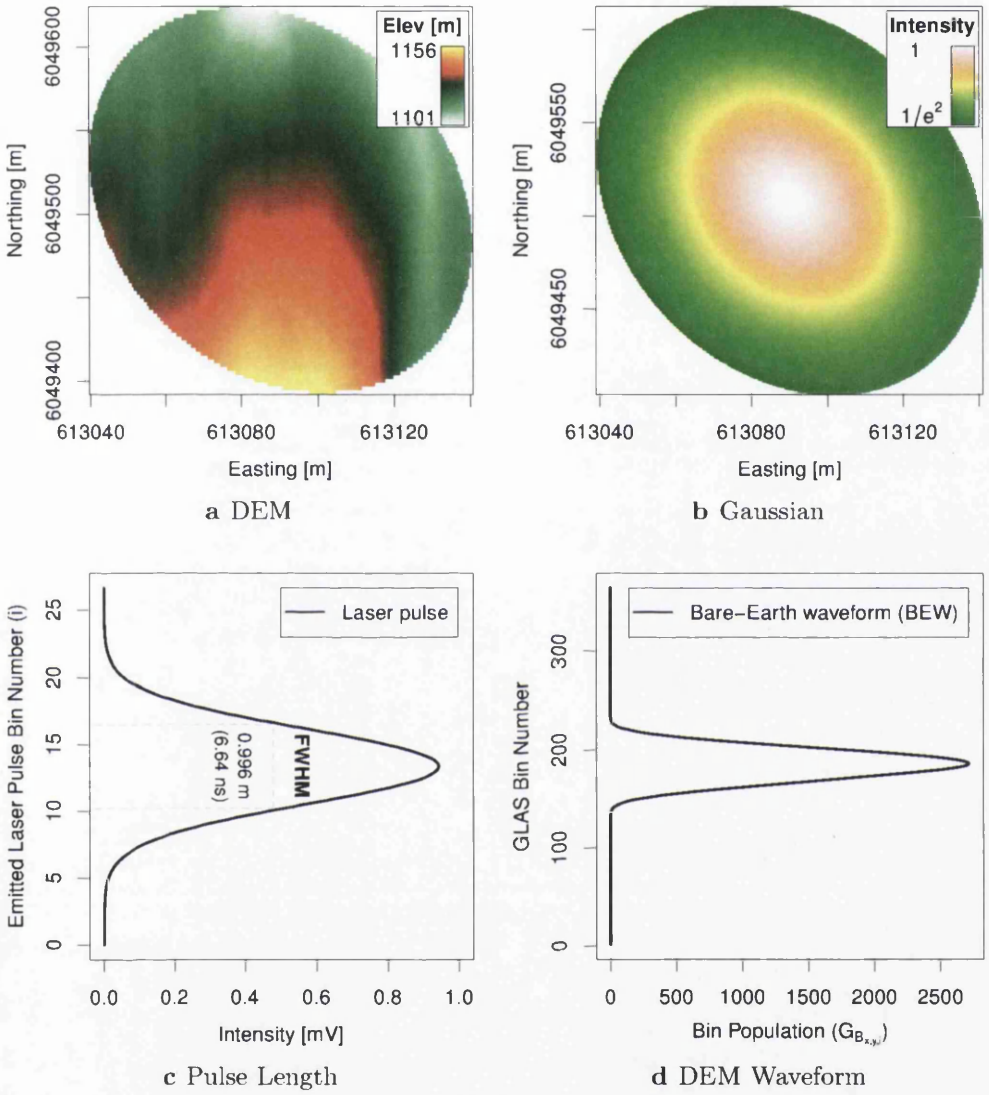
$B_{max} =$ Maximum number of bins

$R_{bin} =$ Resolution of bins ($= 0.15$ m)

$Z_{min}, Z_{max} =$ Minimum and maximum elevations within DEM

In order to obtain an exact waveform representation of the ground, spatially represented by the DEM, Equation 4.5 is modified to determine which elevation values belong to which bin; Z_{max} is replaced sequentially by the elevation that occurs at x and y positions throughout the previously established Easting, Northing grid. Each 0.15 m vertically established bin is populated by the number of elevation occurrences that fall within each specific bin elevation range.

This produces a waveform that is not representative of corresponding GLAS waveforms as the pulse length of the emitted laser pulse is not accounted for. For GLAS,



a DEM **b** Gaussian

c Pulse Length **d** DEM Waveform

Figure 4.2: Nadir view of (a) airborne LiDAR derived bare-Earth DEM, and (b) Gaussian pulse energy for an example GLAS footprint. (c) Laser pulse length fired by GLAS, and (d) convolved bare-Earth DEM waveform for same example footprint.

the emitted pulse is (1-dimensional) Gaussian shaped, where for this particular footprint the emitted laser pulse has a FWHM (or pulse length) of 0.996 m/6.64 ns. The consideration of this pulse length in calculations redistributes elevation bin populations according to the relative intensities of the 2D Gaussian and pulse width intensity. This accounts for features more central within the DEM footprint being more intensely reflected than those at the radial edges (as is true with GLAS; Rosette et al., 2008).

The energy distribution (P) of this 1-dimensional Gaussian pulse (Figure 4.2c) is given by Equation 4.6, where Gaussian amplitude (A) and width parameters (σ_x) are taken from the GLAS GLA14 data product (i_parmTr).

$$P = A \cdot \exp\left(\frac{(x - x_0)^2}{2\sigma_x^2}\right) \tag{4.6}$$

where:

$P =$ Gaussian energy/intensity at positions of x
$A =$ Gaussian maximum amplitude (i_parmTr: parameter 2)
$x =$ Location along x-axis
$x_0 =$ x position of A
$\sigma_x =$ Gaussian width in x (i_parmTr: parameter 4)

To account for the emitted laser pulse length we modify the bins in which elevations occur as a function of the number of 0.15 m bins that constitute the pulse length itself. These bins now represent the bin position that would be seen by GLAS. Equation 4.7 shows how bin position changes according to the position in the pulse length (represented by i).

$$B_{x,y,i} = \underbrace{\overbrace{\left\lfloor \frac{Z_{x,y}}{R_{bin}} \right\rfloor}^{\beta} - \left\lfloor \frac{N}{2} \right\rfloor}_{\text{constant}} + i \tag{4.7}$$

where:

$B_{x,y}$ = Bin position for point [x,y] within DEM
$Z_{x,y}$ = DEM elevation at [x,y]
N = Number of bins in pulse length
i = $1, \cdots, N$

Equation 4.7 highlights that, neglecting the i^{th} dimension, $B_{x,y}$ will cover a range of bins, related to N by the following:

$$\left(\beta - \left\lfloor \frac{N}{2} \right\rfloor + 1\right) \leq B_{x,y} \leq \left(\beta + \left\lfloor \frac{N}{2} \right\rfloor\right) \tag{4.8}$$

Populating each bin individually is done by cumulatively adding the product of the intensity of the 2D Gaussian at the corresponding x and y positions and the intensity of the pulse length at the i^{th} position. This gives rise to a waveform that is representative of GLAS, Equation 4.9 highlights how each bin is populated. $G_{B_{x,y,i}}$ represents bin populations for vertical elevation bins as would be seen by GLAS, which is initially set as zero (representing zero return intensity from a target surface) and is populated sequentially (as x, y, and i change; see Equation 4.9).

$$G_{B_{x,y,i}} = G_{B_{x,y,i}} + (Gz_{x,y} \cdot P_i) \tag{4.9}$$

where:

$G_{B_{x,y,i}}$ = GLAS bin according to $B_{x,y,i}$ (Equation 4.7)
$G_{Z_{x,y}}$ = 2D Gaussian intensity at [x,y]
P_i = Emitted laser pulse intensity at i

Accounting for the pulse length alteration to bin position in $B_{x,y,i}$ allows an extension of the derived GLAS waveform beyond the hard limit of where intensity returns exceed an imposed noise threshold, hence no data are lost in the leading and/or trailing edges of the calculated return (BEW). Figure 4.2d shows the accumulation of Equation 4.9 performed over all values of x, y, and i for an example GLAS footprint.

As a note on uncertainty, the accuracy of the airborne LiDAR and associated products are paramount in the generation of the BEW by the following method. The airborne LiDAR data are subject to inherent, altitude dependent, instrument

SWANSEA LIBRARY UNIVERSITY

vertical uncertainty (ΔI_v; ~ 0.08 m at 1 km above ground altitude here) which propagates through data processing techniques. This is compounded by interpolation uncertainty which is dependent on the spatial sampling rate of LiDAR points (S_R) per grid cell in the DEM used to generate the BEW. S_R is dependent on the LiDAR collection instrument (and post-processing in this case). For this particular dataset the mean S_R is high enough such that at least one LiDAR returned ground point should be found in each grid cell (chosen at 1 m). Also, ΔI_v in this case is less than the chosen (GLAS) vertical resolution at which the BEW is generated at. Considering these DEM influencing values, the produced BEW is somewhat optimised with respect to the grid cell and vertical bin resolutions chosen.

4.1.3 Shadow Waveform

In order to check the validity of suspected shadowing effects in GLAS waveforms, a basis for comparison is required, against which shadow presence can be tested. This basis is established as the "shadow waveform" (SW), constructed from airborne LiDAR point cloud data that occupy individual GLAS footprints; an example of airborne LiDAR point cloud data is shown in Figure 2.3 in Section 2.6.1. Airborne LiDAR data were used to this end as they enable in depth detail of vegetation and ground component structure to be observed. The process of constructing the SW is outlined in this section.

Considering vegetation points only, the airborne LiDAR point cloud representation of each GLAS footprint was gridded along the x and y spatial axes. Grid resolution was set at 1 m, such to match the resolution of the previously derived airborne LiDAR bare-Earth DEM.

The number of airborne LiDAR points (in z dimension) within each x, y grid cell were noted ($N_{x,y}$) with their spatial attributes noted. $N_{x,y}$ where then raised as powers to an imposed transmission coefficient (τ), thus yielding arbitrary (scaled according to τ) intensity reductions ($I_{R_{x,y}}$), which originate from vegetation sub-components (branches, shoots, leaves, etc.) from within each x, y grid cell. Note: $\tau = 0.6$ was chosen in this instance, however, the value of τ is irrelevant (provided it obeys $0 \leq \tau \leq 1$), as it is a scaling factor only, affecting shadow intensity not location (elevation).

In reality, τ is dependent on vegetation characteristics such as: type, density, leaf area, and leaf orientation. It is noted that true values of τ are obtainable through intensity information from the small footprint full waveform LiDAR data that were used to generate the point cloud from which the BEW, and consequently, the SW were inferred. For further consideration, radiative transfer models could also be employed to infer more realistic values for τ than the arbitrary constant employed in the current study. However, defining τ as a constant was by design, such to simplify the nature of this work. By enforcing this definition does not inform on the severity of signal attenuation (shadowing) in any given return, but does however, allow for the vertical location (elevation) at which shadowing occurs to be found, in conjunction with the aims and objectives of this study.

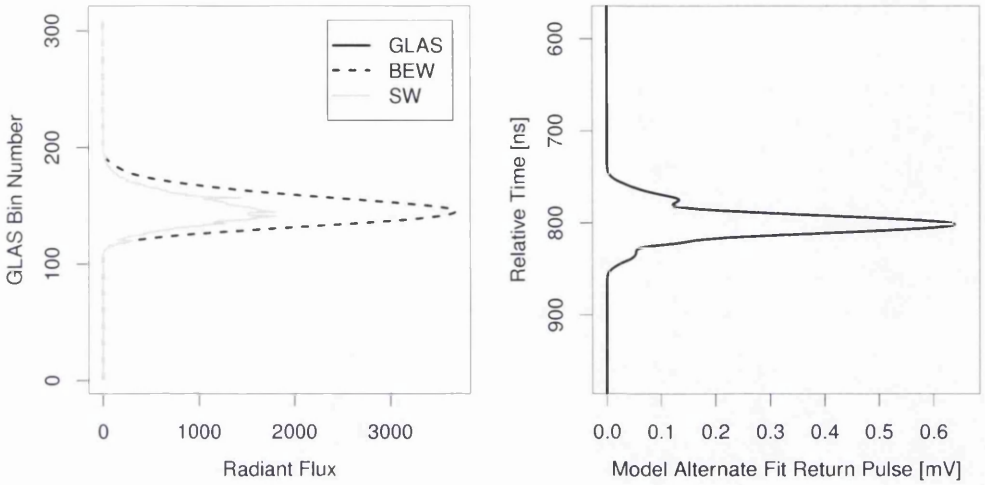
The x,y location of each vegetation sub-component is projected to the ground such to identify the elevation bin (resolved at 0.15 m as before) within which, each vegetation sub-component will cast its shadow; this assumes a nadir viewing geometry. $I_{R_{x,y}}$ is then summed for unique elevation bins (z; according to Equation 4.10) to yield the fraction of the original BEW intensity that is present in the shadow waveform (I_R) as a function of unique elevation.

$$(I_R)_z = \sum_{z=1}^{n} \left(\tau^{N_{x,y}}\right)_z = \sum_{z=1}^{n} \left(I_{R_{x,y}}\right)_z \tag{4.10}$$

Multiplying the BEW by I_R for unique elevation bins yields the SW, which is defined as the theoretical representation of a waveform ground component over a vegetated surface. Each vegetation sub-component reduces the BEW intensity at specific ground elevations. These reductions occur at elevations corresponding to the elevations of the vegetation sub-components if they were on the ground surface over which they are vertically located i.e. if their heights above the ground were 0 m.

An example of the SW and corresponding BEW, derived by the above described methods, is shown for an example waveform in Figure 4.3a. Note that the SW exhibits the same general shape as the BEW from which it was derived, but with reductions in intensity at specific elevations. These reductions are, by definition, caused by vegetation sub-components obstructing incident light, thus casting shad-

ows on the ground i.e. occlusion.



a GLAS Waveform

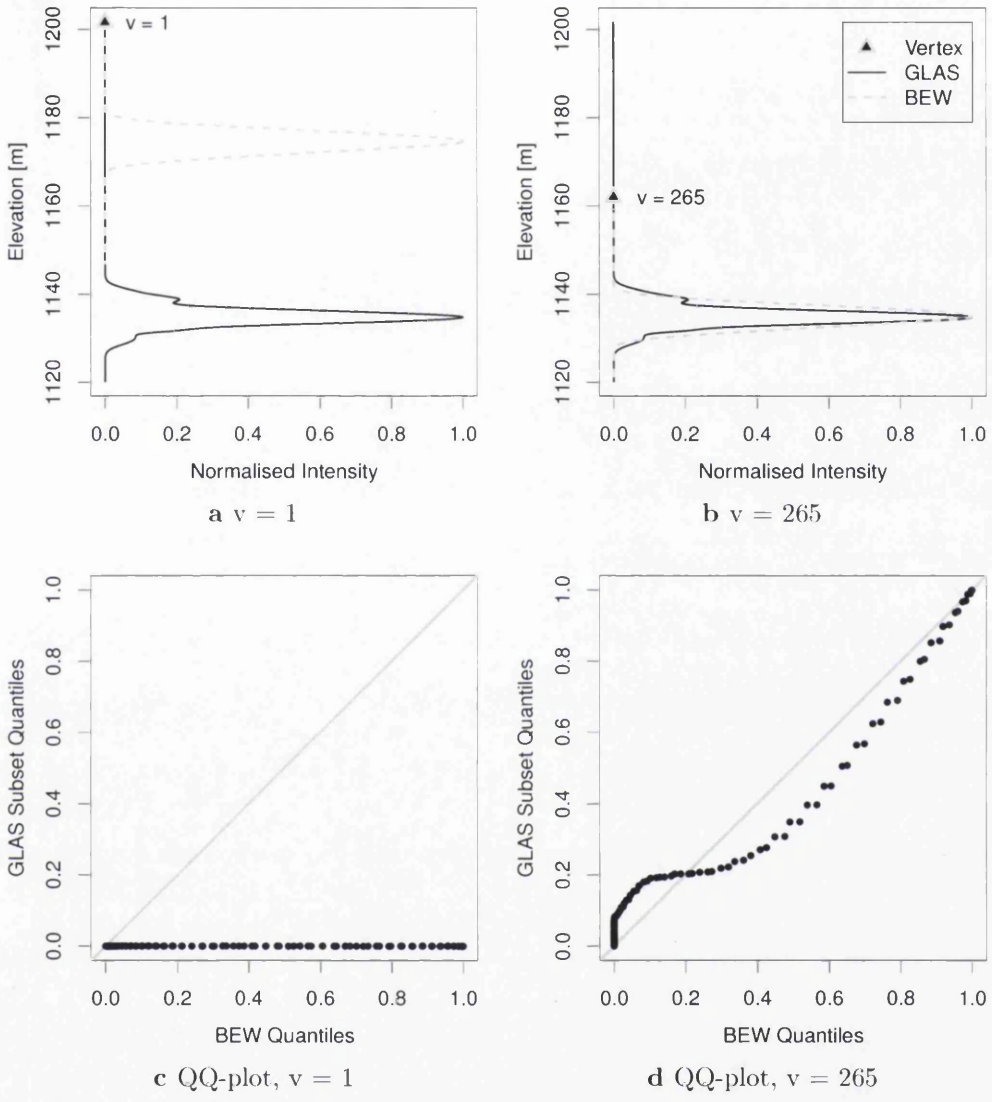
b GLAS Waveform

Figure 4.3: (a) Bare-Earth DEM waveform (BEW; black dashed), and shadow waveform (SW; grey), from airborne LiDAR data, displayed as a function of vertical elevation bins that correspond to GLAS bins (with regards to size). (b) Corresponding model alternate fit (black; from GLA14 parameters) based on physical observations by GLAS, displayed as a function of GLAS observed relative time.

4.1.4 Alignment with GLAS Waveforms

As the derived BEW (Figure 4.2d) is representative of the ground surface only, and GLAS waveforms (Figure 4.3b) are recorded from a finite height above the expected ground to the ground itself, a direct comparison between waveforms is not possible without some alignment and/or cropping. These different recording times dictate that the BEW and GLAS waveforms each exhibit a different number of elevation bins (and thus record different relative time ranges), hence they cannot be aligned by this information.

In order to directly compare the BEW and GLAS waveforms, they must align with respect to identical features common in both waveforms, i.e. their ground components. A method was developed to accomplish this alignment between the BEW and GLAS waveforms; an explanation of which is given below.



a v = 1

b v = 265

c QQ-plot, v = 1

d QQ-plot, v = 265

Figure 4.4: Alignment process of corresponding BEW (dashed line) and GLAS (solid line) waveforms starting at (a) vertex v = 1, BEW is shifted along the GLAS waveform sequentially with increasing v, to v = v_{max}. (b) Illustration of optimal alignment between waveforms, where v = 265 produced greatest correlation ($R^2 = 0.92$). (c and d) Quantile-quantile plots of BEW and (equivalent subset of) GLAS intensity distributions (as a function of elevation) corresponding to vertices from (a), and (b) respectively. (c) Illustrates poor alignment between the BEW and GLAS waveforms, whereas (d) shows relatively good alignment.

For forest imaging, returned GLAS waveforms exhibit multiple peaks, each originating from vegetation and/or ground interactions. Where possible the ground component (defined as in Section 5.1.1) of each GLAS waveform was isolated, where it is assumed that this represents the GLAS observed equivalent of the BEW. These two waveform representations of the same ground surface, from different data sources, can be aligned to inform on shadow presence.

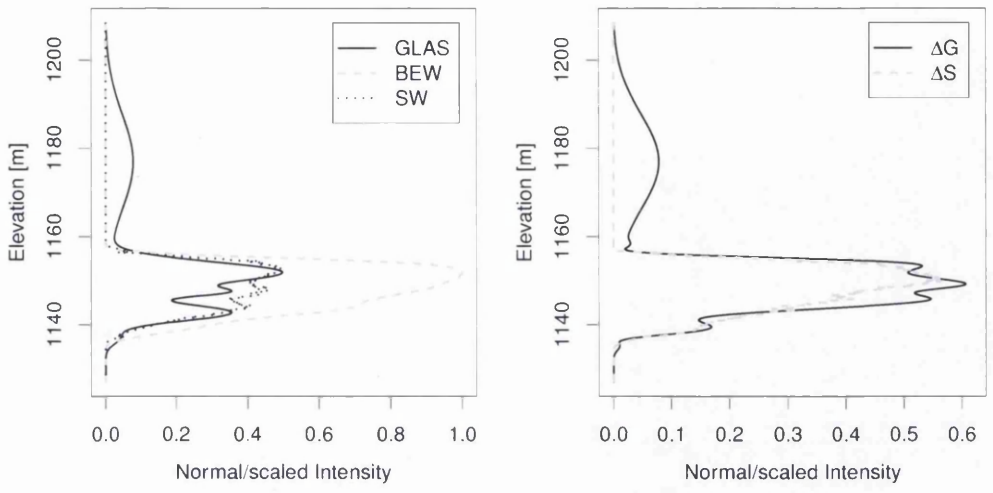
Ground component alignment is achieved by first normalising the intensity of both waveforms such that they equal 1. This is required as the bin population units between the BEW and GLAS waveforms differ (noted between Figures 4.3a and 4.3b), hence normalising achieves parity between respective waveform intensity maxima.

Considering the constituent discrete vertices (v) of individual GLAS waveforms, starting at $v = 1$ (highest elevation above the ground; Figure 4.4a), the BEW and GLAS waveforms are overlayed. The intensity of these two waveforms are compared, where data from both waveforms are available, and the Pearson correlation coefficient (R^2) between them noted. Typically the GLAS waveform consists of more bins than the BEW as it represents both ground and vegetation returns, whereas the BEW represents ground returns only. Comparing waveforms where data from both are available ensures consistent comparisons are made.

This comparison process is repeated such that the BEW is shifted along the GLAS waveform vertical axis from $v = 1$ to $v = v_{max}$, and subsequently resulting R^2 values noted. The assumption being that the greatest value of R^2 corresponds to when the BEW and GLAS waveform intensities are most alike (Figure 4.4b), indicating the GLAS ground component has been located and optimally aligned with the BEW.

Figure 4.4 describes this approach visually. Additionally quantile-quantile (QQ) plots illustrate that the distributions of the BEW and GLAS waveform ground components differ when R^2 is low (poorly aligned; Figure 4.4c), whereas a linear trend appears more prominently as R^2 increases (well aligned; Figure 4.4d). When both waveforms are optimally aligned the excess (in the tails) of the BEW are removed and the intensity at all elevations that exhibit no data (no plot information in Figures 4.4a and 4.4b) are set to zero.

Once optimally aligned (Figure 4.4b), the SW (as described in Section 4.1.3) is



a Overlayed Waveforms

b Overlayed Differences

Figure 4.5: (a) Overlay comparison between bare-Earth waveform (BEW) and GLAS waveforms, with corresponding shadow wave (SW), and (b) differences between the BEW and GLAS (ΔG), and BEW and SW (ΔS) respectively. Note: reductions in the intensities of both the SW and GLAS waveforms occur at almost identical elevations, indicating shadow presence (see Section 4.1.5.

derived, which by definition, is scaled such that its maximum intensity is less than that of the BEW.

In order to minimise the differences between the SW and GLAS waveforms, the aligned GLAS waveform is scaled such that its ground component maximum intensity is identical to the maximum intensity of the SW (see Figure 4.5a). This minimises complication in comparing the differences (as a function of elevation) between both the SW and GLAS waveforms with the BEW (see Figure 4.5b).

With all three waveforms (GLAS, shadow, and bare-Earth) aligned, a comparison between each, as a function of elevation, allows the retrieval of shadow information in GLAS waveforms.

4.1.5 Detecting Shadows

A visual method for identifying shadows in GLAS waveforms is preferred to an automated approach as the likelihood of such a method being able to successfully

identify shadows in all waveforms is low. This is a product of inherent alignment and intensity difference subtleties, which allows some degree of flexibility which is typically difficult to implement computationally. Although there are methods available suited to similar application, these are not within the scope of this study. The visual method of identifying shadows in GLAS waveforms is described below. For any footprint, the differences between the BEW and corresponding GLAS and shadow waveforms (Figure 4.5b), defined as Δ_G and Δ_S respectively, are derived. Relevant ground component data from GLAS waveforms were retained for comparison with the BEW and SW, which by definition represent ground returns only. All non-ground waveform features were removed from analysis as they have no reference against which to be compared.

The implication of comparing ground components is that if the GLAS waveform intensity profile is similar (as a function of elevation) to that of the BEW, the signal is believed to be reflected from the ground only, hence no shadow effects from vegetation components are induced. However, if its profile follows that of the SW, the incident light from GLAS is believed to have been obstructed by vegetation components, and hence shadows cast in its waveform return.

The magnitude of intensity differences between the SW and GLAS waveforms (with shadows) are not expected to be identical. This is due to waveform intensities being normalised, and the SW being obtained using an arbitrary transmission coefficient (τ; Section 4.1.3) that is constant across the entire waveform. In reality τ would vary throughout the footprint with regards to vegetation characteristics such as canopy density, and associated optical properties, in addition to the properties of the incident light (wavelength, power, etc.) from the instrument.

Unlike magnitude, the elevation at which intensity reductions occur in both the SW and GLAS waveform (with shadows) are expected to be very similar as no difference in vegetation elevation is expected. Subtle differences of intensity minima between waveforms are a legacy of minor misalignment, and post-processing smoothing effects.

Figure 4.5 illustrates a waveform that exhibits shadowing effects according to the method described above. Figure 4.5a particularly shows that intensity reductions occur at similar elevations for both SW and GLAS waveforms, although not at the same magnitude.

4.2 Results and Discussion

The process described from Section 4.1.2 to Section 4.1.5 was executed for each waveform, where reliable airborne LiDAR data were available, at the Tumbarumba site. Only Tumbarumba was considered as it is the only site of the seven sites in this study that exhibits a large variance in sloped terrain, which is expected to be key in inducing shadows in waveform returns.

4.2.1 Shadow Presence in Waveforms

In order to determine the presence of shadowing effects in waveform LiDAR, a total of 565 GLAS waveforms from Tumbarumba were visually inspected (for the criteria noted in Section 4.1.5). Statistical analysis was performed to inform on the frequency of this shadowing effect.

Table 4.1: Statistical summary of shadow presence in analysed waveforms from Tumbarumba. Note: T_N = total number of waveforms in slope bin, N = number of shadow affected waveforms in slope range, % = percentage of shadow affected waveforms in slope range, C_N = Cumulative total of shadow affected waveforms, $C_{\%}$ = cumulative percentage of shadow affected waveforms. Where $T_N \leq 4$ additional data is required at these slopes to confirm the observed trend (see Figure 4.6 for visual representation).

Slope (θ) [°]	$\mathbf{T_N}$	**N**	**%**	$\mathbf{C_N}$	$\mathbf{C_{\%}}$
$0 \leq \theta < 5$	34	0	0	0	0
$5 \leq \theta < 10$	136	5	4	5	1
$10 \leq \theta < 15$	144	7	5	12	2
$15 \leq \theta < 20$	122	20	16	32	6
$20 \leq \theta < 25$	60	15	25	47	8
$25 \leq \theta < 30$	26	8	31	55	10
$30 \leq \theta < 35$	21	16	76	71	13
$35 \leq \theta < 40$	12	6	50	77	14
$40 \leq \theta < 45$	4	2	50	79	14
$45 \leq \theta < 50$	3	2	67	81	14
$50 \leq \theta < 55$	1	1	100	82	15
$55 \leq \theta < 60$	2	2	100	84	15

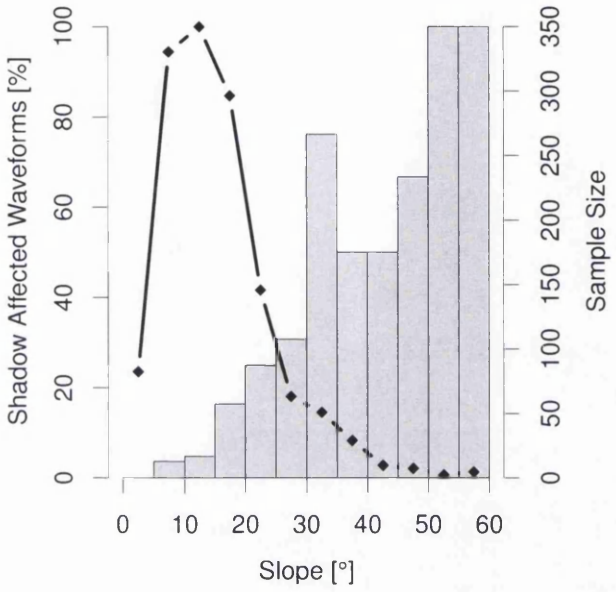


Figure 4.6: Histogram of data from Tumbarumba site indicating the magnitude of the presence of shadows (as a relative percentage of waveforms) within 5° slope bins (grey bars, left y-axis), and the within slope bin sample size (black line, right y-axis). Shadows appear to occur more frequently for higher slopes. See Table 4.1 for additional details.

Analysis revealed that for the forest characteristics at Tumbarumba (see Section 3.1), approximately 15% of GLAS waveforms suffered shadows cast by vegetation subcomponents (branches, shoots, leaves, etc.). The majority of shadow afflicted waveforms are noted over greater slopes (see Figure 4.6 and Table 4.1); approximately 94% of shadow affected waveforms occur at $\geq 10°$, and 62% occurring at $\geq 20°$.

Figure 4.6 illustrates how the magnitude of the presence of shadows (signal attenuation) in waveforms (relative percentage of sample size within each 5° bin; grey bars, left y-axis) varies as a function of 5° slope intervals (bins), the sample size within each bin (black line, right y-axis) is also noted. Figure 4.6 suggests that magnitude of shadow presence is independent of sample size (within each slope bin), as the magnitude of shadow presence is not proportional to the corresponding bin population. However, this is unconfirmed for slopes $\geq 40°$ as very small sample sizes are found here, which may skew results.

The presence of shadows at Tumbarumba was expected, however due to the nature of the vegetation at the site (Section 3.1), the detection of shadows was expected to be considerably more difficult. This site is dominated by Eucalyptus, the leaves of which are vertically orientated, pointing towards the ground, thus effectively reducing Leaf Area Index (LAI). This, by definition, enforces that the fraction of incident light penetrating the canopy and reaching the ground is higher, hence the effect of shadowing from the leaf vegetation subcomponent is diminished.
Furthermore, it is noted that this study does not account for variation in vegetation canopies (within the same species at Tumbarumba) as a function of terrain slope i.e. changing density etc. characteristics as slope becomes increasingly steep. The implication being that as canopy characteristics change, the severity of signal attenuation will react accordingly, however, this study focuses on the retrieval of the elevation at which shadowing occurs, not the severity of the effect. This vegetation complexity provides additional reasoning for defining τ (Section 4.1.3) as a constant, particularly to ensure methodological simplicity.
Speculatively, based on results from Tumbarumba, similar terrain conditions with another species of vegetation (higher LAI) may induce shadowing effects to a higher magnitude. However, this theory remains to be tested due to a lack of data available at present.

4.2.2 Methodological Assumptions

For both the BEW and SW (see Sections 4.1.2 and 4.1.3 respectively) no multiple scattering mechanism is enforced. Neglecting this mechanism in processing still assumes a realistic waveform representation of 3-dimensional spatial data with regards to elevation ranges. However, intensity values would likely differ (with respect to those waveforms produced) if such a mechanism was implemented; this was noted for FLIGHT radiative transfer model waveforms by North et al. (2010). This scattering effect is expected to be minimal for both the BEW and SW, a more significant effect is expected where the density of light obstructing objects increases. Multiple scattering effects are neglected for this study as focus is on the elevations at which shadows occur, not the magnitude of the shadow.
Additionally, the assumption of a realistic waveform representation of 3-dimensional

spatial data assumes that the 3-dimensional data are also reliable. However, as discussed in Section 4.1.2, these DEM data are subject to multiple sources of uncertainty. Such uncertainty is capable of influencing the elevation at which shadowing effects occur, however, the magnitude of this effect is limited, according to source data (airborne LiDAR), interpolation accuracies, and methodological uncertainty propagation. These 3-dimensional data uncertainties can be represented as waveforms that are offset (in elevation) from the BEW by an amount equal to the uncertainties themselves.

For the SW, the magnitude of intensity reductions are to a certain degree arbitrary, based on a defined transmission coefficient (τ) in Section 4.1.3. In reality, τ would rely on a combination of vegetation characteristics, both physical (canopy density, LAI), and optical (absorption and reflective properties), as described in Section 4.1.3.

Furthermore, the technique for creating the BEW and SW assumes vertical incident light. More realistic scenarios with angled illumination geometries are implementable at some time expense. It is expected that with the implementation of such a change, shadows are elongated and are horizontally displaced from directly beneath shadow inducing vegetation subcomponents (shoots, branches, leaves, etc.).

In consideration of such assumptions, with particular regard to uncertainties, cautious conclusions must be drawn when examining results. Shadow occurrence results, whilst based on technically and scientifically valid assumptions, are subject to subtleties, some of which are not explored in this study. Of particular note are the elevations at which shadows occur within a generated SW, which are subject to some uncertainty, and the magnitude(s) of such occurrences are simplistically represented. Such considerations were not explored in-depth here, as the focus of this study was to identify shadows in large footprint waveforms, and their origins.

4.2.3 Vegetation Height Discrepancies

The presence of shadows in waveforms have potential to induce uncertainty in vegetation height estimates (V_H) from waveform LiDAR returns, particularly where a Gaussian decomposition method is used for its derivation. Some examples of Gaus-

sian decomposed GLAS waveforms, where shadows are present, are illustrated in Figure 4.7. This technique is favoured by many for retrieval of V_H from such waveforms despite the illustrated uncertainties (Los et al., 2012; Chen, 2010b; Rosette et al., 2008).

As mentioned (in Section 4.1.3), shadows reduce waveform intensity at specific elevations dependent on within footprint vegetation location. Such reductions can be significant enough to cause "phantom" peaks in waveforms, which can be misinterpreted as physically observed features. This becomes particularly important when deriving V_H, as the waveform peak identified as the ground could be a phantom peak. This can yield an incorrect elevation for the ground, an accurate assessment of which is required for accurate V_H.

The uncertainties in V_H are dependent on the method employed to obtain it, for example, here the method of Rosette et al. (2008) is used, which employs the Gaussian decomposition of waveforms. Particularly, V_H is defined as the difference between the signal start (see Section 3.4), and the location of the Gaussian exhibiting the greatest amplitude of the two that are least elevated ($G_{A_{1,2}}$), which is assumed to represent the ground. If phantom peaks exist in a waveform as a consequence of shadow effects, the locations of fitted Gaussians can be misinterpreted, hence the ground elevation determination is unreliable, inducing great uncertainty in V_H.

Figure 4.7 illustrates some of the issues discussed here, with respect to deriving V_H by the method of Rosette et al. (2008). Of particular interest are the least elevated of the fitted Gaussians, one of which represents the ground by this method ($G_{A_{1,2}}$; **thick** red line in Figure 4.7). The location of $G_{A_{1,2}}$ is discussed with respect to the BEW (grey dashed line in Figure 4.7), which is assumed to be the "true" (GLAS) waveform representation of the ground without shadow effects. Any differences between these two ground representations indicates $G_{A_{1,2}}$ is incorrect, which leads to GLAS derived V_H being under or overestimated. A good Gaussian representation of the true ground is defined where there is little difference between $G_{A_{1,2}}$ and the centroid of the BEW.

Figures 4.7a, 4.7b, and 4.7c indicate shadow presence in the ground component of each respective waveform, induced by vegetation sub-components (see Section 4.2.1). $G_{A_{1,2}}$ for each of these waveforms appear at a lower elevation than that of

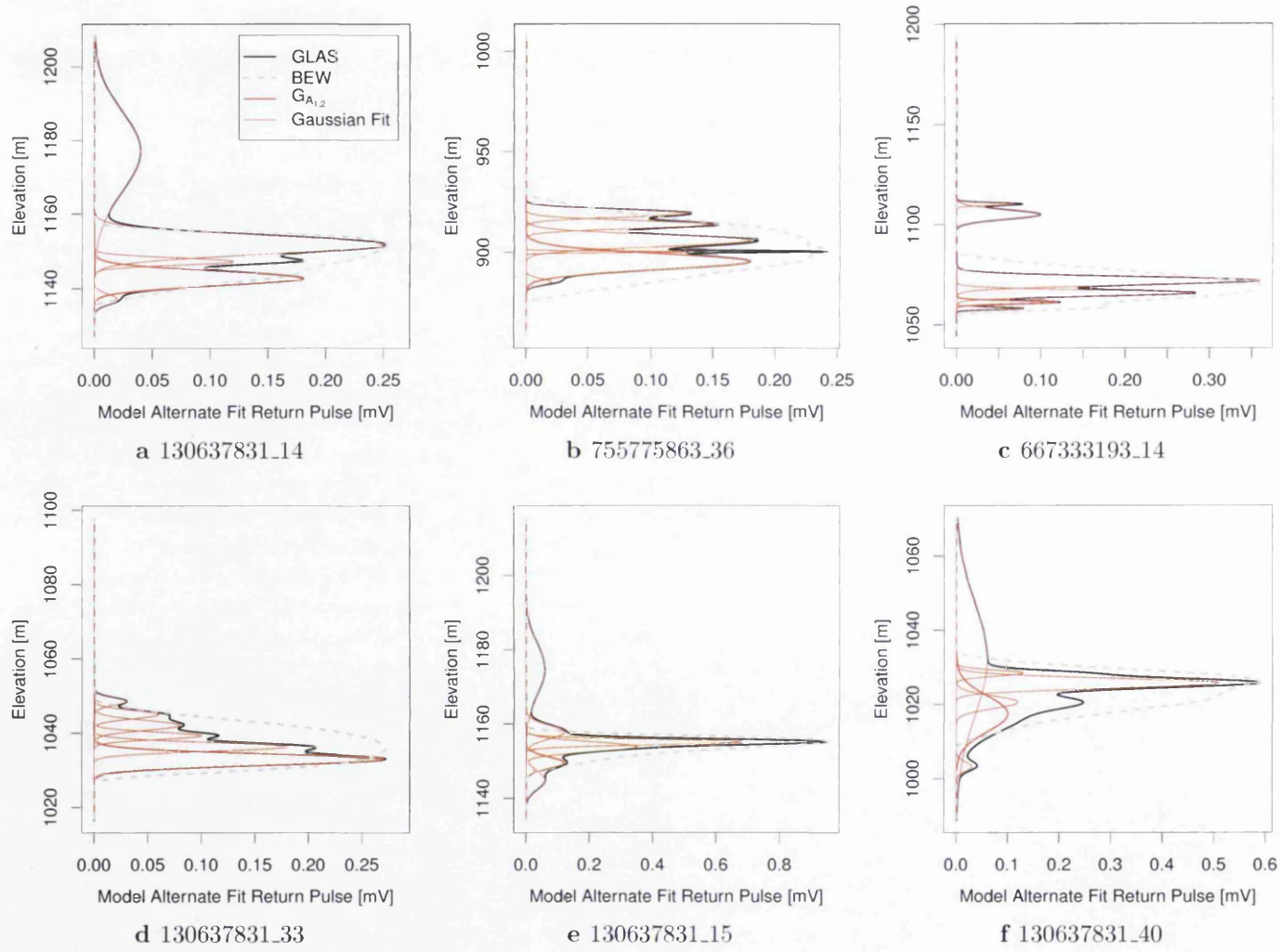


Figure 4.7: Examples of inaccurate V_H, obtained by the Gaussian decomposition method of Rosette et al. (2008). The BEW (dashed line) and GLAS waveforms (black solid line) are shown, in addition to the decomposed Gaussians (red), where the ground representing Gaussian ($G_{A_{1,2}}$) is given by the **thick** red line.

the centroid of the BEW, and if used to derive V_H by the method of Rosette et al. (2008), would induce an overestimate. In these particular cases, Gaussian number three (third least elevated) would better represent the true ground. This equates to an approximate difference between $G_{A_{1,2}}$ and the true ground of 4.63 m, 4.87 m, and 4.54 m for each respective waveform.

For Figure 4.7d, Gaussian 1 (least elevated) is identified as $G_{A_{1,2}}$, the difference between this and the true ground is approximately 3.49 m. In this specific case, a more appropriate representation of $G_{A_{1,2}}$ would be Gaussian 2, however, intensity reductions by shadows in the leading edge of this waveform dictates that this is not the case.

Figures 4.7e and 4.7f suffer with shadowing on their trailing edge; this appears more prominently for Figure 4.7f. As a result the intensities of both Gaussian 1 and 2 are reduced, such that neither represent the true ground, in both cases a more representative $G_{A_{1,2}}$ would be Gaussian 4. For each respective waveform, differences between $G_{A_{1,2}}$ and the true ground were noted at 4.29 m, and 5.65 m. For this study approximately 50% of all shadow afflicted waveforms misidentify $G_{A_{1,2}}$, leading to overestimates of V_H. This equates to approximately 7.5% of all analysed waveforms at Tumbarumba producing inaccurate results. If these estimates are true for all other vegetated regions with similar terrain characteristics, significant portions of global estimates of V_H may be overestimated, inducing uncertainty in terrestrial biosphere knowledge.

Current analysis suggests that uncertainties in V_H occur more frequently over increasingly sloped terrain (as suggested by Figure 4.6 and Table 4.1). This further enforces the requirement of understanding the influence of slope when imaging vegetation by direct waveform LiDAR techniques, an area of this discipline that is still relatively poorly understood, hence the common use of slope filtering techniques (Simard et al., 2011; Los et al., 2012).

4.3 Implications of Shadow Presence

There are many implications of the presence of shadows in waveform LiDAR returns, these can be both positive (refining current techniques/developing new methods to obtain new vegetation metrics), and negative if ignored (as discussed

in Section 4.2.3). Situations where shadow presence can be utilised in a positive manner are discussed in this section.

The knowledge of shadows in waveforms have potential to allow the derivation of previously unobtainable (from waveform LiDAR) vegetation parameters, in addition to refining current techniques, and vegetation parameter accuracies. Briefly, it is suggested that information (and/or refinement) can be obtained by employing shadow knowledge for the following parameters: within footprint tree location (elevation), improved ground detection, and vegetation crown width; also pulse shape knowledge (which is not discussed here as it is more of an engineering application). Many of these parameters will be of use in Above Ground Biomass (AGB) estimates, which are often related to vegetation height (e.g. Lefsky et al., 2005).

The way in which shadow information relates to each of the aforementioned parameters that have potential for derivation and/or refinement are briefly discussed in the following sections.

4.3.1 Tree Identification

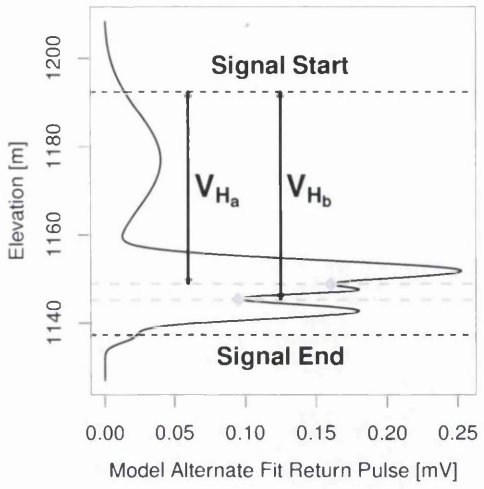
GLAS footprints exhibit large areas, encompassing numerous individual trees, which are not distinguishable for (large footprint) waveform systems. The isolation of individual trees within footprints will allow the identification of individual tree vegetation heights (V_H), and more information on within footprint maximum/minimum V_H.

Isolating individual trees within waveform LiDAR footprints relies on the presence of shadows. Shadows are cast by vegetation subcomponents (shoots, branches, and leaves) which constitute the vegetation crown, the shadow of which (assuming a nadir angle of incident light) will be recorded on the ground directly beneath the vegetation itself. The greatest intensity reduction (for individual shadows) is assumed to correspond to the symmetric centre of each individual tree (which is assumed to grow exactly vertically and exhibit a symmetric crown with uniform density). This assumption is made as the greatest density of light obstructing objects exist here (due to the addition of the trunk to the usual light obstructing vegetation subcomponents). A visual illustration of this, with regards to returned

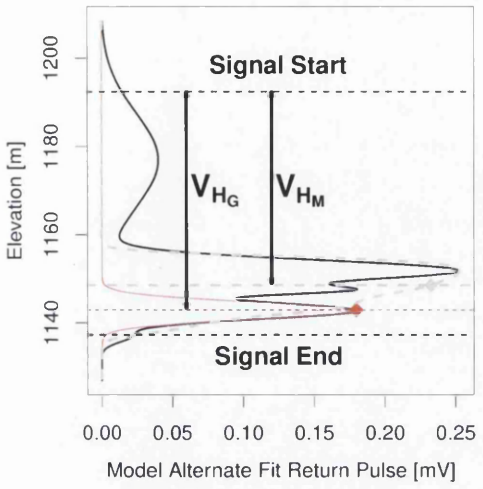
waveforms, is given in Figure 4.8a.
The elevation of the shadows (grey diamonds in Figure 4.8a) are equivalent to the elevations at which the individual trees that cast them grow at. Hence allowing the derivation of individual, within footprint, tree V_H, denoted as V_{H_a} and V_{H_b} in Figure 4.8a.
Limitations by this method are expected to arise from vegetation exhibiting similar V_H, where the above technique will not be indicative of individual trees, but rather tree clusters. In such a situation individual vegetation components in the waveform signal become mixed, inducing difficulty in retrieving individual tree information.

4.3.2 Ground Elevation Refinement

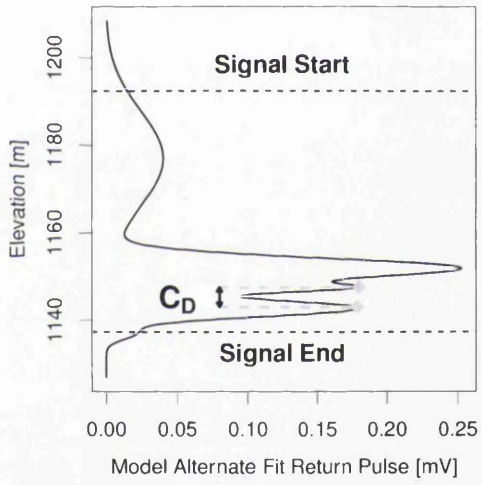
Vegetation height (V_H), by the method of Rosette et al. (2008), is estimated as the difference between the canopy top (i_SigBegOff) and the **mean** ground elevation (assumed as $G_{A_{1,2}}$ from Section 3.9.2); this yields a mean value of V_H (given by V_{H_M}) for the entire footprint. However, this elevation can be misinterpreted when shadows are present in waveforms, as phantom peaks can cause Gaussian decomposition misfitting. This can cause $G_{A_{1,2}}$ (assumed to be the Gaussian that represents the mean ground elevation) to represent elevations lower than that of the mean ground (see Figure 4.8b; red solid line). This can yield overestimates of V_H as interpreted by GLAS (V_{H_G} in Figure 4.8b).
Rectifying this problem is relatively simple, namely, local minima induced by shadows can be eradicated by fitting an envelope to the waveform ground component. This will represent the ground component negating the effects of shadows; this is represented by the BEW (grey dashed line in Figure 4.8b) in this case. Performing the Gaussian decomposition technique to the waveform should allow $G_{A_{1,2}}$ to more accurately represent the mean ground elevation.
Without external data, such as airborne LiDAR derived DEMs (as used here), at present it is not possible to definitively distinguish shadows from physically observed features within any waveform. However, with the use of multivariate regression model techniques linking waveform features to physical vegetation and terrain conditions, it is suggested that probability of finding shadows within indi-



a Tree Location



b Mean Ground



c Crown Width

Figure 4.8: (a) GLAS alternate model fit waveform (black line; see Section 3.4) illustrating shadow elevations (grey diamonds), which equate to within footprint tree elevations, allowing the derivation of individual tree heights (V_{H_a} and V_{H_b}). (b) GLAS waveform (black line) highlighting the improved detection of the **mean** ground elevation (grey diamond) according to the BEW (grey dashed line); $G_{A_{1,2}}$ centroid (red diamond) represents the lower part of the ground due to a shadow induced phantom peak. V_{H_G} and V_{H_M} are vegetation heights derived using the $G_{A_{1,2}}$ centroid, and mean elevation as respective ground locations. (c) GLAS waveform (black line) where the vertical range of the shadow induced intensity reduction is assumed to represent individual tree crown diameters (C_D); for the illustrated case $C_D \simeq 3.45$ m.

vidual waveforms may be possible. However, a multitude of data (high resolution DEM, and waveform LiDAR), with varying vegetation and terrain information would be required in order to form such models, allowing such assumptions. Such a study would highlight the effectiveness of the use of the Gaussian decomposition V_H retrieval method over sloped terrain.

4.3.3 Crown Width

Shadow occurrences in waveforms allow a direct derivation of crown diameter (C_D) since any shadow cast on the ground is caused by the vegetation held in the crown of the corresponding tree. This assumes that the incident light originates from a nadir geometry, and also that the light obstructing vegetation belongs to a single crown.

C_D is deciphered in shadow inflicted waveforms as individual local intensity reductions (see Figure 4.8c). A width measure of this intensity reduction is assumed to be a direct measure of the width of the shadow cast by the crown, and thus a measure of C_D itself; in this case C_D is approximately 3.45 m.

The approximation of C_D arises from the uncertainty to and from which the shadow is identified as starting and ending. Further investigation can quantify this boundary more precisely, yielding a robust framework from which C_D can be obtained reliably. It is noted that in the current example, multiple shadows are present, but only one has been outlined for simplicity. For multiple occurrences within waveforms or throughout homogeneous stands when considering multiple footprints, a mean crown diameter can be estimated provided at least one estimated of C_D is retrievable.

Limitations arise in this technique where vegetation is densely packed, where measures of C_D are less likely to be for individual trees, but rather a tree cluster. This can be useful provided some site information (vegetation type) is known, as some trees grow in clusters dependent on conditions.

4.4 Summary

In summary this chapter highlights the potential of the misinterpretation of returned LiDAR waveforms over vegetated surfaces induced by shadowing effects (where incident light is obstructed by an object), particularly with regards to estimates of vegetation height. Employing the work Hancock et al. (2012) as a basis, physically observed GLAS waveform LiDAR returns were tested for shadowing effects, believed to be induced by vegetation subcomponents (shoots, branches, leaves, etc.) from the crown.

Waveform representations were constructed from 3-dimensional spatial data from a discrete return airborne LiDAR system (Section 4.1) which enabled the identification of shadows in approximately 15% of all waveforms tested at the Tumbarumba site. It was also noted that the frequency at which shadows occur in waveforms increase with more severely sloped terrain (Section 4.2.1).

Implications of the presence of shadows within waveform returns such as: individual tree identification, ground elevation refinement, and crown width retrieval, were also noted in Section 4.3. Examples extended to identifying individual tree elevations within footprints, and the potential of deriving corresponding crown widths. In addition, the influence of shadows has been noted as a key cause in potentially misinterpreting waveform geometry, resulting in incorrect physical feature representation by decomposed Gaussians and thus, incorrect estimates of vegetation height.

Chapter 5

Slope from Full Waveform LiDAR

Waveform LiDAR footprints retrieved over sloped terrain are often spurious, therefore the quantification of such information is beneficial in understanding waveform interactions with such surfaces, and will allow steps to be taken towards reducing uncertainties induced by sloped terrain.
Two methods are developed and tested in order to obtain slope predictions from full waveform LiDAR returns. The first method uses an indirect approach, employing previously validated, relevant airborne LiDAR data metrics to create a look up table (LUT) for slope predictions. The second method builds upon a major concept of the first. However, slopes are derived directly from the waveform itself, negating the use of any other external data.
Slope predictions for both methods are first tested and validated with FLIGHT simulated waveforms, and subsequently applied to GLAS waveforms which are validated against airborne LiDAR equivalent data. The better performing of the two methods was tested further against slope derived from British Ordnance Survey data for the principality of Wales in the United Kingdom.

5.1 Slope Screening Model

The Slope Screening Model (SSM) is the first of two discussed methods developed in this project. The SSM was developed in two stages: initial set-up (Section 5.1.1) and execution/prediction (Section 5.1.2).

The set-up initialises a relationship between quantifiable features in GLAS waveforms and slope, the latter derived from validated external data (airborne LiDAR in this case). This establishes a regression model linking measurable waveform features and slope, which acts as a basis from which slope predictions from other waveforms can be made.
Post set-up, slope predictions can be made on the basis of the width of waveform ground components, via the established LUT.

5.1.1 LUT Relationship

For the purpose of this method, waveforms measured against range (not relative time) are required. Considering GLAS waveforms, the alternate model fit waveform (as described is Section 3.4.2) is employed.
The waveform extent is established based on the waveform signal start (i_SigBegOff) and signal end (i_SigEndOff) locations, as described by (Lefsky et al., 2007). Within the waveform extent, the elevation of the ground component's maximum amplitude is located (Figure 5.1a: filled diamond), defined as the final inflection point within the waveform extent where intensity is ≥ 0.2 V. From here on this location is referred to as the ground peak (G_p).
The location of G_p allows the isolation of the waveform ground component itself (defined as the last peak with intensity ≥ 0.2 V within the waveform extent), required for slope derivations. In order to isolate the ground component, local minima are established (every 10 vertices) within the waveform extent, as a function of return intensity. These are analysed such to obtain two local minima of interest; one exhibiting greater, and the other exhibiting lower elevation than the ground component maximum-amplitude elevation (Figure 5.1b; respective filled triangles). In addition, these minima must simultaneously exhibit the least intensity, and closest proximity to G_p.
This ground component isolation process is automated, executed by first screening the array of local minima within the waveform extent and reporting those that are less and more elevated than G_p as sub-arrays. Considering each minima sub-array individually, the elevations of the start (E_s) and end (E_e) of the ground component can be constrained.

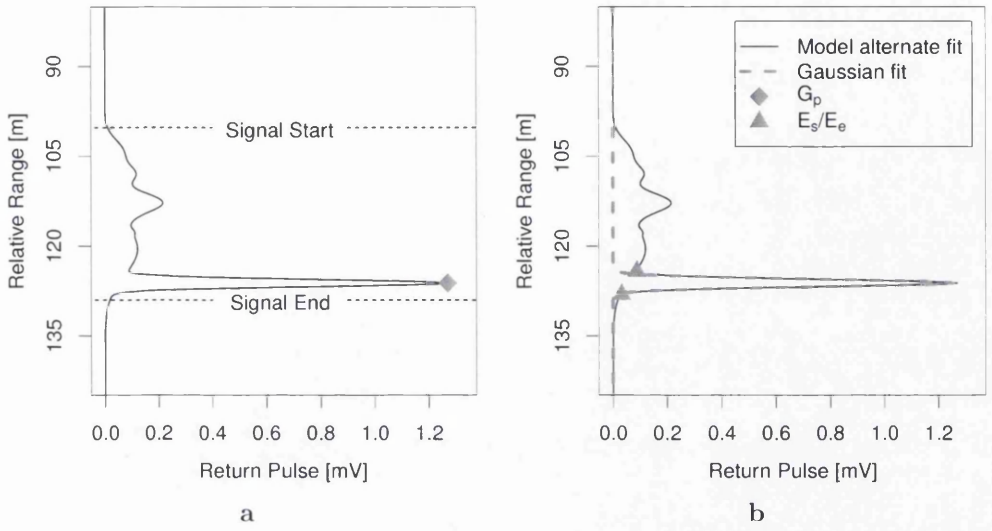


Figure 5.1: Example waveform analysed by SSM. (a) shows the signal start and end boundaries, and the elevation/relative range of G_p (filled diamond), and (b) E_s and E_e, the most and least elevated of the ground component boundaries (filled triangles), and Gaussian fit to ground component (dashed line).

Obtaining E_s and E_e utilises an iterative process based on initial primary (p_m) and secondary (s_m) minima. p_m and s_m are defined as the closest and second closest local minima to G_p, respectively. If the intensity difference (ΔI) between p_m and s_m exceeds a defined sensitivity threshold (T_s), the original selection of p_m is replaced by the original selection of s_m, where s_m is shifted to the neighbour of the newly selected p_m. This iterative process is repeated until the condition $\Delta I \leq T_s$ is met, allowing p_m to be accepted as a ground component boundary (Figure 5.1b: filled triangles). This process is repeated for both sets of minima sub-arrays, yielding minimum and maximum elevation ground component boundaries (both sides of G_p).

A single Gaussian is fitted to the isolated ground component using the method described by Duong et al. (2006), where elevation restrictions (on the y-axis), between E_s and E_e are imposed. The quality of the fit between the Gaussian and the waveform ground component is tested by means of a Kolmogorov-Smirnov (KS) test (see Massey Jr, 1951). Waveforms that exhibit poor fits (where KS-test D-statistic < 0.8; described in Appendix 9.3) are excluded from further analysis.

The width of the well-fitted Gaussians (D-statistic ≥ 0.8) are measured by the standard measure, full width at half maximum (FWHM). As a next step, these waveform associated width measures are paired with slope calculations derived from validated data sources of corresponding footprint area (airborne LiDAR or OS data for this study; see Section 3.10). This allows a link between waveform ground component widths and (relatively) high resolution slope angle information to be established.

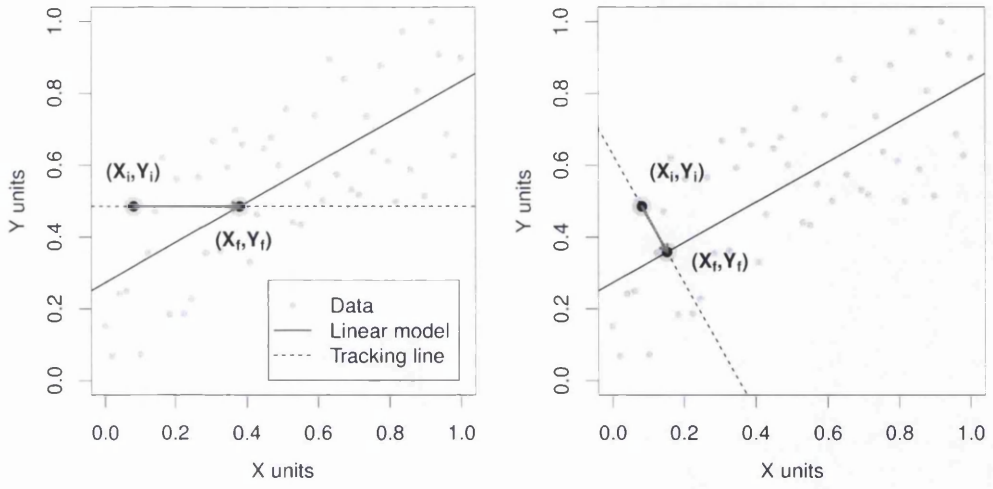
For multiple waveforms linked by these parameters a regression model (and LUT of observations) can be formed; thus allowing slope angle information to be extracted for any given waveform ground component width.

5.1.2 Slope Prediction

Predicting slope from waveforms independently requires the use of processes described in Section 5.1.1, up to the calculation of the ground component width. Employing the waveform ground component width as an input, the previously established LUT is searched for the nearest corresponding width value to that of the input value. The (LUT) width and associated slope value are recorded, giving an input x, y position. This position is tracked perpendicularly to the regression model, formed from the LUT data, as explained below. An example of this process is summarised in Figure 5.2, for mock data (randomly generated on a normal distribution) and an associated linear regression model.

This lookup method is considered more appropriate than that of the standard linear lookup method, which does not account for any variability in its input parameter which can affect the accuracy of its corresponding output (Figure 5.2a). Whereas the least squares method does account, somewhat, for the variability present in its input (ground component width in this case) which will allow the retrieval of more accurate outputs (slope angle; Figure 5.2b).

Due to the nature of this model, inputs that exist below the LUT regression model (on the y-axis), and are located close to zero (on the x-axis), can be tracked (by the least squares method) across the y-axis, yielding a negative output value. This can occur as the least squares look up method tracks input points to the regression model by a line perpendicular to that of the regression model line itself. Thus if



a Linear **b** Least squares

Figure 5.2: Example of (a) linear and (b) least squares look up methods. The latter is employed by SSM to predict slope information from input waveform ground component widths. Here, these look up methods are applied to a single point for a mock dataset (randomly generated on a normal distribution) with an associated regression model (black line), and "tracking" line (dashed line). Note: (X_i, Y_i) = initial coordinates, and (X_f, Y_f) = final (model) coordinates

the "tracking" line exhibits a negative enough gradient, whilst the input point is sufficiently far away from the regression model line (on the y-axis), and near zero (on the x-axis), the y-axis can be crossed, yielding a negative result. Output values that represent negative slope values are unrealistic and are therefore removed from further analysis.

The lookup method requires the regression model coefficients (gradient m_r, and intercept c_r) and some initial values (X_i, Y_i) for FWHM and slope (defined as a function of input FWHM) respectively. Using simple line geometry, tracking perpendicularly to the regression model from input values, an intersection point on the regression model itself can be identified, allowing the derivation of ground slope angle for such a ground component width. For a single input, the intersection coordinates (X_f, Y_f) on the regression model can be found by Equations 5.1 and 5.2, where m_p and c_p are the gradient and intercept of the line perpendicular to the original regression model (dashed line in Figure 5.2b) respectively. Here

Equation 5.2 (Y_f) will inform on modelled slope predictions (given that the x and y axes represent FWHM and slope respectively).

$$X_f = \frac{c_p - c_r}{m_r - m_p} \tag{5.1}$$

$$Y_f = m_r x_f + c_r \tag{5.2}$$

In summary, SSM slope predictions are found by the following steps:

- Input ground component FWHM
- Project point to (previously established equivalence) relationship linear model by a perpendicular regression model (arrow in Figure 5.2b)
- Intersection points contain ground slope information (Y_f), related to original FWHM input

5.2 Independent Slope Prediction Model

The Independent Slope Prediction Model (ISM) is the second of the two developed methods. This method takes key concepts from the SSM, with regards to inferring slope from waveform LiDAR features, however these are applied in a more direct and robust manner.
The concepts of Section 5.1.1 are applicable to the ISM, a summary of the methodological steps are:

- Establish waveform extent according to Lefsky et al. (2007), for a waveform in relative range units
- Locate elevation of ground component maximum amplitude (G_p)
- Establish and screen sub-arrays of local minima established either side of G_p to isolate the waveform ground component boundaries (E_s, and E_e; following the method described in Section 5.1.1
- Fit single Gaussian within the confines of E_s and E_e
- Screen Gaussian fit for fit quality with respect to observed data (KS-test)

Following the steps above in the methodology, similarities to the SSM continue, as a width measure of the fitted Gaussian is recorded. However, the width measured here is derived at the base of the Gaussian, at a threshold where the Gaussian's amplitude exceeds an intensity of $T_G = 0.001$ V. This value of T_G is somewhat arbitrary, but is well suited as the Gaussian fit is performed such that its extruding tails have an almost zero amplitude (fluctuating down to approximately 10^{-10} V) and are therefore not considered when recording it's width. The measured width of the Gaussian at this threshold represents, in reality, the difference between the maximum and minimum ground elevations recorded by the LiDAR footprint.
Slope predictions from waveforms (θ_w) are directly calculated from this width measure/elevation difference (ΔE). In particular, slope is calculated from a variation of Equation 3.5, where $\Delta E = E_{95} - E_{05}$ is represented by this directly measured elevation difference, and can thus be rewritten as:

$$\theta_w = \left(\frac{\Delta E}{\bar{X}_{fp}}\right) \tag{5.3}$$

Where $\bar{X}_{fp}$ is the mean footprint diameter, as in Equation 3.5.

5.3 Model Assumptions

Both the SSM and ISM will work in almost any situation, however, in some instances results are unable to be produced. Both models require the identification of the ground component within waveforms to make meaningful predictions. However, for waveforms retrieved over very complex terrain, the probability that a distinguishable ground component is identified within the waveform decreases. If no ground component is identifiable, no slope predictions can be made, and are thus not assigned.
All slope predictions with meaningful values, represent the *average* planar/linear slope across a given LiDAR footprint with known (average circular) dimensions. This assumes that the maximum and minimum within footprint elevations are separated by a fixed distance represented by these imposed footprint dimensions. In reality these elevations may be separated by a lesser or greater distance, but this information is unobtainable from waveform LiDAR returns. By assumption of an

average planar/linear slope, ground roughness is also not accounted for, however, this information is also unavailable from LiDAR waveforms.
Additionally, airborne LiDAR information is regarded as "truth" in this study, however, in reality it is subject to inherent uncertainty as discussed in Section 4.2.2. It is noted that uncertainties in airborne LiDAR data are capable of distorting elevation data, key in validating model derived slope information. However, for this study the minor discrepancies between field collected data and airborne LiDAR derived equivalents are considered acceptable, justifying airborne LiDAR data as a validation source.

5.4 Filters

Filters were applied to the SSM and ISM methods and associated validation processes respectively. Filters described in Sections 5.4.1 to 5.4.4 were applied during execution of both methods. The filter in Section 5.4.5 was applied to the SSM only, whereas the filter described in Section 5.4.6 was applied to the ISM.

5.4.1 Goodness of Fit

The Kolmogorov-Smirnov test produces a D-statistic that gives a measure of the cumulative difference between an observed and predicted dataset (see Appendix 9.3).
This method was used as a means of testing the quality of the fitted Gaussian to the model alternate fit waveform (from GLA14 products). The tests resulting D-statistic reflects the quality of the fit, where $D = 0$ is a perfect fit, and $D = 1$ is a very poor fit. A poor Gaussian fit generally misrepresents the width of the identified waveform ground component (be it measured at FWHM or at T_G), which, in turn, will lead to spurious slope predictions.
Throughout SSM set-up, SSM prediction, *and* all of the ISM method, the maximum value of D permitted was 0.2, values greater than this were deemed to belong to poorly fitted Gaussians, and hence these waveforms were discarded from excluded analysis.

5.4.2 Saturated Waveforms

The i_satElevCorr GLAS quality flag indicates if a waveform has saturated the detector; if this is the case, these waveforms are removed from further analysis. Saturation typically originates from the most reflective surfaces which produce intense returns. On occasion, the reflection intensity will be greater than that detectable, hence intensities greater than this limit convolve to the detector's permissible upper limit and flatten the peak. This alters the waveform shape, which as a consequence of the methodology of both outlined slope retrieval methods, has the potential to produce spurious results.

5.4.3 Elevation

The GLAS waveform reference elevation (i_elev) was adjusted to match the SRTM ellipsoid by Equation 3.3, and was subsequently compared with SRTM 4.1 elevations (i_DEM_elev), where it is assumed that large elevation differences between the two, indicate problems in either dataset (Los et al., 2012; Simard et al., 2011). As reported by CGIAR-CSI (2013), the 95% confidence interval of the SRTM DEM 4.1 data is 8 m globally, hence any data with an elevation difference between Equation 3.3 and i_DEM_elev exceeding this 8 m limit were excluded from further analysis.

5.4.4 Maximum Vegetation Height

Waveforms were analysed for maximum vegetation height according to Rosette et al. (2008), those that exceeded the documented maximum vegetation height for their respective validation area were excluded from results. For example, when independently testing these methods over Wales (see Section 3.6), the maximum permissible vegetation height was set at 65 m, as the tallest tree recorded within Wales was 63.8 m (Forestry Commission, 2012b).

Although the previous filters should eliminate noise and elevation misalignment issues, this filter was established for the rare occasion that both waveform noise is high and the computed D-statistic (from Section 5.4.1) is low.

5.4.5 Point Density

SSM LUT data were binned in both x and y by intervals of 0.02 m for ground component widths (y) and 0.5° for slope (x), respectively. Data were considered in x and y individually, and a point density threshold was created. This threshold ensured that slope predictions were only made from the LUT, where sufficient data were present. This minimises the effect of predicting slopes from areas where not enough data is present to predict results with high confidence.

The thresholds (in x and y) were defined as $1/e^2$ (13.5%) of the maximum number of points found within any bin. In this study, by this definition, the threshold of the minimum number of data points required in any width bin (in y) was found to be 6. Whereas the minimum number of points required within any slope bin (in x) was found to be 4 at the $1/e^2$ threshold.

5.4.6 Minimum Measurable Slope

Returned waveform ground components (where identifiable) will always exhibit a finite width, because of the duration of the emitted signal, and associated atmospheric attenuation. The implication is that for completely flat surfaces a difference between the highest and lowest within footprint ground elevation (ΔE) is found which translates to finite slope measurements according to Equation 5.3.

An approach developed by Los et al. (2012) allowed for this effect, by relating vegetation height to the 5% values of the area under the first Gaussian (least elevated), as fitted to the waveform according to Duong et al. (2006). This allowed a linear regression model to be established, providing minimum measurable vegetation height estimates as a function of the area under Gaussian 1, which was subtracted from all vegetation height estimates derived from LiDAR waveforms.

A similar refinement can be applied here, relating the FWHM (in metres) of the source emitted laser pulse, to the maximum of the returned waveform ground component amplitude (A_{max}). In this case, 1% values (rather than 5% values) of the emitted pulse width were related to the maximum amplitude per 0.01 mV intervals. This allows ΔE, and consequentially slope estimates, to be adjusted according to:

$$\theta_m = a + bA \tag{5.4}$$

where:

θ_m = Minimum measurable slope [°]
A = Maximum amplitude of waveform ground component [V]
b = 0.166 [m mV^{-1}]
a = 0.712 [m]

Figure 5.3 illustrates Equation 5.4 (black points and dashed line; 1% values of FWHM), showing an increase in θ_m with increased A. The fitted coefficients a, and b, from Equation 5.4 were obtained from the analysis of $\sim$ 100000 waveforms. θ_m is subtracted from *all* slope estimates obtained by ISM.

5.5 Model Performance Parameters

Statistical measures were employed to assess the performance of the slope retrieval models with regards to observed or known data. Measures of correlation (R/R^2), significance (p), and root mean square error (RMSE) are often utilised in such data comparisons. In addition, here we apply some less frequently used measures, these are: goodness of fit (D), factor of two test (F_2), fractional bias (F_B), and normalised mean error (ε_{nme}).
Typical magnitudes of these measures have been summarised by Chang and Hanna (2004) based on extensive experience of evaluating data models for many field datasets. It was concluded for this study that the following model performance values were acceptable: $F_2 \geq 0.5$, $|F_B| < 0.3$, and $\varepsilon_{nme} < 2$. Note, these are not hard limits and it is necessary to consider all performance measures with regards to model acceptance. Explanations of each model performance test are given in further detail in Appendices 9.3 to 9.7.

5.6 Initial Theory Testing

Both slope retrieval techniques (SSM and ISM) were tested with FLIGHT simulated waveforms, and GLAS observed waveforms.
Following results found by Rosette et al. (2010), FLIGHT was chosen to see how changes in waveform shape vary with the effect of ground slope angle. FLIGHT is

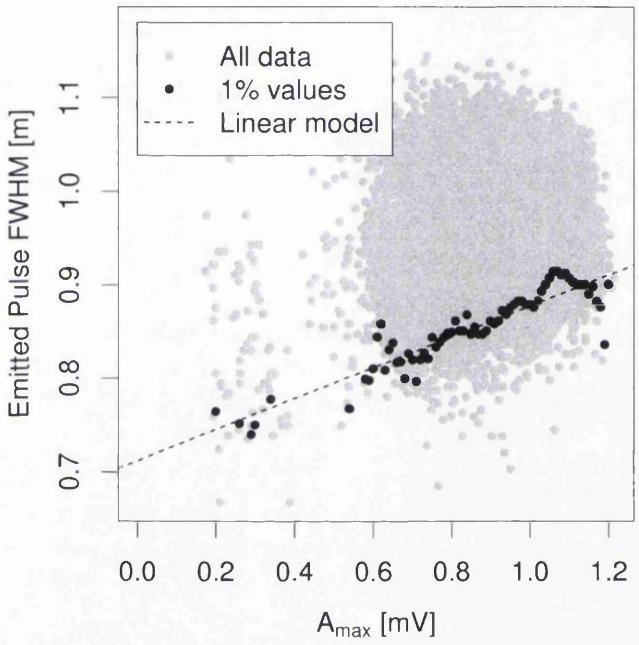


Figure 5.3: Emitted laser pulse width (FWHM) as a function of waveform ground component maximum amplitude. A regression model is formed by use of the 1% values of the emitted laser pulse width as a function of waveform ground component maximum amplitude (A_{max}) per 0.01 mV intervals (black points).

a radiative transfer model based on Monte Carlo simulation of photon transport developed by North (1996). FLIGHT is capable of simulating a forest scene based on forest structural and optical inputs (North et al., 2010), listed in Table 5.1. Incident photons are simulated from an initial geometry, and interactions with the simulated forest are calculated (scattering, reflection, absorption), and recorded at a final detector geometry. This allows a simulated waveform to be built from detected photons, representing the interaction of light with the simulated forest canopy.

FLIGHT simulated waveforms were used to determine if a relationship between the waveform ground component widths and ground slope angles (defined by FLIGHT input parameters) existed. FLIGHT waveforms provide a good platform to test theories as they are somewhat idealised, with controlled slope (no roughness variations), and no atmospheric signal attenuation distorting features. Hence the resulting relationship between the ground component width and slope angle should be be subject to minimal uncertainty.

A total of 84 simulations, each with 20000 photons, were completed in the three dimensional case where the simulated forest consisted of elliptical crowns (as representative of *mature* Scots pine, Forestry Commission, 2012b), with all parameters held constant (see Table 5.1), apart from those of interest; these were F_{cov}, LAI, and ground slope (Table 5.1).

LAI and F_{cov} (marked with ‡ in Table 5.1) covaried within realistic bounds; low values of LAI concur with low values of F_{cov} (row a), and high values of LAI concur with high values of F_{cov} (row c). The parameters in Table 5.1 marked with † were varied yielding a broad array of simulations to maximise the coincidence with waveforms observed in reality, most notably by GLAS.

SSM and ISM were tested for GLAS waveforms that fell within the bounds of each of the study sites used in this project. Resulting slope angle predictions were tested against airborne LiDAR derived equivalent values.

Table 5.1: Summary of parameters and values used in FLIGHT simulations. Note: ‡ denotes that values for these variables must be chosen from corresponding lines only (either a, b, or c). † denotes only a singular value from this list may be chosen per execution; all values are used in combination over 84 executions.

Variable	Units	Description	Value Range
Mode	-	Scattering simulation method	"l"
Dimensions	-	Number of dimensions modelled	3
Zenith	°	Illumination and view zeniths	0, 0
Azimuth	°	Illumination and view azimuths	0, 0
Wavebands	-	Number of simulated wavebands	2
N_{photon}	-	Number of simulated photons	20,000
LAI‡	-	Leaf area index	1 (a)† 2, 3, 4 (b)† 6 (c)†
Foliage composition	-	Fraction of foliage that is: green, shoot, bark	1, 0, 0
LAD	-	Leaf area distribution in 10° bin fractions	0.015, 0.045, 0.074, 0.1, 0.123, 0.143, 0.158, 0.168, 0.174
S_r	-	Soil roughness	0
AOD	m	Aerosol optical depth (< 0 direct beam only)	-1
Leaf size	m	Leaf radius (modelled as circular disc)	0.01
F_{cov}‡	-	Fractional cover	0.1, 0.2, 0.3 (a)† 0.4, 0.5, 0.6 (b)† 0.7, 0.8 (c)†
Slope	°	Planar slope angle of ground	0, 1, 2, 5, 10, 20†
C_{shape}	-	Crown shape (e = ellipse, c = conical, f = filled)	"e"
C_{dim}	m	Crown dimensions (radius and height)	5.25, 16.00
First branch heights	m	First branch height distribution between defined minimum and maximum	0.50, 1.81
DBH	m	Diameter at breast height	0.15

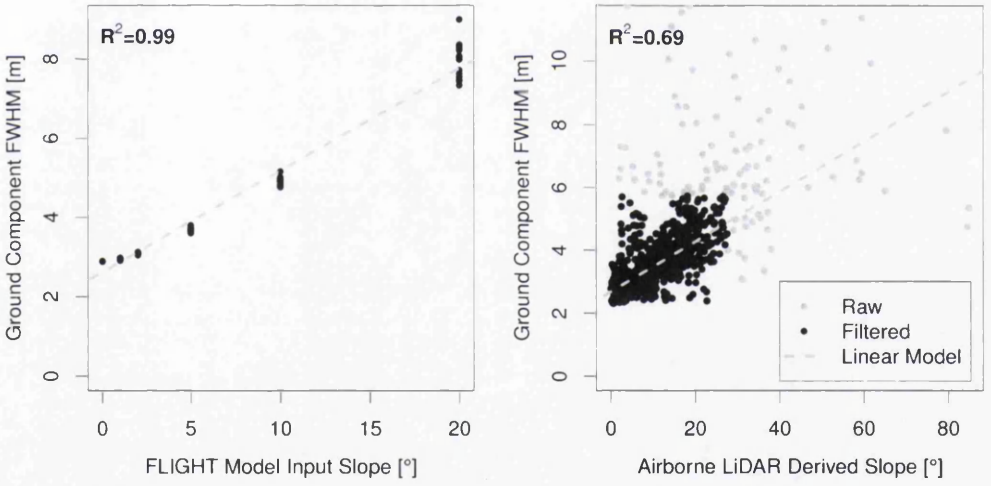
5.7 Results and Discussion

5.7.1 SSM - Look Up Table Establishment

FLIGHT and GLAS waveforms were employed to test SSM slope predictions. Before predictions could be made a lookup table (LUT) needed to be established (see Section 5.1.1) for both data types. For both data types, waveform ground component widths were used to obtain slope angle information through this LUT, where results were compared to reference equivalent values.

For FLIGHT, 84 simulated waveforms were analysed, and slope estimates compared with model input slope values. GLAS derived slope values on the other hand were compared with airborne LiDAR equivalent values made via Equation 3.5 for all footprints within all study sites.

The results for FLIGHT and GLAS equivalence relationships are shown in Figure 5.4, and are briefly discussed below.



a FLIGHT b GLAS

Figure 5.4: Lookup data and fitted linear models (dashed line) between waveform ground component widths and known ground slope angle for (a) 84 FLIGHT simulated waveforms, with changing variables according to Table 5.1, and (b) GLAS footprints within airborne LiDAR coverage regions for all study sites (Section 3.1). Grey points are filtered from the look up data (see Sections 5.4.1 to 5.4.5), hence the relationship consists of the remaining data (black points) only.

Testing SSM LUT with FLIGHT

FLIGHT waveform ground component FWHM widths were linked to FLIGHT slope value inputs (Figure 5.4a), forming a LUT relationship. This was formed over a range of varying vegetation parameters (LAI and F_{cov}), and slope ($0° - 20°$) permutations (Table 5.1). The relationship was confirmed for linearity by the rainbow test of Utts (1982), and exhibits an increased variability in ground component widths with increased slope angle. This strongly suggests that FLIGHT waveform ground component FWHM width (W_F) and ground slope angle (θ_F) exhibit a linear relationship that appears to be (generally) independent of changes in vegetation structure.

The changes in W_F with increased θ_F appear to be slope driven (first order effect), as the vegetation parameters were varied identically for unique values of slope. Vegetation structure changes are a second order effect, that become more dominant for higher slopes. W_F for these higher slope angles vary according to within footprint vegetation position (with regards to elevation), explaining the observed increase in variability, as FLIGHT generates vegetation positions within a scene randomly.

Testing SSM LUT with GLAS

Similarly as with FLIGHT, GLAS waveform ground component FWHM widths (W_G) were compared with validated slope data (Figure 5.4b). In this case airborne LiDAR derived slope values (θ_G) via Equation 3.5 were employed.

Figure 5.4b illustrates that the highest density of data are held within the (black) highlighted trend. Data become increasingly sparse with distance from the formed regression model. The LUT relationships constituent data exhibit a relatively large range for W_G and θ_G. Thus for prediction purposes, any FWHM width value used as an input to this LUT relationship will be attributed to multiple values of slope, and vice-versa.

In addition to the applicable filters described in Section 5.4, the LUT relationship data were scrutinised for the following:

- Understorey presence: dense understorey (present at SOA and Tharandt) mixes low vegetation and ground component signals due to the understorey

vegetation height range, thus resulting in (large) unrepresentative FWHM measures (see Appendix 9.8). Waveforms with understorey presence were found by the use of airborne LiDAR data and removed from further analysis.

- Complex terrain: slopes $\geq 8°$ are generally indicative of complex terrain (most notable at Norunda, Tharandt, and Tumbarumba), where ground returns originate from a range of elevations, which can result in signal mixing, which leads to unrepresentative large FWHM measures (see Appendix 9.8). Waveforms found to exhibit signal mixing (as suggested from airborne LiDAR data) were removed.

- Severe slopes: Airborne LiDAR derived slopes $\geq 50°$ are deemed inaccurate, or to originate from footprints that cover two elevation extremes such as cliffs (Tumbarumba), or with similar effect, water edges (Canadian sites). Neither are representative of forest structure or offer reliable linear slope measurements (see Appendix 9.8), and were removed from further analysis.

These filters explain the grey points in Figure 5.4b. However, regardless of the applied filters, some spurious data remained in the LUT relationship. This is believed to occur because, regardless of passing the goodness of fit test (and other filters), some data may be inherently anomalous. For example, heavily distorted waveforms which may not follow the established relationship can still exhibit good quality Gaussian fits.

Further variability is expected to arise from the differences between FLIGHT and GLAS waveform representations. FLIGHT models slope linearly, not accounting for within footprint topographic variability, whereas GLAS is subjected to such variability, hence this is reflected in returned waveforms. A greater, natural within footprint variability is expected for GLAS waveforms which accounts for some variability illustrated here.

Post filtering, it was concluded that the GLAS LUT relationship results support those suggested by FLIGHT, suggesting a linear relationship exists between W_G and θ_G, however, greater variability is present. This is somewhat expected due to GLAS waveforms being subject to more influences of noise induction such as atmospheric attenuation and within footprint topographic variability. In addition,

the (LUT) data available in this study dictate that SSM slope predictions can be made up to a maximum of 27°. Beyond this, predictions are possible but unreliable, hence any predictions that are $> 27°$ are discarded. However, if data pertaining to slopes $> 27°$ were available for use in the LUT, the SSM predictive range could be extended.

Comparison

Normalising FLIGHT and GLAS waveforms such that the area under the waveforms are equal to 1, allows an assessment of how alike each data source is, and inform on whether a unique LUT relationship is required for each dataset or if one relationship is suitable for all waveform dataset (up to and beyond FLIGHT and GLAS).

Normalised waveforms for these data sources means that the gradients of each linear model fitted to each respective data source's LUT relationship should theoretically be identical. This is not the case, the gradient (b) and intercept (a) of the FLIGHT fitted and GLAS fitted linear models agree to the order of 10^{-3} and 10^{-2} respectively. This is expected to occur from inherent subtleties present between data sources, such as the way FLIGHT doesn't account for within footprint topographic variability and atmospheric attenuation of the signal etc.

This claim can be analysed with regards to within footprint topographic variability, or slope, by employing waveform signal-to-noise ratios (SNR, Equation 5.5). Analysis of waveform SNRs as a function of slope (Figure 5.5), suggests that slope is a key driver in inducing noise in waveforms. As slope increases SNR generally decreases due to increased signal scattering from multiple terrain interactions; this effect has also been noted by Simard et al. (2011). As SNR is not related to slope by a perfect inverse linear relationship it is suggested that the observed difference between the FLIGHT fitted and GLAS fitted LUT data linear models is indirectly a function of slope through SNR.

$$\mathrm{SNR} = \frac{P_s}{P_n} \tag{5.5}$$

where:

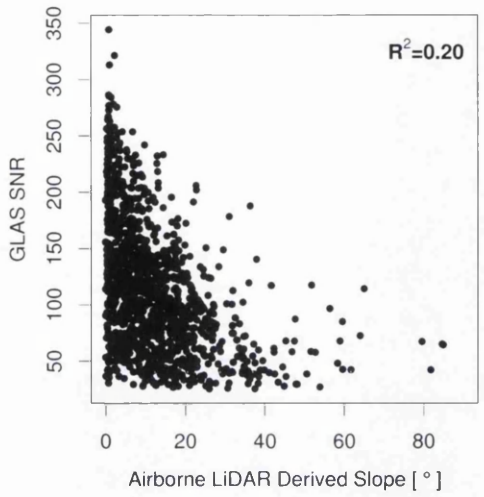


Figure 5.5: Change in signal-to-noise ratio (SNR; Equation 5.5) as a function of airborne LiDAR derived slope angle.

P_s = Signal power (i_maxRecAmp)
P_n = Noise power (i_sDevNsOb1)

The differences in the coefficients of the linear models fitted to FLIGHT and GLAS LUT data suggest that SSM requires unique LUT relationships to be established (Section 5.1) for each data source. A single LUT relationship is not capable of allowing accurate slope predictions across multiple data sources.
Note that the SSM is applicable to vegetated and non-vegetated waveforms, as little difference is expected in the ground component between these two waveform types. However, it is noted that dense canopies ($F_{cov} > 0.85, LAI > 5$) can disrupt ground returns, thus altering waveform geometry and limiting the capability of SSM to infer slope.

5.7.2 SSM - Slope Predictions

The previous section explained the results of establishing a LUT relationship for both FLIGHT and GLAS data. This allows slope predictions to be obtained for both data sources. The following section will present and discuss these inferred slope predictions by use of these previously established LUT relationships.
SSM slope predictions were made for FLIGHT and GLAS waveforms. Initially, a

Table 5.2: Summary of SSM performance parameters for individual sites, SOA is not listed as slope predictions made were outside FWHM confidence region (for reasons described in Section 5.1). Note: n = sample size, R = Pearson correlation coefficient, p_R = fractional probability of no correlation (significance), D = Kolmogorov-Smirnov statistic, F_2 = fraction of predictions within a factor of 2 of observed values, F_B = Fractional bias, ε_{nme} = normalised mean error, and RMSE = root mean square error.

Site	n	R	p_R	D	F_2	F_B	ε_{nme}	RMSE[°]
Tumbarumba	514	0.96	$< 10^{-10}$	0.04	0.97	0.01	0.10	1.79
Loobos	57	0.88	$< 10^{-10}$	0.18	0.67	-0.14	0.31	1.38
Norunda	57	0.67	$O(10^{-8})$	0.35	0.33	0.63	0.96	2.30
Tharandt	109	0.92	$< 10^{-10}$	0.11	0.84	-0.14	0.25	2.51
SOBS	40	0.96	$< 10^{-10}$	0.25	0.70	0.04	0.30	0.90
SOJP	19	0.90	$O(10^{-7})$	0.63	0.26	0.68	1.04	1.12
FLIGHT	186	1.00	$< 10^{-10}$	0.17	0.83	0.00	0.05	0.48

dataset of FLIGHT simulated waveforms were employed to test SSM slope predictions. This consisted of additional LAI settings of 0.15 and 5 and additional F_{cov} settings of 0.25, 0.35, 0.45, 0.55, 0.65, and 0.75, to the existing data present in establishing the LUT relationship (Section 5.6). As noted previously (Section 5.6), these values covaried such that they realistically represented forest structure, and were cyclically combined with the existing permutations of slope from Section 5.6 also.

SSM predictions were made for a total of 186 FLIGHT simulated waveforms, employing the LUT relationship to infer slope values from each waveform's ground component FWHM width with encouraging results (Table 5.2). Results are measured by means of the performance parameters described in Section 5.5. These results were used as a "pathfinder" for GLAS measurements, indicating on possible results for GLAS analysis.

SSM predictions were also made for GLAS waveforms from each study site and compared to airborne LiDAR equivalent data. These results however are biased as the waveforms from which slope is inferred feature in the LUT relationship, from which they are predicted. These data were analysed in this way as an exercise to yield what (theoretically) should be the best results possible for the model, consisting of this precise LUT data. Results were generally encouraging, with inter-site

variance present (Table 5.2).

SSM Predictions with FLIGHT

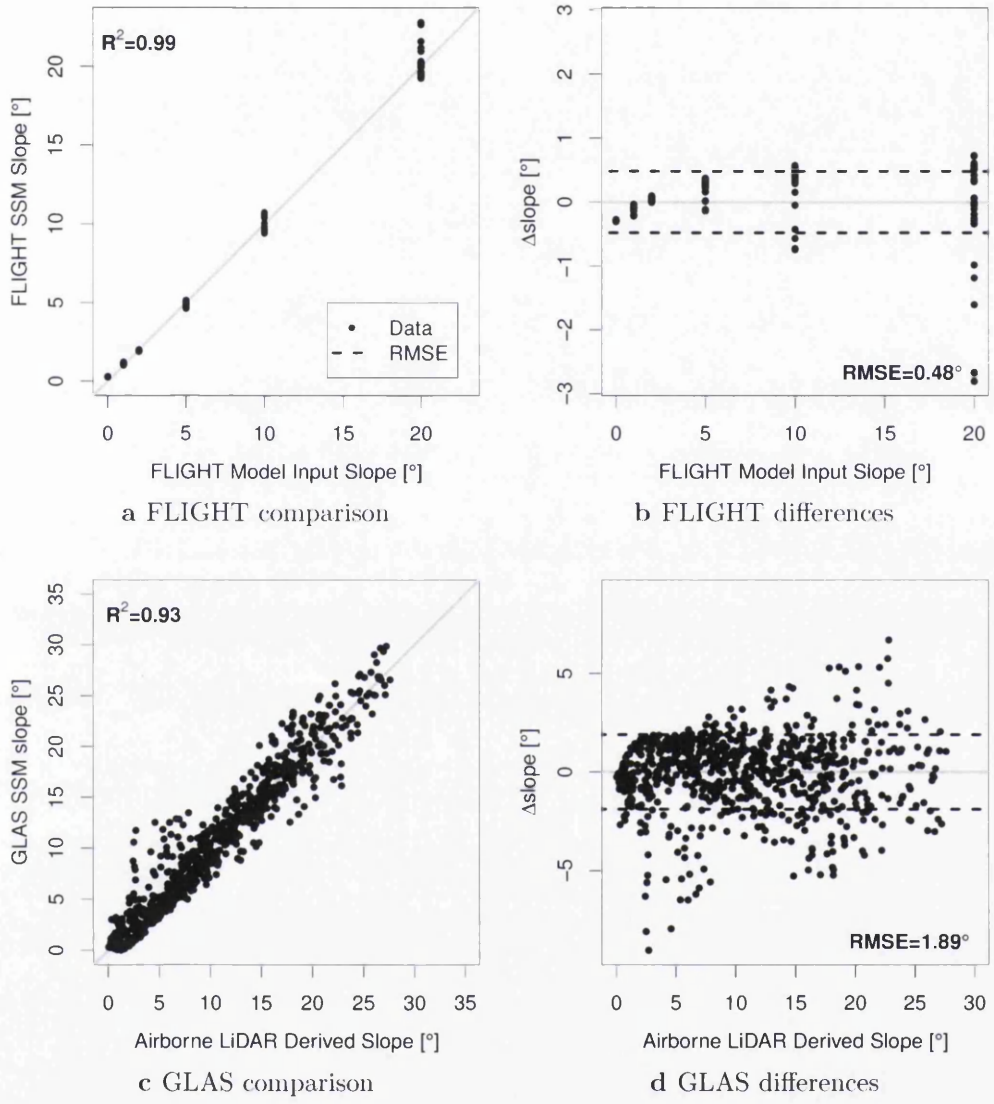
SSM slope predictions for FLIGHT were compared to the model inputs, where direct slope comparisons and corresponding differences between the two slopes are illustrated in Figure 5.6a and 5.6b, respectively. Note that approximately 45% of data employed for prediction purposes were employed in initial establishment of the corresponding FLIGHT LUT. Model performance parameters (shown in Table 5.2) suggest that SSM performs very well for FLIGHT waveforms with these specific vegetation structure arrangements.

Little variability is present over low slope ranges, and increases with increasing slope. The variability present for greater slopes ($\geq 10°$) is attributed to the second order effect of the combination of variable LAI and F_{cov} with sloped terrain, inducing complex scattering events at similar elevations. This suggests that changes in vegetation structure do not disrupt model performance with great magnitude, rather that slope severity is the first order cause of this variability.

SSM Predictions with GLAS

SSM slope predictions for GLAS waveforms were compared to 1 m spatial resolution airborne LiDAR derived equivalent results, the point by point slope comparisons, and corresponding differences are illustrated in Figure 5.6c and 5.6d, respectively. Model performance is measured in the same way as previously stated in Table 5.2. These results are based on 100% of data used to establish the GLAS LUT to infer results in Section 5.7.1 and hence are heavily biased because of site specific tuning. However, these results provide an example of what can be achieved by SSM slope predictions for a combination of variables within GLAS footprint parameters such as: terrain complexity, ground roughness, and vegetation structure.

Greater variability is inherent in these results with regards to those obtained from FLIGHT. This is expected as FLIGHT waveforms suffer no within footprint variability, where GLAS waveforms do, with a particular focus on topography in this case. In addition GLAS waveforms also encounter ground roughness, and multiple



a FLIGHT comparison

b FLIGHT differences

c GLAS comparison

d GLAS differences

Figure 5.6: (a, c) Comparison and (b, d) differences, with RMSE (dashed lines), between known and SSM predicted slope values for FLIGHT simulated and GLAS waveforms respectively.

storey (understorey) canopies, which can effect the accuracy of slope predictions by SSM (Section 5.7.1).

Predictions are best at Tumbarumba, this is expected to be due to the greater sample size included in the LUT for this site. Whereas the poorer performance at Norunda and SOJP is believed to be attributed to the majority of footprints from laser campaign L1A, which exhibits the greatest eccentricity, possibly inducing a greater directional bias in measurements. Such measurements somewhat explain the higher variability present in the LUT relationship.

Considering results from all study sites, a root mean squared error (RMSE) of 1.89° was found, which in general improves equivalent RMSE values for GLAS derived vegetation height by Lefsky et al. (2005); Rosette et al. (2008); Duncanson et al. (2010); Lefsky (2010); Simard et al. (2011) and Los et al. (2012).

5.7.3 ISM - Slope Predictions

Similar to SSM predictions, ISM predictions were made for FLIGHT and GLAS waveforms. The same data (waveforms) as in Section 5.7.2, for both FLIGHT and GLAS, were predicted by ISM and compared against model inputs and all available airborne LiDAR equivalent results for all study sites respectively.

Slope predictions were made using the direct retrieval method outlined in Section 5.2, negating the need for LUT data as is required for SSM predictions. Predictions are based on the width of waveform ground components at an intensity threshold of $T_G = 0.001$ V, rather than the FWHM measure employed for SSM predictions.

Model success is evaluated by model performance parameters, as in the previous section; results from which are shown in Table 5.3 for FLIGHT and GLAS predictions. Results for FLIGHT and GLAS are encouraging, where inter-site variance is present as noted previously.

ISM Predictions with FLIGHT

As done with SSM, ISM predictions for FLIGHT were compared to model inputs, a direct comparison and difference plot are illustrated in Figure 5.7a and 5.7b respectively.

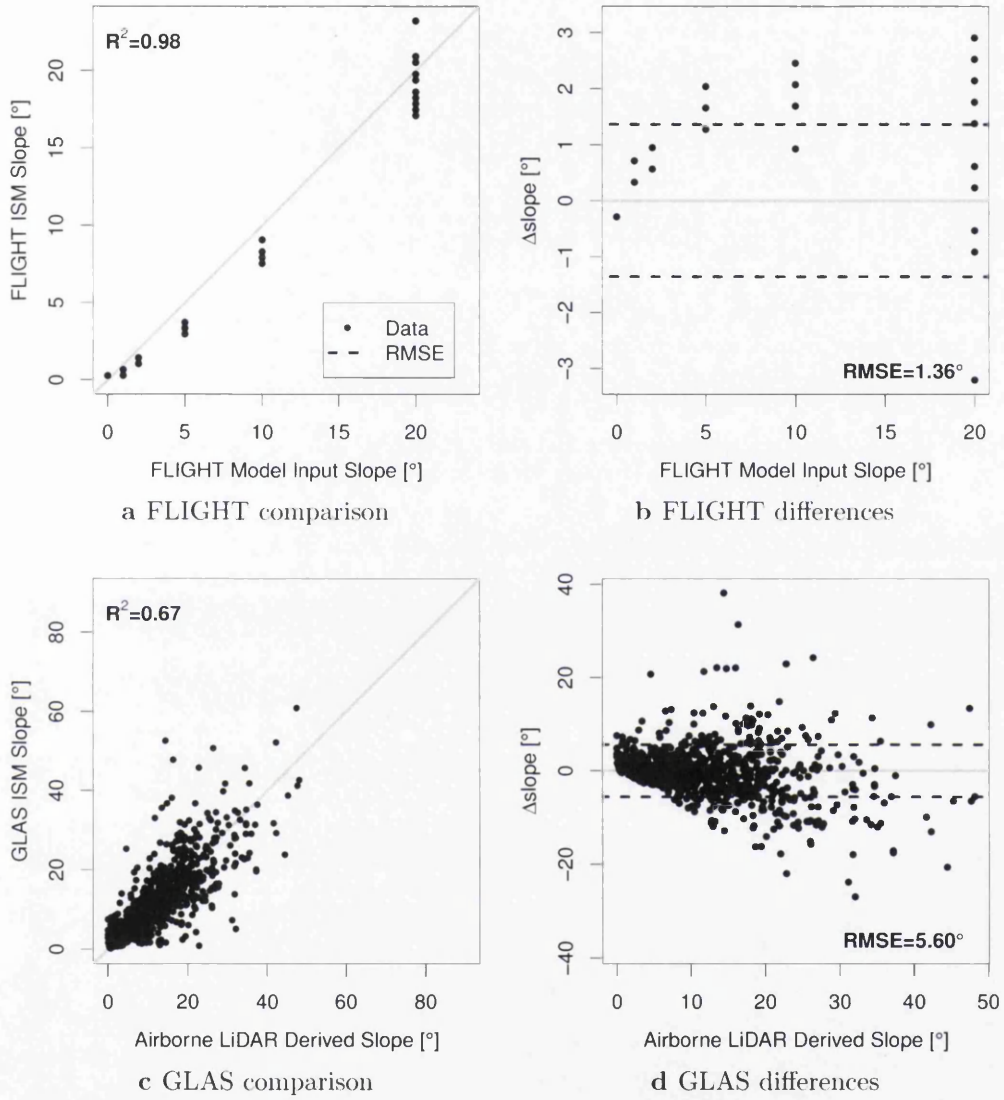
Table 5.3: Summary of ISM performance parameters for individual sites, SOA is not listed as slope predictions were unassigned due to poor quality Gaussian fitting (Section 5.2). Note: n = sample size, R = Pearson correlation coefficient, p_R = fractional probability of no correlation (significance), D = Kolmogorov-Smirnov statistic, F_2 = fraction of predictions within a factor of 2 of observed values, F_B = Fractional bias, ε_{nme} = normalised mean error, and RMSE = root mean square error.

Site	n	R	p_R	D	F_2	F_B	ε_{nme}	RMSE[°]
Tumbarumba	593	0.77	$< 10^{-10}$	0.07	0.86	-0.04	0.28	6.05
Loobos	68	0.54	$< 10^{-10}$	0.18	0.46	0.26	0.72	2.92
Norunda	39	0.42	< 0.05	0.26	0.38	0.41	0.81	1.91
Tharandt	125	0.39	$< 10^{-10}$	0.18	0.49	-0.18	0.62	6.41
SOBS	44	0.63	$< 10^{-10}$	0.25	0.52	0.32	0.85	2.88
SOJP	21	0.76	$< 10^{-10}$	0.62	0.33	0.70	1.07	1.39
FLIGHT	186	0.99	$< 10^{-10}$	0.17	0.82	-0.15	0.17	1.36

Resulting model performance parameters (Table 5.3) suggest that the ISM does not perform as well as SSM equivalent results for these simulated waveforms. However, the results produced by equivalent SSM results are biased (see Section 5.1.2), hence results are not typical of those expected from independent predictions.
Model performance parameters suggest that ISM predictions perform to a high standard, although some variability in results is present. Variability increases as a function of slope (as was noted previously), however, second order vegetation effects appear to be more prominent in this method, where slopes as little as 5° exhibit notable variability. As previously, this is attributed to the combination of variable vegetation structure parameters (LAI and F_{cov}) and slope, which induces complex scattering events causing affected waveforms to be more difficult to interpret.

ISM Predictions with GLAS

GLAS predictions from ISM were compared with airborne LiDAR equivalent slope information as before, where Figure 5.7c and 5.7d illustrate the comparison and differences between ISM predictions and airborne LiDAR derived slope respectively. Similarly to ISM results from FLIGHT waveforms, GLAS ISM predictions are less consistent than corresponding SSM GLAS results, however, these results



a FLIGHT comparison

b FLIGHT differences

c GLAS comparison

d GLAS differences

Figure 5.7: (a, c) Comparison and (b, d) differences, with RMSE (dashed lines), between known and ISM predicted slope values for FLIGHT simulated and GLAS waveforms respectively.

are inherently biased (Section 5.7.2).

Table 5.3 summarises model performance parameters obtained for ISM results, indicating that the ISM is performing to a high standard, producing a combined site RMSE = 5.16°. This is equivalent to typical RMSE values for vegetation height metrics derived from GLAS data (Lefsky et al., 2005; Rosette et al., 2008; Lefsky, 2010; Duncanson et al., 2010; Simard et al., 2011; Los et al., 2012). Similarly to SSM, it is noted that Norunda and SOJP exhibit more marginal performance with ISM. This is suspected to originate from possible directional bias that may exist for the L1A campaign, coverage by which is most frequent at these sites. To clarify, directional bias originates from the footprint semi-major axis being much greater than the semi-minor axis, hence the calculation of the mean footprint diameter ($\bar{X}_{fp}$ in Equation 3.5) is biased towards to the semi-major axis. This axis has a direction associated with it, and hence means that within footprint slope (θ) is biased in this direction.

5.8 SSM and ISM Independent Testing

Due to the potential for obtaining slope predictions by each retrieval method (SSM and ISM), independent validation of each method was investigated, focussing on removing bias effects (from site specific tuning). Each method's predictive capabilities were tested for an independent waveform dataset, where slope predictions were validated against an external source. Such a test is required as all airborne

Table 5.4: Model performance assessment of results obtained from SSM and ISM when compared against OS DEM slope data. Note: n = sample size, R = Pearson correlation coefficient, p_R = fractional probability of no correlation (significance), D = Kolmogorov-Smirnov statistic, F_2 = fraction of predictions within a factor of 2 of observed values, F_B = Fractional bias, ε_{nme} = normalised mean error, and RMSE = root mean square error.

Model	n	R	p_R	D	F_2	F_B	ε_{nme}	RMSE[°]
SSM vs OS DEM	7081	0.51	$< 10^{-10}$	0.10	0.42	-0.04	0.62	7.69
ISM vs OS DEM	7463	0.74	$< 10^{-10}$	0.14	0.59	-0.20	0.43	6.27

LiDAR data (which would be preferable for methodological testing/validation) were employed in the set up of one slope retrieval method (SSM). Hence, in order to independently test both developed slope retrieval methods (SSM and ISM) an independent data source and test site are required to determine which method performs best. British Ordnance Survey (OS) 10 m resolution DEM data (see Section 3.6) were the best (with regards to vertical resolution) available data for testing purposes at the time of this study. OS DEM data were acquired for the principality of Wales, in the United Kingdom, as a means of validating slope predictions. OS DEM slope was calculated via a modified version of Equation 3.5, according to Section 3.10.2.

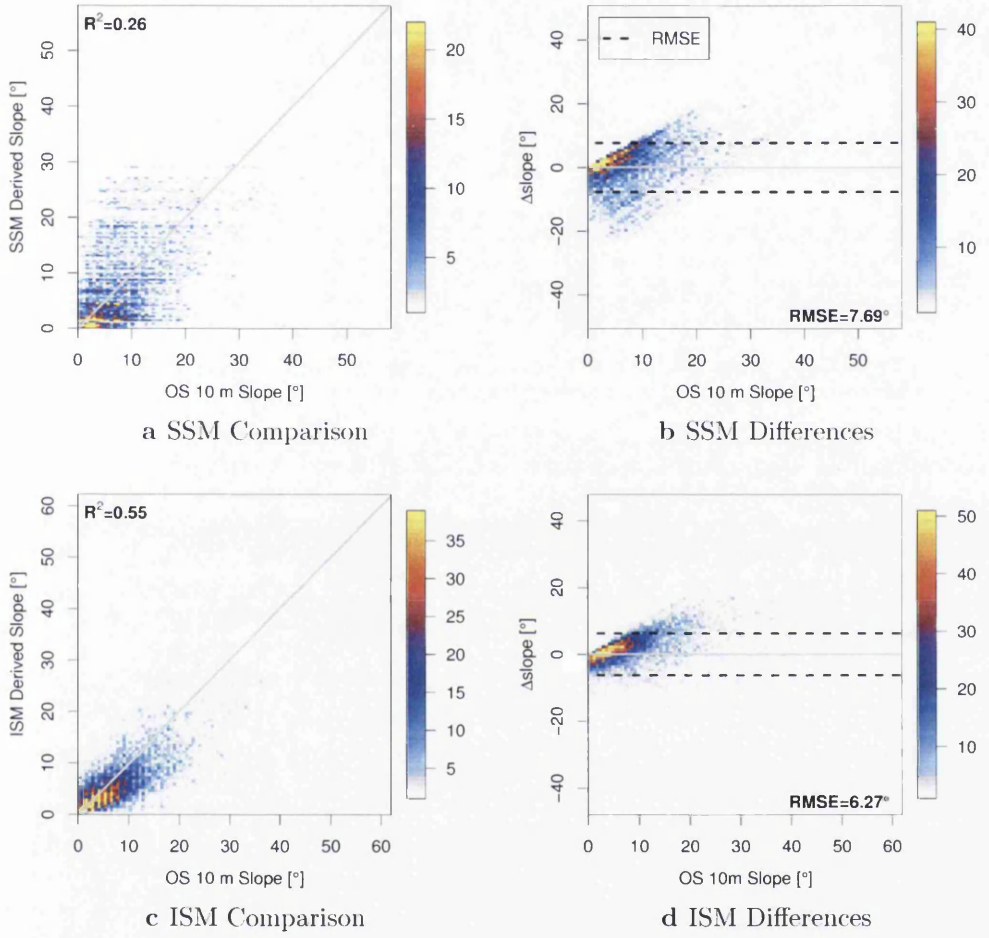
The comparison of SSM predictions and OS derived slopes, and associated differences, are illustrated in Figure 5.8a and 5.8b respectively; due to the number of points, colour density plots were chosen for easier interpretation. Model performance parameters for SSM predictions are given in Table 5.4, where it is noted that SSM performance for an independent waveform dataset is not equivalent to data analysed at the site level.

For ISM, colour density plot comparison and associated difference plots are given in Figures 5.8c and 5.8d respectively, and associated model performance parameters are given in Table 5.4. For ISM model performance is more consistent with that found at the site level, where RMSE increased by 0.67° only.

A comparison of SSM and ISM results show that SSM is limited by its LUT data, from which slope predictions are inferred. This limitation is indicated in Figure 5.8a where no predictions $\geq 30°$ are present. This is a legacy of the LUT data employed here for predictive purposes. ISM appears to have no limitations to this end, capable of inferring slope beyond 30°.

The majority of variability present in both slope comparisons is attributed to environmental factors such as signal attenuation, within footprint topographic variability, and, in some cases, vegetation canopies with multiple layers. These are known to increase the probability of signal mixing in waveforms, which are not accounted for under model assumptions (Section 5.3).

Further complicating data comparison are the noted uncertainties in British OS DEM data (see Section 3.6). In these data interpolation effects, and mountain or moorland classified areas are known to have greater uncertainty.



a SSM Comparison

b SSM Differences

c ISM Comparison

d ISM Differences

Figure 5.8: Colour density plots for (a, c) comparison of OS 10 m DEM derived slope against SSM and ISM slope predictions respectively. (b, d) Differences between OS 10 m DEM derived slope and SSM, and ISM predictions respectively.

From this analysis it was concluded that the ISM slope retrieval method is most generally applicable. As a result, further analysis only considers ISM predictions and disregards inferences made by SSM.

5.9 Uncertainty Analysis

Systematic uncertainties associated with ISM slope predictions can be attributed to two major components used in the calculation of slope itself. These are the within footprint elevation difference (E) and its associated uncertainty ΔE, and the mean radius of the footprint $\bar{s}$ and its associated uncertainty Δs. These components are further attributed to systematic or random uncertainty from respective measurements.

The uncertainty ΔE originates from two unique positions on the model alternate fit waveform, which are affected by noise (Δn) sourced from the raw waveform. This random uncertainty is represented by the standard deviation of the back ground noise (GLA05 product i_sDevNsOb1) of individual waveforms. ΔE is also influenced by the system vertical resolution (ΔR); for GLAS this is constant for all waveforms at 0.15 m. Hence the induced uncertainty in range measurements are given as $\Delta R \simeq 0.15/2$ m.

Δs is influenced by the definition of $\bar{s}$, defined as the mean of the footprint radial components, semi-major (S_{major}) and semi-minor (S_{minor}) axes. Hence the uncertainty in $\bar{s}$ is dependent on the uncertainty in the associated uncertainties of S_{major} and S_{minor}.

For GLAS, only S_{major} information is available, however, as elliptical eccentricity (e) is also given, this allows the calculation of S_{minor} by means of Equation 5.6.

$$S_{minor} = S_{major}\sqrt{1 - e^2} \tag{5.6}$$

Associated measurement uncertainties for these stated parameters (ΔS_{major} and Δe) are supplied by the NSIDC (2012a), thus allowing the derivation of ΔS_{minor}.

Propagating these uncertainties through the calculations that allow the derivation of slope predictions will yield the final uncertainty with respect to individual slope predictions ($\Delta\theta$). An expression for $\Delta\theta$ is given by Equation 5.7, the derivation of which is given in Appendix 9.11.

$$\Delta\theta = \sqrt{\Delta E^2 \left(\frac{\bar{s}^2}{E^4 + \bar{s}^4}\right) + \Delta\bar{s}^2 \left(\frac{E^2}{E^4 + \bar{s}^4}\right)} \tag{5.7}$$

5.9.1 Instrument and Model Uncertainty

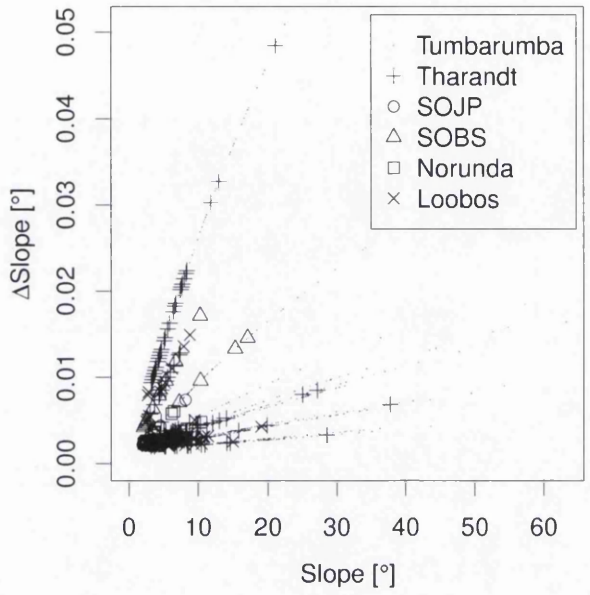
For ISM, analysis of predicted slope uncertainty (Equation 5.7) was investigated on a study site (Figure 5.9a) and a GLAS laser campaign basis (Figure 5.9b; Section 3.4).

Figure 5.9a in conjunction with 5.9b suggests that the uncertainty in slope predictions increases with slope itself. The rate of increase in slope uncertainty is dominated by laser campaign, rather than site geolocation. Hence Figure 5.9a suggests that GLAS stated positional uncertainties from NSIDC (2012a) are stable over time; however this investigation does not offer conclusive evidence to this end.

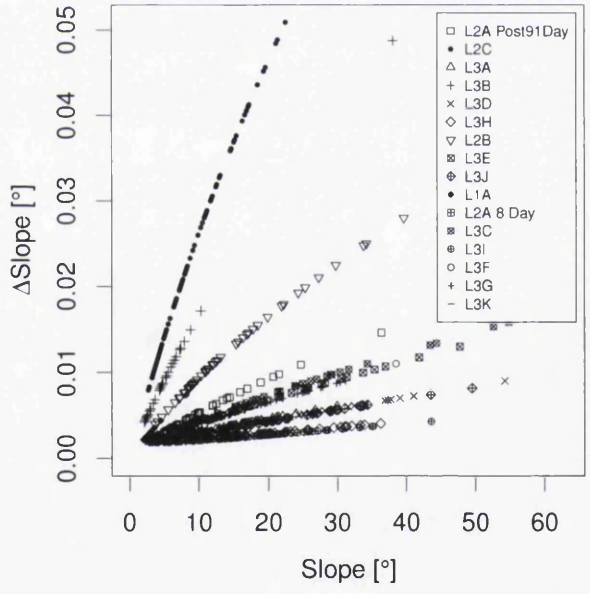
Figure 5.9b indicates that laser campaigns L1A, L2B, L2C, and L3B suffer with the greatest uncertainties of all the sixteen laser campaigns available for testing from all study sites. These four relatively early GLAS campaigns each suffer some subtleties pertaining to laser energy abnormalities, footprint positional and dimensional accuracy, and/or footprint extreme dimensions (Abshire et al., 2005; NSIDC, 2012a).

Energy abnormalities were identified for all four of these campaigns (see Table 5.5). In particular, it was noted that L1A, L2B and L3B all suffered an approximate 30% loss of laser energy over approximately one month long periods (Abshire et al., 2005). L2C suffered the greatest loss in laser energy of approximately 75% over the period of its operational lifetime (Abshire et al., 2005). These laser campaigns also suffer the greatest uncertainties with regards to positional and dimensional (semi-major and semi-minor axes) accuracy. The lowest positional and dimensional accuracies are associated with L2C, and generally improve subsequently with L3B, L2B and L1A respectively (see Table 5.5).

Campaign L1A is reported as the most elliptical of all GLAS footprints, with semi-major and semi-minor axes reported as 148.6 m and 58.2 m respectively. Based on how ISM slope predictions are inferred, assuming a circular footprint when in reality the footprint is heavily elliptical may induce a directional bias effect. This may lead to slope over estimation, as the assumed footprint diameter (from which slope is inferred; Equation 3.5) may be considerably smaller than the true distance over which slope should be inferred.



a Study Site



b Campaign

Figure 5.9: Behaviour of ISM slope uncertainty as a function of ISM slope itself on (a) a site by site basis, and (b) a laser campaign basis.

The above subtleties can induce uncertainty in ISM inferred results, where it is suggested that direct and indirect energy loss (via positional and dimensional uncertainty) are the key drivers of such effects. The magnitude of these effects is

Table 5.5: Summary of uncertainties associated with specific GLAS laser campaigns. Note: ΔE_L = laser energy loss during campaign period, ΔS_M = semi-major axis accuracy, Δ_V = vertical pointing knowledge accuracy, and Δ_H = horizontal pointing knowledge accuracy. The context of this data is shown by Table 3.5, where similar information for all GLAS campaigns is shown.

Campaign	ΔE_L[%]	ΔS_M[m]	Δ_V[m]	Δ_H[m]
L1A	30	148.6 ± 9.8	0.60 ± 3.30	0.79 ± 4.77
L2B	33	89.8 ± 5.1	0.70 ± 2.60	0.93 ± 3.73
L2C	75	88.4 ± 19.1	0.30 ± 7.20	0.37 ± 10.30
L3B	30	79.3 ± 11.6	0.10 ± 2.93	0.07 ± 4.20
source: Abshire et al. (2005), NSIDC (2012a)				

related to the rate at which the slope uncertainty increases, as reflected in Figure 5.9b, where the four highest rates of increasing slope uncertainty belong to L2C, L3B, L2B, and L1A respectively, as noted in Table 5.5. Context of these uncertainty increase rates are given in Table 3.5, with respect to associated uncertainties of semi-major axis accuracy (ΔS_M), vertical pointing knowledge accuracy (Δ_V), and horizontal pointing knowledge accuracy (Δ_H) for all GLAS laser campaigns.

5.9.2 Environmental Variability

Uncertainty in laser footprint positions and dimensions has the ability to effect the accuracy of derived waveform products, as the footprint area is not necessarily illuminating the location that is expected. This will induce further uncertainty, however these uncertainties are related to environmental variability, particularly with regards to the atmosphere, ground surface, and vegetation structure.
Atmospheric attenuation of the emitted signal, by definition, causes intensity reductions in measured returns by means of absorption and scattering events. This effect is dependent on the local atmospheric constituent components, such as cloud, and aerosol presence. Exact atmospheric effects are not quantifiable by GLAS and can be attributed to a source of random uncertainty.

The large footprint nature of GLAS dictates that the illuminated footprint area at the Earth's surface is subject to large within footprint variability. With larger footprints comes an increased surface area sample, and thus an increased probability of measuring multiple land cover types originating from multiple elevations. Across large footprints this effect can cause signal mixing, thus making returned waveforms difficult to interpret. This interpretation process affects the accuracy at which key parameters (waveform ground components) can be identified, which consequentially affect the accuracy of ISM (and SSM) slope predictions.
Similarly the large footprint nature of GLAS can induce variability in slope predictions via vegetation structure. This effect originates from there being a range of vegetation heights within any (vegetated) footprint. This induces, to a lesser degree, some signal mixing, however this is generally only applicable for multiple layer canopies. Particularly for understorey presence, vegetation returns can occur over a range between the top of the understorey layer and the ground, thus making the understorey and ground components difficult to isolate. This vegetation variability can thus induce further uncertainty in slope predictions.

5.10 Summary

In summary this chapter highlights the potential of employing waveform LiDAR as a novel way to derive slope information. Two methods were developed, SSM and ISM, where ISM was found more generally applicable due to its direct retrieval approach, negating the use of look up tables (LUTs), on which SSM relies.
Results reflect the favoured slope derivation technique, with consistently good correlation and model performance parameters (Section 5.7). These results further justify the choice of the ISM for deriving slope information in Chapter 6, where it is used to correct vegetation height estimates from GLAS waveforms.
ISM predictions were analysed for systematic uncertainty, where it was noted that instrument and estimation errors are low. This suggests that major discrepancies originate from environmental variability factors such as atmospheric attenuation and, more likely, within footprint topographic and (vegetation) structural variability. The latter two items are of great interest, as ISM (and SSM) measurements are assumed to be of planar/linear slope between two points in a particular footprint,

whereas in reality 3-dimensional variability is present.

It was concluded that slope derivations from GLAS waveforms are possible, and offer great potential in mapping the Earth's surface, with particular application to filtering, and/or correction processes.

Chapter 6

Refined Vegetation Height Estimates

Sloped terrain is known to induce spurious data from waveform LiDAR returns, hence the removal of footprints recorded over slopes is often executed, resulting in more accurate final findings. However this process can diminish the use of such data. Here methods are tested such to avoid the removal of vast quantities of data, whilst preserving accuracy.

Based on findings from Chapters 4 and 5, refinements to regional to global vegetation height estimates are possible, especially over sloped terrain. Focussing particularly on vegetation height correction for slope, the existing correction method of Lee et al. (2011) (in the following denoted: Lee) is evaluated, and its limitations are highlighted. Furthermore, a new method based on Lee (in the following denoted: modified) is developed and assessed for its feasibility to correct vegetation height estimates for sloped terrain.

These two height correction methods were applied to the global vegetation height product of Los et al., 2012 (in the following denoted: control). As a first step, each height distribution is compared to airborne LiDAR equivalent results for each study site. As a second step, the vegetation height estimates are compared with NDVI data at the near global scale.

6.1 Methods

6.1.1 Filters

GLAS vegetation heights were derived using the method of Rosette et al. (2008) (see Section 3.9.2) for each footprint within each of the seven study sites. Typically, spurious measurements are inherent within GLAS measurements due to a combination of instrument and environmental subtleties. These accumulate physically as: atmospherically attenuated signals, weak emitted and/or reflected laser energies and difficult to interpret waveforms, originating from extreme surface environments (such as very dense canopies).

Filtering techniques were developed in order to eliminate spurious data; a set of stringent parameters have been developed and described by Los et al. (2012). A similar filtering approach was adopted for this study, with one modification, namely the filtering process for high slope was not applied (which is typical when employing waveforms over vast areas of sloped terrain; Simard et al., 2011; Los et al., 2012). Slope afflicted data were retained as the correction of vegetation heights for slope is one of the objectives of this chapter.

The filters developed by Los et al. (2012) were based on a $5° \times 5°$ tile of GLAS data from the Algerian desert, between $20° - 25°$ N and $0° - 5°$ E. This subset of approximately 51000 footprints was collected over 41 days between February and March 2003, during the Laser 1A campaign period (Los et al., 2012). This site was chosen as vegetation presence in the desert is expected to be very low, hence high estimates of vegetation height (V_H) were hypothesised to indicate problems within the GLAS data.

For this study, further filters are applied employing the GLAS GLA14 product parameters (Table 3.8), particularly elevation, and waveform decomposition (of the GLAS signal) parameters. A summary of applied filters is provided below.

Elevation

The elevation filter applied here is identical to that described in Section 5.4.3. The GLAS reference elevation (i_elev) was adjusted to match the SRTM ellipsoid and then compared to the SRTM 4.1 DEM data (i_DEM_elv). Absolute differences

between corresponding data points exceeding 8 m, corresponding to 95% confidence interval, were excluded. Figure 6.1a shows the results of this filter when applied to the GLAS data employed in this study.

Area Under First Gaussian

Los et al. (2012) noted that for low values of the area under Gaussian 1 the spread in vegetation height estimates was large, where the higher values are likely to be unrealistic. The likely cause was suggested as being that low values of the area under Gaussian 1 indicates weak signal strengths possibly caused by atmospheric attenuation, low emitted energy, or a weak return signal due to dense canopy or complex terrain interactions causing multiple scattering events (Los et al., 2012). A threshold was applied to eliminate low values of the area under Gaussian 1, aimed at removing those returns that suffer with atmospheric or instrumental issues, but accounting for tall, dense vegetation interactions (Los et al., 2012). As a compromise of this, a threshold value of 1 V ns was selected, Figure 6.1b shows the elimination of those footprints deemed to potentially cause spurious results by this mechanism.

Amplitude of First Gaussian

A low amplitude of the first Gaussian indicates a data quality issue similar to the small area under the first Gaussian (Los et al., 2012). The ability to distinguish the "true" signal start and end from background noise is diminished due to signal strength issues (Rosette et al., 2013). In order to further fortify against such problems, a 0.05 V minimum amplitude threshold was implemented as a criteria for the first Gaussian. Results are shown in Figure 6.1c.

It was also noted by Los et al. (2012) that a number of outliers were present over the entire range of first Gaussian amplitudes with regards to vegetation height. A second filter was applied, removing the highest 0.1% of vegetation height estimates per 0.1 V amplitude interval; these are also shown in Figure 6.1c.

Sigma of All Gaussians

It was noted that Gaussians (fitted to the raw LiDAR waveform) with a large spread in elevation range (5%–95% values $\geq$ 80 m) are unlikely to originate from vegetation, which only in exceptional circumstances reach these heights (Los et al., 2012). A test was applied to all Gaussian fits in order to remove those waveforms that exhibit the highest 0.1% of sigma values. The thresholds for this test were based on the frequency distributions of the original, unfiltered data. The results of this application are shown in Figure 6.1d.

Neighbour Test

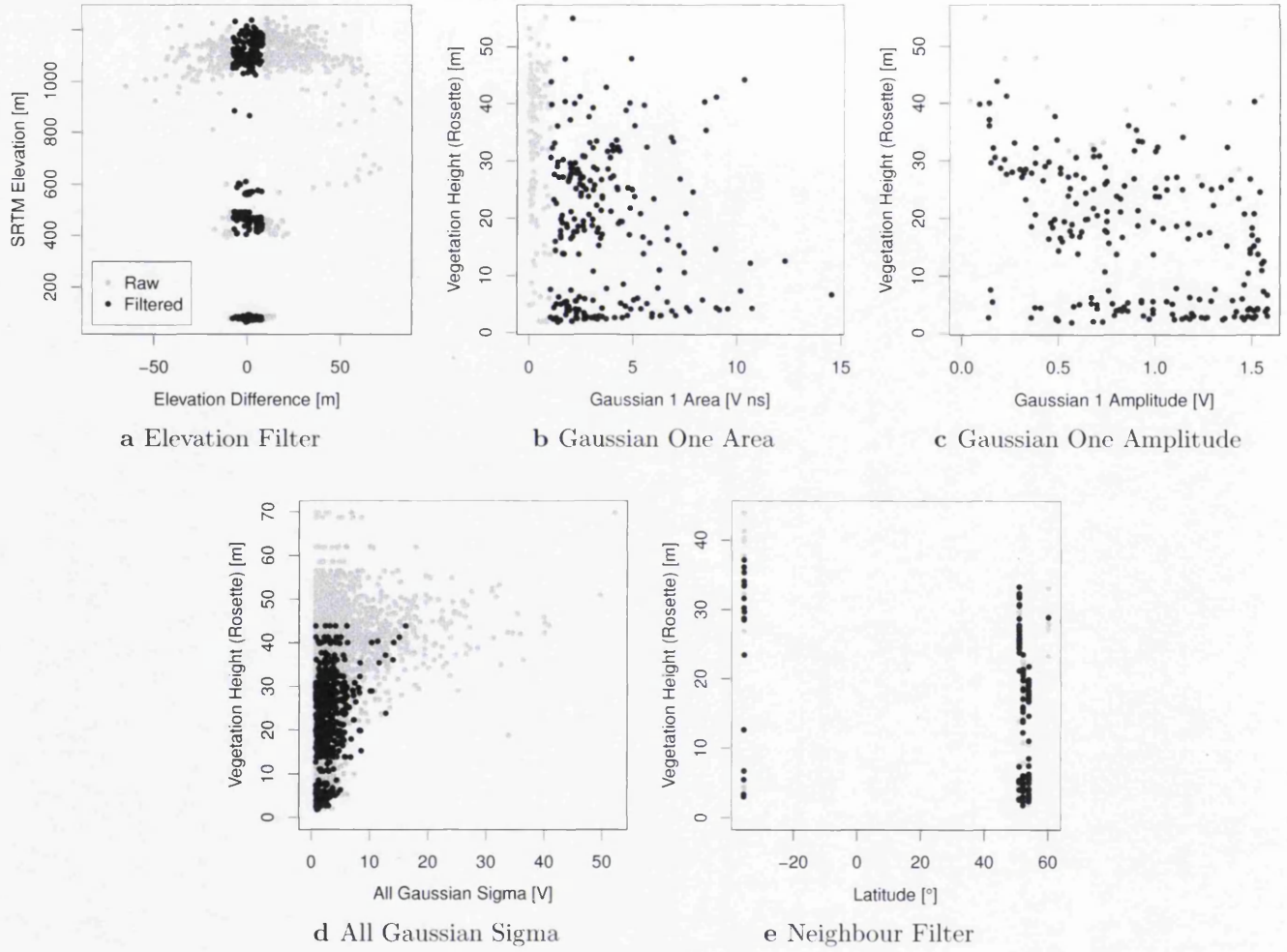
The final filter test adopted from Los et al. (2012) was the neighbour test. This test was developed as a precaution, removing the neighbouring along track footprints of any footprints eliminated by the previous filtering techniques. The results of this final test are shown in Figure 6.1e.

GLAS Quality Flags

In addition to the above filters adopted from Los et al. (2012), the applicable data quality flag filters, supplied in GLAS ancillary data products, were applied to the remaining dataset; of particular interest was the cloud presence filter (i_FRir_qaFlag). Details regarding this quality flag can be found in NSIDC (2012b). For the purposes of this study, data with flag values other than 15 (no cloud presence) were eliminated from further analysis. No figure is present illustrating the magnitude of data filtering by this mechanism as no data required any further elimination (in this case) after the application of the filters of Los et al. (2012).

6.1.2 Height Model Refinement

Vegetation height estimates in this study are based on the method developed by Rosette et al. (2008), where the difference between the signal start (i_SigBegOff) and the centroid range (i_gpCntRngOff) of either the first or second Gaussian is considered.



a Elevation Filter **b** Gaussian One Area **c** Gaussian One Amplitude

d All Gaussian Sigma **e** Neighbour Filter

Figure 6.1: Selection of filters applied to waveforms recorded over study sites, Grey dots represent the raw data, whereas black dots represent those data that remain post filtering. Filters were originally developed by Los et al. (2012), and are applied in the same order for consistency. Note: vegetation heights were calculated by the method of Rosette et al. (2008).

GLAS waveforms always exhibit a finite measurable width (or elevation difference; Section 5.4.6) even for completely flat terrain. Similarly to the refinement applied in Section 5.4.6, a model refinement is applicable to vegetation height estimates also. Los et al. (2012) noted that the area under the first Gaussian (units: V ns) could be used as an indicator of the magnitude of this mandatory finite width effect. It was assumed that the 5% values of the vegetation height distributions (after slope and elevation filtering) per 0.1 V ns intervals could account for the effect of this finite elevation difference. Corrections of the above were based on a regression model of the form given by Equation 6.1.

$$V_{H_{0.05}} = aA + b \tag{6.1}$$

where:

$V_{H_{0.05}}$ = Minimum measurable vegetation height [m]
A = Area under Gaussian one [V ns]
b = 1.91 [m V^{-1} ns^{-1}]
a = 0.11 [m]

In the current study, the derived vegetation height refinement model (values of a and b above) of Los et al. (2012), based on 1400 5% values, was applied (Equation 6.1), as after initial elevation filtering of the available data, the amount of data retained (< 200 points) would not be sufficient to form this described refinement reliably. The value of $V_{H_{0.05}}$ (Equation 6.1) was subtracted from all GLAS vegetation height estimates.

6.1.3 Slope Corrections

Lee et al. (2011) developed a correction to adjust waveform measured vegetation height (V_H; Equation 6.2) for slope (θ) such to obtain true vegetation height (V_{Ht}). The apparent overestimate in V_H (denoted by V_{Hc}), caused by slope for a mean footprint diameter of L, is given by Equation 6.3. In this method it is assumed that V_H is a measurement of vegetation located at the within footprint mean ground elevation (equivalent to elevation at the horizontal centre of the footprint, hence

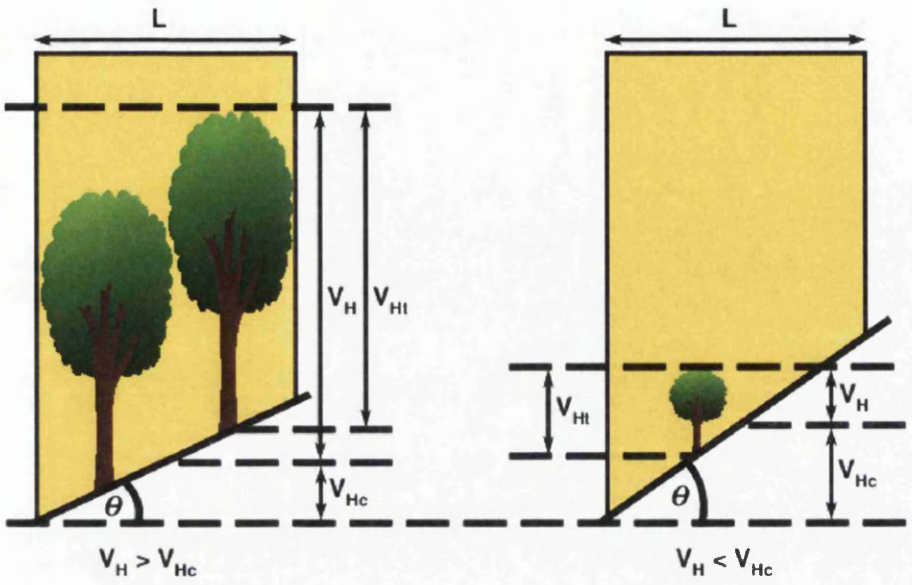


Figure 6.2: Example of the vegetation height correction method of Lee et al. (2011) successfully applied (left), and unsuccessfully applied, causing a negative value of V_{Ht} (right).

introducing a factor of 1/2).

$$V_H = V_{Ht} + V_{Hc} \tag{6.2}$$

$$V_{Hc} = \frac{L}{2} \cdot \tan\theta \tag{6.3}$$

The current investigation tests the feasibility of the outlined height estimate correction technique (Equation 6.3). A potential limitation of this method is that it's not generally applicable when within footprint vegetation location varies. This method assumes that within footprint vegetation is spread homogeneously from the footprint centre up to (aligning perfectly with) the footprint perimeter. This is rarely the case, where vegetation typically does not correspond with the footprint perimeter. This assumption requires that this method finds the signal end and derived ground elevation to be identical.

Due to this subtlety, it is possible that V_{Hc} is greater than the originally inferred

vegetation height estimate (V_H); thus when correcting V_H by V_{Hc}, the corrected vegetation height (V_{Ht}) can become negative. This spurious result is hypothesised to hold for all footprints exhibiting linear slope, with observed vegetation located at an equivalent elevation less than at the footprint horizontal centre (Figure 6.2). Negative values of vegetation height are unrealistic, and are thought to be an artefact of the Lee method assumptions, which do not always hold true for large footprint diameters. Additionally, this effect is believed to become more frequent with steeper slopes.

For this method to consistently work, the exact fraction of the footprint diameter at which height estimates are measured is required. Hence Equation 6.3 is modified to Equation 6.4, where α (derived in the following section) is a fractional percentage (between 0 and 1) of where the measured vegetation falls within the footprint diameter, L.

$$V_{Hc} = \alpha L \cdot \tan\theta \tag{6.4}$$

Figure 6.3 visually represents the importance of α in generalising the Lee vegetation height correction method. In Figure 6.3 V_{Hc} has been modified to account for α, and W_{Ext} illustrates the vertical representation of the waveform extent as defined by Lefsky et al. (2007).

Modified Method

The method of Lee et al. (2011) was modified in this study, with the view of accounting for the within footprint horizontal location of the elevation at which vegetation was observed (E_v). This adapted method (known here on as the "modified" method) also dictates that V_{Hc} accounts for E_v not being located at the footprint perimeter. This modified method is described in conjunction with Figure 6.4, which allows visual definitions of applicable parameters.

Theoretically, between the signal start and signal end for a waveform recording of a vegetated surface, the vegetation components and the entirety of the ground (most and least elevated points) are recorded (Figure 6.4). Using this logic, and the assumption that the Lee method finds the signal end and derived ground elevation to be identical (mentioned previously), leads to over-corrections of V_H as a function of slope (Equation 6.3). This is due to the magnitude of Equation 6.4

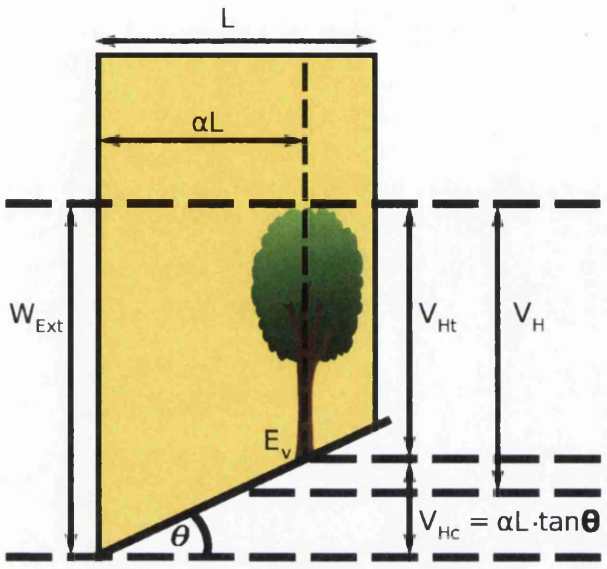


Figure 6.3: Visual example of the modified vegetation height correction method of Lee et al. (2011), accounting for the within footprint elevation of the measured vegetation (E_v). This within footprint horizontal position of this elevation affects the severity of the vegetation height correction (V_{Hc}).

being proportional to the horizontal distance at which E_v is located (αL), with respect to the least elevated point of the ground.

The modified method, defined in Equation 6.5, accounts for E_v through its horizontal location within any given footprint, where the location of E_v is assumed to be obtainable from waveform information (Figure 6.4). This is possible by findings in Chapter 5 i.e. the fact that the ground elevation range is recorded in any footprint. This allows the assumption that the difference between the signal end and the derived ground ($G_{A_{1,2}}$) represents the vertical displacement of E_v from the lowest ground point (signal end; Figure 6.4).

$$V_{H_c} = \beta L \cdot \tan\theta = (1-\alpha)\, L \cdot \tan\theta \tag{6.5}$$

Equation 6.5 requires the proportion of the footprint radius (α) to represent the equivalent horizontal distance of the elevation displacement between E_v and the

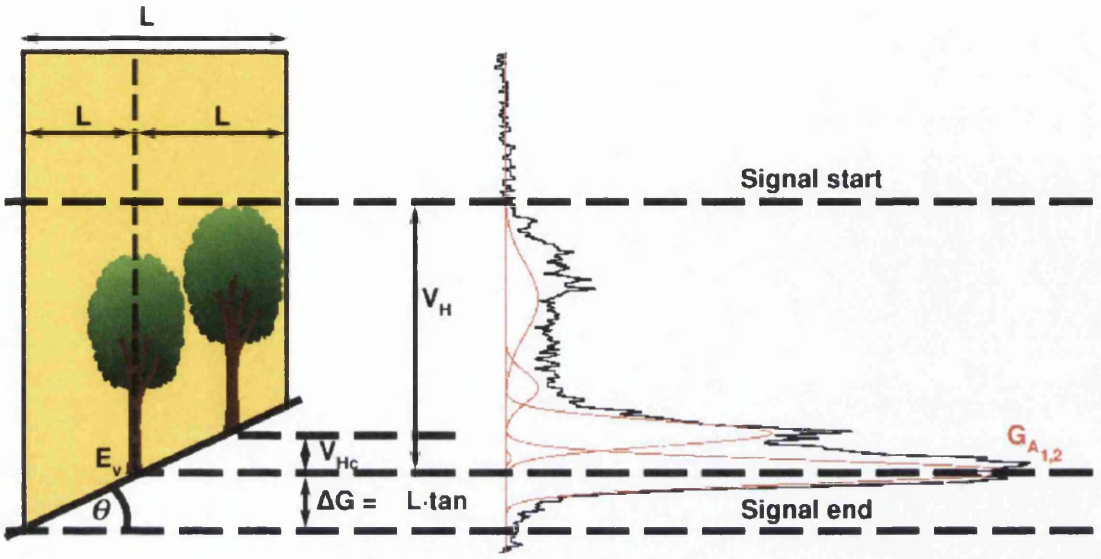


Figure 6.4: Visual example of the hypothesised method in conjunction with a mock LiDAR waveform, fitted with six Gaussians. The lowest of the two Gaussian fits with the greatest amplitude ($G_{A_{1,2}}$) is assumed to correspond to E_v. This defines the new footprint size denoted by βL, from which the modified version of V_{H_c} is calculated (Equation 6.5).

waveform signal end, defined as ΔG ($= \alpha L \cdot \tan\theta$). This fraction of L is obtainable through Equation 6.6, where ΔG, expressed as waveform quantities, is the vertical displacement between $G_{A_{1,2}}$ (i_gpCntRngOff) and the signal end (i_SigEndOff).

$$\alpha = \frac{\Delta G}{L \cdot \tan\theta} \tag{6.6}$$

This definition of α allows Equation 6.5 to be expressed as a function of waveform parameters (through ΔG), the footprint diameter (L), and slope angle (θ):

$$V_{H_c} = L \cdot \tan\theta - \Delta G \tag{6.7}$$

Correcting vegetation height estimates using the modified method will account for the measured vegetation not aligning with the footprint radial perimeter, moreover the footprint diameter is essentially modified by a factor of $\beta = 1 - \alpha$. This removes the over-correction noted in the Lee method.

Interestingly, when $\theta \rightarrow 0$, $V_{Hc} \rightarrow 0$ also (because $\Delta G = 0$ when $\tan(0) = 0$), dictating that height estimates over steeper slopes require larger correction, and conversely shallow slopes requires less correction. When $\theta = 0$, no correction is required, and estimated vegetation height reverts to that defined by the method of Rosette et al., 2008 (Section 3.9.2).

6.2 Results and Discussion

GLAS waveforms were filtered according to the methods described in Section 6.1.1 (Figure 6.1), to provide an accurate distribution of vegetation height estimates (control dataset), obtained by the method of Rosette et al., 2008 (see Section 3.9.2). The control dataset was corrected by both the Lee (Equation 6.3), and modified (Equation 6.7) height correction techniques separately, and investigated for general applicability.

Comparisons of vegetation height estimates of the control, Lee, and modified datasets were made against airborne LiDAR vegetation height information for each study site (see Section 3.1) as a proof of concept. Additionally, further correction application was made to all footprints from GLAS campaign L1A.

6.2.1 Study Site Application

At the site level the control data were compared with (spatially equivalent by footprint; Section 3.8.2) airborne LiDAR vegetation heights. The comparison, and differences between the control and airborne LiDAR equivalent data are visualised in Figure 6.5. Here a general under-estimation of V_H is visible by the method of Rosette et al. (2008). Although results appear non-linear in part, a 1:1 line is illustrated in Figure 6.5a such to indicate where theoretically *perfect* results would be expected. As both airborne and GLAS data are measuring identical parameters from the same region a linear relationship is implied. The same reasoning applies for Figures 6.6a and 6.6c.

Correction Application

The Lee (Equation 6.3), and modified (Equation 6.7) methods were applied individually to control vegetation height data (Figure 6.5). The comparison and corresponding differences for the Lee, and modified methods are shown in Figures 6.6a and 6.6b, and Figures 6.6c and 6.6d, respectively. Additionally, Table 6.1 shows associated model performance parameters (see Section 5.5 for explanations) for the control, Lee, and modified height datasets. Where the Lee correction has been applied increased magnitude under-estimations of V_H are induced. However, for the modified correction already existing underestimates are accounted for, supported by Table 6.1 where the fractional bias $F_B = 0.00$.

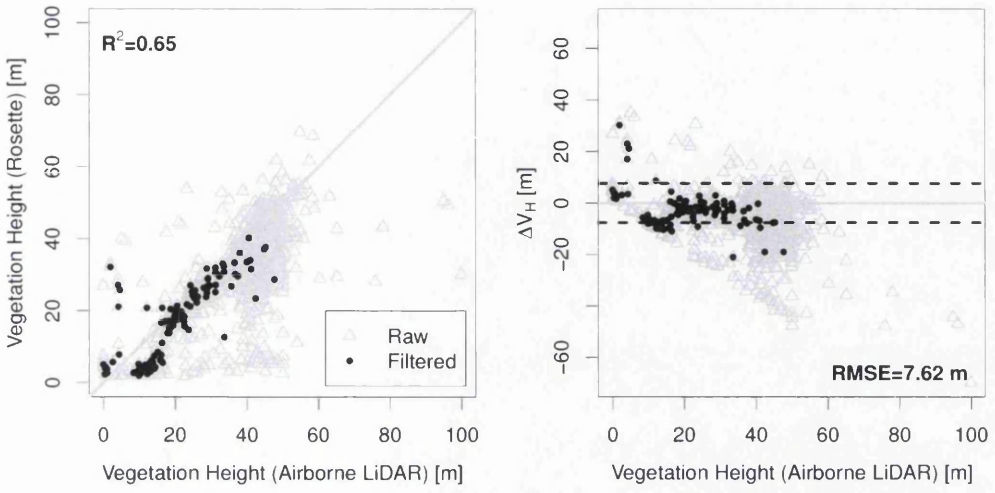
Figures 6.7a and 6.7b show the distribution of the differences between the airborne and GLAS vegetation height estimates (ΔV_H) per 2 m intervals. The difference distribution for the control data is shown with an overlay of the difference distributions for the Lee, and modified results respectively. Explicitly, Figure 6.7a overlays the Lee difference distribution, and Figure 6.7b the modified difference distribution.

Figures 6.7c and 6.7d illustrate the mean differences between the airborne and GLAS vegetation height estimates ($\Delta\overline{V_H}$) as a function of airborne LiDAR derived slope (by 2° intervals). Similarly to above, control $\Delta\overline{V_H}$ values are visualised in addition to the Lee (Figure 6.7c), and modified (Figure 6.7d) corrected equivalent results, respectively.

In general both Lee, and modified methods show a more representative comparison between airborne and GLAS V_H estimates (Figure 6.6), according to the slightly

Table 6.1: Summary of model performance parameters applied to GLAS derived vegetation heights (filtered according to Section 6.1.1) with respect to airborne LiDAR equivalent data for all sites. Statistics for the control, Lee (Equation 6.3), and modified (Equation 6.7) methods are shown.

Status	n	R	p_R	D	F_2	F_B	ε_{nme}	RMSE[m]
Control	110	0.81	$< 10^{-10}$	0.23	0.58	-0.16	0.28	7.62
Lee	110	0.81	$< 10^{-10}$	0.24	0.55	-0.26	0.33	8.74
Modified	110	0.81	$< 10^{-10}$	0.17	0.65	0.00	0.23	6.89



a Comparison b Difference

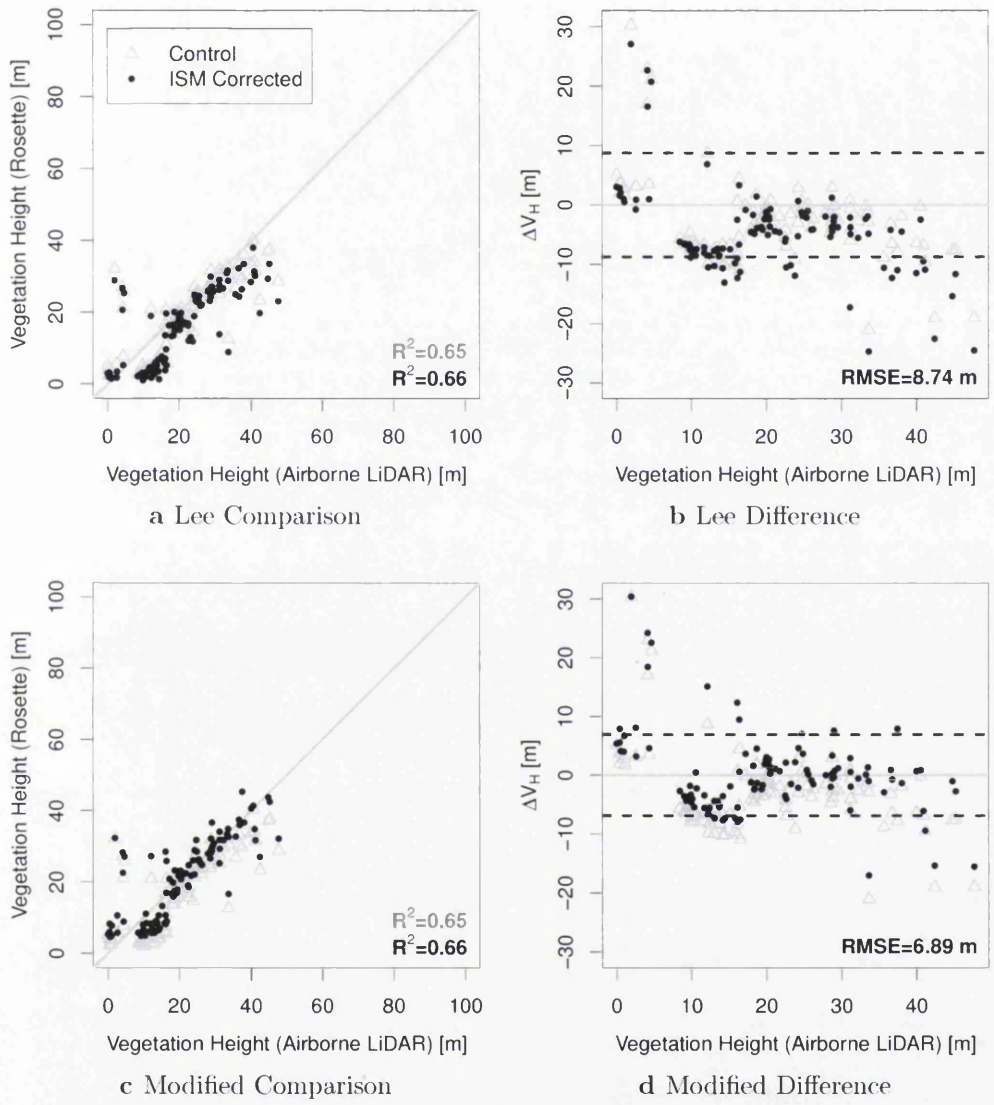
Figure 6.5: (a) Comparison and (b) difference, with RMSE (dashed lines) between airborne and GLAS (control) vegetation height estimates respectively. Note: solid grey line represents the 1:1 line, grey points represent data before the application of filters from Section 6.1.1, whereas black dots represent retained (control) data, post filtering. Rosette indicates V_H was derived by the method of Rosette et al. (2008).

improved Pearson correlation coefficient (R^2). This is further confirmed in Figures 6.7a and 6.7b, indicating less spread difference distributions.

It is noted that the Lee dataset performs poorly, as reflected in corresponding model performance parameters, where the modified dataset shows improved parameters (Table 6.1). This is highlighted in Figure 6.7b, where the difference distribution between airborne and corrected GLAS vegetation height estimates are centred around zero. This is further confirmed in Figures 6.7c and 6.7d, where $\Delta\overline{V_H}$ is more closely centred around zero for all slopes.

Interestingly, control heights tend to be underestimated with regards to airborne LiDAR equivalent information, this is typically the case where GLAS is concerned (Lefsky et al., 2007). However, for heights retrieved over sloped terrain, overestimations are expected (Rosette et al., 2013). This is not the case for the data from the study sites employed in this study, however, it is noted that the majority of "control" data are sourced from footprints with $\leq 10°$ slope.

The general underestimation of control heights highlights the refinement required



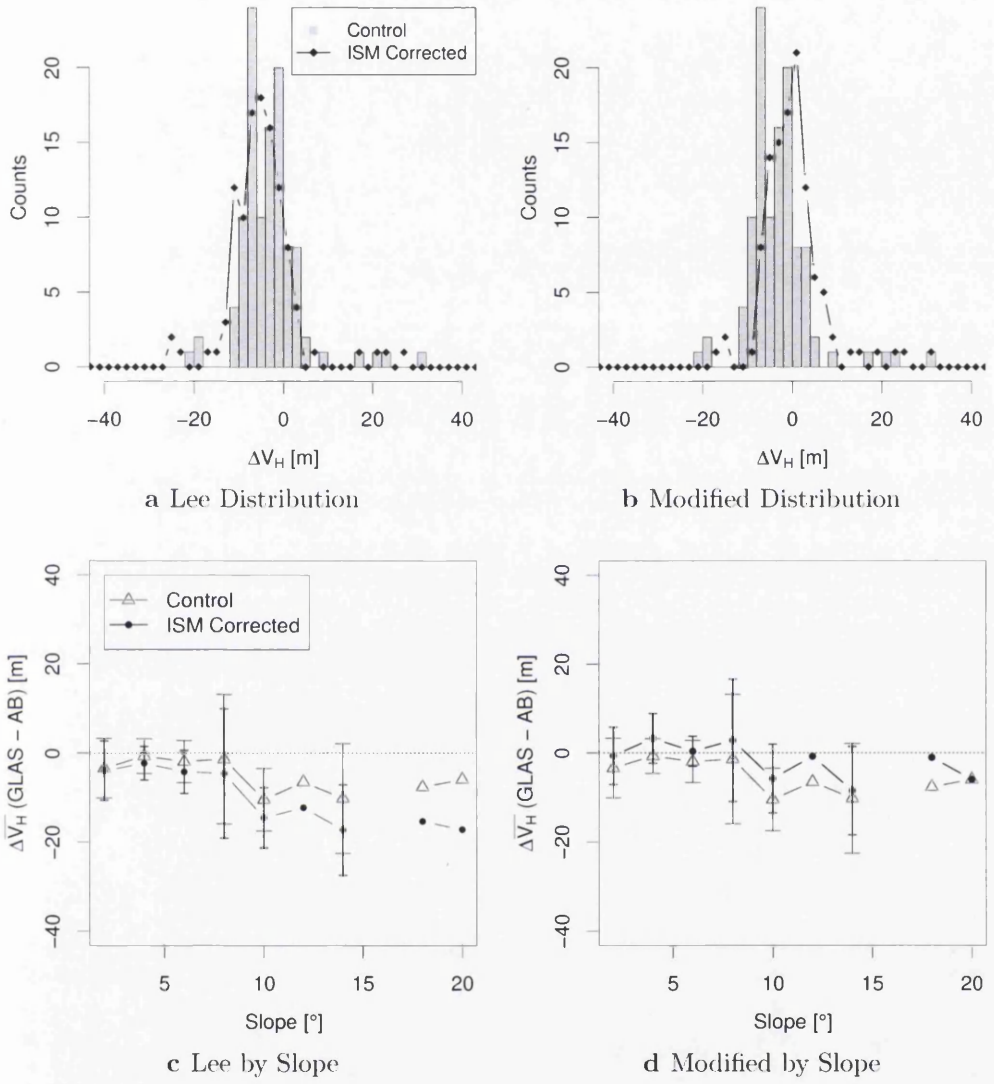
a Lee Comparison

b Lee Difference

c Modified Comparison

d Modified Difference

Figure 6.6: (a, c) Comparison between airborne and GLAS derived vegetation height estimates for control data (grey triangles; filtered according to Section 6.1.1). Lee, and modified corrected control (Equations 6.3 and 6.7 respectively) are given for respective plots (black points). (b, d) Differences and RMSE (dashed lines) for (a) and (c) respectively.



a Lee Distribution

b Modified Distribution

c Lee by Slope

d Modified by Slope

Figure 6.7: (a, b) Difference distributions between airborne (AB) and GLAS vegetation height estimates. The control (grey bars), and corrected distributions (black line) are shown for the (a) Lee, and (b) modified correction methods. (c, d) Mean difference between the airborne and GLAS vegetation height estimates as a function of slope, binned at 2° intervals. The control (grey line), and corrected (black line) by the (c) Lee, and (d) modified methods are shown.

for GLAS to correctly interpret the canopy top of low density crowns, which has been suggested as a cause of the general under-estimation of vegetation height from (large footprint) GLAS waveforms (Lefsky et al., 2007).

6.2.2 GLAS Laser 1A Application

Vegetation height estimates were made for GLAS waveforms recorded during the L1A campaign between 20/02/2003–21/03/2003. All data were subject to the filtering process in Section 6.2.1, where the retained dataset was defined as the "control". GLAS campaign L1A was selected as it was found to provide generally representative results of vegetation and terrain parameters as found in Chapter 5, particularly Figure 5.9 where slope prediction uncertainty is investigated. In addition, justifications for choice of this laser campaign are given under the followimg heading.

Data displayed in Figure 6.8 are displayed on a $0.2° \times 0.2°$ pixel basis, where each pixel assigned value represents the 99^{th} quantile vegetation height. This 99% (V_H) value was obtained for all waveforms within each corresponding pixel's spatial extent. This result is essentially the same as obtained by Los et al. (2012) for GLAS campaign L1A, where the the only differences arise from the $0.2° \times 0.2°$ sampling, and 99% (rather than 90% in Los et al., 2012) V_H values.

The spatial distribution of vegetation height in Figure 6.8 agrees with literature distributions from both Simard et al. (2011) and Los et al. (2012). Distribution agreements with the global product produced by Lefsky (2010) are poorer, where some questions have been raised with regards to the accuracy of the vegetation heights themselves (Simard et al., 2011; Los et al., 2012).

Comparison Choice

The vegetation height distribution products derived in this study were compared with the equivalent product of Los et al. (2012) as this product is a direct measurement product with stringent filters (similar to those in this study) applied to eliminate spurious data. Additionally GLAS L1A was employed for comparison purposes at the near global scale as it represents a "middle ground" between the best and worst of GLAS campaigns with regards to footprint dimensional, and positional accuracies.

The dataset of Simard et al. (2011) was not used for comparison because the only direct measurements employed were those from GLAS laser campaign 3C, which were subsequently slope filtered with a 5° threshold (derived in the same way

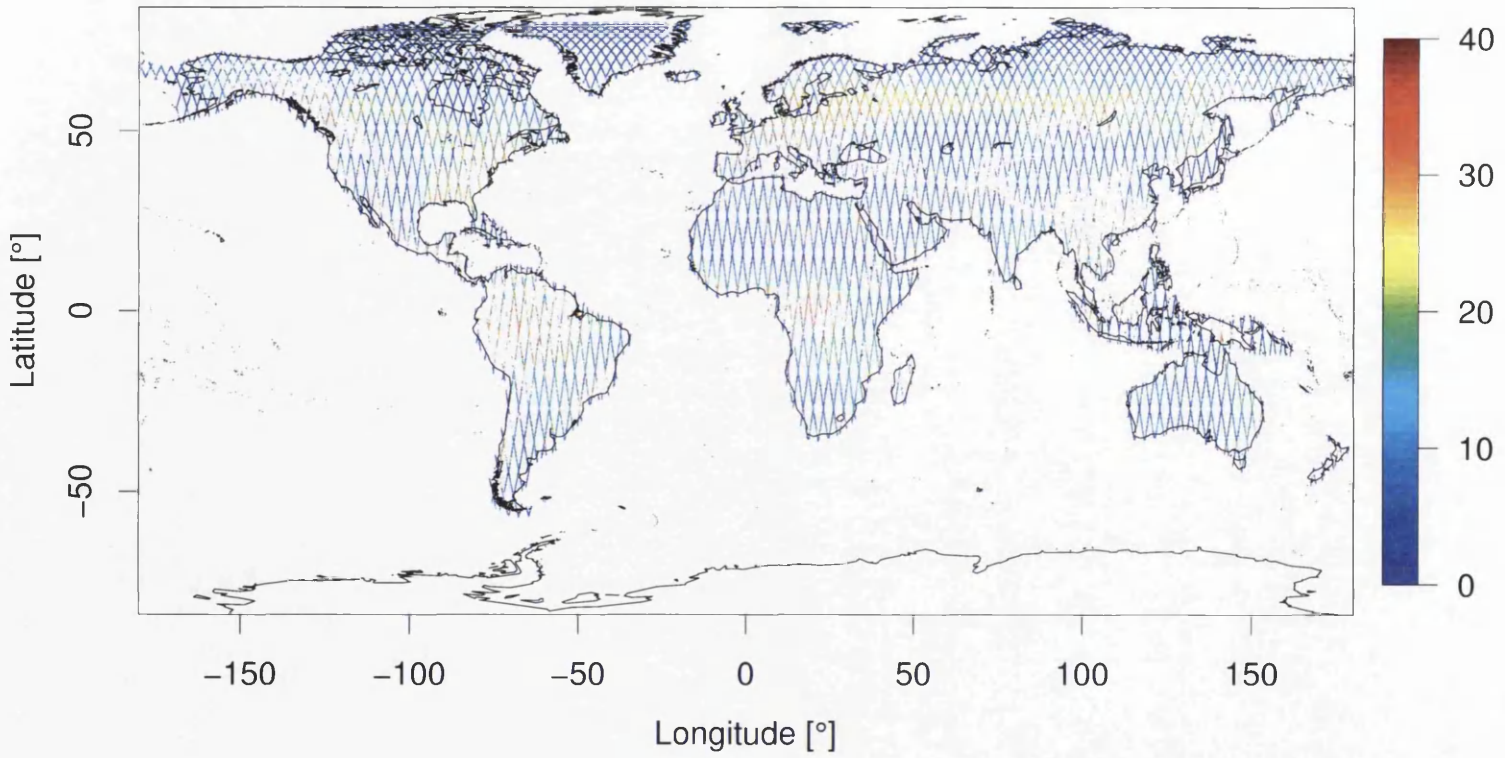


Figure 6.8: Vegetation height (V_H; in metres according to colour) as derived using the method from Rosette et al. (2008). V_H values are the 99^{th} quantiles of all estimated vegetation heights from all waveforms that fall with each corresponding $0.2° \times 0.2°$ pixels spatial extent. Data are shown in orbital lines as a consequence of satellite orbital path, lines would appear more dense if more data from other GLAS laser campaigns were employed.

as in Los et al., 2012). These data were used as a basis (in conjunction with 7 other climate and terrestrial ancillary variables) for the use of the regression tree method, Random Forest (RF; Breiman, 2001) to model vegetation heights where no GLAS coverage was present. This produced model results based on a single laser campaign, different to measurements used in this study, hence comparisons of vegetation height products are unverifiable by direct data.

Correction Application

The two previously discussed vegetation height correction methods, Lee (Equation 6.3), and modified (Equation 6.5), were applied to the control dataset, and defined as the Lee, and modified datasets, respectively.

A mean global difference was noted between the control vegetation height distribution and both Lee, and modified distributions, supported in Table 6.2, by additional statistical difference tests. The mean global difference between the control and Lee distributions was -0.91 m, whereas the equivalent difference between the control and modified distributions was 0.99 m. Similar improvements are noted in respective RMSE values between the control and corrected (Lee and modified) distributions with regards to equivalent airborne LiDAR information.

It is noted, due to the nature of the method employed to extract slope directly from waveforms (Chapter 5), some slope estimates are irretrievable. Hence, for such occurrences no vegetation height correction is applied. These uncorrected vegetation heights are still present in the difference calculation between the control and Lee, and modified height distributions. These zero differences bias the true effect of the applied correction methods, yielding unrepresentative results with respect to the severity of vegetation height change.

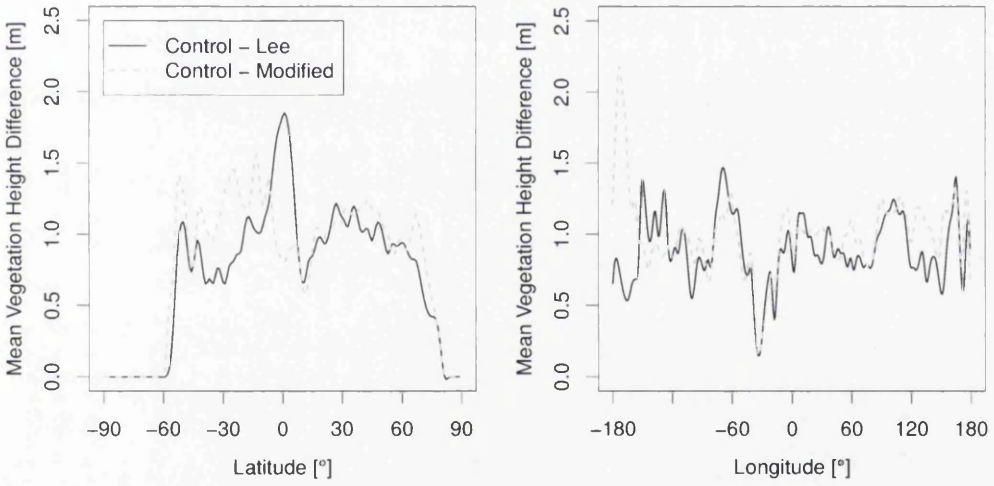
A more representative effect of the slope corrections and the corresponding differences between respective height distributions is realised by excluding data where no slope information is available from the relevant statistical tests. Mean differences between the control and Lee height distributions, and control and modified distributions were found to be -1.29 m and 1.83 m respectively (see Table 6.2), when data with no slope information were removed from analysis.

The absolute mean differences between the control–Lee, and control–modified

Table 6.2: Summary of statistical parameters testing the statistical differences between the control, Lee, and modified data distributions for footprints within the GLAS L1A campaign. Statistics for these differences are shown for both the inclusion (top two lines) and exclusion (bottom two lines) of footprints where slope information is irretrievable by the Independent Slope Method (Section 5.2). Note: n = sample size, M_D = mean difference in vegetation height, R = Pearson correlation coefficient, T = T-test statistic, F = F-test statistic, p_X = fractional probability of null hypothesis being true (significance), e.g. p_R = significance of R.

	Correction	**n**	**M_D [m]**	**p_{M_D}**	**R**	**p_R**	**T**	**p_T**	**F**	**p_F**
Inclusion	Lee	65307	-0.91	$< 10^{-10}$	0.96	$< 10^{-10}$	851.46	$< 10^{-10}$	724989.49	$< 10^{-10}$
	Modified	65307	0.99	$< 10^{-10}$	0.98	$< 10^{-10}$	1326.71	$< 10^{-10}$	1760161.68	$< 10^{-10}$
Exclusion	Lee	45805	-1.29	$< 10^{-10}$	0.96	$< 10^{-10}$	851.46	$< 10^{-10}$	724989.49	$< 10^{-10}$
	Modified	45805	1.83	$< 10^{-10}$	0.98	$< 10^{-10}$	1326.71	$< 10^{-10}$	1760161.68	$< 10^{-10}$

height distributions are shown in Figure 6.9 as a function of latitude and longitude (at 1° intervals). Latitudinal analysis indicates mean height differences follow that of the tree line, explaining zero difference in the extreme latitudes, in addition to greater differences at the tropics, a suspected function of increased vegetation presence. Longitudinally, a more consistent height difference is noted, as expected, due to differences originating from a broad range of forests from all latitudes. Greater differences noted at extreme longitudes are a consequence of the disproportional contribution of the boreal forests of Alaska and the Russian peninsula to these longitude bands, as the majority of these longitudes are occupied by the Pacific ocean otherwise.
All statistical calculations in Table 6.2 were based on the data (each pixel) displayed within Figure 6.8, not the full number of waveforms employed to create each $0.2° \times 0.2°$ pixel.



a Latitude **b** Longitude

Figure 6.9: Zonal mean vegetation height difference for control - Lee, and control - modified distributions, as a function of (a) latitude, and (b) longitude, at 1° intervals.

6.3 Effect of Surface Type

In order to test the general applicability of the Lee, and modified correction techniques (Equation 6.3 and 6.5 respectively) "tiles" of data, recorded over different surface types were selected. Tiles of dimensions $5° \times 5°$ were selected based on coordinates that encompassed one the following surface types: Boreal forest (BF), Desert (D), Temperate forest (TF), and Tropical forest (TR); these surface types were translated from SiB2 vegetation classes 4, 11, 1 and 2, and 3 respectively (see Sellers et al., 1996). The locations of respective tiles are given in Table 6.3. Each unique tile is employed to test the degree of variability of the Lee and modified height distributions with respect to the control. Table 6.3 shows the statistical mean differences (and other test statistics) for each tile, where unaltered (as a result of irretrievable slope information) data are included and excluded (indicated by E) in separate calculations. All statistical values stated in the following section are based on retained data after exclusion of data with irretrievable slope.

In general, tile statistical tests appear less significant than at the near global scale (Table 6.2), this is likely due to the decreased number of observations in each respective tile. The greatest magnitude difference between the control and corrected distributions (Lee and modified) appear consistently for tropical forests. Over the Amazon basin the mean difference between control–Lee corrections range between -6.65 and -4.37 m, and 1.72 and 3.32 m for control–modified, however these values are not statistically significant. Relative to corresponding values from other tiles such corrections are large.

The greater difference observed for the tropical forest originates from derived slope values, which are of poor quality due to potential interference from multiple canopy layers present in the tropical forest; such characteristics induce difficulty in retrieving slope (Section 5.3). These subtleties have potential to induce slope overestimates, and thus over-corrections to vegetation heights.This effect is much more severe in Radar observations, as noted for SRTM measurements over tropical regions (Rodriguez et al., 2005; Jarvis et al., 2008).

The smallest mean difference between vegetation height distributions is noted in the Saharan desert. Here the mean difference ranges from -1.34– -1.99 m for the control–Lee comparison, and between 0.92–1.87 m for the control–modified com-

Table 6.3: Tile locations, and statistical parameters summarising the differences between the control, Lee, and modified vegetation height distributions for GLAS campaign L1A. 5° × 5° tiles were selected as a function of surface type, uniquely representing: Boreal forest (BF), Desert (D), Temperate forest (TF), and Tropical forest (TR). Statistics, for within tile differences, are shown for datasets that include and exclude (indicated by E) footprints where slope information is irretrievable. Note: ψ, λ = latitude, longitude, n = sample size, M_D = mean difference in vegetation height, R = Pearson correlation coefficient, T = T-test statistic, F = F-test statistic, p_X = significance of corresponding test statistic, e.g. p_R = significance of R.

Site	ψ, λ [°]	Correction	n	M_D [m]	p_{M_D}	R	p_R	T	p_T	F	p_F
BF	55–60 N, 50–55 E	Lee	334	-2.17	< 0.05	0.97	$< 10^{-10}$	101.28	$< 10^{-10}$	10256.79	$< 10^{-10}$
		Lee (E)	217	-2.58	< 0.05	0.97	$< 10^{-10}$	101.28	$< 10^{-10}$	10256.79	$< 10^{-10}$
		Modified	334	1.30	> 0.05	0.95	$< 10^{-10}$	83.86	$< 10^{-10}$	7033.10	$< 10^{-10}$
		Modified (E)	217	1.99	> 0.05	0.95	$< 10^{-10}$	83.86	$< 10^{-10}$	7033.10	$< 10^{-10}$
D	20–25 N, 0–5 E	Lee	422	-1.34	$< 10^{-6}$	0.88	$< 10^{-10}$	54.89	$< 10^{-10}$	3012.45	$< 10^{-10}$
		Lee (E)	207	-1.99	$< 10^{-6}$	0.88	$< 10^{-10}$	54.89	$< 10^{-10}$	3012.45	$< 10^{-10}$
		Modified	422	0.92	< 0.05	0.91	$< 10^{-10}$	65.87	$< 10^{-10}$	4338.89	$< 10^{-10}$
		Modified (E)	207	1.87	< 0.05	1.87	$< 10^{-10}$	65.87	$< 10^{-10}$	4338.89	$< 10^{-10}$
TF	35–40 N, 85–90 W	Lee	261	-2.25	< 0.05	0.92	$< 10^{-10}$	56.34	$< 10^{-10}$	3174.76	$< 10^{-10}$
		Lee (E)	173	-2.36	< 0.05	0.92	$< 10^{-10}$	56.34	$< 10^{-10}$	3174.76	$< 10^{-10}$
		Modified	261	1.27	> 0.05	0.96	$< 10^{-10}$	77.88	$< 10^{-10}$	6064.58	$< 10^{-10}$
		Modified (E)	173	1.92	> 0.05	0.96	$< 10^{-10}$	77.88	$< 10^{-10}$	6064.58	$< 10^{-10}$
TR	0–5 S, 60–65 W	Lee	87	-6.65	< 0.05	0.78	$< 10^{-10}$	17.19	$< 10^{-10}$	295.48	$< 10^{-10}$
		Lee (E)	45	-4.37	< 0.05	0.78	$< 10^{-10}$	17.19	$< 10^{-10}$	295.48	$< 10^{-10}$
		Modified	87	1.72	> 0.05	0.93	$< 10^{-10}$	34.81	$< 10^{-10}$	1211.61	$< 10^{-10}$
		Modified (E)	45	3.32	> 0.05	0.93	$< 10^{-10}$	34.81	$< 10^{-10}$	1211.61	$< 10^{-10}$

parison. Little correction is expected in such a region, as little vegetation is present. Relative to the mean differences for each of the other tiles, these values align well with the mean difference observed for both the control–Lee, and control–modified comparisons at the near global scale (Table 6.2).

Both the boreal region of the central Siberian plateau, and the temperate forests of the east coast of the United States, exhibit greater corrections than applied at the desert regions, but less than those applied at the tropical region, as expected. These regions experienced almost identical corrections for both the control–Lee, and control–modified comparisons; a mean difference between -2.58 and -2.20 m, and 1.30 and 1.90 m was noted as a consequence of the application of these two techniques respectively.

In summary, the greatest effects of the correction techniques are at tropical regions, where only the Lee correction is statistically significant. The smallest correction is found in the desert regions, where both correction applications appear statistically significant. Both the boreal and temperate forest regions produced very similar results, where only the Lee correction application produced any statistically significant results. This however, does not imply that application of the Lee correction improves estimates of V_H (see Figures 6.6 and 6.7).

6.4 Comparison with NDVI

The $0.5° \times 0.5°$ resolution mean annual Normalised Difference Vegetation Index (NDVI) product from the Moderate Resolution Imaging Spectroradiometer (MODIS) was used as an indirect way to validate the global vegetation height distributions produced in this study by the Lee (Equation 6.3), and modified (Equation 6.5) correction techniques. MODIS data were selected from 2003 only, corresponding to the acquisition year of GLAS campaign L1A (NSIDC, 2012a); a global wall-to-wall mean annual NDVI map for all vegetation cover fractions is shown in Appendix 9.12.

The NDVI is near linearly related to the fraction of photosynthetically active radiation (fPAR) absorbed by vegetation canopies for photosynthesis. This is linked to the amount of CO_2 absorbed by the vegetation (Sellers et al., 1996), which is allocated to various vegetation components, such as leaves and woody biomass,

above and below ground (Potter et al., 1993). For these reasons a positive relationship is expected to exist between mean annual NDVI and vegetation height, as demonstrated by Los et al. (2012). Following this reasoning, NDVI can be used to evaluate vegetation height products produced in this study.

Only vegetated areas were selected, as inclusion of non-vegetation objects (e.g. buildings) will induce uncertainty, as NDVI is a vegetation measure only. A minimum threshold of 40% vegetation cover was imposed following Los et al. (2012), allowing a comparison for vegetated surfaces only. The vegetation cover fraction (VCF) of the MODIS continuous field products (Hansen et al., 2003; Townshend et al., 2011) was used to remove surfaces with $\leq 40\%$ vegetation at this scale. The VCF is a measure of the fraction of vegetation (expressed as a percentage) present within a pixel projected on to the Earth's surface. In this case, the 500 m resolution pixel product was selected, and aggregated to $0.5° \times 0.5°$ pixels (matching the NDVI grid); A global map of NDVI for pixels that exhibit VCF $\geq 40\%$ is shown in Figure 6.10.

The resulting correlations for comparisons of vegetation height and NDVI, for pixels that exhibit VCF $\geq 40\%$ were calculated for the control, Lee, and modified height distributions. Correlations of R = 0.26, 0.18, and 0.26 were found for NDVI comparisons with the control, Lee, and modified distributions, respectively. Colour density plots illustrated by Figure 6.11 support these relationships. Two high density concentrations are visible (yellow) in Figures 6.11a to 6.11c, from varying V_H, the more constrained these concentrations are the greater correlation is expected. Figure 6.11a exhibits the most constrained density concentration for shorter trees, whereas Figure 6.11c shows similar for taller trees (although less clearly).

It is noted that, for almost identical conditions, the best correlation between NDVI and the control height distribution (essentially the same as Los et al., 2012) in this study does not match that derived by Los et al. (2012). Similarities may be expected as almost identical height retrieval and filtering procedures were executed, however, a discrepancy is possible for two reasons. First, additional data is available for analysis beyond the 60° latitude parallel here (see Section 6.5 for further details). Second, only vegetation height data available from GLAS laser campaign L1A were employed for the comparison in this study whereas Los et al. (2012)

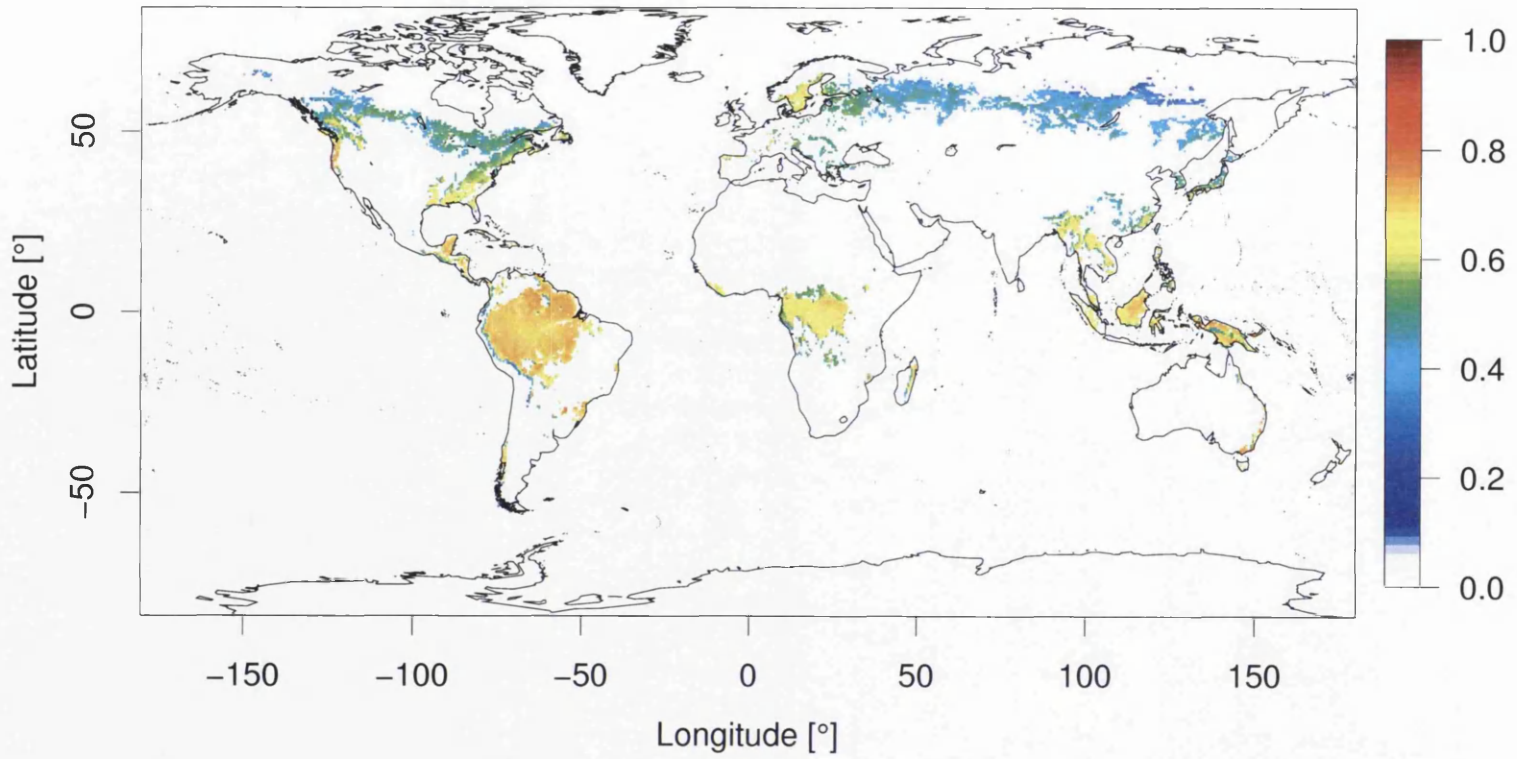
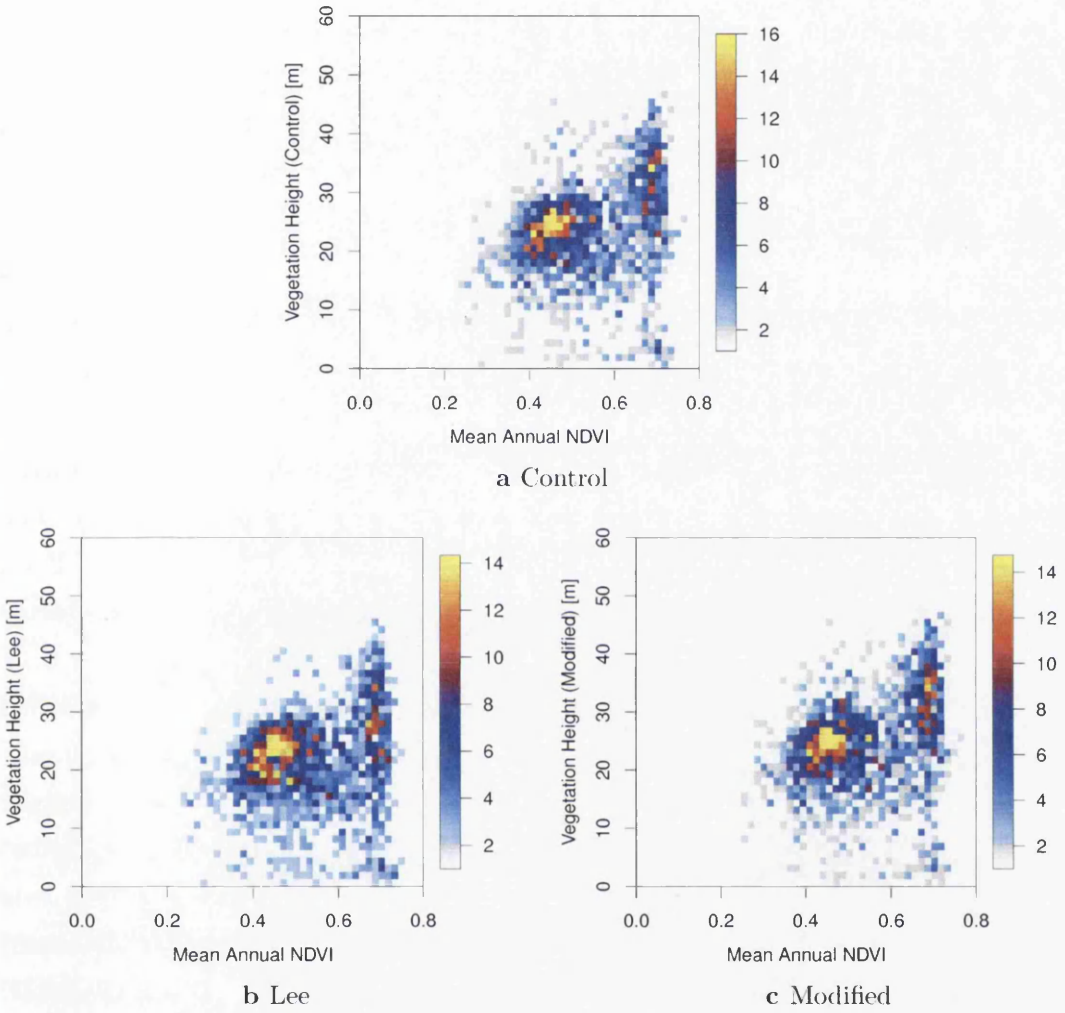


Figure 6.10: Global map of Normalised Difference Vegetation Index (NDVI; according to colour) from MODIS $0.5° \times 0.5°$ tiles, with the MODIS derived vegetation cover fraction (VCF) restricted to $\geq 40\%$. A global map of NDVI for all surfaces (VCF between 0–100%) is illustrated in Appendix 9.12.

employed results from all laser campaigns.

Current results suggest the difference between NDVI–vegetation height correlations for the control and modified height distributions is non-existent. Hence clear conclusions with regards to which distribution best represents reality at this spatial scale cannot be drawn. Equivalent correlation results for the Lee height distribution indicates a decrease in R with respect to the other distributions, suggesting a less realistic representation of reality.

Furthermore, it is noted that poor correlations found in this section are (somewhat) a consequence of the choice of NDVI data. The ideal choice of NDVI data would be those recorded in the same time frame as when the GLAS L1A data were collected, hence allowing a more direct comparison. Alternatively, maximum NDVI (rather than mean NDVI) would have been a more valid choice for comparison purposes here as these data would relate to leaf-on conditions, whereas mean data relate to a composite of both leaf-on and leaf-off conditions. This is of important note, as whilst V_H would not have changed (significantly) between these two phenological states, NDVI would have. Neither of these datasets were accessible at the time of this study, hence justifying the choice of mean NDVI data used. Hence, to further investigate the integrity of these results, more comparable data are required such to infer more precise conclusions.



a Control

b Lee

c Modified

Figure 6.11: Colour density plots illustrating the global relationship between vegetation height and NDVI where vegetation cover $\geq 40\%$ for (a) control, (b) Lee, and (c) modified vegetation height distributions.

6.4.1 Effect of Vegetation Cover Fraction

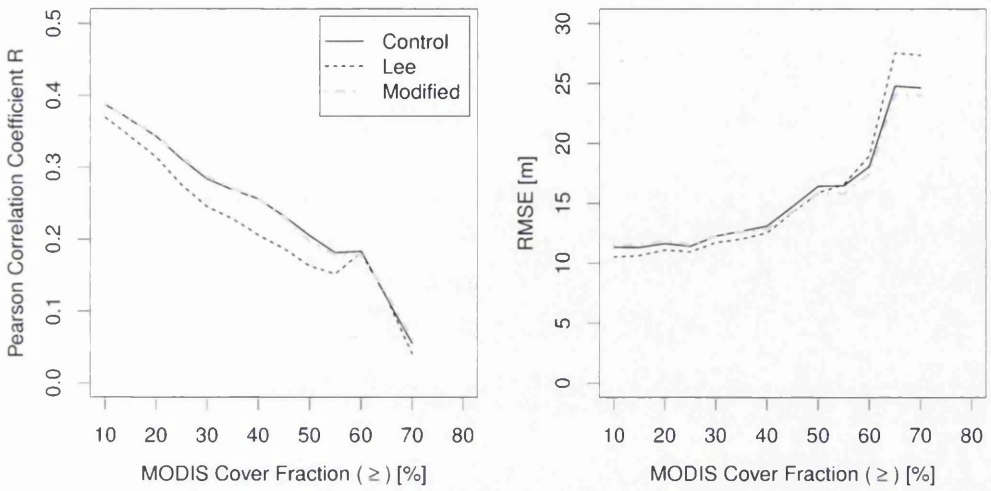
Correlation and RMSE values between NDVI and each respective vegetation height distributions (control, Lee, and modified) were analysed as a function of increasing minimum VCF. The implication being that as minimum VCF increases, the number of available pixels for comparison across the globe will decrease, restricting analysis to increasingly vegetation dominated regions. Analysis was performed from $\geq 10\%$ to $\geq 70\%$ minimum VCF at 5% intervals.

Figures 6.12a, and 6.12b illustrate the effect that increasing minimum VCF had on correlations, and on RMSE values between NDVI and respective height distributions. Figure 6.12a indicates that R decreases as minimum VCF increases for all height distributions. Interestingly, the results for the control and modified datasets are almost identical for all minimum VCF. Results for the Lee dataset consistently exhibit lower values of R up to $\geq 60\%$ VCF, where results for all three datasets are very similar.

The observed overall decrease in R is attributed to the Pearson correlation test itself. By its nature the test requires a broad spread of data in both comparison variables, i.e. NDVI and vegetation height in this case. However, as minimum VCF becomes more restrictive, the range of both NDVI and vegetation height becomes more concentrated thus resulting in lower values of R.

The minor differences between correlations for the control, and modified datasets, suggest that the small mean difference in global vegetation heights between the two datasets (~ 0.99 m; Table 6.2) is masked by the high variance present in NDVI and vegetation height values. Namely, as vegetation height increases by 0.99 m (mean difference between control and modified height distributions), the range of NDVI values for comparison purposes do not change significantly, thus resulting in almost identical results.

Interestingly, the mean global vegetation height difference between the control and Lee distributions occurs at a smaller magnitude, but is negative (~ -0.9 m; Table 6.2). Hence, by the previous logic, little difference in correlation results between these height distributions should be observed. The observed difference here is attributed to the reduction in the mean vegetation height between the two datasets (rather than increase, as noted above), suggesting the Lee height distribution is a



a Pearson correlation

b RMSE

Figure 6.12: Change in (a) Pearson correlation coefficient, R, and (b) RMSE, between vegetation height and NDVI with increasing VCF. Correlations were calculated for control, Lee, and modified vegetation height distributions.

less accurate representation of reality.

Figure 6.12b indicates that RMSE increases with increasing minimum VCF, for similar reasons as explained for the decrease in correlation under similar circumstances. For lower values of minimum VCF (up to $\geq 40\%$), the control and modified data produce very similar vales of RMSE, whereas the Lee data exhibits the smallest RMSE values. Beyond this and up to $\geq 50\%$ VCF, the Lee and modified data are very similar, out-performing the control. From $\geq 55\%$ the Lee data exhibits the greatest RMSE values, where the modified method exhibits the smallest.

From the previous justification of NDVI–vegetation height comparison (Section 6.4), smaller values of VCF will invariably exhibit a greater percentage of shorter vegetation, and higher VCF, a greater percentage of taller vegetation. Hence, by this logic Figure 6.12b suggests the Lee height distribution performs best for shorter vegetation, and the modified method for taller vegetation. However, the differences in RMSE values between height distributions noted here are not statistically significant.

Conversely to the GLAS vegetation height comparison with airborne LiDAR data, it is found that the correlation coefficients are considerably different between the control and modified height distributions, with respect to the Lee distribution. However, at the near global scale other statistics, such as RMSE, do not show such sizeable differences. This suggests a limitation in the use of NDVI as a validation tool.

This limitation of NDVI is a legacy of its derivation from passive optical data, which inherently does not allow full vegetation structure information to be obtained, as is possible with direct LiDAR measurements. This (in combination with environmental variability) explains the broad range of NDVI values present in Figure 6.11 for almost all values of vegetation height. Unfortunately this restricts the confidence at which it can be stated that the modified vegetation height distribution improved on the control (if at all) at this scale, but suggests that it is equivalent.

At the site level results are clearer, where the modified height distribution represents equivalent measures from high resolution airborne LiDAR more accurately than both the control, and Lee height distributions (see Section 6.2.1).

6.5 Additional Comparisons

A legacy of retrieving slope directly from LiDAR waveforms by the direct method outlined in this study (Section 5.2) is two-fold. First, it offers a means of identifying sloped terrain from one single dataset, employing the same points to aid within-footprint slope estimates, negating the need for external datasets (such as SRTM; Simard et al., 2011; Los et al., 2012). This removes discrepancy that is often introduced by employing multiple datasets, particularly as they are often available at different spatial resolutions, which in turn effects within footprint variation. Second, a slope correction allows additional data to be employed as it can be deemed more reliable than before correction. The amount of newly available data is proportional to the number of footprints that encounter sloped terrain. For example, Simard et al. (2011) and Los et al. (2012) filtered data based on slope, and removed large proportions of data in doing so.

It was tested how much more vegetation height data becomes available using the

ISM slope retrieval technique. Analysis was performed with respect to the dataset available from Los et al. (2012), and it was found that approximately 65000 additionally available data points for this laser campaign can be used in analysis (discussed below). The selection for comparison is reasoned in Section 6.2.2.

In the product of Los et al. (2012), a 10° slope filter sourced from a 3×3 matrix of SRTM 4.1 tiles was applied; slope was calculated in much the same way as described in Section 3.10. A 3×3 SRTM tile matrix represents larger areas than GLAS footprints, and overlooks within footprint variation. This translates to underestimates of slope from SRTM (because of its resolution) with respect to those found at the footprint scale. In the case of Los et al. (2012), it is likely that some footprints are retained on the basis of SRTM slope underestimation, where at a less coarse resolution they would normally be excluded.

In addition, SRTM tiles are spatially restricted between ±60° latitude. In the case of Los et al. (2012) this meant it was not possible to calculate vegetation heights outside these bounds, as it was not possible to verify the data. Calculating slope independently of the SRTM elevation model (or any other external data) allows spatial continuity, both in resolution and spatial limits, and allows filtering to continue beyond ±60° latitude.

Additional data points made available by use of the ISM approach are calculated and expressed as the extra percentage of the number of data points available with regards to Los et al. (2012). Two aspects were investigated with respect to additional data; first, data available from slope knowledge $> 10°$ only, and second, data available from no spatial limitations as imposed by SRTM (±60°). Results were calculated for the GLAS L1A campaign, based on the control dataset derived by filtering procedures outlined in Section 6.1.1.

The mean percentages of additionally available data from GLAS campaign L1A from slope $> 10°$ only, and spatial extent expansion were 0.2% and 29.5%, respectively. Figure 6.13 indicates grid cells where additional data points are available by the ISM slope retrieval technique. Note: only data with retrievable slope information are displayed, partially accounting for small data additions for slope only. Additionally, Figure 6.13 illustrates that the majority of additional data points are available beyond 60° latitude, and a much smaller amount of additional data are available between ±60°. Of additional data beyond ±60° latitude, additions

appear to be in the temperate forests of Europe and the USA, as well as the tropical forests of South America, Africa, and south east Asia. Above 60° latitude, a considerable data addition to the boreal forests of Siberia and northern Europe is notable.

Of the additionally available data by the ISM slope retrieval method, approximately 42.3% have usable slope information (see Chapter 5 as to why slope is not available for all waveforms). Using the method interpreted by Simard et al. (2011), all waveforms with only one fitted Gaussian (from the fitting method of Duong et al., 2006) were deemed to be unrepresentative of vegetation, and more likely be a measurement of the ground surface only. By eliminating such waveforms those remaining are interpreted as having recorded vegetated surfaces. Employing this technique it is estimated that approximately 3.4% of additionally available data contributes to the vegetation height distribution derived from the GLAS L1A campaign, this translates to approximately 65,000 extra vegetation measurements. The potential of such a technique when applied to subsequent laser campaigns is considerable, revealing high density coverage of northern forest regions with direct measurements.

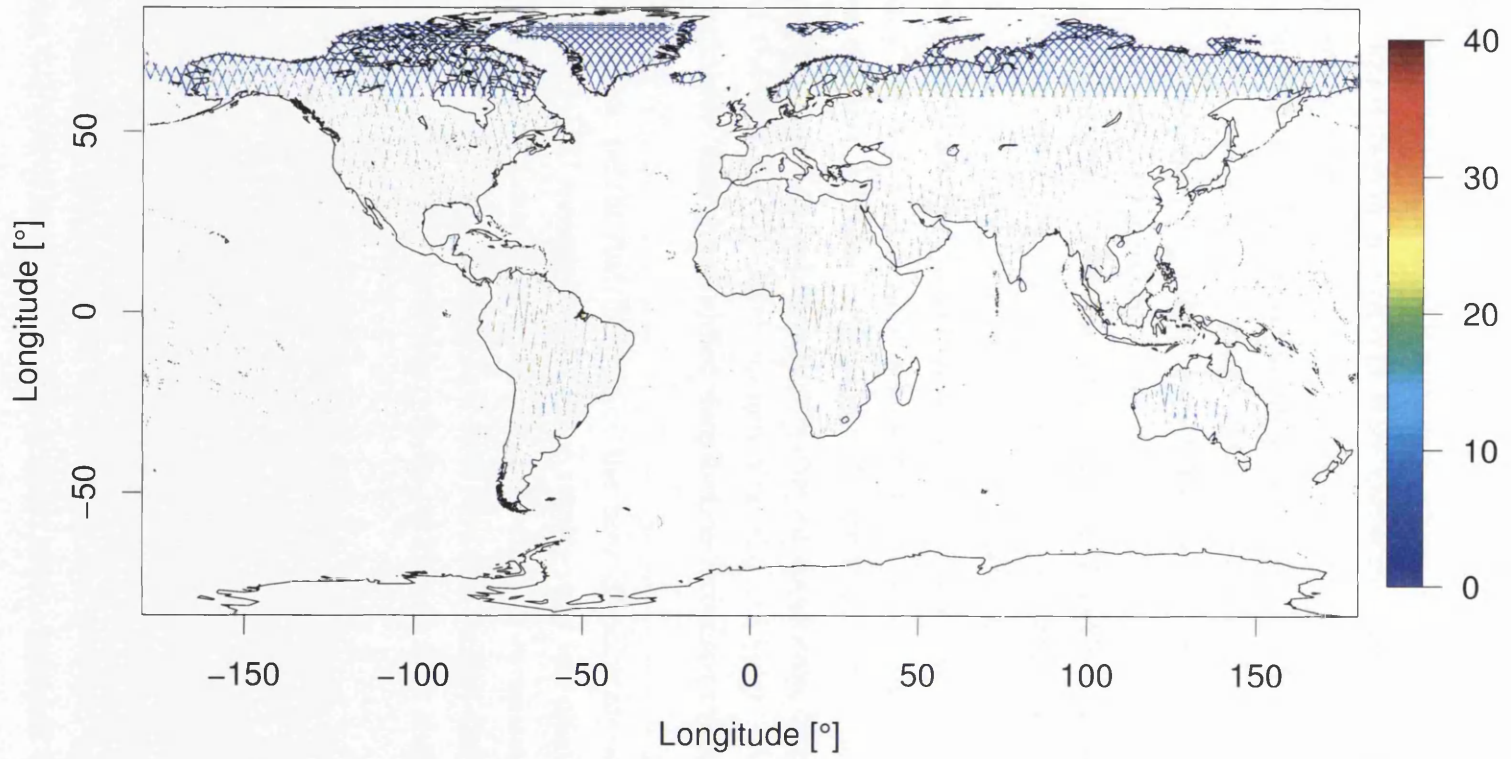


Figure 6.13: Map of the location of additional data points available for the control vegetation height distribution from GLAS L1A campaign where slope is retrieved by ISM. Approximately 29.7% more data are available by this method with respect to the method of slope retrieval employed by Los et al. (2012).

6.6 Summary

In summary this chapter investigated the potential of correcting vegetation height estimates from waveform LiDAR returns as a function of surface slope. Two correction techniques were employed, one from literature (Lee et al., 2011), and another adapted from this.

Investigations found the general applicability of the method outlined by Lee et al. (2011) was flawed as a result of unrealistic assumptions, such as within footprint vegetation location, which effect the severity of the applied correction. This lead to consistent underestimation of the vegetation height distribution derived from GLAS LiDAR returns with respect to airborne LiDAR equivalent measurements at the site level (Section 6.2.1). Conversely the modified height distribution, developed during this study, improved model performance parameters yielding more representative vegetation height estimates at the site level (Table 6.1).

Employing the site level results as a pathfinder, both correction techniques (Lee: Equation 6.3, and modified: Equation 6.5) were applied to the GLAS L1A campaign, offering near global coverage across all continents of the globe. Results mirrored those at the site level, namely that the Lee height distribution reduced height estimates, and the modified distribution increased them, both with statistical significance.

Each vegetation height distribution at the near global scale was compared with mean annual NDVI measurements where results were less conclusive. Correlation and RMSE calculations were made, as a function of increasing vegetation cover fraction (VCF), for the control, Lee, and modified height distributions. Results suggested that the Lee distribution under-performed with respect to the others, which were very similar.

In addition, employing the ISM method of slope retrieval for each GLAS footprint provided access to approximately 29.7% more data points, with respect to the data product of Los et al. (2012). It was estimated that approximately 3.4% of the newly available data were vegetation returns.

Based on the results at the site level, and somewhat on those at the global scale, the new vegetation height estimates yielded by the modified correction technique (Equation 6.5) are of considerable importance, their suggested accuracy cannot be

ignored. Accurate measurements of vegetation height estimates are particularly important in deriving vegetation metrics such as biomass and carbon stocks.

Chapter 7

Discussion

This study has explored the use of waveform LiDAR for use in retrieving terrain and vegetation parameters (namely terrain slope and vegetation height) on regional to large scales, with a particular focus on reducing uncertainty in parameters derived from waveform LiDAR measurements.
Investigations were executed with the goals of retrieving ground slope from vegetated surfaces and refining vegetation height metrics with regards to this error inducing information. In addition, with the long-term aim of vegetation height corrections, shadowing effects within waveform LiDAR returns have also been investigated. The effects of such phenomena were noted with respect to vegetation height estimates, and also with the potential of pioneering new techniques to retrieve previously unidentifiable metrics from LiDAR waveforms.

7.1 Methodological Discussion

Methods were derived such to improve knowledge of waveform LiDAR interactions, and reduce uncertainty in parameters derived from such data. Here respective methods for meeting the outlined aims and objects are assessed, where the problems they address, and performance (with respect to validation data) are discussed.

7.1.1 Shadows in Full Waveform LiDAR

A method for producing a 2-dimensional continuous waveform representation of 3-dimensional spatial data was employed in order to test the fundamentals of detecting vegetation induced signal attenuation (shadowing) effects in physically observed GLAS LiDAR waveforms. Such an exercise allows better understanding of the intricacies that arise when analysing LiDAR waveforms recorded over complex forested surfaces, with regards to deriving terrain information. This is the first test of its kind with regards to large footprint observations, the only other physical shadow observations were noted by Wallace et al. (2012) under controlled conditions; additionally Hancock et al. (2012) noted this phenomena in modelled waveforms from the ray tracer of Lewis (1999). However, neither of these investigations this effects origins, occurrence conditions, and/or likelihood of occurrence, as was the goal of this study.

The presence of shadows in waveforms is a phenomenon that is widely overlooked, in doing so results inferred from shadow affected waveforms are likely inaccurate due to phantom peaks that can be induced in the waveform geometry. Phantom peaks can induce uncertainty in results in many ways, particularly with regards to vegetation height (V_H) estimates, as discussed in Section 4.2.3. Understanding the intricacies of why such phenomena may occur, under what conditions, and with what frequency allows first steps to be taken towards reducing uncertainties in inferred V_H results.

The method employed to test the presence of shadows is generally applicable to all forms of 3-dimensional data (a requirement to form a 2-dimensional waveform), provided footprint size, energy distributions, and pulse width information are available. It informed on the frequency of the presence of shadows, in addition to their origins, and the conditions under which they are likely to occur. In particular, by this method shadows were found to be induced by vegetation subcomponents (shoots, branches, leaves, etc.), and tend to occur more often in the presence of slope ($\geq 10°$).

Taking in to account possible sources of uncertainty in DEM data and uncertainty propagation from interpolation (and other) methodological sources, the severity, and to some extent the elevations, of shadow occurrences can only be realised at

moderate certainty. Hence, whilst methodological principles described in Chapter 4 remain valid, more cautious conclusions have to be considered as a result of these sources of uncertainty. In particular, these pertain to the exact elevation at which shadowing occurs, in addition to the severity at which they occur.

Although subject to some uncertainties, this methodological work is principally accurate, providing insight into the intricacies of incident laser light with complex terrain, particularly in the presence of a vegetation layer. This work can help provide rationale for radiative transfer investigations of conditions under which the shadowing phenomenon is expected to occur, furthering understanding of LiDAR interactions even more. Furthermore, the knowledge of shadows in waveforms allows the refinement of vegetation parameters (such as V_H), and investigations to be made into the retrieval of additional parameters (as discussed in Section 7.3.2) previously unobtainable from waveform LiDAR data.

7.1.2 Slope from Full Waveform LiDAR

Two methods (SSM, Section 5.1, and ISM, Section 5.2) were developed and tested, capable of estimating ground slope angle from all forms of waveform LiDAR for the first time. Such information will allow improved knowledge of terrain complexity, which has been noted as a key parameter in inducing uncertainty in vegetation parameters derived from waveform information as noted by: Lefsky (2010); Simard et al. (2011); Los et al. (2012); and Rosette et al. (2013).

Knowledge of ground slope also offers a means of reducing uncertainties in waveform derived vegetation parameters, particularly V_H, by correcting such parameters as a function of slope. Additionally, slope has been used as a means of filtering waveforms from datasets, where waveforms exhibiting slopes greater than a predefined threshold are removed. Such a filtering technique has been employed by Simard et al. (2011) and Los et al. (2012) for their global vegetation height products, where waveforms with slopes $\geq 5°$ and $\geq 10°$ respectively, were filtered. In these studies slope was derived from SRTM 90 m DEM tiles (3×3 matrix format; Los et al., 2012), chosen for its accuracy and vast coverage ($\pm 60°$). However, due to coarse tile resolution, slopes are often underestimated, hence SRTM data are not ideally suited to this purpose. Furthermore, for waveform footprint filtering,

(near) square SRTM tiles do not offer spatially concurrent coverage of the same areas as such footprints, therefore introducing further uncertainty in SRTM derived slopes.

The methods developed in this study address these issues as data used to derive slope is identical to that used to derive other waveform parameters, and hence spatially concurrent. Hence, slope underestimation is theoretically negated. These methods therefore improve the certainty with which the same spatial region is being measured for slope, particularly in comparison to corresponding SRTM derived slopes, which often cover a larger region than waveform footprint areas.

Of each method, SSM slope estimates were based on lookup table (LUT) data, leading to slope estimation limits based on the range of data available in the LUT (see Section 5.1). In addition, this method produced poor results when independently tested. The inaccuracy in results is expected to occur as a result of limited data available in the LUT, from which all slope predictions are made. This method, whilst not accurate enough for real world application, provided the first steps on which the subsequently developed ISM was built (see Section 5.2).

ISM slope estimates were inferred directly from waveform features, negating the need for LUTs, providing more representative results than SSM slope estimates when compared with airborne LiDAR equivalent data. Additionally, as ISM results are directly reliant on waveform geometries (see Section 5.2) there is no upper limit to its slope estimation capabilities. This method of slope retrieval is applicable to all forms of waveform LiDAR, where results are expected to be similar, if not better (footprint size dependent), than those obtained in this study; greatest results are expected from footprints where diameters are similar to crown diameters.

There are uncertainties associated with GLAS derived slope results, these originate from instrumental limitations and associated uncertainty propagation, in addition to environmental subtleties. Systematic uncertainties are quantified according to Equation 5.7, where variables pertaining to GLAS footprint dimensions, and waveform inferred information (such as ΔE from Equation 5.3) are key in such calculations. Of particular note are the dimensions of the footprint from which slope is inferred, specifically the difference between the semi-major and semi-minor axes. As GLAS footprints are not perfectly circular, but more elliptical (campaign

dependent), directional bias may be introduced to slope inferences, as slope is calculated over the mean (circular) diameter of the footprint. However, whilst accounting for this directional effect, the principles of the slope retrieval method hold but are subject to different magnitude uncertainties according to laser campaign; particularly the difference between reprective campaign's semi-major and semi-minor axes. These differences are noted in Table 3.5, and again in Table 5.5 (and Figure 5.9b) where the campaigns most likely to suffer directional bias in slope retrievals are highlighted.

7.1.3 Vegetation Height Correction

Two methods (Lee and modified; see Section 6.1.3) were tested as a means of refining waveform LiDAR derived vegetation height (V_H) estimates as a function of slope (derived by the ISM method described in Section 5.2). Refinements were applied to waveforms retained by a stringent set of filters (control dataset; see Section 6.1.1), such to reduce discrepancies between GLAS and high resolution airborne LiDAR measures of V_H.

Key vegetation parameters such as timber volume and biomass are often derived using V_H, where inaccurate assessment of V_H yields inaccurate child parameters. V_H is often derived directly from waveform LiDAR waveforms, however where slope is present waveform geometries can become distorted dictating that V_H retrieval becomes more difficult. Refinement of V_H for slope takes steps in reducing uncertainties in such parameters, which are of paramount importance in the assessment of forest productivity and carbon sequestration rates (Allouis et al., 2012).

Of the two correction techniques applied in this study, the Lee correction was not generally applicable as it made some assumptions that were unrealistic (such as vegetation occurring at the footprint perimeter). The modified correction was developed to be less sensitive to stringent assumptions as noted in the Lee method, offering an improved method for refining V_H for slope. This was reflected in V_H results, which became more representative of airborne LiDAR equivalent data than Lee corrected V_H estimates. However, clear conclusions could not be drawn from the near global correction application, partially due to the choice of data against which V_H estimates were compared (see Section 6.4).

7.2 Site Complexity

Performance of models developed and tested within this study is variable, dependent on specific terrain and vegetation characteristics, namely terrain slope angle and roughness, and vegetation structural parameters such as: number of canopy layers, tree density, and leaf area index (LAI), etc. The diversity of sites (see Section 3.1) employed here allowed an assessment of method performance as a function of environmental variability, highlighting method pros and cons where applicable.

Assessment has been carried out for shadow identification, slope retrieval, and vegetation height corrections.

7.2.1 Conditions for Shadowing

Results suggest that the occurrence of shadows in LiDAR waveforms is related to terrain and vegetation characteristics (see Section 4.2.1). Of the sites included in this study, Tumbarumba is the only site suitable for identifying shadows in waveforms due to its complex terrain (up to $\sim 30°$; see Section 3.1).

Eucalyptus dominates at Tumbarumba, which exhibit leaves that hang vertically towards the ground, thus effectively reducing the leaf area index (LAI) found throughout the forest. Hence by the definition of LAI (and the assumptions listed above), this dictates that this vegetation offers a lesser area capable of obstructing incident photons, thus reducing the magnitude of the shadowing effect. As a consequence of the proved existence of shadows under these conditions, it is not unreasonable to expect this phenomenon to occur with greater severity for areas where vegetation canopies exhibit greater densities (and similar terrain conditions). However, it is not possible to test this assumption at present due to lack of applicable data.

7.2.2 Terrain and Vegetation Parameter Sensitivity

Specific environmental characteristics can induce uncertainty in results (supported by Tables 5.2 and 5.3), understanding these causes can help refine future slope retrieval. Site characteristics encountered within this study are discussed with re-

gards to SSM and ISM predictions from GLAS waveforms.

Predictions at **SOJP** are least representative of airborne LiDAR equivalent reference data. This is not explained by site specific terrain or vegetation characteristics, but by the small sample size and small slope range (between $0° - 7°$) present. This is illustrated for SSM and ISM in Figures 9.3b and 9.4b respectively.

More representative predictions for SSM and ISM were found at **SOBS**, partially explained by an increased sample size and data range, however, increased variability in slope predictions was noted (particularly for ISM). This is caused by the relatively thick canopy at this site, where tree density is the highest of all sites (Section 3.1). This reduces the capability of incident light penetrating the vegetation layer to the ground, hence the accuracy of predictions can be somewhat compromised. This is also noted in Chapter 6, where vegetation height corrections (as a function of slope) appear to be sporadic for dense tropical forests.

Predictions for SSM and ISM at **SOA** were unsuccessful as adequate quality checks with regards to Gaussian fitting (see Section 5.1.1) were not met. This is an expected consequence of the presence of a dense multilayer canopy (understorey) at this site, which distorts the returned waveform ground component due to signal mixing (explained in Section 5.7.1). This makes the resulting waveform difficult to interpret, as it represents two physical features as one, hence slope estimates here would not be realistic even if predictions were able to be made.

It is noted that many forests exhibit similar understorey conditions, at present the discussed methods are unable to adequately assess resulting waveforms recorded for such conditions. However, with varying terrain, and other vegetation conditions, slope predictions can be made for the majority of waveforms (provided a methodologically valid ground peak is found). This is reflected in results from Tharandt where dense understorey is noted in places.

Loobos exhibits generally flat topography (with some minor undulation), with a relatively open canopy in places (Section 3.1), explaining the majority of SSM and ISM results between $0° - 12°$ (Appendices 9.9 and 9.10, respectively). Model performance parameter quality varies between SSM and ISM results, however, both perform to a good standard, with variability in predictions originating from the minor surface undulation (Section 3.1). This suggests that for (large footprint) GLAS waveforms such conditions are able to be over-looked when employing these

slope retrieval methods.

For Loobos and **Tharandt**, SSM model performance parameters are similar, whereas equivalent ISM measures diverge. The spread of data (for SSM and ISM respectively) at Tharandt (Appendices 9.9 and 9.10, respectively) differ over a range of $0° - 25°$. This is a suspected effect of site specific tuning (for SSM) and the interpretation of relatively complex waveform geometries (for ISM). Such complexities are suspected from the combination of the sloped terrain and dense understorey present at this site (Section 3.1).

Norunda GLAS slope predictions are least representative of airborne LiDAR equivalent slope for the European sites for SSM and ISM. This is partially explained by the small predictive range of results ($0° - 8°$) at this predominantly flat site. However, terrain roughness complexities (boulders; Section 3.1) are expected to contribute to this relatively poor performance.

Tumbarumba was best represented (by GLAS slope predictions) of all sites for SSM and ISM, partially due to having the greatest amount of data for analysis. Tumbarumba has the largest slope range of all the sites, ranging between $0° - 27°$ (LUT data limitation; see Section 5.1) for SSM, and $0° - 50°$ for ISM, the latter performing consistently at these high slopes. The variability present in predictions are from the terrain complexities that exist here, inducing signal mixing, resulting in difficult to interpret waveforms.

Slope estimates from ISM (see Section 5.2) were employed to correct (control) V_H estimates, hence the subtleties that disrupt slope retrieval from LiDAR waveforms have a direct influence on the accuracy of corrections applied via the Lee (Equation 6.3) and modified (Equation 6.7) correction techniques. This is reflected particularly in tropical regions where multi-layer, dense canopies induce uncertainty in slope predictions, which dictates that V_H corrections as a function of slope are also unreliable.

7.3 Research Implications

Implications of the research presented in this thesis are discussed here, with respect to furthering knowledge of waveform LiDAR behaviour, and reducing uncertainty in waveform derived vegetation parameters. Implications are discussed specifically

with regards to the retrieval of previously unobtainable (from waveform LiDAR) terrain and vegetation parameters, and the refinement of existing vegetation parameters.

7.3.1 Terrain Parameter Retrieval

Provided a waveform ground component can be accurately identified, **ground slope angle** (θ_w) can be obtained (assuming model quality checks are met). Slope angle knowledge currently acts as a means of filtering waveform data of high slopes, as with Simard et al. (2011) and Los et al. (2012). Additionally this knowledge allows corrections of vegetation parameters such as height (V_H), discussed in Section 7.3.3.

Predictions of θ_w from the two methods developed in this study are improved relative to those from previous studies (namely Simard et al., 2011 and Los et al., 2012), where SRTM 90 m tiles are employed to infer slope. Due to resolution discrepancies between GLAS LiDAR footprints and the SRTM tiles, slope representation across footprints are often unrealistic, as SRTM tiles represent a far greater area than corresponding footprints causing slope underestimation. This can be particularly problematic when correcting V_H for slope. By definition θ_w represents the slope for the exact same area of LiDAR footprints, negating any discrepancies that would otherwise be introduced by spatially mismatching datasets. Hence, this allows a more accurate representation of slope within any footprint, the knowledge of which will aid the understanding of the influence of terrain on waveform returns.

Moreover, more accurate slope information will allow studies that employ such slope filtering techniques to yield more accurate final results (vegetation height information etc.) by excluding slope induced spurious data. Additionally, slopes derived from SRTM may allow footprints with terrain slope greater than an imposed threshold beyond the filtering procedures, due this data underestimating slopes. For example, for a true slope of 12°, SRTM derived slope can be less than a set threshold of 10° due to the spatial discrepancy between the area/footprint of interest and the area represented by the SRTM data. For the developed methods presented in Chapter 5 such spatial discrepancies are negated.

Predictions of θ_w unlike those from SRTM, allows slope retrieval beyond $\pm 60°$ latitude (Rodriguez et al., 2005; Jarvis et al., 2008). The region beyond 60° N contains a considerable portion ($\sim$ 30%) of the boreal forest, and tundra regions. Understanding the terrain of which is important in deciphering waveform behaviour as a function of these land types. Furthermore, by θ_w predictions, V_H derivations and corrections can be applied in this region with greater confidence (discussed in Section 7.3.3).

Furthermore, considering shadow presence (which has been found to occur most commonly over sloped terrain) slope predictions could be incorrect. This is possible in cases where phantom peaks (caused by shadow presence) are identified, as ground returns from sloped terrain are often difficult to interpret. Hence accounting for this change in waveform geometry from shadows will lead to more accurate slope estimations. Hence, in future (if possible) shadow affected waveforms should be corrected before slope retrieval is attempted.

Providing more accurate slope information than that currently available allows more accurate vegetation height retrievals to be made. This will contribute to the work of multiple government agencies striving to accurately account biomass, and carbon stocks within national/continental/global forests. This will allow for more accurate monitoring of forest vegetation, in addition to the potential of more sustainable forest management programs being created.

7.3.2 Vegetation Parameter Retrieval

The presence of shadows in returned LiDAR waveforms present opportunities for extracting unobtainable (from waveform LiDAR by previous methods) vegetation parameters such as within footprint individual tree location/elevation (Section 4.3.1), and individual crown diameters (C_D; Section 4.3.3). Time limitations restricted quantitative investigations of these parameters, however, qualitative reviews of their retrieval are discussed (see Section 4.3). All descriptions should be considered with caution, subject to uncertainties outlined in Chapter 4.

Within footprint **vegetation location/elevation**, if accurately quantified, offers the opportunity of extracting vegetation height (V_H) estimates of individual trees within any footprint (subject to vegetation configuration; e.g. unclustered).

This allows opportunities of investigating sub-footprint resolution vegetation parameters (see Figure 4.8a), potentially allowing a more realistic representation of spatially distributed vertical data, rather than a footprint resolution generalisation. This is particularly pertinent to vegetation information in predictive models. However, it is noted that all estimates of regarding elevation information are subject to source data (DEM) uncertainties.

Vegetation **crown diameter** are expected to be retrievable by the use of shadows in waveform LiDAR. The dimension of shadows cast within any waveform footprint are expected to be relatable to incident light obstructing objects, namely vegetation subcomponents within the crown (assuming a nadir viewing geometry), C_D is retrievable. Figure 4.8c provides a visual example of how such a parameter may be extracted from a shadow afflicted waveform. It is noted that the estimated C_D will be dependent on some thresholds which define the positions at which the shadow is believed to start and end. Such information is useful in (leafy) biomass estimates of forest biomes, and modelling impacts (discussed in Chapter 8).

C_D particularly is used in biosphere modelling approaches, such to validate current knowledge and/or run predictions based on current vegetation conditions. These parameters will add to the current knowledge base of vegetation conditions if obtainable at large scales (beyond regional). This will invariably yield better understanding of the role of these parameters in temporal biosphere evolution and yield more accurate biosphere predictions from models.

The certainty to which C_D can be retrieved is dependent on the severity of shadowing, which is directly dependent on the transmission coefficient (τ), chosen at 0.6 for simplicity in this study. If an unsuitable choice of τ is employed to create a shadow waveform (SW), significant signal attenuation may not be recorded, hence C_D will be irretrievable. Hence, caution must be taken when choosing inherent methodological parameter values. Further work may allow more robust conclusions to be drawn, however, this cannot be stated with certainty from this study alone.

If deemed suitably accurate by the methods discussed, the addition of more accurate V_H will further fortify our knowledge of forest stocks (as discussed in Section 7.3.1). Where the addition of C_D at a global scale would allow the disambiguation of vegetation vertical structure, knowledge of which is key in inferring accurate

assessments of forest biomass. Although, by the presented method, C_D would only be retrievable where shadows are present (not globally), the opportunity to assess the use of such a vegetation parameter retrieved in this way is presented.

7.3.3 Vegetation Parameter Refinement

Vegetation height (V_H) estimates from shadow affected waveforms can be spurious as the identified ground elevation within the waveform is incorrect as a direct consequence of phantom peaks that arise from shadowing. Accurately identifying shadows in waveforms has been suggested to have potential to allow more accurate representations of V_H from such waveforms.

V_H estimates have potential to be improved as a consequence of the removal of shadows in LiDAR waveforms as improved estimates of the ground elevation are possible (see Section 4.3.2). This will be of significant use in reducing uncertainty in maximum within-footprint V_H estimates over relatively high slopes ($\geq 10°$). Removing shadow induced intensity reductions in a waveform (as described in Section 4.3.2; Figure 4.8b) allows the representation of observed physical features only, neglecting optical subtleties (shadows), which are noted as having potential to induce uncertainty in estimates of V_H (see Section 4.2.3).

Additionally, $\mathbf{V_H}$ estimates can be further refined as a function of ground slope angle (θ_w), as demonstrated in Chapter 6. Similarly to the effect of shadows, the developed corrections allow for more accurate assessment of vegetation heights, particularly over slope terrain, but for waveforms with little slope interference also. Additionally, the correction techniques allow more accurate corrections for V_H as there is no spatial discrepancy between data employed for deriving V_H itself, and slope (θ_w), as mentioned in Section 7.3.1.

Also with reference to Los et al. (2012) (and Section 7.3.1), V_H estimates can be directly, accurately obtained beyond 60° N as slope information is obtained from the LiDAR waveform data itself. This was not possible by the method of Los et al. (2012), where limitations were imposed by the external SRTM data.

Explicitly, by the methods described in this study, V_H information can be obtained and corrected between ±86° latitude, a limitation of GLAS range. The importance of V_H information beyond 60° N is accentuated by a large percentage

of the boreal forest inhabiting the region. The neglection of which, with regards to forest inventory and sustainability assessment, cannot be overlooked. Furthermore, with global temperature set to rise over the coming decades (IPCC, 2007), the tree line may extend further north, exacerbating the requirement of Earth surface knowledge (terrain parameters such as slope) in more northerly regions.

7.4 Modelling Contributions

Satellite data holds a unique position in that it allows vegetation parameter retrieval up to the near global scale, providing predictive models with input data. In addition satellite data can allow further understanding of waveform interactions over a vast array of surface types. The contributions are discussed below.

7.4.1 Predictive Models

Predictive models (Dynamic Global Vegetation Models, DGVMs and Land Surface Models, LSMs; see Section 1.2.3) can employ data from this study as input data such to predict environmental evolution over variable time scales. In particular, satellite LiDAR provides direct measurements of the vertical dimension, informing on vertical vegetation structural properties. Used in combination with land cover maps this information allows estimates of biophysical parameters which are often used in model feedback processes to drive evolutionary ecosystem changes. The accuracy of such predictive models and their outputs are correlated to the quality of data inputs, hence the pursuit of input data accuracy is of notable importance. Vertical vegetation parameter information sampled by satellite observations act as input data for predictive models (such as CASA or JULES) from which almost spatially continuous coverage can be obtained by extrapolation over different cartographic classifications of vegetation type and cover. Such coverage allows for a better representation of global land cover maps, illustrating the variability of biophysical parameters present with unique vegetation cover classes.
Up to now, vertical information was either assigned as a constant values (dependent on land class type) or inferred by indirect measurements from passive optical sensors, both of which accommodate large uncertainty. By the methods developed

in this study, uncertainties in (footprint sampled) direct vertical vegetation parameters, such as V_H are reduced by almost 1 m in RMSE (at the study site scale according to Table 6.1). Additionally terrain parameter knowledge (such as slope) is also of benefit as this is known to effect gas exchange processes at the surface (Baldocchi et al., 1988).

7.4.2 Radiative Transfer Models

In addition to DGVMs and LSMs, the results from methods developed in the current study can be used in radiative transfer models to further validate our understanding of waveform interaction with variable surface types. The FLIGHT model was employed in this research as a pathfinder to assess the feasibility of deriving slope information from waveform LiDAR returns. The application of such models can be used to further assess the influence of terrain on the retrieval of accurate terrain and vegetation parameters and their associated uncertainties (North et al., 2010; Rosette et al., 2013). Furthermore, the application of such models under appropriate simulated conditions will aid in furthering knowledge of shadowing effects in LiDAR waveforms. The current study suggests that such effects present in waveforms increase the probability of deriving spurious vegetation parameters, where further modelling work may be able to confirm this.

7.5 Summary: Reducing Uncertainty

The research presented throughout this thesis focussed on better understanding waveform LiDAR returns in conjunction with reducing uncertainty in parameters derived from such data. The presence of shadows in waveforms has been investigated, along with methods developed such to derive slope information from waveforms for filtering purposes, or correcting vegetation height (V_H) estimates. Following previous research (Hancock et al., 2012; Wallace et al., 2012), a method was developed with the view of identifying shadows in GLAS observed waveforms. Results suggest that shadowing typically occurs in waveforms from vegetated surfaces with slopes $\geq 10°$. Noted implications of shadow presence in waveforms were for identifying within footprint individual vegetation elevation, and individ-

ual tree crown diameter (C_D), both of which can be used to infer more accurate vegetation structural information, important in forest inventory and sustainability assessment.

At present the identification of shadows without external 3-dimensional data (as used in Section 4.1.2) is not possible. However, the development of further methods (see Section 8.3) may inform on waveforms that are more likely to suffer with shadow afflictions.

In addition to shadowing effect assessment, methods were developed to retrieve slope information directly from LiDAR waveforms, with the view of employment as a filtering technique or to correct estimates of V_H. Slope information was derived by two methods, where methodological assessment via airborne LiDAR and independent ordnance survey validation processes concluded that the Independent Slope Model (ISM) provided superior representations of slope. Where, assumptions associated with this method are reflected in results, highlighting that the model performs less accurately where complex waveform geometries are encountered, which are noted to be induced by multi-layer canopies, atmospheric effects, and within footprint terrain variability.

Accounting for model assumptions, the ISM is suitable for retrieving slope information from all waveform data sources, not just GLAS, as tested in this study. This information has previously been unobtainable through waveform returns, demonstrating that waveform LiDAR still has great potential in physical parameter retrieval within the Earth Observation scientific community.

ISM Slope information was employed for waveform derived V_H correction by two methods: Lee and modified (see Section 6.1.3). Corrections were applied to a control V_H dataset formed from GLAS waveforms filtered according to Section 6.1.1, where V_H was calculated by the method of Rosette et al. (2008) for the filter retained waveforms.

The control, Lee, and modified V_H datasets were analysed with respect to airborne and GLAS LiDAR equivalent information for each study site, and near globally, respectively. Results suggested the modified dataset provides a better representation of airborne LiDAR derived V_H, where a (near) global mean difference between the modified and control was noted at -1.83 m. All near global V_H datasets were compared with normalised difference vegetation index (NDVI) information as a

function of increasing vegetation cover fraction (VCF). Changes in correlation and RMSE between V_H estimates and NDVI with increasing VCF suggested that the control dataset performs best for less dense/shorter vegetation, whereas the modified datasets is best for more dense/taller vegetation. However, uncertainty in these result are noted.

In summary, the three topics investigated in this Thesis link to a common theme with the prospect of refining biomass, and timber volumes from local to near global scales. The refinement of such parameters will take steps towards narrowing discrepancies in the carbon cycle, specifically with regards to carbon allocation within vegetation. Additionally, such refinements can possibly be employed in reducing uncertainties in the water cycle also. With particular focus on vegetation water use efficiency etc.

Information from each of the presented research topics in conjunction with radiative transfer models allows the better understanding of waveform LiDAR behaviour for complex vegetation and terrain, providing rationale for such models. Combining these highly variable characteristics for in-depth analysis has allowed the identification of shadow effects in waveform returns, a phenomena that is typically overlooked in waveform analysis. Additionally, previously unobtainable (by waveform LiDAR) terrain and (potential) vegetation parameters can be obtained reducing the number of unquantified environmental variables across the globe. The combination of these findings allows for a more concise representation of our global terrestrial environment, with reduced uncertainties.

Finally, this study provided results that clearly demonstrate the potential of waveform LiDAR as a vegetation imaging tool, demonstrating that further potential remains in its investigation. The use of GLAS was somewhat chosen for its capability of providing almost global coverage of waveform data; coverage to this extent is exceptionally useful for inferring local to global changes in the terrestrial biosphere. On this basis, the scientific community would certainly benefit from investment in an additional waveform LiDAR instrument capable of global imaging of the Earth's biosphere.

Chapter 8

Conclusions

This chapter provides a summary of the research undertaken in this project, and underpins the relevance of key findings to the fields of quantitative forestry applications and modelling prospects. Conclusions are drawn for the developed techniques within this study, their applications, and prospective potential. Future research directions are suggested with respect to the limitations of the current developed techniques, and alterations considered such to accommodate proposed future missions and highlight the demand for future waveform sensors.
The thesis concludes by summarising how project aims and objectives have been met, in addition to highlighting how increased understanding of waveform LiDAR behaviour aids in reducing uncertainty in derived terrestrial parameters.

8.1 Research Summary and Contribution

This research has highlighted the already existing contribution of LiDAR technologies to Earth Observation applications, and explored it further. The potential of waveform LiDAR systems for the derivation of further (previously unobtainable by LiDAR) terrestrial parameters, and their refinement has been highlighted. A summary of each chapter within the thesis is given with specific scientific contributions highlighted.
Chapter 1 provides an introduction to the proposed research, illustrating scientific rationale and research context. **Chapter 2** reviews available literature, highlight-

ing previous advancements in LiDAR technologies and their uses for retrieving terrestrial biophysical parameters at different spatial scales. Gaps in existing research are highlighted and discussed. **Chapter 3** describes the study sites, field work, and appropriate data employed in this study. Data processing methods are described for airborne and spaceborne LiDAR data, in addition to specific methods used relatively frequently. Contributory research is presented in **Chapters 4–6** (and discussed in **Chapter 7**), the principle contributions of which are discussed in the following paragraphs.

Chapter 4 presents findings on a method (see Section 4.1) developed to identify the presence of shadows in physically observed, large footprint GLAS waveform LiDAR returns following shadows noted in previous studies (Hancock et al., 2012; Wallace et al., 2012). Applying this method at the Tumbarumba site, approximately 15% of all 565 analysed waveforms were noted to suffer with shadow afflictions capable of distorting vegetation parameters if derived from such waveforms. Of these waveforms approximately 94% occur where terrain slope is $\geq 10°$. Shadows in waveforms have potential to allow the derivation of vegetation parameters that were unobtainable from waveform LiDAR by previous methods, these are: individual within footprint vegetation elevation, and crown diameter.

It was concluded that shadows do occur in physically observed, large footprint GLAS waveforms, originating from vegetation subcomponents (shoots, branches, leaves, etc.) disrupting incident laser light. Shadows identified here exhibit sufficient magnitude such that they can effect the accuracy to which vegetation parameters can be obtained from such waveforms.

Chapter 5 presents findings with regards to the development and testing of two new methods with the aim of deriving ground slope angle (θ_w) information from waveform LiDAR returns. Both methods made slope predictions from GLAS waveforms, which were compared to equivalent airborne LiDAR derived slopes. The first method (Slope Screening Model; SSM) provided site fitted results indicating a successful representation of true slope (see Table 5.2). Similarly the second method (Independent Slope Model; ISM) provided successful results, even without site fitted results (see Table 5.3). Each method was independently tested (no site fitted results) against ordnance survey slope information, where it was concluded that the ISM is the better of the two developed methods for slope retrieval from

LiDAR waveforms (see Table 5.4).

It was concluded that slope information can be obtained from waveform LiDAR returns, the reliability of which is dependent on environmental variabilities, such as: terrain and vegetation structural characteristics.

The ISM for slope retrieval developed in Chapter 5 was applied in **Chapter 6** to correct vegetation height (V_H, as derived by Rosette et al., 2008) for slope. Two methods were employed for this purpose, the first was developed by Lee et al. (2011), and the second is a further development of this, proposed during this study. Both correction methods were applied to a control dataset (established from GLAS waveforms retained after the application of filters described in Section 6.1.1) for data at each of the study sites, and GLAS campaign L1A independently.

At the site level the modified V_H dataset shows better agreement with airborne LiDAR derived equivalent measures than both the control and Lee datasets. At the global scale V_H datasets were compared with spatially corresponding MODIS normalised difference vegetation index (NDVI) information as a function of increasing vegetation cover fraction (VCF), where correlation is expected (as described in Section 6.4). Here correlation and RMSE measures suggest that the control V_H data correlate better with NDVI for less dense/shorter vegetation, whereas the modified V_H data correlate better with more dense/taller vegetation.

It was concluded that GLAS derived V_H estimates can be corrected for slope with reasonable success based on site level results. The Lee correction method was found to be less generally applicable, whereas the modified method provided more general applicability. No clear conclusion can be made with regards to which V_H dataset is most representative of reality from NDVI comparisons at the global scale. This is due to this comparison comprising of different parameters, which are obtained by active and passive remote sensing instruments respectively.

8.2 Prospects of Future Missions

The GLAS instrument provided the first opportunity to directly measure global topographic and above-surface elevation data, and vegetation information over an extended time scale (2003–2009; see Section 3.4). GLAS afforded many opportunities with regards to vegetation imaging despite not being optimally optically

configured for use with vegetation. From before the last data acquisition of GLAS, the remote sensing community have petitioned for a similar mission be configured such to accommodate vegetation applications only (Rosette et al., 2013).
DESDynI (see Section 2.7.1), was such a mission, set to be capable of providing the most dense coverage of the Earth's surface of any spaceborne LiDAR system. Rosette et al. (2013) have shown the potential of such a LiDAR system for vegetation imaging, further endorsing the need for such a system.
At present the possibility of two future missions exists: ICESat-2, and LIST (see Section 2.7.3; provided the LiDAR portion of DESDynI development is not reinitiated). ICESat-2 is the follow up mission of ICESat, primarily focussed on cryosphere applications, set to use (currently under test) micro pulse technologies. LIST is a specific vegetation monitoring mission due for launch before 2020 that proposes continuous global LiDAR coverage. This extreme density of coverage with medium size footprint is likely to offer surface evaluations that resemble airborne LiDAR coverage which should aid in further reducing uncertainties associated with assumptions made for LiDAR profiling (as noted in this study), in addition to allowing a finer resolution depiction of the natural environment.
Findings of this study suggests that to reduce the influence of environmental variability in LiDAR measurements, all future missions should consider smaller footprint dimensions. Also all missions should be capable of reaching sufficient latitudes, such to encompass all major forest regions. Ideally future missions will offer potential for the development of new techniques from new technologies such to retrieve relevant environmental parameters that are irretrievable by current technology/and or methods, thus allowing a more in-depth understanding of the terrestrial biosphere and beyond.

8.3 Implications for Future Work

As a direct result of the findings of this study, further investigations can be pursued with the view of retrieving new vegetation parameters (from a large footprint system), and reducing uncertainties in measured parameters. Such investigations are discussed here.

8.3.1 Independent Shadow Identification

It has been noted that shadows can disrupt waveform geometries and shapes, such that spurious V_H estimates can be identified, and/or previously unobtainable (by waveform LiDAR) vegetation parameters can be retrieved. Such results are only possible in the presence of 3-dimensional supplementary data (as discussed in Chapter 4), which is typically unavailable upwards of regional scales. For results beyond regional scales to be achieved, a method of independently identifying waveforms where shadows are present is required.

Such a method for shadows at beyond regional scales is suggested by employing multivariate regression models, linking shadows with slope, and vegetation species information. This is theorised to inform on the likelihood that a waveform is affected by shadowing based on terrain and vegetation characteristics. Such an investigation is recognised to require appropriate 3-dimensional data from a multitude of differing surface and vegetation types to provide any verifiable results.

8.3.2 GLAS Geolocation Accuracy

The geolocation accuracy of GLAS waveforms has been questioned (Hopkinson, 2013, personal communication). It was suggested that the geolocational accuracy of footprints are dependent on their geographical position on the globe.

Uncertainties in positional accuracy can lead to footprint measurements of objects from different locations than are expected. Hence, information may be misinterpreted and waveform interactions appear less well understood for the expected target.

A proposed investigation into such claims for forests can be performed by comparing spatially coherent GLAS and airborne LiDAR footprints that vary in position by the associated GLAS positional uncertainty limits. Analysing for correlation variations between multiple parameters (such as V_H and slope) within positional variations can inform on possible positional uncertainty issues.

8.3.3 Implications

If results were obtained from the methods outlined in the previous two sections, further uncertainty reductions can be realised in the presented "modified" global V_H results.

Global knowledge of shadows in LiDAR waveforms would allow more accurate estimates of V_H by the refined identification of the ground elevation (see Section 4.3.2) from waveforms that suffer shadow-induced phantom peaks. Further refinements can be made by applying waveform ground peak refinements before obtaining and correcting for slope by the appropriate methods (see Sections 5.2 and 6.1.3, respectively). This is important as refining the waveform ground peak will change its geometry, on which slope derivations, and hence correction, directly depend.

Accurate assessments of GLAS positional accuracies will allow more accurate spatial assessments of vegetation vertical profiles. Inconsistency, if present, between waveform footprints makes derived results less useful especially if used in such a context where the spatial location of vertical measurements is important (such is the case if used for finding the source of flux tower measurements).

8.4 Concluding Remarks

The spaceborne LiDAR mission ICESat/GLAS was the first mission of its kind, providing direct measurements of the Earth's vertical dimension. This system provided an opportunity to assess the potential of continuous waveform satellite LiDAR profiling for vegetation imaging purposes. Mission success has stimulated the scientific community and governmental bodies to partition for future missions dedicated to vegetation analysis to reassess, and further support information gathered by this innovative satellite mission.

This study was pursued with the focus of improving waveform interpretation understanding, and reducing uncertainty in waveform derived parameters. To this end, the large footprint GLAS system, previously on board ICESat was employed for investigative purposes, allowing near global LiDAR coverage from a multitude of different land types. Research was investigated in three chapters, each of which are summarised and concluded here.

Research found that shadow presence was noted in approximately 15% of 565 analysed GLAS waveforms from the Tumbarumba study site in New South Wales, Australia. It was concluded that the source of this effect was coupled to slope angle ($\geq 10°$), and vegetation presence (density dependent). In particular the frequency of shadows increased with increasing slope up to $\sim 30°$, beyond which at least half of the analysed waveforms suffered shadow afflictions. It is theorised that the magnitude of shadow effects will increase as a function of vegetation canopy density (particularly higher LAI), as these vegetation components provide a greater mass to cast shadows by intercepting incident light.

The implications of identifying shadows in waveforms were noted, where methods for the identification of spurious V_H estimates, individual within footprint vegetation elevation, and individual crown diameters are described, providing appropriate supplementary data are available. Without such data, independent identification of shadow effects is not yet possible.

An additional investigation was pursued such to develop and evaluate methods for retrieving slope information from LiDAR waveforms. Slope information was estimated from FLIGHT and GLAS waveforms, where results were validated against FLIGHT model inputs, and airborne LiDAR equivalent information (validated by in-situ field data) respectively. Of the two developed methods, the Independent Slope Model was concluded to provide the most consistently accurate results: $R^2 = 0.67$ and RMSE $= 5.60°$ for airborne LiDAR equivalent slope comparison, and $R^2 = 0.55$ and RMSE $= 6.27°$ for Ordnance Survey equivalent slope comparison. RMSE values noted here are equivalent to those found for V_H measures from other studies (Rosette et al., 2008; Los et al., 2012).

In conjunction with slope retrieval, refinements were made to GLAS waveform estimates of V_H as a function of slope itself. For this purpose, two methods were tested, that of Lee et al., 2011 (Lee), and the other a further development of this method (modified). These corrections were applied to a control V_H dataset for all study sites (site scale) and to all waveforms in the GLAS L1A campaign (near global scale).

At the site scale results concluded that the modified correction method, yielded V_H information that was most representative of the airborne LiDAR equivalent data ($R^2 = 0.65$, RMSE $= 6.89$ m). At the near global scale control, Lee, and Modified

V_H estimates were compared with MODIS normalised difference vegetation index (NDVI) data as a function of increasing vegetation cover fraction (VCF), where results were less conclusive. However, the Lee V_H distribution resulted in the poorest correlation consistently. Comparison results suggested that the control dataset provided the most accurate V_H estimates for shorter/more dense vegetation, and the modified V_H estimates most accurate for taller/more dense vegetation.

Furthermore, by these methods of slope and parameter refinement, accurate, direct waveform measures of V_H are available beyond 60° N, which were previously unavailable as slope information (used as data filter) beyond this was not quantifiable (as SRTM did not extend here; see Section 6.5). The requirement of data beyond this point is paramount, as a large portion of the worlds second largest forest (boreal) exists here.

Results from this study have demonstrated the potential for vegetation parameter refinement (V_H), and the identification of terrain (slope) and vegetation (vegetation elevation and crown diameter) parameters using continuous waveform LiDAR. Such information fulfils the aims and objectives set out in this project and fills research gaps through providing additional information that allows the understanding of waveform interpretation, and reduces uncertainties in waveform derived parameters. These improvements can be propagated to estimates of other biophysical parameters such as timber volume, biomass, and carbon stocks estimates, in addition to reducing uncertainty in DGVM and LSM outputs, by constraining predictions through more precise input data. Future missions, better suited to vegetation analysis will further exploit the potential of obtaining increasingly refined parameters. Coupled with predictive modelling, this reduces the variability of scenarios through which the biosphere evolves, aiding in providing a sustainable strategy for forest inventory management and carbon sequestration.

Furthermore, this work informs on the continued use of waveform LiDAR data from a satellite mission that was decommissioned almost 5 years ago, in late 2009. In particular this work has demonstrated the use of large footprint satellite LiDAR in retrieving vegetation parameters, in addition to retrieving parameters that have previously been unobtainable from such data. Moreover, such parameters were retrieved with considerable success at a wide scale, indicating the continuing uses of such satellite LiDAR data. The addition of a similar mission in the current

day would allow for further observations of vegetation, allowing further potential for methodological developments to be made such to better understand global biosphere behavior. Such a mission, coupled with current ICEsat/GLAS observations would constitute the longest time line of near global LiDAR observations of vegetation that have ever been previously achieved.

Chapter 9

Appendices

9.1 Field Data Results

Results are shown for six of the seven study sites employed throughout this study. Figure 9.1 illustrates the accuracy of the collected airborne LiDAR for each site, with little difference in accuracy across each site. This accuracy is also supported by Table 9.1, which summarises appropriate performance parameters for all sites when data are combined.

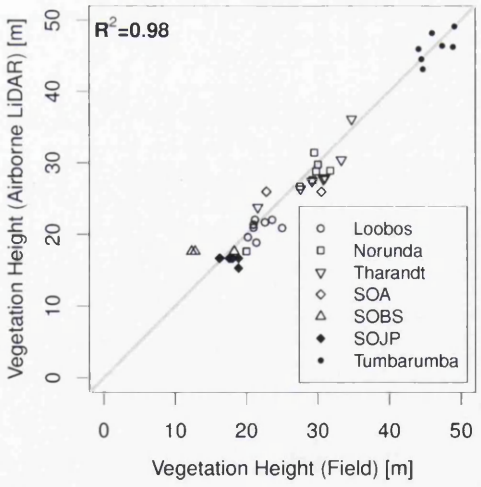


Figure 9.1: Validation of airborne LiDAR through accuracy of vegetation height estimates with respect to equivalent data collected in the field.

Table 9.1: Summary of appropriate performance parameters for the comparison of field, and airborne LiDAR derived vegetation height estimates. Note: n=sample size, R=Pearson correlation coefficient, p_R=significance of R, D=Kolmogorov-Smirnov statistic, F_2=fraction of predictions within a factor of 2 of observed values, F_B=Fractional bias, ε_{nme}=normalised mean error, and RMSE=root mean square error; here airborne data are the predictions, whereas field data are observations.

Parameter	Value
n	42
R	0.98
p_R	$< 10^{-10}$
D	0.16
F_2	0.89
F_B	-0.02
ε_{nme}	0.06
RMSE [m]	2.35

A sample size of 42 data points is relatively small, however, even with such a small sample adequate results are still obtained. The small sample may explain some variability in some parameters, such as D, F_2, and RMSE. The resulting performance parameters are encouraging, suggesting that the airborne LiDAR is capable of representing "true" vegetation parameter values appropriately.

9.2 Pearson Correlation Coefficient

Pearson's correlation coefficient (R) is a linear measure of how well two variables (x and y) are related. Results range from ± 1, where the greater the value of R, the more correlated x and y are. Equation 9.1 expresses R mathematically.

$$R = \frac{N(\Sigma xy) - \Sigma x \Sigma y}{\left(N\Sigma x^2 - (\Sigma x)^2\right)\left(N\Sigma y^2 - (\Sigma y)^2\right)} \tag{9.1}$$

9.3 D-statistic

Highlighted in Section 5.4.1, the D-statistic is a measure of the difference between the cumulative fraction of observed ($F(x)$) and predicted ($G(x)$) datasets, hence a

smaller value of D represents a better fit of the predicted to the observed dataset. The D-statistic is defined as the maximum difference between F(x) and G(x), summarised in Equation 9.2; for further information regarding the KS-test see Massey Jr (1951).

$$D = \max|F(x) - G(x)| \tag{9.2}$$

9.4 Factor of Two

The factor of two test (F_2) is a measure of how often predicted (X_{mod}) data fall within a factor of two of observed (X_{obs}) data. This measure requires the number of observations (N), in each dataset to be the same. Equation 9.3 expresses F_2 mathematically.

$$F_2 = \frac{N_{|X_{obs}-X_{mod}|<X_{obs}/2}}{N_{X_{obs}}} \tag{9.3}$$

9.5 Fractional Bias

The fractional bias (F_B) is a measure of the mean difference between an observed (X_{obs}) and predicted (X_{mod}) dataset, expressed as a fraction, rather than as an offset between the two datasets (known as the bias). F_B (Equation 9.4) is a normalised measure of model performance, allowing easy evaluation of how close X_{mod} measurements are to X_{obs}. Note: $\bar{X}$ represents the mean of X.

$$F_B = \frac{2(\bar{X}_{mod} - \bar{X}_{obs})}{(\bar{X}_{mod} + \bar{X}_{obs})} \tag{9.4}$$

9.6 Normalised Mean Error

The normalised mean error (ε_{nme}) is a measure of the mean error fraction present between observed (X_{obs}) and predicted (X_{mod}) datasets. ε_{nme} is expressed as a fraction of the sum of the difference between X_{mod} and X_{obs} and the sum of X_{obs} (Equation 9.5).

$$\varepsilon_{nme} = \frac{\Sigma|X_{mod} - X_{obs}|}{\Sigma X_{obs}} \tag{9.5}$$

9.7 Root Mean Squared Error

The Root Mean Squared Error (RMSE) informs on the differences between observed (X_{obs}) and predicted (X_{mod}) values. The RMSE is the square root of the quotient of the summation of the square of the differences between the observed and predicted values and the number of observations (Equation 9.6). This is not a unitless parameter, instead this value takes the units of X_{obs} and X_{mod}, which are always identical.

$$\text{RMSE} = \sqrt{\frac{\Sigma \left(X_{obs} - X_{mod}\right)^2}{N_{X_{obs}}}} \tag{9.6}$$

9.8 SSM Interpretation Subtleties

Subtle changes in waveform ground component FWHM widths effect the integrity of the SSM LUT relationship, and hence predictions also. Three examples of effects on FWHM widths are discussed in the following.

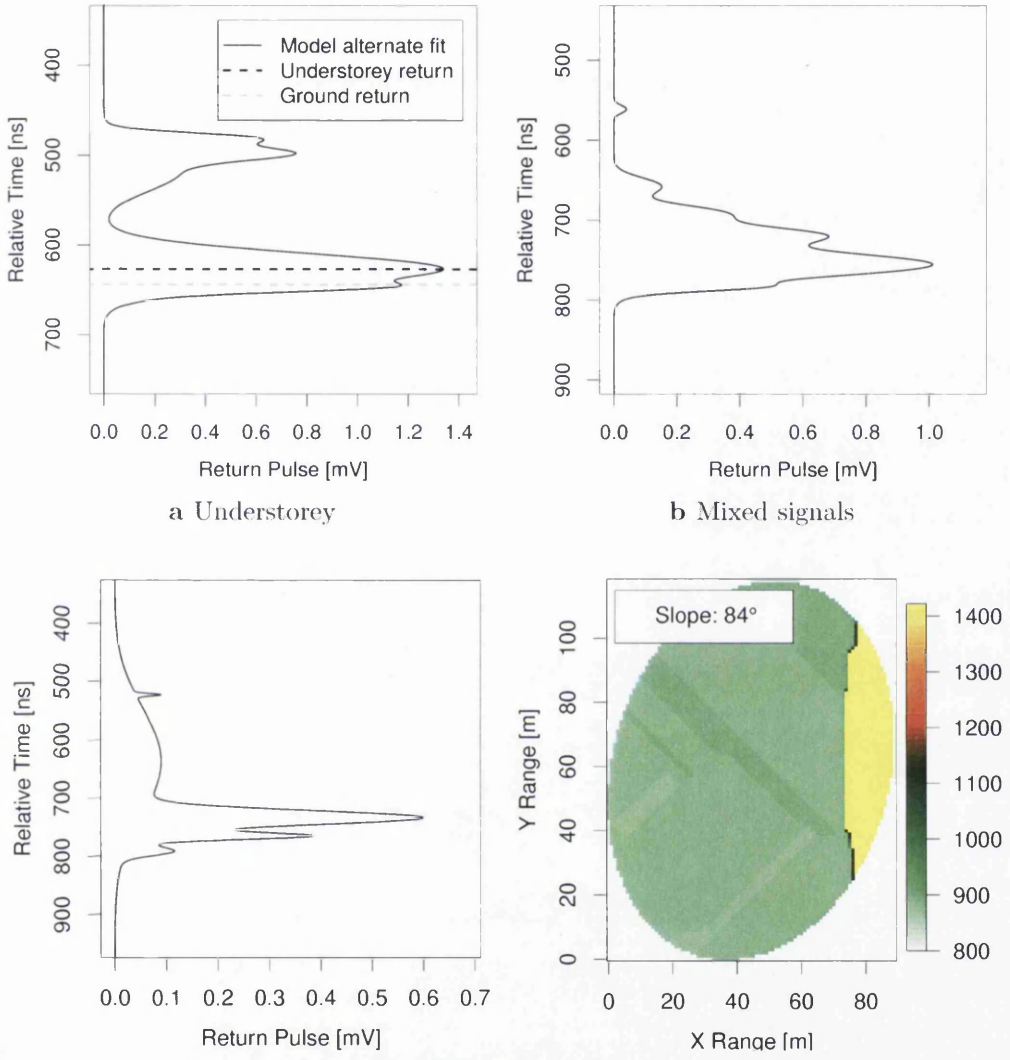
Understorey is short vegetation, that in waveform returns would appear as a peak, however, when the elevation difference between the understorey top reflection and the ground is small, signal mixing can occur (Figure 9.2a). This occurs because the vertical range of vegetation present in the understorey typically (over the vast area illuminated by GLAS footprints) ranges from the mean ground elevation up to the undertorey maximum height. Hence clearly distinguishable peaks for the ground and understorey and not visible, and appear mixed.

For similar reasons, sloped terrain can cause this effect also, however, this is not caused by understorey but rather by vegetation and ground returns occurring at similar elevations. This can lead to waveforms that have no clearly distinguishable vegetation or ground component(s). These two components appear mixed, as in Figure 9.2b, a waveform retrieved over a slope of 19°.

Incorrect interpretation of the external slope data (airborne data in this case) can cause a poor relationship between the waveform ground component width and slope. This is because the slope between the minimum and maximum elevation values of the airborne DEM is assumed to be linear. This assumption introduces some uncertainty in results, however, it is deemed acceptable in non extreme cases (within footprint elevation difference $\leq$ 25 m). In extreme cases, (elevation difference $\geq$ 25 m; Figure 9.2c right panel), this assumption of linearity breaks down, resulting in small ground component widths (Figure 9.2c left panel) corresponding to large values of slope. This is a spurious result, as the within footprint vertical elevation difference (represented by ground component width) is strongly related to (linear) slope, by:

$$y = mx + c \tag{9.7}$$

Where m is the gradient, c the y-intercept, x represents slope, and y represents the within footprint vertical elevation difference.



a Understorey

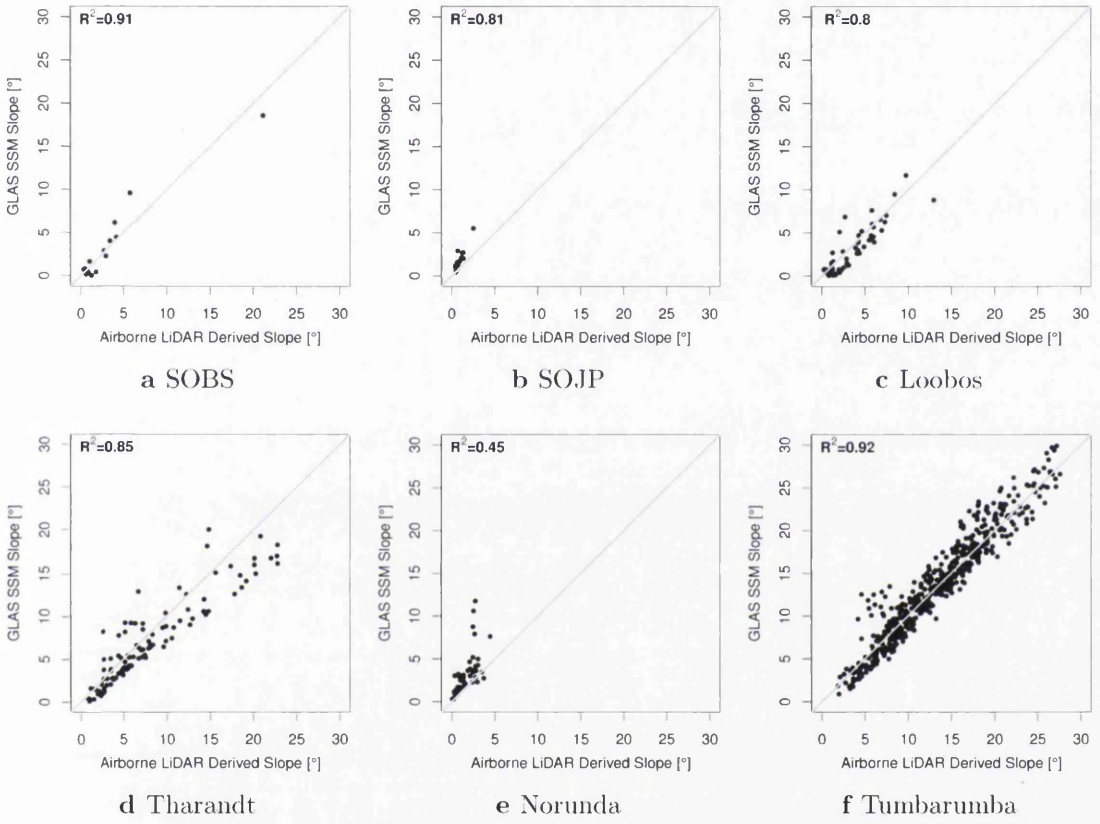
b Mixed signals

c Severe airborne slope

Figure 9.2: Illustrations of waveform ground component FWHM subtleties (for SSM) with regards to (a) understorey presence, (b) vegetation and ground signal mixing due to sloped terrain, and (c) airborne LiDAR over estimation of slope due to extreme elevation differences from a cliff edge.

9.9 SSM Site Results

Figure 9.3 illustrates the comparison of SSM slope predictions and airborne LiDAR derived slope values for each site employed throughout this study. SOA data is not present as no valid predictions where made at this site.



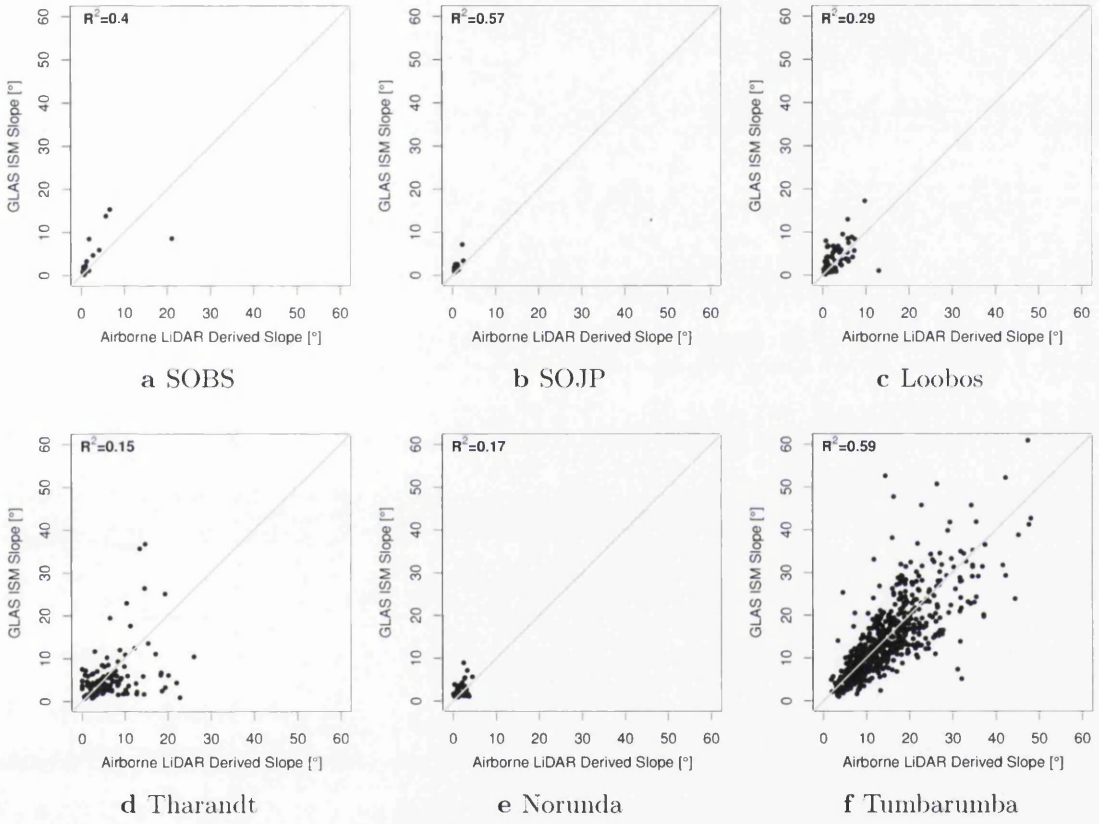
a SOBS b SOJP c Loobos

d Tharandt e Norunda f Tumbarumba

Figure 9.3: Comparison of (site tuned) SSM slope predictions with respect to airborne LiDAR derived slope information for each study site employed throughout this study.

9.10 ISM Site Results

Figure 9.4 illustrates the comparison of ISM slope predictions and airborne LiDAR derived slope values for each site employed throughout this study. SOA data is not present as no valid predictions where made at this site.



a SOBS b SOJP c Loobos

d Tharandt e Norunda f Tumbarumba

Figure 9.4: Comparison of ISM slope predictions with respect to airborne LiDAR derived slope information for each study site employed throughout this study.

9.11 Derivation of ISM Uncertainty

Employing the commonly used Taylor uncertainty analysis formula (Equation 9.8; example with function F(x,y,z)), applying to ΔE and Δs (from Section 5.9) initially, $\Delta\theta$ can then be calculated with further propagation.

$$\Delta F = \sqrt{\Delta x^2 \left(\frac{\partial F}{\partial x}\right)^2 + \Delta y^2 \left(\frac{\partial F}{\partial y}\right)^2 + \Delta z^2 \left(\frac{\partial F}{\partial z}\right)^2} \tag{9.8}$$

Starting with the calculation of ΔE, considering how the within footprint elevation difference (E) is formed (equivalent to $E_{95} - E_{05}$ in Equation 3.5), and substituting appropriate variables into Equation 9.8, ΔE can be expressed as:

$$\Delta E = \sqrt{\Delta E_{95}^2 \left(\frac{\partial E}{\partial E_{95}}\right)^2 + \Delta E_{05}^2 \left(\frac{\partial E}{\partial E_{05}}\right)^2} \tag{9.9}$$

Here, $\frac{\partial E}{\partial E_{95}} = \frac{\partial E}{\partial E_{05}} = 1$. Considering Δn and ΔR from the definitions of E_{95} and E_{05}, the two constituent uncertainties are identical for both ΔE_{95} and ΔE_{05} i.e. $\Delta E_{95} = \Delta E_{05}$, allows ΔE to be expressed as:

$$\Delta E = \sqrt{2E_{95}^2} \tag{9.10}$$

However, substituting for Δn and ΔR, the uncertainties associated with ΔE_{95}, according to Equation 9.8 take the form:

$$\Delta E_{95} = \sqrt{\Delta n^2 + \Delta R^2} \tag{9.11}$$

Hence, substituting Equation 9.11 in to Equation 9.10, ΔE is given by:

$$\Delta E = \sqrt{2\left(\Delta n^2 + \Delta R^2\right)} \tag{9.12}$$

Now considering Δs, the uncertainty associated with the footprint semi-minor axis is required. Considering the form of how this variable is calculated (Equation 5.6),

and substituting in to Equation 9.8, ΔS_{minor} takes the following form:

$$\Delta S_{minor} = \sqrt{\Delta S_{major}^2 \underbrace{\left(\frac{\partial S_{minor}}{\partial S_{major}}\right)^2}_{\partial S_{major}} + \Delta e^2 \underbrace{\left(\frac{\partial S_{minor}}{\partial e}\right)^2}_{\partial e}} \tag{9.13}$$

Where ∂S_{major} from Equation 9.13 is equivalent to Equation 9.14, where $u = 1 - e^2$:

$$\partial S_{major} = \left(\frac{\partial\left(\partial S_{major}\right)}{\partial u} \cdot \frac{\partial u}{\partial e}\right)^2 \tag{9.14}$$

Evaluating Equation 9.14 and substituting, ∂S_{major} is found to be:

$$\partial S_{major} = \left(\frac{S_{major}}{2\sqrt{1-e^2}} \cdot (-2e)\right)^2 \tag{9.15}$$

Similarly for ∂e:

$$\partial e = \left(\sqrt{1-e^2}\right)^2 \tag{9.16}$$

Substituting, Equations 9.15 and 9.16 back into Equation 9.13 allows ΔS_{minor} to be expressed as:

$$\Delta S_{minor} = \sqrt{\Delta e^2\left(1-e^2\right) - \Delta S_{major}^2\left(\frac{e^2 S_{major}^2}{1-e^2}\right)} \tag{9.17}$$

With an expression for ΔS_{minor} derived, the uncertainty associated with the average footprint radius $\Delta\bar{s}$ is now obtainable. Repeating the same process employing Equation 9.8, $\Delta\bar{s}$ can be expressed as:

$$\Delta\bar{s} = \sqrt{\Delta S_{major}^2 \underbrace{\left(\frac{\partial\bar{s}}{\partial S_{major}}\right)^2}_{=\frac{1}{4}} + \Delta S_{minor}^2 \underbrace{\left(\frac{\partial\bar{s}}{\partial S_{minor}}\right)^2}_{=\frac{1}{4}}} \tag{9.18}$$

Finally, employing both ΔE and $\Delta\bar{s}$, an associated uncertainty can be derived for ISM slope predictions. Using that slope predictions are calculated from Equation 3.5, where $E_{95} - E_{05} = E$ and $\bar{X}_{fp} = \bar{s}$ in this instance, and substituting for associated uncertainties in to Equation 9.8, the uncertainty in ISM slope predictions

$(\Delta\theta)$ can be expressed as:

$$\Delta\theta = \sqrt{\Delta E^2 \underbrace{\left(\frac{\partial\theta}{\partial E}\right)^2}_{\partial E} + \Delta\bar{s}^2 \underbrace{\left(\frac{\partial\theta}{\partial\bar{s}}\right)^2}_{\partial\bar{s}}} \tag{9.19}$$

Evaluating Equation 9.19, and substituting, ∂E is found to be:

$$\partial E = \left(\frac{\bar{s}}{E^2 + \bar{s}^2}\right)^2 \tag{9.20}$$

Similarly for $\partial\bar{s}$:

$$\partial\bar{s} = \left(\frac{-E}{E^2 + \bar{s}^2}\right)^2 \tag{9.21}$$

Substituting Equations 9.20 and 9.21 in to Equation 9.19, $\Delta\theta$ can finally be expressed as:

$$\Delta\theta = \sqrt{\Delta E^2 \left(\frac{\bar{s}^2}{E^4 + \bar{s}^4}\right) + \Delta\bar{s}^2 \left(\frac{E^2}{E^4 + \bar{s}^4}\right)} \tag{9.22}$$

9.12 Global Mean Annual NDVI (2003)

Figure 9.5 illustrates the mean annual NDVI distribution at a $0.5° \times 0.5°$ resolution globally for 2003. The value of NDVI is coloured according to the colour ramp inlay in Figure 9.5. As expected low NDVI is shown for desert and polar regions, whereas tropical, temperate, and boreal regions are accented with greater NDVI.

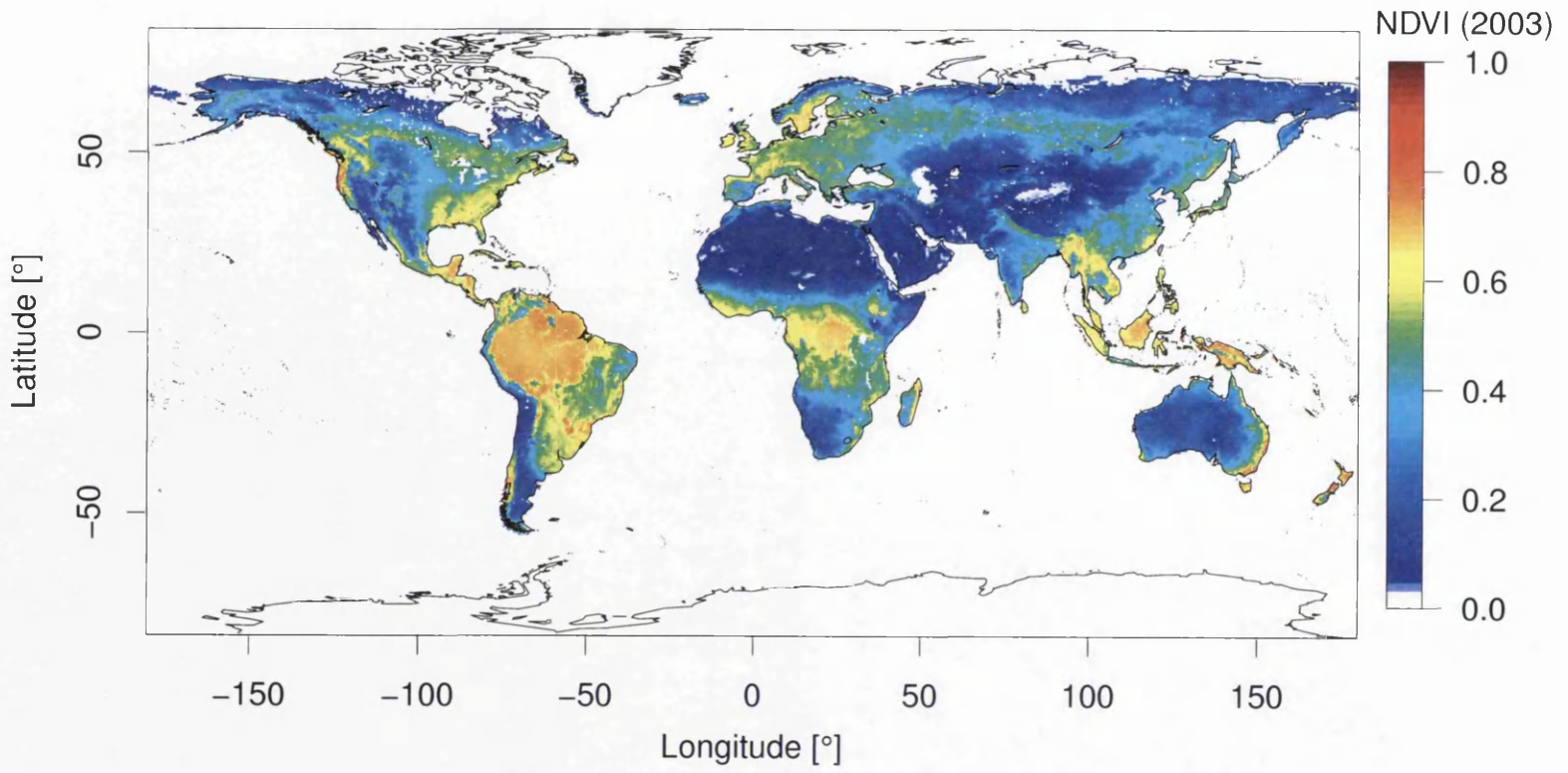


Figure 9.5: Global map of mean annual Normalised difference Vegetation Index (NDVI) from MODIS $0.5° \times 0.5°$ tiles for 2003.

9.13 Global Vegetation Cover Fraction (2003)

Figure 9.6 illustrates the mean annual Vegetation Cover Fraction (VCF) distribution at a $0.5° \times 0.5°$ resolution globally for 2003. The value of VCF is measured as a percentage, coloured according to the colour ramp inlay in Figure 9.6. As in Appendix 9.12, low VCF is shown for desert and polar regions, whereas tropical, temperate, and boreal regions exhibit greater VCF percentages.

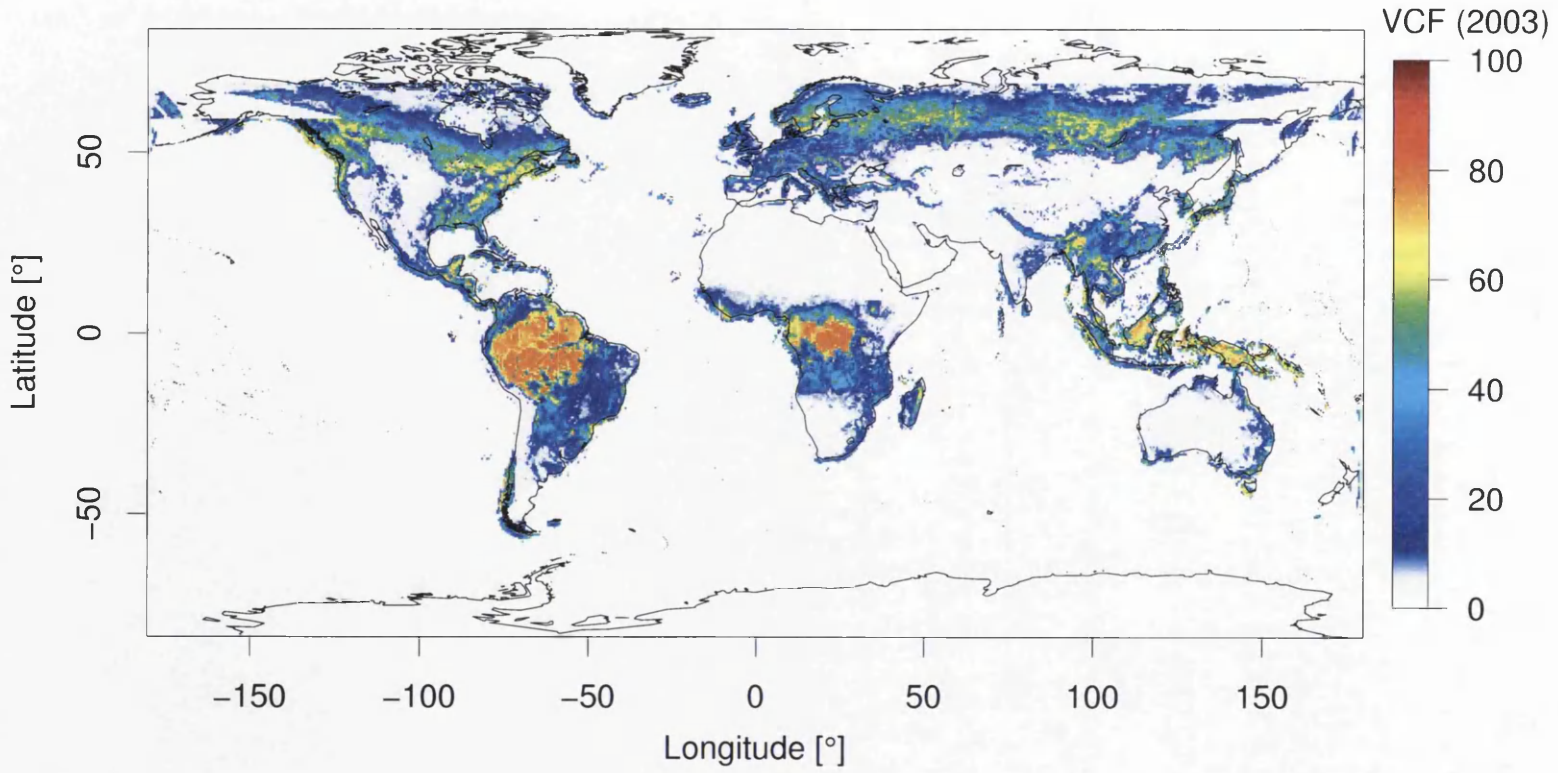


Figure 9.6: Global map of mean annual Vegetation Cover Fraction (VCF) aggregated to $0.5° \times 0.5°$ tiles from the MODIS MOD044B 500 m resolution data product from 2003.

References

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Harding, D., Kwok, R., Lefsky, M., Markus, T., et al. (2010). The ICESat-2 Laser Altimetry Mission. *Proceedings of the IEEE*, 98(5):735–751.

Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., Yi, D., and Liiva, P. (2005). Geoscience Laser Altimeter System (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters*, 32(21):L21S02.

Allouis, T., Durrieu, S., Véga, C., and Couteron, P. (2012). Stem Volume and Above-Ground Biomass Estimation of Individual Pine Trees From LiDAR Data: Contribution of Full-Waveform Signals. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5):924–934.

Andersen, H., McGaughey, R. J., and Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using LIDAR data. *Remote sensing of Environment*, 94(4):441–449.

Anderson, J., Martin, M. E., Smith, M. L., Dubayah, R. O., Hofton, M. A., Hyde, P., Peterson, B. E., Blair, J. B., and Knox, R. G. (2006). The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. *Remote Sensing of Environment*, 105(3):248–261.

AndúJar, D., Rosell-Polo, J. R., Fernández-Quintanilla, C., Dorado, J., et al. (2013). Potential of a terrestrial LiDAR-based system to characterise weed vegetation in maize crops. *Computers and Electronics in Agriculture*, 92:11–15.

Antonarakis, A. S., Richards, K. S., and Brasington, J. (2008). Object-based land cover classification using airborne LiDAR. *Remote Sensing of Environment*, 112(6):2988–2998.

ARSF (2013). Airborne Research Survey Facility: Leica ALS50-II LIDAR System. [online]. URL: http://arsf-dan.nerc.ac.uk/trac/wiki/Sensors/LeicaLIDAR.

ASPRS (2010). The American Society for Photogrammetry and Remote Sensing: LAS Specification Version 1.3. [online]. URL: http://www.asprs.org/a/society/committees/standards/LAS_1_3_r11.pdf.

Axelsson, P. (1999). Processing of laser scanner dataalgorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2):138–147.

Baldocchi, D. D., Hincks, B. B., and Meyers, T. P. (1988). Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology*, pages 1331–1340.

Baltsavias, E. P. (1999). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2):199–214.

BCAL (2011). BCAL LiDAR Tools. version 1.3.0. Idaho State University, Department of Geosciences, Boise Center Aerospace Laboratory (BCAL), Boise, Idaho. URL: http://bcal.geology.isu.edu/envitools.shtml.

Betts, R. A. (2000). Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. *Nature*, 408(6809):187–190.

Blair, J. B., Rabine, D. L., and Hofton, M. A. (1999). The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2):115–122.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., and Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112(10):3876–3890.

Boyd, D. S. and Hill, R. A. (2007). Validation of airborne lidar intensity values from a forested landscape using hymap data: preliminary analyses. In *Proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007 Espoo*, pages 12–14.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brenner, A. C., Zwally, H. J., Bentley, C. R., Csathó, B. M., Harding, D. J., Hofton, M. A., Minster, J. B., Roberts, L., Saba, J. L., Thomas, R. H., and Yi, D. (2003). *Geoscience Laser Altimeter System (GLAS) Algorithm Theoretical Basis Document 4.1: Derivation of Range and Range Distributions From Laser Pulse Waveform Analysis for Surface Elevations, Roughness, Slope, and Vegetation Heights.* NASA Goddard Space Flight Center.

Brovkin, V., Ganopolski, A., and Svirezhev, Y. (1997). A continuous climate-vegetation classification for use in climate-biosphere studies. *Ecological Modelling*, 101(2):251–261.

Brunt, K. M., Neumann, T. A., Markus, T., Brenner, A. C., Barbieri, K. A., Field, C. T., Sirota, J. M., and Hart, W. D. (2011). ICESat-2 Science Project Office at NASA/GSFC. [online]. URL: http://icesat.gsfc.nasa.gov/icesat2/data/mabel/mabel_docs.php.

Bufton, J. L., Harding, D. H., and Garvin, J. B. (1999). Shuttle Laser Altimeter: Mission Results and Pathfinder Accomplishments. In *Shuttle Small Payloads Symposium, NASA Conference Publication CP-1999-209476*, pages 337–345.

Bye, I. (2013). Personal communication.

Carabajal, C. C., Harding, D. J., Luthcke, S. B., Fong, W., Rowton, S. C., and Frawley, J. J. (1999). Processing of shuttle laser altimeter range and return pulse data in support of SLA-02. *International Archives of Photogrammetry and Remote Sensing*, 32(3214):65–72.

CGIAR-CSI (2013). The CGIAR Consortium for Spatial Information: SRTM 90 m Digital Elevation Data. [online]. URL: http://srtm.csi.cgiar.org/.

Chambers, J. Q., Higuchi, N., Tribuzy, E. S., Trumbore, S. E., et al. (2001). Carbon sink for a century. *Nature*, 410(6827):429–429.

Chang, J. C. and Hanna, S. R. (2004). Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1):167–196.

Chase, A. F., Chase, D. Z., Weishampel, J. F., Drake, J. B., Shrestha, R. L., Slatton, K. C., Awe, J. J., and Carter, W. E. (2011). Airborne LiDAR, archaeology, and the ancient Maya landscape at Caracol, Belize. *Journal of Archaeological Science*, 38(2):387–398.

Chen, J. M., Govind, A., Sonnentag, O., Zhang, Y., Barr, A., and Amiro, B. (2006). Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology*, 140(1):257–268.

Chen, Q. (2010a). Assessment of terrain elevation derived from satellite laser altimetry over mountainous forest areas using airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):111–122.

Chen, Q. (2010b). Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry. *Remote Sensing of Environment*, 114(7):1610–1627.

Chen, X., Vierling, L., Rowell, E., and DeFelice, T. (2004). Using lidar and effective LAI data to evaluate IKONOS and Landsat 7 ETM+ vegetation cover estimates in a ponderosa pine forest. *Remote Sensing of Environment*, 91(1):14–26.

Clark, M. L., Clark, D. B., and Roberts, D. A. (2004). Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, 91(1):68–89.

Cobby, D. M., Mason, D. C., and Davenport, I. J. (2001). Image processing of airborne scanning laser altimetry data for improved river flood modelling. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(2):121–138.

Comstock, J. M., Ackerman, T. P., and Mace, G. G. (2002). Ground-based lidar and radar remote sensing of tropical cirrus clouds at Nauru Island: Cloud

statistics and radiative impacts. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 107(D23):AAC–16.

Dai, R., Gong, W., Xu, J., Ren, X., and Liu, D. (2004). The edge technique as used in Brillouin lidar for remote sensing of the ocean. *Applied Physics B*, 79(2):245–248.

Dassot, M., Constant, T., and Fournier, M. (2011). The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges. *Annals of forest science*, 68(5):959–974.

de Noblet-Ducoudré, N., Claussen, M., and Prentice, C. (2000). Mid-Holocene greening of the Sahara: first results of the GAIM 6000 year BP experiment with two asynchronously coupled atmosphere/biome models. *Climate Dynamics*, 16(9):643–659.

de Noblet-Ducoudré, N. I., Prentice, I. C., Joussaume, S., Texier, D., Botta, A., and Haxeltine, A. (1996). Possible role of atmosphere-biosphere interactions in triggering the Last Glaciation. *Geophysical Research Letters*, 23(22):3191–3194.

Devereux, B., Amable, G., Crow, P., and Cliff, A. (2005). The potential of airborne lidar for detection of archaeological features under woodland canopies. *Antiquity*, 79(305):648–660.

Dickinson, R. E. (1984). Modeling evapotranspiration for three-dimensional global climate models. *Climate processes and climate sensitivity*, pages 58–72.

Dickinson, R. E., Jaeger, J., Washington, W. M., and Wolski, R. (1981). *Boundary subroutine for the NCAR global climate model.* Atmospheric Analysis and Prediction Division, National Center for Atmospheric Research.

Dolman, A. J., Moors, E. J., and Elbers, J. A. (2002). The carbon uptake of a mid latitude pine forest growing on sandy soil. *Agricultural and Forest Meteorology*, 111(3):157–170.

Donnellan, A., Rosen, P., Graf, J., Loverro, A., Freeman, A., Treuhaft, R., Oberto, R., Simard, M., Rignot, E., Kwok, R., et al. (2008). Deformation, ecosys-

tem structure, and dynamics of ice (DESDynI). In *Aerospace Conference, 2008 IEEE*, pages 1–13. IEEE.

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., Chazdon, R. L., Weishampel, J. F., and Prince, S. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, 79(2):305–319.

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., and Hofton, M. (2003). Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography*, 12(2):147–159.

Dubayah, R., Blair, J. B., Bufton, J. L., Clark, D. B., JáJá, J., Knox, R., Luthcke, S. B., Prince, S., and Weishampel, J. (1997). The vegetation canopy lidar mission. In *Proceedings of Land Satellite Information in the Next Decade, II: Sources and Applications. Bethesda (MD): American Society of Photogrammetry and Remote Sensing*, pages 100–112.

Dubayah, R. O. and Drake, J. B. (2000). Lidar remote sensing for forestry. *Journal of Forestry*, 98(6):44–46.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., and Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, 115(G2).

Duncanson, L. I., Niemann, K. O., and Wulder, M. A. (2010). Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*, 114(1):138–154.

Duong, H., Lindenbergh, R., Pfeifer, N., and Vosselman, G. (2009). ICESat full-waveform altimetry compared to airborne LASER scanning altimetry over the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10):3365–3378.

Duong, H., Pfeifer, N., and Lindenbergh, R. (2006). Analysis of repeated ICESat full waveform data: methodology and leaf-on/leaf-off comparison. In *Proceedings: Workshop on 3D Remote Sensing in Forestry*, pages 239–248. URL: http://www.rali.boku.ac.at/3drsforestry.html.

Duong, V. H., Lindenbergh, R., Pfeifer, N., and Vosselman, G. (2008). Single and two epoch analysis of ICESat full waveform data over forested areas. *International Journal of Remote Sensing*, 29(5):1453–1473.

EDINA (2013). PROFILE DEM [ASC geospatial data], Scale 1:10000. EDINA Digimap Ordnance Survey Service. Updated: November 2009. Downloaded: Thu Jan 17 12:15:59 GMT 2013. URL: http://edina.ac.uk/digimap.

Einstein, A. (1917). On the Quantum Theory of Radiation. *Physikalische Zeitschrift*, 121(18):167–183.

Environment Agency (2013). Environment Agency Geomatics Group Integrated spatial data. [online]. URL: https://www.geomatics-group.co.uk/geocms/Homepage.aspx.

FAO, UNDP, and UNEP. (2008). United Nations Collaborative Programme on Reducing Emissions from Deforestation in Developing Countries. URL: http://www.undp.org/content/dam/aplaws/publication/en/publications/environment-energy/www-ee-library/climate-change/un-redd-programme-brochure/unredd.pdf.

Feigenwinter, C., Mölder, M., Lindroth, A., and Aubinet, M. (2010). Spatiotemporal evolution of CO_2 concentration, temperature, and wind field during stable nights at the Norunda forest site. *Agricultural and Forest Meteorology*, 150(5):692–701.

Fluxnet (2013). Fluxnet: A Global Network. Integrating Worldwide CO_2, Water and Energy Flux Measurements. [online]. URL: http://fluxnet.ornl.gov/.

Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A. (1996). An integrated biosphere model of land surface processes,

terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles*, 10(4):603–628.

Forestry Commission (2012a). Forestry Statistics 2012 and Forestry facts & Figures 2012. [online]. URL: http://www.forestry.gov.uk/forestry/infd-8xfcwy.

Forestry Commission (2012b). Scots pine (pinus sylvestris) information at the Forestry Commission. United Kingdom Forestry Commission. URL: http://www.forestry.gov.uk/forestry/INFD-5NLFAP.

Freeman, A., Rosen, P., Jordan, R., Johnson, W. T. K., Hensley, S., Sweetser, T., Loverro, A., Smith, J., Sprague, G., and Shen, Y. (2009). DESDynI–A NASA Mission for Ecosystems, Solid Earth, and Cryosphere Science. In *Proceedings of the Pol-InSAR Workshop, Frascati.*

Friend, A. D., Stevens, A. K., Knox, R. G., and Cannell, M. G. R. (1997). A process-based, terrestrial biosphere model of ecosystem dynamics (Hybrid v3. 0). *Ecological Modelling*, 95(2):249–287.

Fukushima, Y., Hiura, T., and Tanabe, S. I. (1998). Accuracy of the MacArthur-Horn method for estimating a foliage profile. *Agricultural and Forest Meteorology*, 92(4):203–210.

Garvin, J., Bufton, J., Blair, J., Harding, D., Luthcke, S., Frawley, J., and Rowlands, D. (1998). Observations of the Earth's topography from the Shuttle Laser Altimeter (SLA): laser-pulse echo-recovery measurements of terrestrial surfaces. *Physics and Chemistry of the Earth*, 23(9):1053–1068.

Gaumont-Guay, D., Black, T. A., McCaughey, H., Barr, A. G., Krishnan, P., Jassal, R. S., and Nesic, Z. (2009). Soil CO_2 efflux in contrasting boreal deciduous and coniferous stands and its contribution to the ecosystem carbon balance. *Global Change Biology*, 15(5):1302–1319.

Gibbs, H. K., Brown, S., Niles, J. O., and Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2(4):045023.

Goodwin, N. R., Coops, N. C., Bater, C., and Gergel, S. E. (2007). Assessment of sub-canopy structure in a complex coniferous forest. In *ISPRS Workshop on Laser Scanning and SilviLaser 2007, Espoo, September 12–14, 2007, Finland.*

Grünwald, T. and Bernhofer, C. (2007). A decade of carbon, water and energy flux measurements of an old spruce forest at the Anchor Station Tharandt. *Tellus B*, 59(3):387–396.

Hancock, S., Lewis, P., Foster, M., Disney, M., and Muller, J. P. (2012). Measuring forests with dual wavelength lidar: A simulation study over topography. *Agricultural and Forest Meteorology*, 161:123–133.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, 7(10):1–15.

Hansen, M. C., Stehman, S. V., and Potapov, P. V. (2010). Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences*, 107(19):8650–8655.

Harding, D. J., Blair, J. B., Rabine, D. L., and Still, K. (1998). SLICER: Scanning lidar imager of canopies by echo recovery instrument and data product description. Technical report, NASA Goddard Space Flight Center, Greenbelt, MD.

Harding, D. J. and Carabajal, C. C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical research letters*, 32(21):L21S10.

Harding, D. J., Dabney, P., Valett, S. R., and Kelly, A. (2010). SIMPL Laser Altimeter Measurements of Lake Erie Ice Cover: a Pathfinder for ICESat-2. In *American Geophysical Union, Fall Meeting 2010.* Abstract available from: http://adsabs.harvard.edu/abs/2010AGUFM.C54B..08H.

Harding, D. J., Gesch, D. B., Carabajal, C. C., and Luthcke, S. B. (1999). Application of the Shuttle Laser Altimeter in an accuracy assessment of GTOPO30, a

global 1-kilometer digital elevation model. *International Archives of Photogrammetry and Remote Sensing*, 32:81–85.

Harding, D. J., Lefsky, M. A., Parker, G. G., and Blair, J. B. (2001). Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, 76(3):283–297.

Hawbaker, T. J., Keuler, N. S., Lesak, A. A., Gobakken, T., Contrucci, K., and Radeloff, V. C. (2009). Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, 114(G2).

Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., Neumann, K., Riedel, T., and Schröter, K. (2005). Global biomass mapping for an improved understanding of the CO_2 balance–the Earth observation mission Carbon-3D. *Remote Sensing of Environment*, 94(1):94–104.

Hill, R. A. (2007). Going undercover: mapping woodland understorey from leaf-on and leaf-off lidar data. In *ISPRS Workshop on Laser Scanning and SilviLaser 2007, Espoo, September 12–14, 2007, Finland.*

Hill, R. A. and Thomson, A. G. (2005). Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, 26(17):3763–3779.

Hinsley, S. A., Hill, R. A., Bellamy, P. E., and Balzter, H. (2006). The Application of Lidar in Woodland Bird Ecology: Climate, Canopy Structure, and Habitat Quality. *Photogrammetric Engineering and Remote Sensing*, 72(12):1399–1406.

Hodgson, M. E. and Bresnahan, P. (2004). Accuracy of airborne lidar-derived elevation: empirical assessment and error budget. *Photogrammetric engineering and remote sensing*, 70(3):331–340.

Höfle, B. (2010). LiDAR Research Group. [online]. URL: http://www.laserdata.info/index.php?pageref=bericht&sid=589&fid=2143#pix.

Hofton, M. A. and Blair, J. B. (2002). Laser altimeter return pulse correlation: A method for detecting surface topographic change. *Journal of Geodynamics*, 34(3):477–489.

Hofton, M. A., Minster, J. B., and Blair, J. B. (2000). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4):1989–1996.

Hofton, M. A., Rocchio, L. E., Blair, J. B., and Dubayah, R. (2002). Validation of vegetation canopy lidar sub-canopy topography measurements for a dense tropical forest. *Journal of Geodynamics*, 34(3):491–502.

Holmgren, J., Persson, Å., and Söderman, U. (2008). Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*, 29(5):1537–1552.

Hopkinson, C. (2013). Personal communication.

Hopkinson, C. and Chasmer, L. (2007). Using discrete laser pulse return intensity to model canopy transmittance. *The Photogrammetric Journal of Finland*, 20(2):16–26.

Hopkinson, C. and Chasmer, L. (2009). Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, 113(1):275–288.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, 83(1):195–213.

Huete, A., Didan, K., van Leeuwen, W., Miura, T., and Glenn, E. (2011). MODIS vegetation indices. In *Land Remote Sensing and Global Environmental Change*, pages 579–602. Springer.

Hyyppä, J., Kelle, O., Lehikoinen, M., and Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39(5):969–975.

IPCC (2003). Good practice guidance for land use, land-use change and forestry. institute for global environmental strategies (iges) for the intergovernmental panel on climate change. [online]. URL: http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_contents.html.

IPCC (2007). *Climate Change 2007: The Physical Science Basis.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA and Melbourne, Australia. Contribution of Working Group I to Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Soloman, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Avert, K. B., Tignor, M., and Miller, H. L. (eds.)].

Irish, J. L. and Lillycrop, W. J. (1999). Scanning laser mapping of the coastal zone: the SHOALS system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2):123–129.

Jansma, P., Mattioli, G., and Matias, A. (2001). SLICER laser altimetry in the Eastern Caribbean. *Surveys in geophysics*, 22(5-6):561–579.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, A. (2008). Hole-filled SRTM for the globe version 4. Available from the CGIAR-CSI SRTM 90 m database. Available at: http://srtm.csi.cgiar.org (accessed 25/07/2012).

JULES (2013). Joint UK Land Environment Simulator, Science in JULES - Vegetation. [online]. URL: https://jules.jchmr.org/model-description/vegetation.

Kljun, N., Black, T. A., Griffis, T. J., Barr, A. G., Gaumont-Guay, D., Morgenstern, K., McCaughey, J. H., and Nesic, Z. (2007). Response of net ecosystem productivity of three boreal forest stands to drought. *Ecosystems*, 9(7):1128–1144.

Koenig, L., Martin, S., Studinger, M., and Sonntag, J. (2010). Polar airborne observations fill gap in satellite data. *Eos, Transactions American Geophysical Union*, 91(38):333–334.

Koukoulas, S. and Blackburn, G. A. (2005). Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and

multi-spectral remotely sensed data. *International Journal of Remote Sensing*, 26(3):431–455.

Krabill, W., Thomas, R., Jezek, K., Kuivinen, K., and Manizade, S. (1995). Greenland ice sheet thickness changes measured by laser altimetry. *Geophysical Research Letters*, 22(17):2341–2344.

Kurtz, N. T. and Farrell, S. L. (2011). Large-scale surveys of snow depth on Arctic sea ice from Operation IceBridge. *Geophysical Research Letters*, 38(20).

Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D. (2009). Thinning and volume loss of the Arctic Ocean sea ice cover: 2003–2008. *Journal of Geophysical Research: Oceans (1978–2012)*, 114(C7):doi:10.1029/2009JC005312.

Lagergren, F., Eklundh, L., Grelle, A., Lundblad, M., Mölder, M., Lankreijer, H., and Lindroth, A. (2004). Net primary production and light use efficiency in a mixed coniferous forest in Sweden. *Plant, Cell & Environment*, 28(3):412–423.

Lancaster, R. S., Spinhirne, J. D., and Palm, S. P. (2005). Laser pulse reflectance of the ocean surface from the GLAS satellite lidar. *Geophysical research letters*, 32(22).

Lee, S., Ni-Meister, W., Yang, W., and Chen, Q. (2011). Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA. *Remote Sensing of Environment*, 115(11):2776–2785.

Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, 37(15):L15401.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., and Harding, D. (1999a). Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment*, 70:339–361.

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., and Gower, S. T. (2002a). Lidar remote sensing of above-ground biomass in three biomes. *Global ecology and biogeography*, 11(5):393–399.

Lefsky, M. A., Cohen, W. B., Parker, G. G., and Harding, D. J. (2002b). Lidar Remote Sensing for Ecosystem Studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 52(1):19–30.

Lefsky, M. A., Harding, D. J., Cohen, W. B., Parker, G., and Shugart, H. H. (1999b). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, 67(1):83–98.

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Espirito-Santo, F. D. B., Hunter, M. O., and de Oliveira Jr, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32(22):L22S02.

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

Leuning, R., Cleugh, H. A., Zegelin, S. J., and Hughes, D. (2005). Carbon and water fluxes over a temperate *Eucalyptus* forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. *Agricultural and Forest Meteorology*, 129(3):151–173.

Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System. *Agronomie*, 19(3-4):185–210.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in physical geography*, 27(1):88–106.

Lindroth, A., Grelle, A., and Morén, A. S. (1998). Long-term measurements of boreal forest carbon balance reveal large temperature sensitivity. *Global Change Biology*, 4(4):443–450.

Los, S. O., Pollack, N. H., Parris, M. T., Collatz, G. J., Tucker, C. J., Sellers, P. J., Malmström, C. M., DeFries, R. S., Bounoua, L., and Dazlich, D. A. (2000). A global 9-yr biophysical land surface dataset from NOAA AVHRR data. *Journal of Hydrometeorology*, 1(2):183–199.

Los, S. O., Rosette, J. A. B., Kljun, N., North, P. R. J., Suárez, J. C., Hopkinson, C., Hill, R. A., Chasmer, L., van Gorsel, E., Mahoney, C., and Berni, J. A. J. (2012). Vegetation height products between 60° N and 60° S from ICESat GLAS data. *Geoscientific Model Development Discussions*, 4:2327–2363.

Luthcke, S. B., Carabajal, C. C., and Rowlands, D. D. (2002). Enhanced geolocation of spaceborne laser altimeter surface returns: parameter calibration from the simultaneous reduction of altimeter range and navigation tracking data. *Journal of Geodynamics*, 34(3):447–475.

MacArthur, R. H. and Horn, H. S. (1969). Foliage profile by vertical measurements. *Ecology*, pages 802–804.

Maclean, G. A. and Krabill, W. B. (1986). Gross-merchantable timber volume estimation using an airborne LIDAR system. *Canadian Journal of Remote Sensing*, 12:7–18.

Maiman, T. H. (1960). Stimulated optical radiation in ruby. *Nature*, 187(4736):493–494.

Mallet, C. and Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1):1–16.

Mallet, C., Bretar, F., and Soergel, U. (2008). Analysis of full-waveform lidar data for classification of urban areas. *Photogrammetrie Fernerkundung GeoInformation (PFG)*, 5:337–349.

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J., and Vehmas, M. (2004). Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment*, 90(3):319–330.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

McCaughey, J. H., Barr, A. G., Black, T. A., Nesic, Z., Morgenstern, K., Griffis, T., and Gaumont-Guay, D. (2002). Carbon and energy exchanges at three boreal forest sites in BERMS study region in 2000 and 2001. In *Proceedings: 25th Conference on Agricultural and Forest Meteorology.* URL: https://ams.confex.com/ams/pdfpapers/38372.pdf.

Means, J. E. (2000). Comparison of large-footprint and small-footprint lidar systems: design, capabilities, and uses. In *Proceedings of the Second International Conference on Geospatial Information in Agriculture and Forestry. Lake Buena Vista, Florida*, pages 85–192.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., Harmon, M. E., and McKee, W. A. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, 67(3):298–308.

Molero, F. and Jaque, F. (1999). The laser as a tool in environmental problems. *Optical Materials*, 13(1):167–173.

Morsdorf, F., Meier, E., Kötz, B., Itten, K. I., Dobbertin, M., and Allgöwer, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, 92(3):353–362.

Morsdorf, F., Nichol, C., Malthus, T. J., Patenaude, G., and Woodhouse, I. H. (2008). Modelling multi-spectral LIDAR vegetation backscatter assessing structural and physiological information content. In Hill, R., Rosette, J., and Suaŕez,

J., editors, *Proceedings of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory*, pages 257–265. URL: http://geography.swan.ac.uk/silvilaser/SilviLaser_2008_Proceedings.pdf.

Muirhead, K. and Cracknell, A. P. (1986). Airborne lidar bathymetry. *International Journal of Remote Sensing*, 7(5):597–614.

Mullen, L., Herczfeld, P. R., and Contarino, V. M. (1996). Hybrid lidar-radar ocean experiment. In *Summaries of papers presented at the Conference on Lasers and Electro-Optics, 1996*, page 98. IEEE.

Myneni, R. B., Nemani, R. R., and Running, S. W. (1997). Estimation of Global Leaf Area Index and Absorbed Par Using Radiative Transfer Models. *IEEE Transactions on Geoscience and Remote Sensing*, 35(6).

Næsset, E. (1997). Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 61(2):246–253.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1):88–99.

NAP (2008). *Satellite Observations to Benefit Science and Society: Recommended Missions for the Next Decade.* The National Academies Press. Authored by the Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future. Editor: Henson, R., National Research Council.

NASA (1994). NASA: Clementine Project Information. [online]. URL: http://nssdc.gsfc.nasa.gov/planetary/clementine.html.

NASA (1996). NASA: Mars Global Surveyor Mission Information. [online]. Last updated: November, 2010. URL: http://www.nasa.gov/mission_pages/mgs/.

NASA (2006). NASA: Airborne UV Differential Absorption LIDAR (DIAL): Measuring Ozone and Aerosols in the Atmosphere. [online]. Last updated: March, 2006. URL: http://www.nasa.gov/centers/langley/news/factsheets/DIAL.html.

NASA (2007). Report from the ICESat-II Workshop. Technical report, National Aeronautics and Space Administration, Linthicum, Maryland, USA. Contributors: Abdalati, W., Carabajal, C., Csatho, B., Dubayah, R., Fowler, D., Fricker, H. A., Hall, F., Harding, D., Joughin, I., Kimes, D., Kwok, R., Laxon, S., Lefsky, M., Luthcke, S., Martin, S., Minnett, P., Ranson, J., Sauber, J. M., Scambos, T., Spinhirne, J., Thmoas, R., Urban, T., Wiscombe, W., and Zwally, J.

NASA (2008). Report of the DESDynI Applications Workshop. Technical report, National Aeronautics and Space Administration, UC Sacramento Conference Center, Sacramento, USA. Contributors: Donnellan, A., Bawden, G., Rundle, J., Hager, B., Brakenridge, B., Eichelberger, J., Hunsaker, C., Killeen, T., Lu, Z., Parrish, J., Pichel, B., Turcotte, D., and Woodward, N.

NASA (2011). NASA's Earth Science News Team: MABEL Flies Her Maiden Voyage. [online]. URL: http://www.nasa.gov/topics/earth/features/mabel-maiden.html.

NASA (2012). NASA Academy of Program/Project & Engineering Leadership (APPEL) Case Stuies: Vegetation Canopy LiDAR (VCL). [online]. URL: http://www.nasa.gov/offices/oce/appel/knowledge/publications/VCL.html.

NASA (2013a). NASA Goddard Space Flight Center: ICESat-2 Information. [online]. URL: http://icesat.gsfc.nasa.gov/icesat2/mission_objs.php.

NASA (2013b). NASA: Moderate Resolution Imaging Spectroradiometer. [online]. URL: http://modis.gsfc.nasa.gov/about/.

NASA (2013c). NASA Science Missions: Deformation, Ecosystem Structure and Dynamics of Ice (DESDynI). [online]. URL: http://science.nasa.gov/missions/desdyni/.

NASA/GSFC (2013). ICESat-2 Science Project Office at NASA/GSFC: MABEL Documentation. [online]. URL: http://icesat.gsfc.nasa.gov/icesat2/data/mabel/mabel_docs.php.

Nelson, R. (2010). Model effects on GLAS-based regional estimates of forest biomass and carbon. *International Journal of Remote Sensing*, 31(5):1359–1372.

Nelson, R., Krabill, W., and Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote sensing of environment*, 24(2):247–267.

Nelson, R., Oderwald, R., and Gregoire, T. G. (1997). Separating the ground and airborne laser sampling phases to estimate tropical forest basal area, volume, and biomass. *Remote Sensing of Environment*, 60(3):311–326.

Nelson, R., Ranson, K. J., Sun, G., Kimes, D. S., Kharuk, V., and Montesano, P. (2009). Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of Environment*, 113(3):691–701.

Nelson, R., Short, A., and Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. *Scandinavian Journal of Forest Research*, 19(6):500–511.

NERC (2013). Research Programmes at the Natural Environment Research Council. Available from: http://www.nerc.ac.uk/research/programmes/ (accessed 24/07/2013).

Neuenschwander, A. L., Urban, T. J., and Gutierrez, R. Schutz, B. E. (2008). Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: implications for vegetation mapping. *Journal of Geophysical Research*, 113:G02S03. doi:10.1029/2007JG000557.

Ni-Meister, W., Jupp, D. L. B., and Dubayah, R. (2001). Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9):1943–1958.

North, P. R. J. (1996). Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4):946–956.

North, P. R. J., Rosette, J. A. B., Suárez, J. C., and Los, S. O. (2010). A Monte Carlo radiative transfer model of satellite waveform LiDAR. *International Journal of Remote Sensing*, 31(5):1343–1358.

NRC (2007). *Earth science and applications from space: National imperatives for the next decade and beyond.* National Academy Press, Washington, DC.

NSIDC (2012a). ICESat/GLAS Data at NSIDC. National Snow and Ice Data Center. URL: http://nsidc.org/data/icesat/index.html.

NSIDC (2012b). ICESat/GLAS Data Quality Flags at NSIDC. National Snow and Ice Data Center. URL: http://nsidc.org/data/docs/daac/glas_altimetry/data_-dictionary.html.

Optech (2013). Optech Incorporated: Airborne Laser Terrain Mapper Gemini. [online]. URL: http://www.optech.ca/gemini.htm.

Pacala, S. W., Hurtt, G. C., Baker, D., Peylin, P., Houghton, R. A., Birdsey, R. A., Heath, L., Sundquist, E. T., Stallard, R. F., Ciais, P., et al. (2001). Consistent land-and atmosphere-based US carbon sink estimates. *Science*, 292(5525):2316–2320.

Paris, J.-D., Ciais, P., Rivier, L., Chevallier, F., Dolman, H., Flaud, J.-M., Garrec, C., Gerbig, C., Grace, J., Huertas, E., et al. (2012). Integrated Carbon Observation System. In *EGU General Assembly Conference Abstracts*, volume 14, page 12397.

Parker, G. G., Lefsky, M. A., and Harding, D. J. (2001). Light transmittance in forest canopies determined using airborne laser altimetry and in-canopy quantum measurements. *Remote Sensing of Environment*, 76(3):298–309.

Patenaude, G., Hill, R. A., Milne, R., Gaveau, D. L. A., Briggs, B. B. J., and Dawson, T. P. (2004). Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, 93(3):368–380.

Peng, C. (2000). Understanding the role of forest simulation models in sustainable forest management. *Environmental Impact Assessment Review*, 20(4):481–501.

Peterson, B., Dubayah, R., Hyde, P., Hofton, M., Blair, J. B., and Fites-Kaufman, J. (2005). Use of LIDAR for Forest Inventory and Forest Management Application. In *Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium.*

Popescu, A., Schorstein, K., and Walther, T. (2004). A novel approach to a Brillouin–LIDAR for remote sensing of the ocean temperature. *Applied Physics B*, 79(8):955–961.

Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, 31(9):646–655.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A. (1993). Terrestrial ecosystem production: a process model based on global satellite and surface data. *Global Biogeochemical Cycles*, 7(4):811–841.

Qiyang, X., Kuntao, Y., Xinbing, W., et al. (2002). Blue-green lidar ocean survey. *Beijing, National Defence Industry Publishing Company*, pages 36–54.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Reitberger, J., Krzystek, P., and Stilla, U. (2006). Analysis of full waveform lidar data for tree species classification. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 3):228–233.

Reitberger, J., Krzystek, P., and Stilla, U. (2008). Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International journal of remote sensing*, 29(5):1407–1431.

Ridder, R. M. (2007). Global forest resources assessment 2010: options and recommendations for a global remote sensing survey of forests. *FAO For Resource Assessment. Programme Work. Pap*, 141.

Riegl (2013). Riegl Laser Measurement Systems: Airborne Laser Scanning. [online]. URL: http://www.riegl.com/products/airborne-scanning/.

Rodriguez, E., Morris, C. S., Belz, J. E., Chaplin, E. C., Martin, J. M., Daffer, W., and Hensley, S. (2005). An assessment of the SRTM topographic products. Technical report, JPL D-31639, NASA Jet Propulsion Laboratory, Pasadena, California.

Rosette, J., Field, C., Nelson, R., DeCola, P., Cook, B., et al. (2011). A new photon-counting lidar system for vegetation analysis. In *Proceedings of Silvi-Laser 2011, 11th International Conference on LiDAR Applications for Assessing Forest Ecosystems, University of Tasmania, Australia, 16-20 October 2011.*, pages 1–8. Conference Secretariat.

Rosette, J. A. B. (2009). *Biophysical Parameter Retrieval from Satellite Laser Altimetry.* PhD thesis, Department of Geography, Swansea University.

Rosette, J. A. B., North, P. R. J., Rubio-Gil, J., Cook, B., Los, S., Suaŕez, J., Sun, G., Ranson, J., and Blair, J. B. (2013). Evaluating Prospects for Improved Forest Parameter Retrieval From Satellite LiDAR Using a Physically-Based Radiative Transfer Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(1):45–53.

Rosette, J. A. B., North, P. R. J., and Suárez, J. C. (2008). Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing*, 29(5):1475–1493.

Rosette, J. A. B., North, P. R. J., Suarez, J. C., and Los, S. O. (2010). Uncertainty within satellite LiDAR estimations of vegetation and topography. *International Journal of Remote Sensing*, 31(5):1325–1342.

Rutzinger, M., Höfle, B., Hollaus, M., and Pfeifer, N. (2008). Object-based point cloud analysis of full-waveform airborne laser scanning data for urban vegetation classification. *Sensors*, 8(8):4505–4528.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108(24):9899–9904.

Saatchi, S. S., Houghton, R. A., Dos Santos Alvala, R. C., Soares, J. V., and Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13(4):816–837.

Sanz, R., Rosell, J., Llorens, J., Gil, E., and Planas, S. (2013). Relationship between tree row LIDAR-volume and leaf area density for fruit orchards and vineyards obtained with a LIDAR 3D Dynamic Measurement System. *Agricultural and forest meteorology*, 171:153–162.

Sassen, K. and Benson, S. (2001). A midlatitude cirrus cloud climatology from the Facility for Atmospheric Remote Sensing. Part II: Microphysical properties derived from lidar depolarization. *Journal of the atmospheric sciences*, 58(15):2103–2112.

Sato, H., Itoh, A., and Kohyama, T. (2007). SEIB–DGVM: A new Dynamic Global Vegetation Model using a spatially explicit individual-based approach. *Ecological Modelling*, 200(3):279–307.

Schawlow, A. L. and Townes, C. H. (1958). Infrared and optical masers. *Physical Review*, 112(6):1940–1949.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P. (2005). Overview of the ICESat mission. *Geophysical Research Letters*, 32(21):L21S01.

Sellers, P. J., Hall, F. G., Kelly, R. D., Black, A., Baldocchi, D., Berry, J., Ryan, M., Ranson, K. J., Crill, P. M., Lettenmaier, D. P., et al. (1997). BOREAS in 1997: Experiment overview, scientific results, and future directions. *Journal of Geophysical Research*, 102(D24):28731–28.

Sellers, P. J., Mintz, Y. C. S. Y., Sud, Y. C. e. a., and Dalcher, A. (1986). A simple biosphere model (SiB) for use within general circulation models. *Journal of the Atmospheric Sciences*, 43(6):505–531.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S. O., Justice, C. O., Dazlich, D. A., and Randall, D. A. (1996). A revised land surface parametrization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of climate*, 9(4):706–737.

Shi, J., Ouyang, M., Gong, W., Li, S., and Liu, D. (2008). A Brillouin lidar system using F–P etalon and ICCD for remote sensing of the ocean. *Applied Physics B*, 90(3-4):569–571.

Sigma Space (2013). Sigma Space Corporation: Imaging Photoncounting Altimeter. [online]. URL: http://www.sigmaspace.com/index.php/instruments/3d-imaging-lidars/imaging-photoncounting-altimeter.

Simard, M., Pinto, N., Fisher, J. B., and Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*, 116(G4):G04021.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., et al. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9(2):161–185.

Six, J., Callewaert, P., Lenders, S., De Gryze, S., Morris, S. J., Gregorich, E. G., Paul, E. A., and Paustian, K. (2002). Measuring and understanding carbon storage in afforested soils by physical fractionation. *Soil Science Society of America Journal*, 66(6):1981–1987.

Smith, D. E., Zuber, M. T., Frey, H. V., Garvin, J. B., Head, J. W., Muhleman, D. O., Pettengill, G. H., Phillips, R. J., Solomon, S. C., Zwally, H. J., et al. (2001). Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research: Planets (1991–2012)*, 106(E10):23689–23722.

Spinhirne, J. D., Hansen, M. Z., and Caudill, L. O. (1982). Cloud top remote sensing by airborne lidar. *Applied optics*, 21(9):1564–1571.

Streutker, D. R. and Glenn, N. F. (2006). LiDAR measurement of sagebrush steppe vegetation heights. *Remote Sensing of Environment*, 102(1):135–145.

Suárez, J. C., Ontiveros, C., Smith, S., and Snape, S. (2005). Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. *Computers & Geosciences*, 31(2):253–262.

Sun, G., Ranson, K. J., Kimes, D. S., Blair, J. B., and Kovacs, K. (2008). Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of Environment*, 112(1):107–117.

Svelto, O. (2010). *Principles of Lasers*. Springer, 5th edition. ISBN 978-1-4419-1301-2.

Terrasolid (2011). Terrasolid processing laser and images: Terrascan software. [online]. URL: http://www.terrasolid.fi/en/products/terrascan.

Todd, K., Csillag, F., and Atkinson, P. M. (2003). Three-dimensional mapping of light transmittance and foliage distribution using lidar. *Canadian journal of remote sensing*, 29(5):544–555.

Toman, K. (1984). Doppler and the Doppler Effect. Technical report, Defense Technical Information Center Document.

Townshend, J., Hansen, M., Caroll, M., DiMiceli, C., Sohlberg, R., and Huang, C. (2011). User Guide for the MODIS Vegetation Continuous Fields product Collection 5 Version 1. Technical report, University of Maryland, College Park, Maryland, USA.

UNEP (2009). United Nations Environmental Programme: Reducing Emissions from Deforestation and Degradation. [online]. URL: http://www.unep.org/climatechange/UNEPsWork/REDD/tabid/245/language/en-US/Default.aspx.

UNFCCC (1999). The Kyoto Protocol: a guide and assessment. [online]. URL: http://unfccc.int/kyoto_protocol/items/2830.php.

UQU (2013). Umm Al-Qura University: Gaussian Distribution Mathematics. [online]. URL: http://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/4282164/Gaussian%20function.pdf.

Utts, J. M. (1982). The rainbow test for lack of fit in regression. *Communications in Statistics-Theory and Methods*, 11(24):2801–2815.

van Gorsel, E., Leuning, R., Cleugh, H. A., Keith, H., Kirschbaum, M. U. F., and Suni, T. (2008). Application of an alternative method to derive reliable estimates of nighttime respiration from eddy covariance measurements in moderately complex topography. *agricultural and forest meteorology*, 148(6):1174–1180.

Van Loon, B. (2005). Radar 101: Celebrating 101 years of development. *Proceedings of the IEEE*, 93(4):844–846.

Vierling, K. T., Vierling, L. A., Gould, W. A., Martinuzzi, S., and Clawges, R. M. (2008). Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, 6(2):90–98.

Wagner, W., Hollaus, M., Briese, C., and Ducic, V. (2008). 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. *International Journal of Remote Sensing*, 29(5):1433–1452.

Wallace, A., Nichol, C., and Woodhouse, I. (2012). Recovery of forest canopy parameters by inversion of multispectral LiDAR data. *Remote Sensing*, 4(2):509–531.

Waring, R. H., Way, J., Hunt, E. R., Morrissey, L., Ranson, K. J., Weishampel, J. F., Oren, R., and Franklin, S. E. (1995). Imaging radar for ecosystem studies. *BioScience*, 45(10):715–723.

Wehr, A. and Lohr, U. (1999). Airborne laser scanningan introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2):68–82.

Weitkamp, C. (2005). *Lidar: range-resolved optical remote sensing of the atmosphere*, volume 102. Springer.

Woodward, F. I., Lomas, M. R., and Betts, R. A. (1998). Vegetation-climate feedbacks in a greenhouse world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1365):29.

Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., et al. (2011). Measuring forest structure

and biomass in New England forest stands using Echidna ground-based lidar. *Remote sensing of Environment*, 115(11):2965–2974.

Yu, A. W., Krainak, M. A., Harding, D. J., Abshire, J. B., Sun, X., Valett, S., Cavanaugh, J., and Ramos-Izquierdo, L. (2010a). Airborne lidar simulator for the lidar surface topography (LIST) mission. In *NASA Earth Science Technology Forum, Arlington, VA*.

Yu, A. W., Stephen, M. A., Li, S. X., Shaw, G. B., Seas, A., Dowdye, E., Troupaki, E., Liiva, P., Poulios, D., and Mascetti, K. (2010b). Space Laser Transmitter Development for ICESat-2 Mission. In *LASE*, pages 757809–1–757809–11. International Society for Optics and Photonics.

Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., and Gerard, P. D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, 87(2):171–182.

Zwally, H. Schutz, H., Bentley, C., Bufton, J., Herring, T., Minster, J., Spinhirne, J., and Thomas, R. (2011). *GLAS/ICESat L2 Global Land Surface Altimetry Data, V33*. Boulder, Colorado USA: National Snow and Ice Data Center.

Zwally, H. J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., et al. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34(3):405–445.