Biophysical parameter retrieval from satellite laser altimetry.

Rosette, Jacqueline

Biophysical Parameter Retrieval from Satellite Laser Altimetry

Jacqueline Rosette

Department of Geography

Swansea University

Thesis submitted to the University of Wales in fulfilment of the requirements for the degree of Doctor of Philosophy

2009

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 .....15th June 2009.....

Statement 1

This thesis is the result of my own independent work, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ... .(candidate)

Date .....15th June 2009.....

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 .....15th June 2009.....

Summary

Quantifying and monitoring vegetation distribution and change are fundamental to carbon accounting and requirements of national forest inventories. This research explores the potential of the Geoscience Laser Altimeter System (GLAS), launched in 2003 by NASA as the first global Earth surface satellite LiDAR mission. The project study site is the Forest of Dean, Gloucestershire, UK, a highly mixed, temperate forest with varied topography.

Methods are developed to distinguish the regions within waveforms returned from vegetation and ground. When compared with field measurements, estimation of canopy height gives a correlation of R^2=0.92; RMSE=2.81m.

Waveform indices are determined and evaluated with respect to their potential to estimate biophysical parameters. Heights of cumulative energy percentiles within the waveform prove to be significant estimators. When compared to calculations from independent yield models, results show correlations with stand-level top height (R^2=0.76; RMSE 3.9m) and stemwood volume (mixed composition stands dominated by broadleaves: R^2=0.47, RMSE=75.6m^3/ha; conifers: R^2=0.66, RMSE=82.5m^3/ha). Uncertainty analysis is undertaken of both waveform and yield model estimates.

Canopy cover is estimated for the area beneath GLAS waveforms, corrected for differences in reflectance for ground and canopy surfaces. These are assessed against airborne LiDAR estimates, validated using hemispherical photography. The method produces results with R^2=0.63; RMSE=11% for stands with greatest coverage by broadleaves and R^2=0.41; RMSE 16% for conifer-dominated stands.

Small footprint airborne LiDAR (AL) is widely accepted to offer valuable data regarding forest parameters. An evaluation of AL and GLAS results demonstrate that the broad GLAS footprint dimensions allow similar estimation of stand-level parameters (e.g. AL/yield model Top Height: R^2=0.73, RMSE=4.5m). Direct comparison of GLAS with AL shows ground surface identification with mean difference of 0.32m and that elevation profiles correspond well (98^{th} percentiles R^2=0.76, RMSE=3.4m). Finally, prospects for use of LiDAR in carbon accounting, assimilation within models and for forestry applications are discussed.

Acknowledgements

First and foremost I would like to thank Peter North for his guidance, thought-provoking questions and the opportunities offered to me, to Sietse Los for his continuous support throughout the project, insightful feedback and encouragement & to Juan Suárez for his invaluable advice, enthusiastic interest and for generously sharing data and his own findings. Without their help and support, this work would not have been possible.

I also wish to thank the late Mike Barnsley for willingly finding time among his busy schedule; John Armston for making available to me the airborne LiDAR software which he developed; Steve Shaw for his extensive IT support and David Korn, at the National Snow and Ice Data Center (NSIDC) for responding to many data and information requests.

I would like to acknowledge the Forestry Commission, Forest Research, Biometrics Division, Woodlands Surveys Team for hosting me for three invaluable months at the Northern Research Station and the Forest of Dean District Office which kindly gave permission to carry out field work and provided background information, specific to the study site.

To Ewan, Rochelle, Damian and all of the other postgraduates and staff for the many laughs, entertainment and the happy times at Swansea and to my family and friends for supporting me throughout the years in all that I do.

Finally but fundamentally I am indebted to the Natural Environment Research Council (NERC) for having funded this research through the Climate and Land-Surface Systems Interaction Centre (CLASSIC).

Acronyms and Abbreviations

AVHRR	Advanced Very High Resolution Radiometer
BRDF	Bidirectional Reflectance Distribution Function
CHM	Canopy Height Model
CLASSIC	Climate and Land-Surface Systems Interaction Centre
CO_2	Carbon dioxide
DBH	Diameter at Breast Height
DESDynI	Deformation, Ecosystems and Dynamics of Ice mission
DSM	Digital Surface Model
DTM	Digital Terrain Model
FAO	Food and Agriculture Organization of the United Nations
FC	Forestry Commission
FLIGHT	Forest Light interaction model
GCM	General Circulation Model
GLAS	Geoscience Laser Altimeter System
GORT	Geometric Optical Radiative Transfer
GP_{MaxAmp}	Maximum Amplitude of Gaussian Peaks 1 or 2
ICESat	Ice, Cloud and land Elevation Satellite
IPCC	Intergovernmental Panel on Climate Change
JULES	Joint UK Land Environment Simulator
LAI	Leaf Area Index
LASER	Light Amplification by Stimulated Emission of Radiation
LiDAR	Light Detection and Ranging/ Laser-Induced Direction and Ranging

LIST	Laser Imaging for Surface Topography
LVIS	Laser Vegetation Imaging Sensor
NDVI	Normalised Difference Vegetation Index
NERC	Natural Environment Research Council
NPP	Net Primary Production
ns	Nanosecond
NSIDC	National Snow and Ice Data Center
RMSE	Root Mean Square Error
R_{WT}	Regression using Waveform Extent and a Terrain Index
SLA	Shuttle Laser Altimeter
SLICER	Scanning Lidar Imager of Canopies by Echo Recovery
TH	Top Height
TRIFFID	Top-down Representation of Interactive Foliage and Flora Including Dynamics
UNEP	United Nations Environment Programme
UNFCCC	United Nations Framework Convention on Climate Change
VCL	Vegetation Canopy LiDAR

Contents

List of Tables

List of Figures

Chapter 1. Introduction and Research Context

Remote sensing technologies are offering increasingly more advanced opportunities to inform and support the environmental sciences and to increase our understanding of the world in which we live. This expanding awareness and knowledge therefore creates an exciting environment in which to explore possibilities offered by a pioneering satellite sensor which provides a third dimension to remote sensing data that had previously not been possible on such scales.

1.1 Research Environment

This project contributes to current research needs through exploring the potential of an innovative means of reducing uncertainty in vegetation distribution. The work helps to address research questions identified by the Natural Environment Research Council within its Climate Change priority of improving knowledge of the properties of forest carbon sinks.

The project also meets research interests in the field of operational forestry, in particular Forest Research, the research agency of the Forestry Commission. Among its many duties, Forest Research is responsible for conducting the British National Forest Inventory, for informing silviculture practices and improving understanding of

underlying production processes. This research has involved a close collaboration with Forest Research in order to ensure the practical relevance of the work.

In this Chapter, an outline of the need for knowledge regarding vegetation biophysical parameter distribution is discussed in relation to improving our understanding of the Earth's processes and current conditions. Ways in which remote sensing, and in particular LiDAR, may address these are then presented. This provides the motivation for the aims and objectives of this research and the section concludes with a description of the structure of the thesis in order to meet these objectives.

1.2 Scientific Context to Research

The complex characteristics of the Earth's land surface result in great variability in elevation, slope, roughness, reflectance, vegetation height and vegetation structure, often over small spatial scales. These factors are significant in a broad range of Earth science disciplines and are the products of lithospheric, cryospheric, ecological and atmospheric processes whose combined effects have generated and continue to modify the Earth's ecosystems observed today. Identification and characterisation of these landscape parameters are necessary to assist in understanding the interplay between the processes leading to the formation and sustainability of features, to provide boundary conditions for incorporation within biosphere models and to improve our management and preservation of the environment.

Laser altimetry can potentially contribute to landsurface property descriptions which are currently lacking, such as improved representation of global topography,

surface roughness and vegetation heights at different spatial resolutions to meet a variety of needs.

1.2.1 Carbon Accounting and Vegetation Monitoring

The importance of quantifying and monitoring changes in carbon stock has been recognised in international agreements such as the Kyoto Protocol to the United Nations Framework Convention on Carbon Change (UNFCCC) 1997, whereby countries are required to report annually their direct human-induced emissions and removal of carbon dioxide - CO_2 (Broadmeadow and Matthews, 2003; UNFCCC, 2007). Estimating this requires prior knowledge of current carbon distribution. This distribution is currently largely unknown; although it is thought that a considerable store of unaccounted carbon is contained within temperate and boreal forests. Vegetation is an effective carbon sink, but impacts such as disease, wildfires, drought, flooding or anthropogenic actions (felling, pollution, landuse change, afforestation or increased atmospheric CO_2) can produce changes, thereby impacting on ecosystem processes. The significance of the role of forests in mitigating climate change is underlined by global initiatives, policies and incentives such as to encourage practices of Reducing Emissions from Deforestation and Forest Degradation in Developing countries - REDD (FAO *et al.*, 2008; UNEP, 2009) and to report on Land Use, Land Use Change and Forestry – LULUCF (IPCC, 2003).

The terrestrial carbon sink is thought to be greatly reduced and LiDAR therefore offers the opportunity of comprehensively mapping carbon stores within vegetation (if not litter and soils), in addition to observing vegetation regeneration and improving understanding of carbon flux patterns between the land, atmosphere

and oceans over space and time (Hese *et al.*, 2005). In fact, following discussions relating to an international forest monitoring network during the 2008 G8 Summit, as part of the Global Forest Resources Assessment 2010, optical and radar remote sensing are to be adopted as a means of recording distribution of forests and monitoring changes to forest extent (FAO, 2007). LiDAR remote sensing may offer a future contribution to this goal. This monitoring is fundamental as land use, forestry and natural vegetation responses to atmospheric CO_2 have a great effect on greenhouse gas emissions. These in turn can be seen as being not only of global environmental concern but also of economic impact e.g. conservation, fuel stocks, sustained population levels and land use.

1.2.2 Model Improvement

Vegetation plays a significant role in global climate and biogeochemical cycles, particularly concerning carbon, with approximately one quarter of atmospheric carbon dioxide fixed annually as gross primary production. To accurately model this and other land surface processes in General Circulation Models (GCM), properties such as radiation absorption, plant physiology, surface characteristics and climatology are required. These multi-temporal global datasets are only possible from remotely sensed sources (Myneni *et al.*, 1997).

Biosphere Models

Computer-generated models of the biosphere provide a valuable means to improve understanding of the immensely complex interactions between interdependent systems affecting the Earth. By their very nature, models function as

generalisations of reality and a series of component models replicating the interplay of systems often provide input to complex broader-themed Biosphere models.

Dynamic Vegetation Models are particularly valuable in enabling prediction of the carbon balance under changing ecosystem structure and composition brought about by climatic changes (Hese *et al.*, 2005). Vegetation is often represented within each grid cell as generalised Plant Functional Types and climate-driven habitat changes are used to model vegetation succession and plant lifecycle. Examples include the Lund-Potsdam-Jena dynamic global vegetation model (Sitch *et al.*, 2003), the Top-down Representation of Interactive Foliage and Flora Including Dynamics (TRIFFID) which provides inputs to JULES, the Joint UK Land Environment Simulator (Cox and Best, 1999; Cox *et al.*, 1999; Purser and Anderson, 2000; JULES, 2008), and the Simple Biosphere Model modification (SiB2) (Sellers *et al.*, 1996a; Sellers *et al.*, 1996b; Denning_Research_Group, 2005).

Biome-specific model parameters relating to vegetation state and phenology may be obtained directly from processed satellite data (Denning_Research_Group, 2005). For example SiB land cover class biophysical parameters such as canopy height, leaf angle distribution and photosynthesis-related factors are considered static with time and are derived from literature sources and look up tables whilst the biophysical parameters LAI (leaf area index) and fAPAR (fraction of photosynthetically active radiation absorbed by vegetation) are permitted to vary with time and are derived from relationships with normalised difference vegetation index (NDVI) described by Sellers *et al.*, (1996b); Myneni *et al.*, (1997); Los, (1998).

LiDAR may potentially support models by allowing direct satellite input of canopy height, fractional canopy cover and light penetration parameters for example

or by permitting comparison with NDVI-derived LAI and fPAR or relating NDVI to vegetation height (Los *et al.*, submitted). Alternatively LiDAR may provide a time effective, non-destructive means of calculating LAI, a factor which is fundamental to ecosystem modelling due to its contributions to biophysical cycles (Chen *et al.*, 2004; Riano *et al.*, 2004).

Radiative Transfer Models

The ability of laser altimetry to retrieve vertical height profiles of vegetation, additionally lends itself to applications with Radiative Transfer models which simulate remotely sensed data to better understand and reproduce the influence of vegetation structure on radiation scattering and absorption.

Vegetation structure and, from this, LAI and fPAR, are important determinants of vegetation productivity. The structural variation both within and between vegetation canopies and subsequent multi-scattering events necessitates a three-dimensional approach to reflectance simulation. LiDAR may provide a means of comparing modelled and observed three-dimensional vegetation structure and consequent sensitivity of reflectance.

Radiative transfer modelling has been used for six global vegetation classes based on vegetation structure to obtain relationships of LAI and fAPAR from AVHRR NDVI reflectance calculations (Myneni *et al.*, 1997). Satellite LiDAR may offer a further means of validating the distribution of these vegetation canopy structural classes and calculated parameters.

Conversely, radiative transfer modelling may improve understanding of the complex interactions which determine waveform characteristics by exploring the

sensitivity of waveforms to vegetation and ground surface structural and reflectance properties (Ni-Meister *et al.*, 2001; North *et al.*, submitted).

1.2.3 Forest Management

The research during the course of this project seeks particularly to assist the needs of forest management and monitoring. The management of forests in Britain is undertaken with consideration of multiple purposes including production of timber and other commodities, recreation, wildlife habitat conservation, water quality protection, open space preservation and as a future buffer against climate change. This requires reliable data regarding forest resource conditions and trends (Broadmeadow and Matthews, 2003; Forest_Research, year unknown).

Challenges posed by such necessities of accounting and subsequent monitoring of forest stocks include the costs and impracticalities associated with extensive application of conventional field survey methods, where uncertainties would nevertheless remain due to inaccessibility to privately-owned woodlands. In terms of timber production forecast, this may be significant as the private sector accounts for approximately 60% of British woodlands and currently 80% of timber production (Forestry_Commission, 2003; Forest_Research, 2006; Forestry_Commission, 2007).

Currently, the process of National Forest Inventory in Britain relies on large quantities of field measurements and photo interpretation carried out manually in order to delineate forest and non-forest areas (Forestry_Commission, 2003; Broadmeadow and Matthews, 2004; Forest_Research, 2006; Gilbert, 2007). LiDAR may therefore provide a means to remotely quantify vegetation for inaccessible

private woodlands in order to incorporate within the collection of data for the national forest inventory. Vertical information contributed by LiDAR may assist with this classification and the accuracy of assessments.

1.2.4 Opportunities Afforded by Remote Sensing

Earth observation plays a significant role in enhancing the knowledge of vegetation distribution and its biophysical properties. Optical remote sensing has been shown to provide indirect associations using reflectance properties for NDVI (e.g. Defries and Townshend, 1994; Los *et al.*, 1994; Sellers *et al.*, 1994) and LAI (e.g. Los *et al.*, 2000; Myneni *et al.*, 2002) although there are difficulties associated with use of reflectance alone (North, 2002). More direct assessment using the physical properties of vegetation and interactions between canopy structure and photons detected by the sensor have been demonstrated through radiative transfer modelling (North, 1996; Myneni *et al.*, 1997; Ni-Meister *et al.*, 2001; Kotchenova *et al.*, 2003).

Active instruments have been exploited to estimate vegetation height and volume using radar (Balzter *et al.*, 2003; Gaveau *et al.*, 2003; Tansey *et al.*, 2004; Balzter *et al.*, 2007) and, more recently, LiDAR (Light Detection and Ranging) systems which estimate these using physical interactions between intercepted surfaces and photon paths.

In terms of carbon accounting, an important limitation of remote sensing estimates is that direct account is only taken of above ground biomass whilst root systems form a substantial contribution to carbon storage. However, Jenkins *et al.*, (2003) have shown a means to overcome this using USA-wide allometric equations

for estimating above-ground biomass for broad species classes and found generalised relationships accounting for vegetation component biomass (root, foliage, stem bark and stemwood). This potential has also been demonstrated by Fang *et al.*, (1998) who have related stem volume density (m^3/ha) to stand biomass (Mg/ha) comprising both above and below-ground biomass. Such methods may offer a means of more accurately mapping biomass distribution if remote sensing techniques such as LiDAR are able to retrieve appropriate biophysical parameter estimates from forests.

This research therefore aims to develop replicable methods for estimating significant forest biophysical parameters using broad vegetation classes which may be identified using landcover maps (Hill and Smith, 2005), forest inventory interpretive forest types or the Forestry Commission management database and which may permit further stratification or specific relationships to be explored in the future.

1.3 Aims and Objectives

The aim of this thesis is to assess the potential for forest biophysical parameter retrieval offered by satellite LiDAR remote sensing. Limitations are examined with respect to instrument specification, sites suitable for the application of approaches and capabilities presented by related available technologies.

In order to achieve this, the project objectives are:

I. To gather field data with which to assess and validate methods of waveform processing.

II. To develop and evaluate techniques for deriving forest parameters from large footprint, satellite LiDAR waveforms.

III. To assess estimated parameters using field measurements, airborne LiDAR remote sensing data and forestry yield models with management database information.

IV. To explore capabilities of LiDAR remote sensing of vegetation in the context of forest inventory and management.

The research seeks to develop methods which offer means of estimating vegetation parameters of interest to practitioners and modellers which can be replicated and applied to different vegetation types and on regional to national scales. Results are considered with a view to determining how conventional field and model-based methods may be complemented or enhanced. These aspects are explored within the coming Chapters, as outlined below.

1.3.1 Thesis Structure

Chapter 2 introduces the context to the research with respect to current research gaps, information provided by alternative remote sensing systems and the state of the art regarding LiDAR of vegetation.

Chapter 3 discusses the principles underlying the GLAS sensor used within this project and the pre-processing of data which is undertaken prior to their release to the public.

The sources of data and field methods applied to validate LiDAR processing results are presented in **Chapter 4**. Methods developed to address specific research questions are described within the associated sections, Chapters 5-7.

Chapter 5 explores and tests methods of estimating vegetation height from within large footprint LiDAR waveforms. Two techniques of identifying the region of the waveform returned from vegetation are subsequently used throughout the remainder of the project.

These methods are developed further and applied in **Chapter 6** in order to estimate vegetation top height and stemwood volume, two parameters commonly used within forestry.

Chapter 7 presents a comparison of satellite LiDAR results with those obtained using airborne LiDAR data. Situations in which the latter surpasses the capabilities of large footprint LiDAR systems are introduced.

Conclusions to the research are drawn in **Chapter 8** which also looks to further prospects for ICESat/GLAS or successor sensors. Future directions for research and applications of the data are suggested.

Chapter 2. LiDAR Remote Sensing

This Chapter begins with a historical context to LiDAR remote sensing. Principles common to the airborne and satellite LiDAR systems used within this project are introduced. Specific aspects relating to small footprint airborne laser scanning and large footprint, satellite LiDAR profiling are then presented. An overview of previous large footprint systems and future potentialities of satellite LiDAR missions are discussed. Finally, an overview is given of relevant LiDAR applications demonstrated by other authors.

2.1 LiDAR Historical Background

The technique of laser altimetry has a relatively recent history. The American physicist, C. H. Thomas developed the first microwave laser (MASER) in 1954, publishing a paper in 1958 together with A. L. Schawlow, exploring possibilities of laser exploitation in the visible and infrared spectrum. In 1960, fellow American T. H. Maiman developed the first successful optical laser using a ruby crystal. Airborne laser ranging was first attempted in the 1960s, however it wasn't until the early-to-mid 1970s that this began to be developed and, in the late 1980s, GPS developments permitted accurate geolocated range measurements on a larger scale. It was only since the mid 1990s that the first commercial systems began to appear (Wehr *et al.*, 1999; Earth_Science_Office, 2005).

The last fifteen years have seen the application of laser altimetry to the planetary sciences with fundamental impact: Spaceborne laser ranging on the Clementine Mission (1994) produced global, large-scale topographic relief images of the Earth's moon; the Near Earth Asteroid Rendezvous Mission (1996) supported the Shoemaker Laser Rangefinder which provided a complete topographic map of the 433 Eros asteroid to metre-level resolution; whilst the Mars Orbital Laser Altimeter, part of the Mars Global Surveyor mission (1996), generated high resolution maps of Martian topography, surface roughness and surface change, greatly enhancing the understanding of the evolution and composition of the planet (Luthcke *et al.*, 2002b).

However, this spaceborne technology has only begun to be significantly exploited in the last five years to advance our understanding of the Earth and the processes that combine and contribute to shape it. Therefore the exploration of this technology and its applications to improve our knowledge of the three-dimensional distribution of features, may offer great potential to complement and enhance current remote sensing approaches.

2.2 LiDAR Technical Aspects

2.2.1 LiDAR Principles

Laser is an acronym for Light Amplification by Stimulated Emission of Radiation (Federation_of_American_Scientists, year unknown). There are two main classes of laser; pulse and continuous wave lasers (Wehr *et al.*, 1999). The technique of interest to this project utilises pulsed lasers which transmit very short single bursts of high intensity light to a target. The backscattered signal returned from atmospheric

components or reflected from Earth-surface objects allows properties of those intercepted surfaces to be determined. The principle is similar to that of pulsed microwave radar, although it uses optical not microwave components to direct, intercept and detect the emitted radiation. LiDAR is defined as Laser-Induced Direction And Ranging or, more usually, Light Detection And Ranging (Charlton *et al.*, 2003; Ordnance_Survey, date unknown), the shorter wavelength range offers the opportunity of detecting atmospheric constituents which would be beyond the capabilities of radar for example (Molero and Jacque, 1999).

Many of the principles of satellite laser altimetry have developed from twenty years experience of radar altimetry (Brenner *et al.*, 2003). The principles have many similarities although significant differences occur due to different wavelengths and beam width, influencing the height and roughness calculations. The wider radar beam (10-20 km) causes surface slope to have a great effect and surface roughness will broaden the received pulse, meaning this often dominates the shapes of the radar pulse. A further complication is radar's ability to penetrate surfaces such as firn meaning the pulse may be returned from some depth below the surface.

In contrast, laser altimetry has a relatively smaller footprint meaning that returns are reflected from a more identifiable position on the surface (Figure 2.1). It also operates at a higher electromagnetic frequency preventing the signal from penetrating below the surface. This may therefore offer a more appropriate means of estimating topographic and vegetation properties. The characteristics of LiDAR systems and applications of data are discussed later in this Chapter.

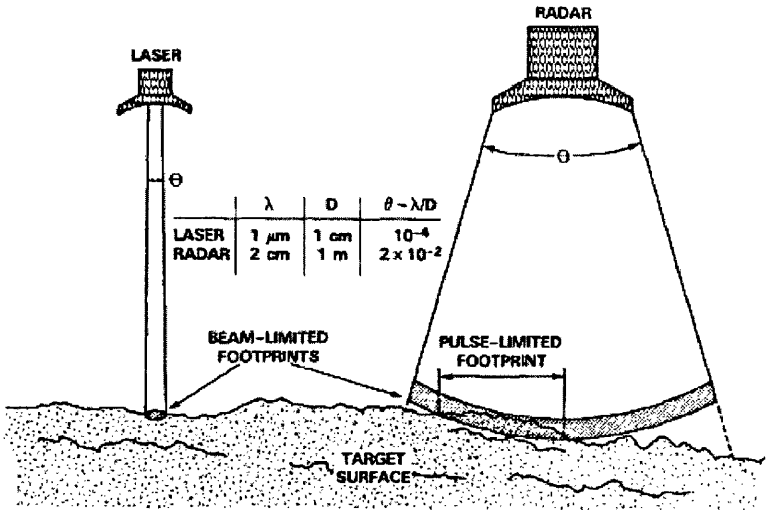


Figure 2.1 Contrast between LiDAR and Radar wavelength, beam divergence and consequent footprint size at the ground surface (Source: Bufton, 1989)

2.2.2 LiDAR Instruments

LiDAR instruments may be divided into DIAL, Doppler and Range Finder systems:

DIAL

Differential Absorption LiDAR (DIAL) can be used to assess concentrations of atmospheric constituents (e.g. ozone, water vapour, pollutants). Two different laser wavelengths are selected of which, for the component of interest, one will be absorbed whilst the other scattered. The concentration of the atmospheric components is therefore determined by the difference in intensity of the two return signals (Earth_Science_Office, date unknown).

Doppler

Doppler LiDAR measures the velocity of the target and is based on the principle that, if the laser pulse intercepts an advancing or receding target, the returned wavelength will be slightly modified (the Doppler shift); the wavelength

will be shorter if the object is advancing and longer if it is receding. One example of such an application is the accurate remote measurement of wind velocity through detection of suspended dust and aerosol particles transported by the wind (Earth_Science_Office, date unknown).

Range Finder

Range Finders form the simplest group and are used to measure the distance between the LiDAR instrument (altimeter) and the target surface(s) (Earth_Science_Office, date unknown). It is this latter group of LiDAR systems which is of interest to this study. Within this group, satellite, airborne and terrestrial platforms are found, successively producing increasing data spatial resolution. The area illuminated by each laser pulse is known as the 'footprint'.

2.2.3 Determining Range and Time

LiDAR range-finding involves the emission of laser pulses from the instrument positioned on a platform, towards a target (e.g. ground), from which it is reflected from the intercepted surfaces. Features further from the sensor will intercept the laser energy later than those closer to it.

The time for the pulse echo to reach the sensor is measured and, using the fact that the laser pulse travels at the speed of light, the total return distance travelled between the sensor and the intercepted surfaces can be calculated. The distance between the altimeter and the intercepted object is therefore half of this value:

$$R = c.t/2 \qquad (2.1)$$

where R is the Range, c is equal to the speed of light ($3x10^8$ ms^{-1}) and t is the time elapsed between the pulse emission and returned signal detection (Baltsavias, 1999; Wehr *et al.*, 1999; Abshire, date unknown).

LiDAR time units are generally in nanoseconds (ns), each being equal to approximately 15cm in one-way distance between the sensor and target. This permits the three-dimensional reproduction of Earth surface relief and above-surface object structures (e.g. vegetation, ice cover, atmospheric aerosols and cloud structure).

Range resolution ΔR is directly proportional to time resolution Δt and can therefore be derived as follows (Wehr *et al.*, 1999):

$$\Delta R=0.5c\Delta t \qquad (2.2)$$

Time is measured by a time interval counter, initiated on emission of the pulse and is triggered at a specific point on the leading edge of the returned pulse. This position is not immediately evident and therefore is set to occur where the signal voltage reaches a pre-determined threshold value. The steepness of the received pulse (rise time of the pulse) is a principal contributory factor to range accuracy and depends on the combination of numerous factors such as incident light wavelength, reflectivity of targets at that wavelength, spatial distribution of laser energy across the footprint and atmospheric attenuation (Baltsavias, 1999). The return pulse leading edge rise time is therefore formed by the strength of the return signal from the highest intercepted surfaces within the footprint. This will vary with the nature of the surface; flat ice sheets producing abrupt returns with fast leading edge rises and multilayered, complex vegetation creating broad returns (Harding *et al.*, 1998; Ni-Meister *et al.*, 2001).

The leading edge of the pulse is one of the factors explored during the course of this research as a contributor to uncertainty within biophysical parameter estimates from waveforms.

To avoid confusion of the pulses arriving at the time interval counter, pulses are not usually transmitted until the previous pulse echo has been received. In theory this could limit flight factors such as the altitude of the altimeter (i.e. for a pulse rate of 25kHz, the maximum range is 6km), although in practice this is rarely limiting. Rather, limits are often formed by laser power and beam divergence; atmospheric transmission; target reflectivity; detector sensitivity; increasing influence of altitude and attitude errors in position accuracy (Baltsavias, 1999).

2.2.4 Data Acquisition Criteria

The minimum detectable object within a laser footprint depends primarily on its reflectivity as opposed to its size. If, for example, a sensor is able to measure the distance to a flat, even surface of area A, with a reflectivity of 5%, then the minimum area of a detectable object with 100% reflectivity would be $A/20$. However, other factors combine to influence object detection, among which are range; laser power; atmospheric conditions; background irradiation; whether target reflectivity is diffuse, specular or a combination; terrain inclination; target three dimensional structure; laser wavelength; laser aperture, detector sensitivity and noise level (Baltsavias, 1999).

GPS receivers and an Inertial Navigation System (INS) on board the platform plus direction in which the altimeter was pointing, together identify the precise

geographical co-ordinates of the footprint. In the case of aircraft-mounted altimeters, scanning can occur at right angles to the direction of travel.

In addition to topographic lasers, emitting light in one waveband to provide elevation data, bathymetric lasers emit in two wavelengths, generally 1064nm and 532nm. This would enable details of the pulse path to be determined e.g. whether atmospheric moisture may be present which would affect the returned signal (Wehr and Lohr, 1999).

The minimum height at which scanning can usually be performed may be determined by platform specifications, flight regulations (e.g. aircraft restrictions over urban areas) and eye safety range. The most sensitive detectors are available between 800nm and 1000nm wavelengths, but at these optical wavelengths, eye safety is still a concern. Therefore if greater pulse energy is needed, wavelengths at which the eye is less sensitive are required. Reflectance response of the target surface of interest at different wavelengths is also an important consideration in relation to the information the mission aims to retrieve e.g. snow and ice reflect weakly at 1535 nm (Wehr and Lohr, 1999).

With all other conditions remaining constant, the maximum range is generally proportional to the square roots of reflectivity and of the laser power. Best range performance is achieved when the atmosphere is cool, dry and clear, as shorter wavelengths experience greater scattering by atmospheric water vapour, CO_2, dust particles or smoke. These conditions may occur at night rather than in bright sunlight (Baltsavias, 1999).

Greater sensor altitude will increase diameter in addition to reducing density of footprints. LiDAR footprint size is important as small footprints may underestimate the true canopy height (Lim *et al.*, 2003b) due to a lower probability

of capturing the canopy top, whilst larger footprints mean the percentage of the total return contributed by the canopy top is small, decreasing waveform leading edge steepness and therefore increasing the possibility of misallocating the beginning of the signal. This may be particularly the case for mature vegetation with large height variation.

Difficulties posed by wide footprints, if greater than the crown size of dominant individuals, are a resulting underestimation of upper canopy variability (Lefsky *et al.*, 1999a) or topographic variation (Hofton and Blair, 2002) although increased point density may counteract this effect (Morsdorf *et al.*, 2004). The GLAS sensor used in this research produces footprints of approximately 64m diameter (Abshire *et al.*, 2005). The ability to retrieve canopy height from waveforms and to identify representative ground elevations are tested during the course of this research. Upper canopy height variability and slope are considered as a potential factors influencing uncertainty in large footprint LiDAR estimates.

Broad footprints (up to crown size) have the advantages of increased probability both of sampling canopy tops and for the laser energy to reach the ground. For airborne LiDAR campaigns, expense is also reduced per unit area due to the larger swaths at higher altitudes (Drake *et al.*, 2002) however height estimate accuracy and area coverage are sensitive to LiDAR campaign specifications (Goodwin *et al.*, 2006; Takahashi *et al.*, 2008).

Many factors therefore combine to determine the accuracy and data specifications of LiDAR campaigns. Laser altimetry provides a relatively economical, time-effective means of generating accurate vertical surface information for both small and large areas.

2.3 Laser Altimetry Systems

This study focuses on large footprint, full waveform, satellite LiDAR profiling though also incorporates a comparison with small footprint, discrete return, airborne laser scanning. The main principles of both are outlined below.

2.3.1 Discrete Return Laser Scanning

Laser scanners are often mounted on aircraft or helicopter platforms and produce a dense coverage of relatively small footprints (generally of sub-metre diameter) in swaths perpendicular to the direction of travel (Figure 2.2).

Footprint spacing across track is established by aircraft altitude and angular separation between successive transmitted pulses, itself affected by aircraft roll. Along-track footprint spacing is determined by pulse repetition rate, number of cross-track footprints, aircraft ground speed and pitch attitude (Harding *et al.*, 1998; Baltsavias, 1999).

Laser altimeters are currently unable to perform several range measurements simultaneously, however between 2000 - 100,000 range measurements per second can be performed, each to a different surface location. Of the pulse returned to the sensor, discrete return devices often record only first and last echoes (or more recently some intermediate echoes). This technique allows dense point clouds to be produced with which processing is undertaken (Figure 2.3).

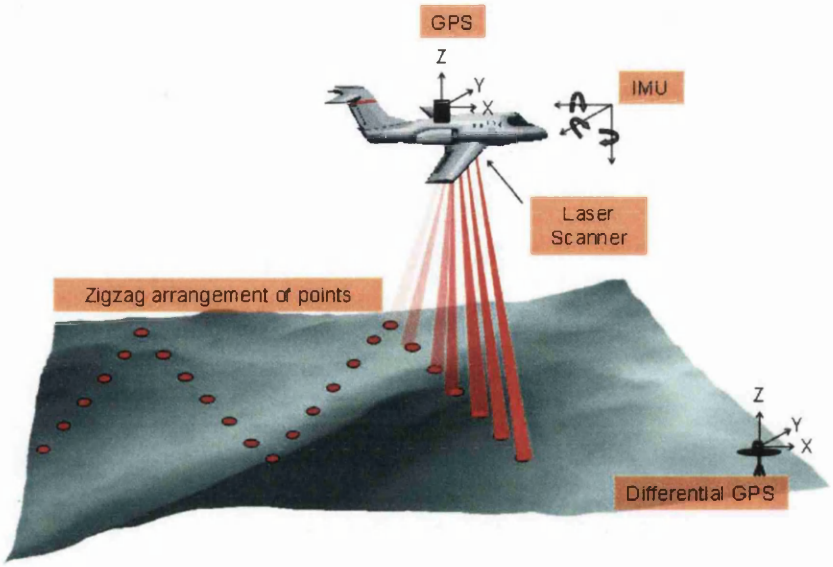


Figure 2.2 Illustration of airborne laser scanning pattern along the direction of travel due to the oscillating scanning mirror plus on-board Geographical Positioning System (GPS) and Inertial Measurement Unit (IMU) which records aircraft motion (speed and attitude – pitch, roll and yaw) to determine accurate sensor geolocation. (Image courtesy of Forest Research)

Intensity of these echoes is also recorded. Intensity is recognised as a complex unitless parameter which is determined not only by target reflectivity but is also strongly influenced by range, object orientation, target structure and scanning angle (Boyd and Hill, 2007).

LiDAR campaign specifications can be modified to accommodate application requirements; e.g. detailed 3D urban reconstruction or coarse resolution digital elevation models. During a survey flight, position and orientation system data, the laser data co-ordinates, ranges and intensities, possibly with associated scan angles, are retrieved.

Figure 2.3 Airborne LiDAR point cloud illustrating point density with up to four recorded echoes per emitted pulse producing density of approximately 25 points per square metre. Data are shown as unprocessed echoes using Forest Research-commissioned data for Glen Affric, Scotland (First, second and third echoes are shown as progressively deeper shades of green and last returns are illustrated in white).

These geographically-registered data can be processed and incorporated within a map co-ordinate system. This results in a cloud of semi-randomly distributed laser points in elevation and horizontal position. The distribution and density of these points is determined by the laser system scanning pattern, platform height, velocity, field of view and sampling frequency as well as target surface topography (Axelsson, 1999; Wehr and Lohr, 1999; Skinner, 2002).

The raw LiDAR data points are therefore irregularly spaced. This is further exacerbated with airborne scanning as this occurs across the direction of travel. The mirror used to guide the outgoing beam slows to change direction at the edge of each swath. This results in a concentration of data points at the end of each scan line (Cobby *et al.*, 2001; Skinner, 2002).

Data are filtered to remove noise, anomalous returns and measurement errors. In the case of vegetation applications, airborne LiDAR point clouds may then be classified into ground (Zhang *et al.*, 2003) and canopy returns above this surface.

Interpolation of data to produce a digital terrain model or digital surface model (of the upper canopy surface) serves to resolve the irregular dispersion of points through producing a regular, resampled grid. However interpolation can be expected to have some uncertainty especially at the centre of the swath where point density is lower and where there is no overlap with adjacent swaths. The work for this project adopted this approach in the processing of airborne LiDAR data for canopy delineation (following Suárez *et al.*, 2008a) as well as utilising raw returned points above the interpolated ground surface for analysis of canopy metrics and percentage cover (Armston *et al.*, in preparation 2008).

Airborne laser scanning offers the opportunity for retrieval of detailed vegetation parameters for entire forests, a task which would not be feasible using traditional field techniques. Generally speaking, cost restrictions limit this form of high point density LiDAR acquisition to relatively small areas suitable for forest-level data capture. The level of detail obtainable permits analysis at both a stand level and individual tree level and the unrivalled ability to 'see' beneath the forest canopy.

2.3.2 Full Waveform LiDAR Profiling

Waveform recording systems follow the same principles as discrete return devices only recording the entire returned signal (waveform), or a 'window' of which in the case of high altitude platforms. Rapid sampling frequencies of 1GHz (1 billion samples per second) permit the entire echo waveform to be stored for analysis, providing breakdown of vertical structure with great accuracy (GeoLas_Consulting, date unknown).

The entire echo waveform of each individual emitted laser pulse can provide information concerning vegetation density by height, surface slope and roughness for example (Figure 2.4), in addition to inferred reflectance characteristics. The latter is in the form of waveform amplitude, but is complicated by issues of complex interactions with intercepted surfaces as outlined for airborne LiDAR intensity values above and, in particular for high altitude platforms, by signal attenuation through the atmosphere.

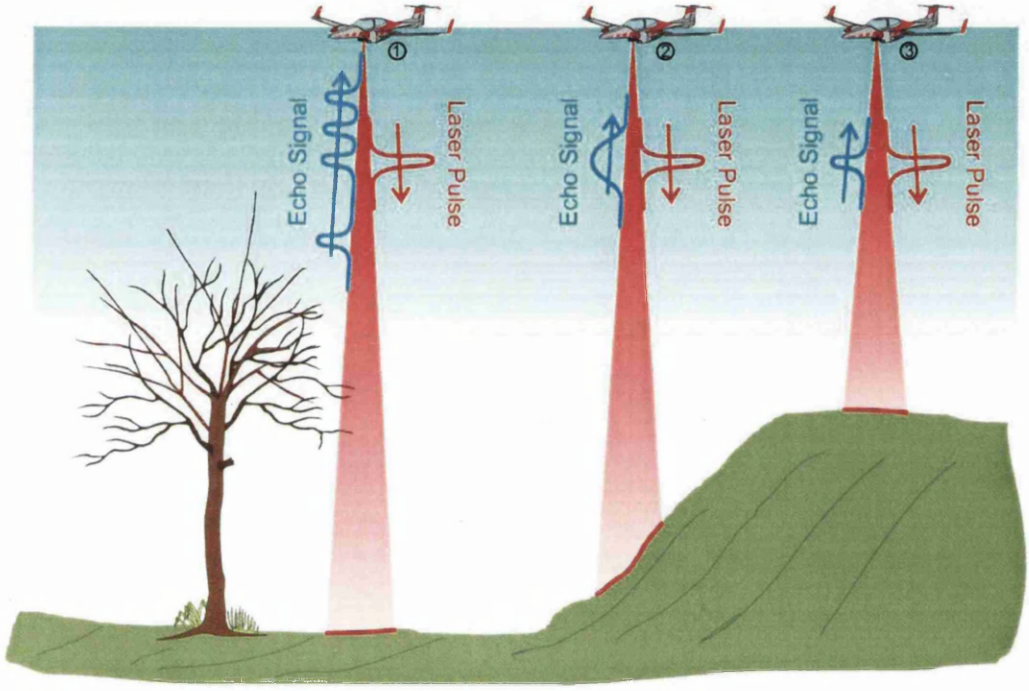


Figure 2.4 Representation of returned echo signals for different target characteristics. (left) Multi-echo returned signal from vegetation surfaces; (centre) returned pulse broadened from sloped surfaces; (right) narrow pulse width from a bare, flat surface. Source: Baltsavias, (2007)

Despite the fact that the power of the return signal diminishes with increased canopy depth, in the case of large footprint diameters, for relatively low relief, ground returns may be expected in almost all footprints (Parker *et al.*, 2001). The reduction in laser energy with depth is due to reflectance and absorption through the canopy. Some studies have applied the MacArthur-Horn algorithm to compensate for this by weighting a normalised cumulative distribution of waveform energy by

-1×ln(1-closure) in order to transform this into a cumulative distribution of canopy area (Harding *et al.*, 2001)

Until recently, full waveform LiDAR systems generally produced coarser mid-to-large footprint diameters of 10-100m spatial resolution compared with small footprints of discrete return devices. Nevertheless, usage was generally restricted to research bodies primarily due to acquisition costs and data storage limitations (Todd *et al.*, 2003). The complex interactions with intercepted surfaces, and infinite variations of spatial distribution of these within footprint areas, create a challenging environment within which to form generalisations of ground and vegetation feature representation within large footprint waveforms (e.g. for surface classification).

Broad LiDAR footprints may be produced in swaths (e.g. SLICER – Section 2.3.2.1) or sequentially along the ground track – LiDAR profiling (e.g. GLAS – Chapter 3). However, most recent advancements have seen the advent of small footprint, full waveform LiDAR (Figure 2.5) and the beginnings of multi-spectral LiDAR systems (Morsdorf *et al.*, 2008), as computer processing power has become more advanced and able to cope with the vast volumes of data produced.

This project focuses on large footprint, LiDAR profiling using the Geoscience Laser Altimeter System (GLAS) and opportunities and limitations offered by this. The evolution of this instrument in terms of previous systems is summarised next.

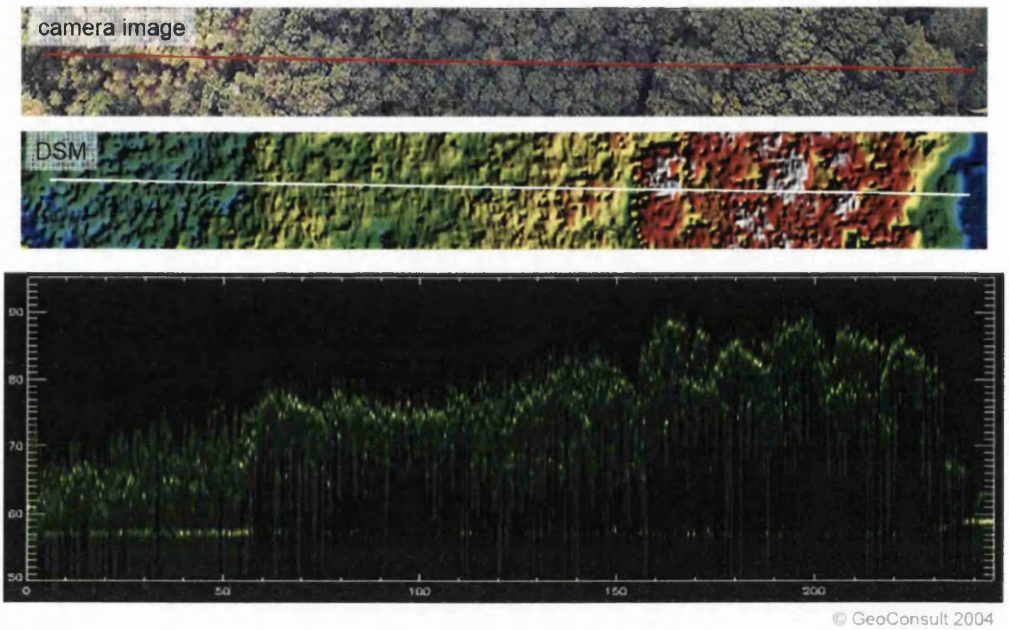


Figure 2.5 Shown from top to bottom, an example of a transect across a vegetated site showing aerial photography of varying crown roughness, a digital surface model (DSM) of the uppermost canopy surface derived from LiDAR data and small footprint, full waveform LiDAR data illustrating laser energy penetration through the canopy to the ground. Image produced by Geoconsult 2004, source: Baltsavias, (2007).

2.3.2.1 ICESat/GLAS Predecessors

The following large footprint, full waveform LiDAR sensors may be regarded as predecessors to the GLAS sensor currently in orbit upon the Ice Cloud and land Elevation Satellite (ICESat). Although specifications may differ, these instruments served to provide proof of concept for biophysical parameter retrieval using such systems.

Scanning LiDAR Imager of Canopies by Echo Recovery

The Scanning LiDAR Imager of Canopies by Echo Recovery (SLICER) is an airborne laser altimeter developed within NASA's Goddard Space Flight Center to support its Topography and Surface Change Program, Terrestrial and Ecology Program and the Boreal Ecosystems Atmosphere Study (BOREAS). Experience

derived from SLICER has been incorporated in the production of the LiDAR Vegetation Imaging Sensor (LVIS) which has since superseded SLICER.

The pulse of SLICER differs from LVIS, the Vegetation Canopy LiDAR proposal (VCL) and GLAS in being of Rayleigh form as opposed to Gaussian. This affects the distribution of laser power across the footprint which is not consistent but decreases towards its edge, in practice decreasing the effective size of the footprint.

SLICER was designed to have a fast leading edge rise time and high peak power in order to achieve high ranging accuracy to multiple targets within the footprint. Aircraft supporting the altimeter generally fly at a constant altitude resulting in footprint diameters which vary with relief of the underlying ground.

Laser Vegetation Imaging Sensor

The Laser Vegetation Imaging Sensor (LVIS) is a medium-altitude airborne sensor. It was developed by the NASA Goddard Space Flight Center in the late 1990s to succeed SLICER and as a trial platform for concepts and instrumentation planned for the proposed spaceborne Vegetation Canopy LiDAR mission. It is capable of operating at altitudes up to 10km and, with its 7° field of view, it can produce swaths up to 1km wide with 25m diameter footprints (although footprints up to 80m are possible). As such, coverage demonstrates the ease of acquisition over large areas and potential for large scale landscape mapping and monitoring. Laser pulses at 1064nm wavelength and of less than 10ns duration, are emitted at repetition rates up to 500Hz.

LVIS differs from previous air and spaceborne LiDAR systems which calculate range to the surface using separate waveform digitisers and time interval units. LVIS uniquely utilises only a detector and oscillator for all roles, thus

eliminating any inconsistencies which may exist between instruments and allowing the return pulse to be located precisely. The receiver is also capable of housing two additional detectors, thereby offering the capability of retrieving data at two wavelengths and polarisations.

As previously mentioned, and especially pertinent to wide footprints, difficulties can be posed by spatially varying energy distribution. The laser transmitter was designed to produce a Gaussian energy distribution and the output pulse is channelled through filters, allowing energy to be modified to optimise the strength of the return signal. Technological developments have therefore attempted to overcome weaknesses identified in preceding altimeters and were due to be incorporated into the spaceborne VCL (Blair *et al.*, 1999).

Shuttle Laser Altimeter

NASA's Shuttle Laser Altimeter (SLA) I aboard STS-72, 11th-20th January and SLA II upon STS-85 which was in orbit between 7th and 18th August 1997 took this technology into space for the Earth sciences (Carabajal *et al.*, 1999; Luthcke *et al.*, 2002a; Frawley *et al.*, 2005).

Few spaceborne LiDAR Earth Observation missions have been realised and the Shuttle Laser Altimeter is a modification of the Mars Observer Laser Altimeter (MOLA-1). Both SLA missions were of short duration, but they succeeded in obtaining coverage of latitudes between ±57°. The SLA produced 100m diameter footprints, 70m apart along the pass ground tracks, providing 342,000 geolocated land and 562,000 ocean surface returns (Carabajal *et al.*, 1999; Frawley *et al.*, 2005).

Despite the large footprint diameter, the SLA mission provided examples of waveforms produced by different surface types. However, their comprehensive use

to derive estimates of vegetation distribution has been limited, largely due to the limited spatial and temporal coverage. However, the mission has also served to improve footprint geolocation accuracy in preparation for GLAS (Luthcke *et al.*, 2002a).

Vegetation Canopy LiDAR

The proposed Carbon-3D space mission aimed for the first time to accurately estimate above-ground biomass on a global scale using satellite LiDAR.

There is great potential offered by the integration of LiDAR and optical remote sensing and the Mission proposed to position the Vegetation Canopy LiDAR (VCL) and obtain measurements of Bidirectional Reflectance Distribution Function (BRDF) on a single platform, thereby providing both vertical and horizontal data, permitting a 3D perspective and quantification of above-ground carbon stocks. Co-located BRDF data would allow the extrapolation of LiDAR footprints for complete spatial coverage and identification of vegetation characteristics, whilst LiDAR waveform analysis would provide fine-scale vertical canopy structure and surface target biophysical properties (through multi-angular optical observations). As 90% of above-ground carbon is contained in tree stems, this would permit biomass to be calculated as a function of tree height (Hese *et al.*, 2005).

The sensor configurations aimed to produce 25m diameter footprints along three tracks with 4km spacing between these using a swath of 8km. The satellite would orbit at 390-410km altitude. The spatial resolution of the BDRF multi-spectral data aimed to be <25m at nadir, corresponding to that of the VCL. The VCL proposal consisted of three near infrared laser beams capable of detecting returned signals from the canopy top and ground, even in dense vegetation cover.

This proposal was not funded, but the principles remain of relevance to this study. A future programme of similar extent and design providing global estimates of carbon distribution, fluxes and stability would have clear benefits to a wide range of applications (Hese *et al.*, 2005).

2.3.3 GLAS Proposed Successors

The success of the current ICESat/GLAS programme, discussed in following Chapters, and merit for continued future research and practical applications has been recognised in proposals for subsequent satellite LiDAR missions.

ICESat II is under development and proposed for launch in approximately 2015, following the original specifications of the first ICESat mission (Dubayah, 2008; Nelson, 2008). This would therefore allow data acquisition for monitoring vegetation change over an extended period of time, though is likely to result in a data gap between the lifespan of the current ICESat and the launch of its successor.

Different sensor specifications may produce variations in estimates which may be due to instrument effects as opposed to actual change. Research would therefore be needed to determine whether results are compatible. However, the following NASA sensors may prove to be more appropriate to vegetation applications than the ICESat missions.

DESDynI (Deformation, Ecosystems, and Dynamics of ICE) proposes to provide global spatially-continuous Earth observations using both multi-beam LiDAR profiling and L-band RaDAR data (Dubayah, 2008; Nelson, 2008). The LiDAR configuration under discussion would produce approximately 25m diameter footprints, with 25-30m spacing, thereby creating near-continuous profiles along

track. Possibilities include 3-5 beams with longitudinal distance between them of 2-5km.

The mission specification for LIST (Laser Imaging for Surface Topography) intends to produce adjoining footprints using swath mapping (Nelson, 2008). Footprints would be in the region of 5m diameter and the sensor would collect continuous global coverage data during its five year intended lifespan.

DESDynI is proposed for launch later than ICESat II with the timeframe for LIST being considerably later still. All missions are planned to be underway before 2020. Investment in these proposals demonstrates the value with which NASA regards obtaining repeat elevation data of the Earth's surface and its overlying features, using satellite LiDAR.

2.4 LiDAR Applications for Forest Inventory and Operational Forestry

The use of LiDAR has been demonstrated for many applications of retrieving information relating to vegetation. The majority of these have been from airborne platforms, although recent years have seen an increasing use of terrestrial and spaceborne LiDAR.

Forest biophysical parameter retrieval for stand level analysis has been demonstrated at local scales (e.g. Hyyppä *et al.*, 2001; Næsset, 2002; Patenaude *et al.*, 2004) and for ground surface relief (Jansma *et al.*, 2001; Hofton and Blair, 2002). Extracting forest floor topographic data can permit archaeological remains to be located which would otherwise not be visible (Crow *et al.*, 2007). Increasingly the benefits of data fusion, combining spectral data/aerial photography and airborne

LiDAR data, have also been shown to aide surface-type classification (e.g. Hill and Thomson, 2005; Suárez *et al.*, 2005a) and for the purposes of individual tree delineation (e.g. Suárez *et al.*, 2005a; Suárez *et al.*, 2008a) leading to its application to assess stand conditions using models (Morsdorf *et al.*, 2004; Suárez *et al.*, 2008a; Suárez *et al.*, 2008b).

The ability to characterise forest structure also lends itself to analysis of habitat conditions and to infer bird and animal distributions (Hinsley *et al.*, 2002; Hinsley *et al.*, 2006). Uniquely, LiDAR analysis also presents the opportunity to identify properties of understorey vegetation (Goodwin *et al.*, 2007; Hill, 2007) and therefore offers prospects for monitoring regeneration processes. Potential for this is also provided by LiDAR gap analysis (e.g. Gaulton and Malthus, 2008) which illustrates potential for the establishment of shrubs or tree saplings.

However, to date, few satellite LiDAR missions of the Earth's surface have been realised (Winker *et al.*, 1996; Zwally *et al.*, 2002; Frawley *et al.*, 2005). Resulting studies have involved the fields of atmospheric sciences (e.g. Spinhirne *et al.*, 2003; Dessler *et al.*, 2006; Yang *et al.*, 2008); glaciology (e.g. Smith, 2003; Herzfeld *et al.*, 2008; Zwally *et al.*, 2008) and topography (e.g. Garvin *et al.*, 1998; Carabajal and Harding, 2006; Atwood *et al.*, 2007). Studies concerning the performance of the ICESat/GLAS satellite LiDAR mission for vegetation analysis are an area of research development (e.g. Lefsky *et al.*, 2005; Lefsky *et al.*, 2007; Nelson, 2008; Simard *et al.*, 2008; Sun *et al.*, 2008) and this field of study is currently without a dedicated LiDAR campaign for this purpose. Indeed GLAS was designed primarily for determining ice volume changes. Such global LiDAR missions offer the opportunity for vegetation biophysical parameter retrieval at unprecedented scales (Hese *et al.*, 2005).

One of the principal attributes of LiDAR is the ability to estimate vegetation height profiles which presents opportunities of deriving further biophysical parameters. Of most interest to this project is the ability of LiDAR to derive woodland stand-level parameters for forestry or inventory purposes, particularly from large footprint, full waveform data. Experiences of relevant studies by other authors are outlined below.

2.4.1 Vegetation Height, Volume and Canopy Cover

Laser altimetry is currently the only technique capable of measuring tree heights in closed canopies and therefore offers a remote and non-destructive means of estimating vegetation volume, biomass or carbon content to account for vegetation distribution. This avoids difficulties posed by inaccessibility, time or cost-intensive field campaigns. Potential has been demonstrated for stand-level estimates in both tropical sites (Drake *et al.*, 2002; Drake *et al.*, 2003) and forests in the USA (Lefsky *et al.*, 1999b; Harding *et al.*, 2001; Breidenbach *et al.*, 2008) with LVIS and SLICER, in the UK using small footprint discrete return LiDAR (e.g. Patenaude *et al.*, 2004) and at a USA State-wide scale using laser profiling (Nelson *et al.*, 2004; Nelson *et al.*, 2008a).

Nelson *et al.*, (2004) and Nelson *et al.*, (2008b) have shown the potential for USA Delaware State-wide documentation of commercially viable volume and total above-ground biomass using first return airborne LiDAR profiling. This approach found best biomass results for vegetation stratified into broad vegetation groupings (mixed wood, $R^2 = 0.28$, RMSE = 49.17 t/ha; hardwood, $R^2 = 0.30$, RMSE = 115.51 t/ha; conifers, $R^2 = 0.44$, RMSE = 68.05 t/ha; and wetlands $R^2 = 0.95$, RMSE = 24.16

t/ha) using simple linear regression for mean height metrics (combined with canopy cover in the case of hardwoods). This method of vegetation type stratification was also employed for the retrieval of biomass using GLAS data (Nelson, 2008) using conifer, deciduous and mixed wood classes for southern Quebec, Canada. Consistency in use of predictive models was found to be necessary if monitoring changes over time as this alone accounted for 7% of apparent biomass differences.

Sun *et al.*, (2008) compared GLAS energy waveform quartiles with the same from LVIS waveforms and found these to be highly correlated (50^{th} percentile R^2 = 0.83; 75^{th} percentile R^2 = 0.82). Using LVIS, these waveform indices have been previously found to be significantly correlated with mean diameter at breast height (DBH), basal area and above-ground biomass (Drake *et al.*, 2003). These metrics are affected both by vertical canopy distribution and canopy cover, as more open canopies would allow greater laser penetration therefore reducing the height at which percentiles fall.

Sun *et al.*, (2008) found tree heights from field measurements were estimated from GLAS waveforms with R^2 = 0.57, RMSE = 4.46m. Estimates of biomass combining evergreen conifers and deciduous trees varied with date of GLAS acquisition (R^2 = 0.78, RMSE = 30Mt/ha for late Autumn; R^2 = 0.59, RMSE = 24.56 Mt/ha captured in June).

LiDAR data were found in some studies to over-estimate observed ground level. The difficulty of obtaining accurate measurements beneath dense canopies is widely recognised and is dependent on sampling density, scanning angle, canopy density and closure as well as processing methods. Leaf litter and understorey vegetation, as opposed to the true ground surface, may contribute to this overestimation (Patenaude *et al.*, 2004; Goodwin *et al.*, 2006).

In contrast, LiDAR canopy tree tops are commonly underestimated and this tendency is more often the case with the smaller or largest footprints as previously mentioned. Several factors contribute to this including canopy structure (e.g. conical or spherical crowns); canopy density (determining leaf or vegetation surface area, position in relation to illumination angle); footprint size (larger footprints have a greater probability of sampling maximum height though will also reduce the proportion of the signal relating to tree tops); strength and wavelength of the emitted signal (which affects the return signal strength); threshold detection selected and signal to noise ratio sensitivity.

Figure 2.6 illustrates how contiguous waveform footprints may uniquely represent canopy cover and volume at depths within a vegetation canopy (Lefsky *et al.*, 1999a).

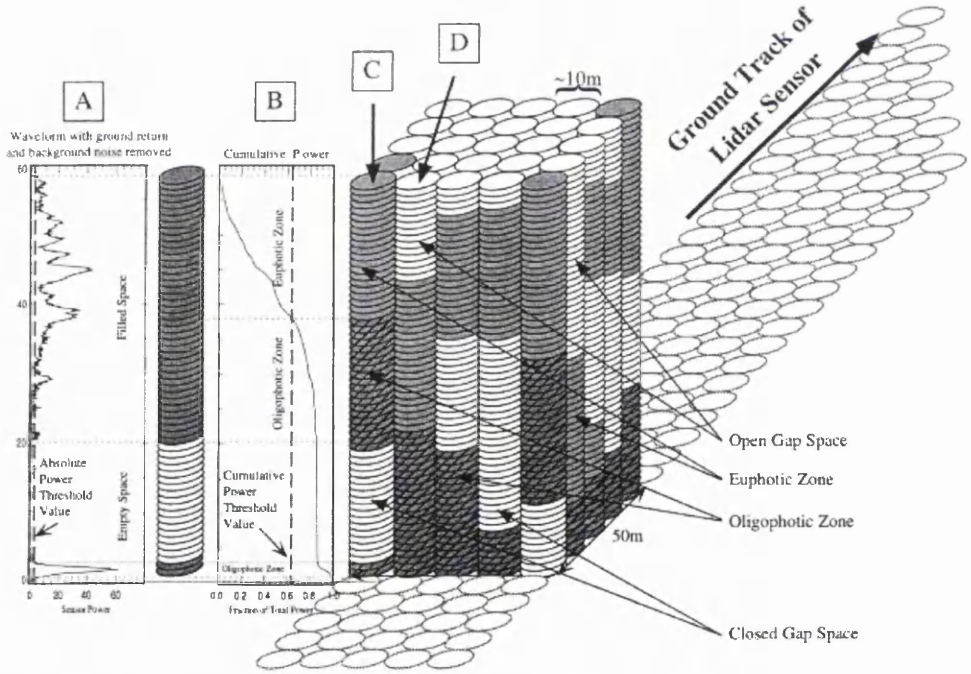


Figure 2.6 Canopy crown model demonstrating the ability to spatially represent crown closure and foliage distribution within canopies using waveform energy within height bins (Lefsky *et al.*, 1999a).

Waveforms were converted to height profiles with thresholds established to allow classification into four canopy zones: Empty Space consisting of no canopy/ ground returns; Euphotic Zone comprising the upper 65% canopy closure elements; Oligophotic Zone forming the remaining elements; and Closed and Open Gap Space, the latter derived from the difference between the waveform extent and the maximum estimated vegetation height for each footprint. This permits the proportions of each class to be calculated (Lefsky *et al.*, 1999a; Lim *et al.*, 2003b) which can be used for volume and canopy cover distribution calculations and to demonstrate canopy light regime. This approach may be relevant for small-mid diameter contiguous footprint coverage such as the planned LIST mission.

LiDAR point clouds and, in particular, waveforms can reveal distinct characteristics relating to canopy structure, volume and canopy cover distribution which can allow assumptions of shade tolerance and competition for light (Parker *et al.*, 2001), indications of stand microclimate or stand developmental stage (Lefsky *et al.*, 1999a; Harding *et al.*, 2001; Skinner, 2002) to be inferred (Figure 2.7). Older canopies or those with high yield class generally have larger gaps and therefore increased probability of a ground or "side" hit. Assessments requiring comparisons of waveform amplitude to infer stand cover would require energy to be normalised for varying energy attenuation though the atmosphere (Parker *et al.*, 2001).

Surfaces intercepted by LiDAR include branches and stems in addition to foliage (Chen *et al.*, 2004; Riano *et al.*, 2004). This provides an estimation of plant area index or canopy cover as a proportion of the area under the entire waveform. However, estimation of transmittance profiles from signal echoes does not directly account for canopy absorption of laser light through the canopy. Nevertheless, at 1064nm for example, absorption by both needle and broadleaf foliage is small

(Parker *et al.*, 2001). Therefore waveforms may provide direct estimates of canopy cover and LAI (e.g. Parker *et al.*, 2001; Todd *et al.*, 2003; Riano *et al.*, 2004).

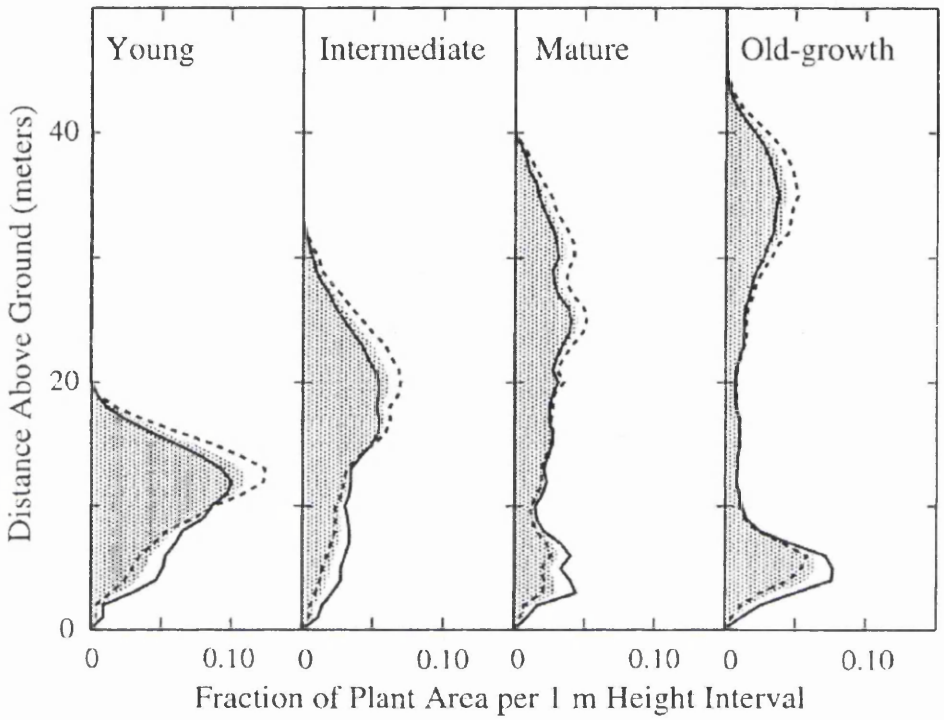


Figure 2.7 Age-related waveforms for deciduous broadleaf vegetation illustrating height variation, changes in foliage distribution and energy penetration to the ground (Harding *et al.*, 2001).

Overview

All volume-based studies by the above authors assessed LiDAR metric performance against field-derived data and found significant correlations. A variety of allometric equations were utilised to estimate volume, biomass or carbon content for the different vegetation types. The investigations suggest that relationships between LiDAR metrics and above-ground biomass differ between bioclimatic zones (Drake *et al.*, 2002; Drake *et al.*, 2003). This implies that consideration of these relationships for distinct vegetation classes may improve biomass prediction capabilities for future global mapping. Significant relationships were also

demonstrated for LiDAR estimation of LAI and light transmission through the canopy.

The studies suggest that a metric often exists of the optimum distribution of intercepted laser pulse energy representing a reliable estimation of a biophysical parameter. The position at which this occurs varies with canopy and crown structure, in addition to the biophysical property being considered (Lim *et al.*, 2003a). The above studies suggest the strongest relationships apply metrics which comprise information regarding distribution of tree and foliage heights whilst attributing greater weighting to taller vegetation; this therefore represents the upper portion of the canopy and indicates that a weighting factor is being applied which is proportional to the stem diameter required to support the canopy (Lefsky *et al.*, 1999b; Harding *et al.*, 2001). Additionally, seasonality may be a factor in determining height estimate accuracy (Sun *et al.*, 2008). The study presented in this thesis uses data captured in October 2005 whilst vegetation was predominantly still in leaf.

This project therefore explores means of identifying indices within the waveform canopy return in order to develop regression relationships for the estimation of vegetation biophysical parameters. The thesis compares GLAS height metrics with small footprint airborne LiDAR data and uses a percentile-based approach to estimate stemwood volume using broad stratification of vegetation classes. As a first step, this research develops a method of estimating canopy cover from large footprint waveforms and evaluates them against those calculated using airborne LiDAR data, validated using hemispherical photography.

2.5 LiDAR Forestry Applications and Inventory Summary

LiDAR data has been shown to provide valuable three-dimensional details of vegetation volume and structural arrangement and to relate these to vegetation life stage. This perspective enables the complexity of vegetation organisation to be better understood, opening the possibility of applying this knowledge, using satellite technology, for larger areas.

LiDAR technology is well-equipped to provide remote, unintrusive, geo-referenced, three-dimensional information regarding surface elevation, vegetation structure and human infrastructure to an unprecedented degree of accuracy and over vast scales, relatively economically and time-efficiently.

Potential applications are broad ranging and span both the natural and human environments. Some of these capabilities are unique to LiDAR technology and as altimeter readings are less sensitive than some other sensors to environmental conditions such as sun angle, leaf state and moderate weather conditions, there is a relatively wide window with regard to LiDAR acquistion opportunities (Airborne_Laser_Mapping, date unknown; GeoLas_Consulting, date unknown).

LiDAR data capture does, however, inevitably present some limitations. When mapping dense vegetation, pulses are scattered by and reflected within the canopy, thereby obstructing the beam penetration and exit route and resulting in a delayed echo response than would occur from the 'true' surface directly. Radiative transfer modelling will help to understand the sensitivity of waveforms to surface properties (Ni-Meister *et al.*, 2001; North *et al.*, submitted). The majority of LiDAR systems emit in the near-infrared band, but some minerals, water (and therefore

cloud and fog), asphalt and tar, absorb these wavelengths causing null or poor return echoes. LiDAR, particularly full waveform data, can also result in large quantities of information and therefore require sufficient processing capabilities to accommodate this (Airborne_Laser_Mapping, date unknown). A further important consideration is that laser altimetry should not be considered an imaging system as, even with high density laser scanning, only samples of the surface are collected (Lim *et al.*, 2003a).

However, this emerging technology offers to improve our understanding of the human and physical environment across a broad range of disciplines. Vertical profiles of vegetation and associated parameter distribution can provide a useful means of improving understanding of this field.

2.6 Conclusion

This Chapter has summarised the principles of LiDAR and identified applied studies using airborne laser scanning and mid-to-large footprint waveform scanning and profiling. Comprehensive Satellite LiDAR applications for biophysical parameter retrieval have been identified as a field of research which has been undergoing recent development. Based on the positive outcomes from airborne platforms, potential is anticipated for diverse vegetation types over a range of scales. Previous and recent experiences of deriving height, volume and canopy cover estimates have been outlined.

Possible inputs of LiDAR-derived data have been identified for dynamic vegetation models and radiative transfer models. On forest management scales, process-based growth models or site condition simulation models (e.g. Edwards and Christie, 1981; Gardiner *et al.*, 2004) require inputs of stand-level vegetation

structural characteristics. LiDAR estimates of vegetation properties may provide a means of determining more representative inputs to drive models or of validating model-generated predictions and consequently produce more accurate simulations which realistically replicate processes.

Recognition is needed of the fact that LiDAR remote sensing is restricted to the calculation of above-ground vegetation volume and, by inference, carbon. However, root and soil systems represent large stores of carbon in forest systems. Therefore to obtain a comprehensive account of carbon stores, LiDAR-derived carbon estimations may need to be supplemented by field studies and modelling techniques in order to produce representative estimates.

Whilst airborne LiDAR data often provides high spatial density information during a single campaign for a targeted area, satellite platforms offer opportunities for systematically monitoring significantly larger areas through relatively frequent repeat coverage (2-3 times annually with the current ICESat mission), albeit on a sampling basis and at coarse resolution. The following Chapter introduces the principles of the GLAS instrument used in this project.

Chapter 3. ICESat/GLAS Principles and Pre-Processing

This chapter aims to provide an overview of the principles of the ICESat/GLAS data which are used within this research and the processing of the data by the GLAS Science Team prior to their release to the public as data products.

3.1 Ice Cloud and land Elevation Satellite

The Ice Cloud and land Elevation Satellite (ICESat) is the first global laser altimetry mission (Hofton and Blair, 2002). The primary purpose is to monitor ice mass balance through changes in ice elevation. However, secondary objectives are of relevance to this project and include retrieval of land surface elevations, land surface roughness and multiple near-surface canopy heights over land. The mission aims to additionally contribute to topographic mapping and digital elevation models as well as detecting changes of elevation greater than one metre per year for selected regions.

3.1.1 Geoscience Laser Altimeter System Overview

The Geoscience Laser Altimeter System (GLAS) is a full waveform LiDAR profiler carried on board ICESat. The satellite orbits the Earth at an altitude of

600km, travelling at a speed of 26,000 km per hour and emitting laser pulses at 40 Hz, thereby sequentially producing footprints, nominally of 70m diameter, with 172m intervals between consecutive footprint centres (Zwally *et al.*, 2002; NSIDC, 2003; Schutz *et al.*, 2005). Figure 3.1 shows a ground track crossing Antartica with surface and cloud elevations from GLAS indicated.

GLAS consists of three lasers which function exclusively and are used successively, one of which currently remains operational (Kichak, 2003; Abshire *et al.*, 2005). All emit pulses at both 532nm and 1064nm wavelengths, simultaneously acquiring details of atmospheric composition (cloud and particular aerosol distribution and density) in addition to surface feature characteristics (glaciological, topographical, oceanographic and vegetative). Quality flags indicate corrections applied and sources of concern such as saturated returns or significant forward scattering from cloud.

Whilst 70m footprints are stated as optimal for ice sheet monitoring, smaller diameters of approximately tree crown diameter (10-25m) are more advantageous for vegetation and land surface observations as less surface variation is present within the area illuminated by the laser pulse. However smaller footprints require higher pulse repetition to acquire adequate sampling of the Earth's Surface which would reduce the laser lifespan and consume more power.

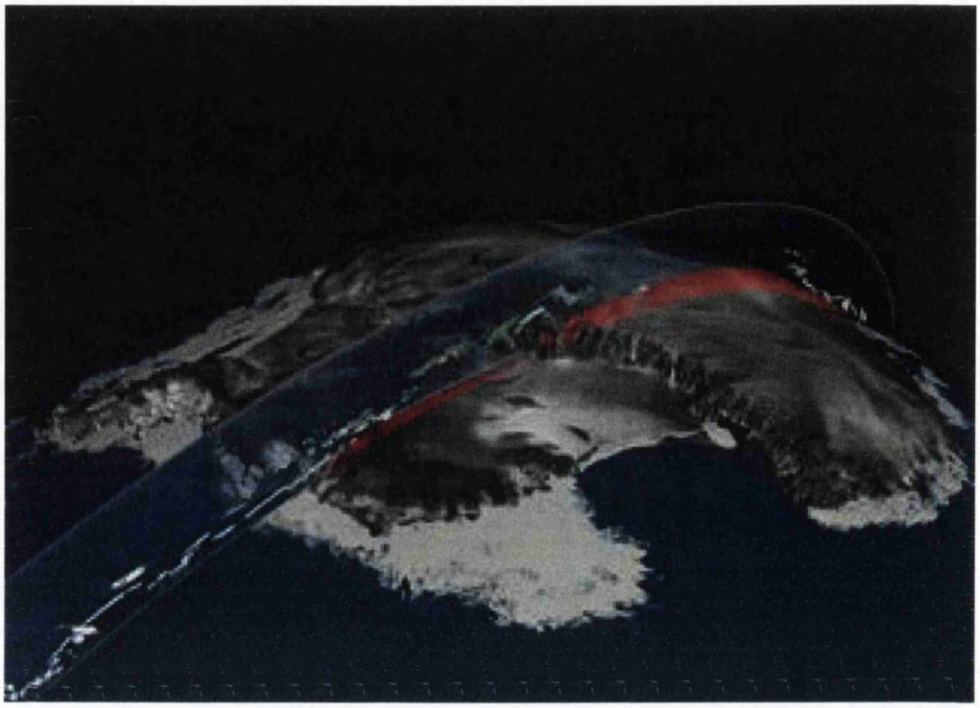


Figure 3.1. Lithograph with fifty times vertical exaggeration, illustrating ICESat-derived profile of coincident surface and cloud elevations over Antartica (Shirah and Kekesi, 2003).

The initial calibration and validation phase of GLAS laser operation involved an orbit which repeated ground tracks every eight days. The mission phase then consisted of a repeat orbit of 91 days, with original plans to subsequently change to 183 day repeat track cycles. This in fact didn't occur for reasons outlined below. The mission specification of an 183 day repeat cycle would have produced reference ground track spacing of 15km at the equator and 2.5km at its maximum latitude of ±86° (Brenner *et al.*, 2003).

Despite the large footprint size and relatively low emitted pulse repetition frequency, the first laser on board ICESat failed after just 74 days of pre-launch operation and 36 days of spaceborne operation. The Independent GLAS anomaly Review Board assessed the failure of GLAS Laser 1 to be attributable to a manufacturing fault in the laser diode arrays which introduced large quantities of indium solder. This, together with the high temperatures associated with high current

requirements of the system, caused the progressive erosion of the gold conductors by the formation of non-conducting gold-indide (Kichak, 2003).

Consequently, a modified mission plan was adopted for the two remaining lasers to maximise operational life, as it was assumed that the fault may also exist with these. Amendments have subsequently been applied to the laser operation protocol to minimise future thermal stresses. The overall solution has been to adapt the mission plan and to operate each remaining laser at limited intervals using a 91 day repeat orbit in order to moderate the effect of temperature and accumulation of gold-indide.

The three lasers forming GLAS are operated sequentially until the end of each laser's life. An inclination angle of 94° is used to prevent saturation due to specular reflection which, particularly over ice surfaces, would result from direct nadir reflection to the 1m diameter Beryllium telescope which captures the returned pulse. The intended mission lifespan was three years' duration with anticipated extension to five years. This objective has therefore been achieved but with significantly reduced data acquisition both in terms of discontinuous data capture and lower ground track density.

Current performance is anticipated to allow laser operation for a further 3-4 data campaigns twice per year from 2009 (David Korn, NSIDC, pers. comms).This aims to coincide as closely as possible with the launch of ICESat II for which the intention is to revert to the original mission specification as current faults will have been addressed. Subsequent missions therefore aim to continue coverage for a combined period of fifteen years, permitting long term interpretation of LiDAR Earth observations. An overview of current ICESat laser campaigns to date is shown in Table 3.1.

Table 3.1. Summary of laser operations and attributes (NSIDC, 2003; NSIDC, 2008). TBD indicates information to be determined by NASA. Laser campaign numbers indicate the laser in operation.

Laser campaign	Start date	End date	Horizontal accuracy (m)	Vertical accuracy (m)
L1A	20.02.2003	29.03.2003	4.6±9.3	0.08±0.16
L2A	24.09.2003	18.11.2003	TBD	TBD
L2B	17.02.2004	21.03.2004	TBD	TBD
L2C	18.05.2004	21.06.2004	37.7±53.4	0.66±0.93
L3A	03.10.2004	08.11.2004	0.0±2.7	0.00±0.05
L3B	17.02.2005	24.03.2005	17.4±22.8	0.30±0.40
L3C	20.05.2005	23.06.2005	TBD	TBD
L3D	21.10.2005	24.11.2005	TBD	TBD
L3E	22.02.2006	27.03.2006	TBD	TBD
L3F	24.05.2006	26.06.2006	TBD	TBD
L3G	25.10.2006	27.11.2006	TBD	TBD
L3H	12.03.2007	14.04.2007	TBD	TBD
L3I	02.10.2007	05.11.2007	TBD	TBD
L3J	17.02.2008	21.03.2008	TBD	TBD

3.1.2 Data overview

Near global coverage of between ±86° latitude is provided by ICESat/GLAS acquired during its 91 day repeat orbit cycle of which GLAS is operated at intervals which initially captured measurements for three seasons each year and, since 2008, during Spring and Autumn: generally for 33 to 55 day periods during February-March, May-June and October-November. Data coverage for the UK is illustrated within Figure 3.2. Longitudinal spacing between tracks at the latitude of the UK is approximately 36km.

Figure 3.2. National Inventory of Woodland and Trees (NIWT1) classes shown as black shading overlaid by ICESat/GLAS ground track passes for Britain throughout the ICESat mission. Figure courtesy of J. Suárez, Forest Research.

Table 3.2. Numbers of footprints intercepting Interpretive Forest Type classes within the first National Inventory of Woodland and Trees (NIWT1). Table courtesy of J. Suárez, Forest Research.

Interpretive Forest Types from NIWT1	ICESat footprints
Broadleaves	3362
Coniferous	5426
Coppice	19
Coppice with standards	10
Felled	465
Ground prepared for planting	717
Mixed	1469
Shrub	151
Young trees	1959

This intermittent operation of the lasers aims to repeat the same sub-cycle during each laser campaign. Repeat ground tracks or cross-over points of tracks therefore potentially allow changes over time to be detected (Shuman *et al.*, 2005).

Except where stated, the following account is largely made with reference to Brenner *et al.*, (2003) in their Algorithm Theoretical Basis Document.

The laser footprint is nominally 70m diameter on the surface across its longest axis. The transmitted laser pulse has a width of 5ns, equivalent to 75cm in surface elevation. Assuming a Gaussian transmitted pulse of 5ns duration, the expected return from a flat, highly reflective surface would also be Gaussian in form of 5ns duration. The returned pulse, however, will be broadened by the distribution of intercepted surface heights within the footprint, if the surface is sloped, rough or a combination of both.

Footprints are approximated in this study as circular although reports show an elliptical footprint (which varies between lasers) with average equivalent circular area diameter of 64m (Abshire *et al.*, 2005). For the laser operation used in this study (L3D), near infrared (NIR) wavelength footprints are produced with major and minor axes of 52m±1.1 and 44.4m diameter respectively (NSIDC, 2008). As laser energy decreases with distance from the footprint centre, diameter is defined as the point at which energy has fallen to $1/e^2$ of peak amplitude (pers. comms. GLAS Science Team via NSIDC). The result of this diminishing energy towards the footprint margins is that returned waveforms are most representative of the features closest to the footprint centre. Figure 3.3 illustrates energy distribution of the transmitted pulse (Laser Profiling Array - LPA) as observed at the laser reference camera. The LPA provides pointing knowledge of the transmitted pulse and is a component of the Stellar Reference System which, combined with precision attitude determination (PAD), allows the calculation of footprint location accuracy (Schutz, 2001; Bae and Schutz, 2002).

Footprint horizontal geolocation (Schutz, 2002) is yet to be determined by NASA for L3D, however is expected to vary between 0.0 ± 2.7 metres (L3A) and 17.4 ± 22.8 metres (L3B) (NSIDC, 2008). Field measurements for this study aim to take these factors into account (following Carabajal and Harding, 2001).

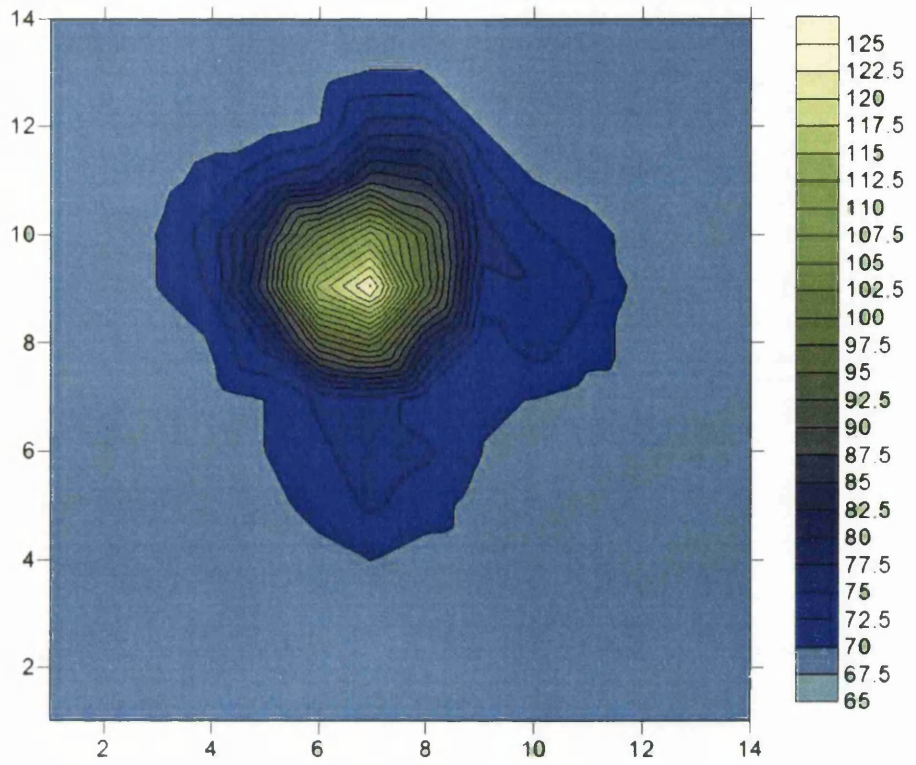


Figure 3.3. Transmitted pulse laser profiling array (LPA) for footprint 885917496_0 (product GLA04). X and Y axes represent pixel numbers (pixel size in far field beam divergence angle is 3.388 arcsec/pixel). Scale of the LPA image projected to the Earth's surface (as also reported by Harding and Carabajal, 2005) can be considered as Laser3D mean axis dimensions of approximately 52x44m. The colour key shows pixel intensity indicating relative beam energy distribution across the footprint, where the highest value represents peak amplitude and the lowest value is zero energy. (Pixel size and unit of measure: pers. comms. GLAS Science Team via NSIDC).

3.1.3 Waveform Structure

Assuming a minor effect of forward scattering by clouds and aerosols, the waveform echo is determined by the range distribution within the footprint, reflectivity of the surfaces (the average of which is calculated as the ratio of transmitted and received energy after it has been scaled for range) and pattern of the incident beam.

Three key measurements are therefore produced; the range between the satellite and surface footprint, the shape of the reflected waveform and the laser

power returned from the surface. Bufton, (1989) identifies the total area under the received pulse, being proportional to the pulse energy, as a means of measuring surface albedo. Kotchenova *et al.*, (2003) also recognise that, for vegetated areas having surfaces with similar reflective properties and vertical structure within a footprint, a larger amplitude of the return indicates a greater volume of canopy material.

The effect of slope (assumed to be a tilted, planar surface) or surface roughness is to broaden the returned waveform; however the relative contribution of each cannot be distinguished. Extracting estimates of surface characteristics from broad footprint LiDAR waveforms can therefore be challenging.

3.1.3.1 Uniform surfaces

For footprints without vegetation or complex features such as buildings, the interpretation of surface elevation (defined as the surface without overlying features) is relatively straightforward as the returned waveform (after processing discussed above) is simply representative of the vertical distribution of intercepted surfaces, the reflectance of the surfaces and the spatial distribution of laser energy across the footprint.

It is assumed that the return will closely resemble a Gaussian and therefore a Gaussian pulse is used to fit the waveform. The centroid of the Gaussian pulse is used to calculate the range to the mean surface, which is then corrected for atmospheric delays. The corrected range, position of the satellite above the footprint and the off-nadir pointing position are then used to calculate the surface elevation.

3.1.3.2 Complex surfaces

Multi-modal Waveforms

Footprints which contain vegetation and other features above a sloped surface can produce waveforms in which signals produced by surface roughness and topography are combined (Figure 3.4).

Experience derived from airborne LiDAR suggest footprints over complex surfaces would produce multi-modal waveforms with each peak corresponding to distinct heights within the footprint; the final peak representing the ground level and the uppermost signal being that returned from the highest canopy layer for example.

With increasing surface relief (slope and roughness) and footprint size (creating greater variation within the footprint), the contributions to the waveform shape due to relief and overlying features, become increasingly more difficult to distinguish (Figure 3.5). In addition, with larger footprints, the problem is exacerbated as multiple vegetation layers are spatially averaged across the footprint and furthermore, representative vegetation height is harder to identify meaningfully. Independent knowledge is therefore necessary for direct interpretation of laser footprints of this extent.

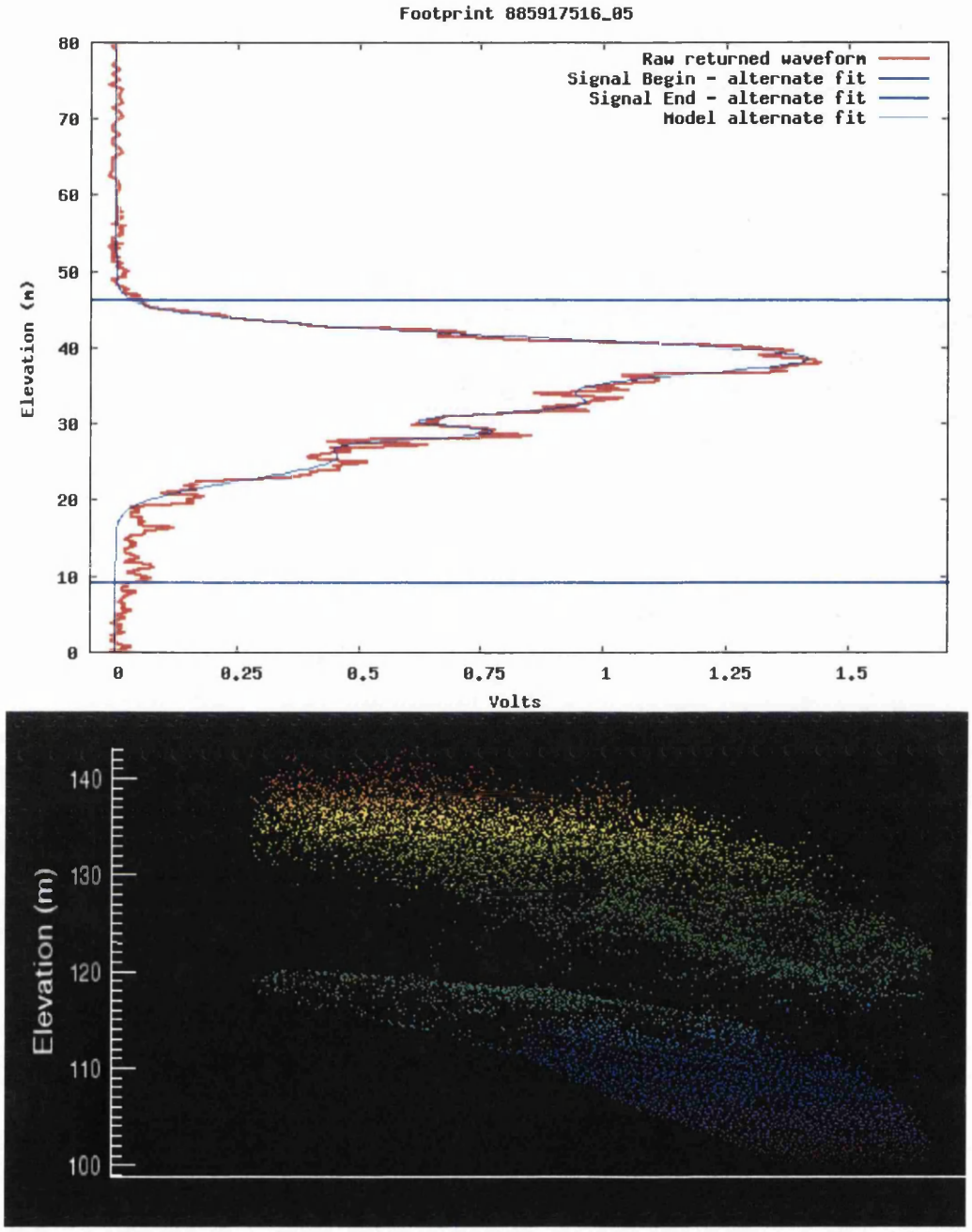


Figure 3.4 Returned waveform combining signal from ground and vegetation surfaces (above). Coincident airborne LiDAR point cloud for illustrative purposes (below). This figure illustrates the difficulty posed by identifying a ground elevation within waveforms for which terrain and vegetation are at similar elevations. Extent of airborne LiDAR point cloud is 70m diameter.

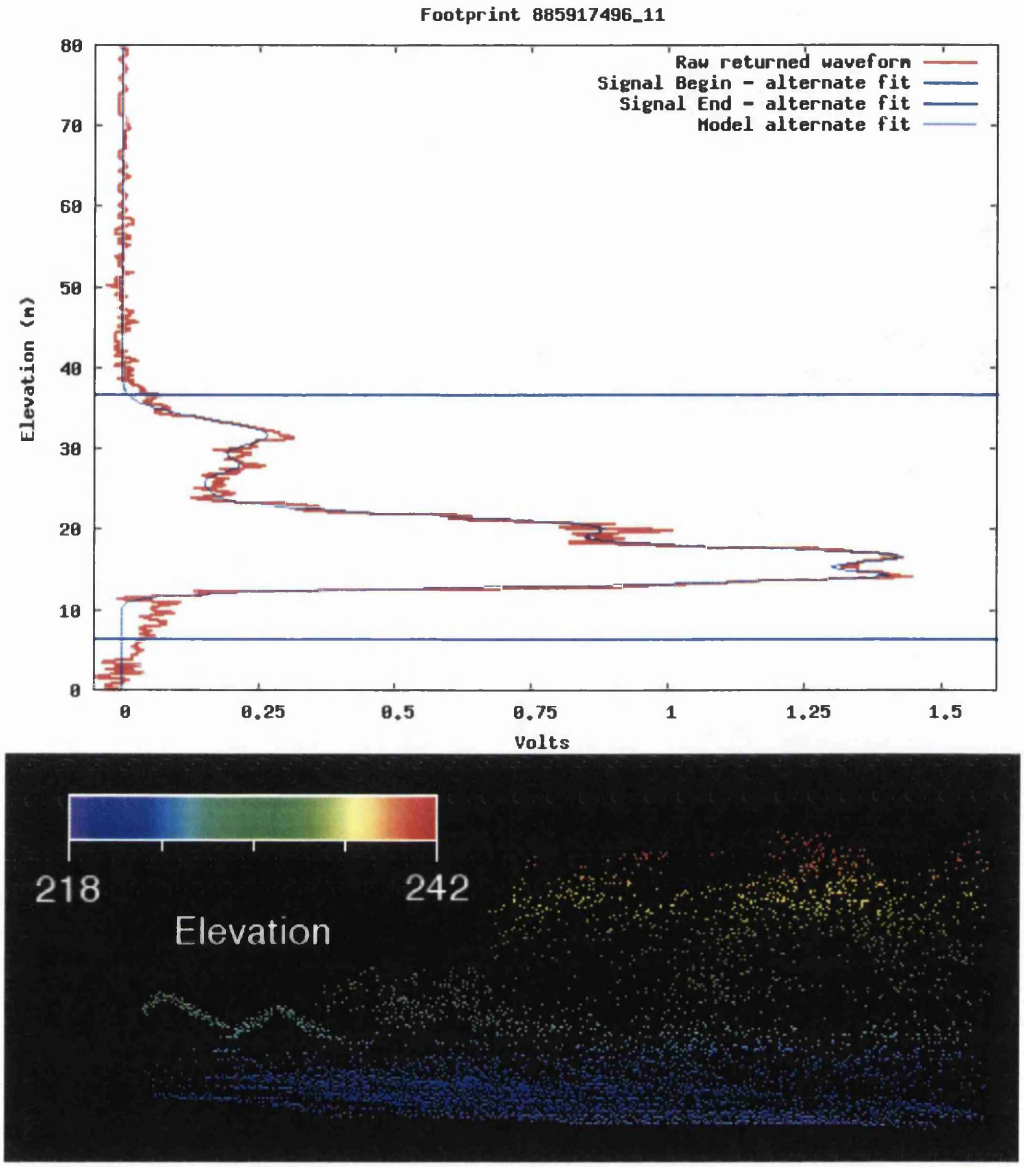


Figure 3.5 GLAS waveform example for complex intercepted surfaces and airborne LiDAR point cloud illustrating spatial distribution of intercepted features (70m diameter subset coincident with the GLAS footprint). Spatial separation within waveforms of contributing surfaces is not possible. Local supplementary knowledge is necessary to determine whether waveforms are formed by purely natural or human infrastructure features.

Multiple scattering from elements within the canopy will also lengthen the photon paths thereby delaying the return pulse and affecting the shape of the waveform. This is dependent on laser energy wavelength and transmission between the foliage (determined by vegetation density and three-dimensional structure of canopy components). Brenner *et al.*, (2003) additionally suggest that literature

evaluating LiDAR performance for retrieval of vegetation biophysical parameters has assumed the effects of multiple scattering on LiDAR-derived vegetation heights to be minimal.

Notwithstanding this, Kotchenova *et al.*, (2003) recognise that failure to take multiple scattering into account may lead to interpretation errors at the lower portions of waveforms for near infrared wavelengths as the path for multiple scattered photons is lengthened, thereby delaying their return and resulting in an apparent large volume of low vegetation shown within the waveform.

Time dependent radiative transfer theory has been used to model LiDAR pulse scattering for both coniferous and closed canopy deciduous vegetation. Kotchenova *et al.*, (2003) suggest that understanding and knowledge of vegetation type and characteristics within the footprint area is necessary for accurate interpretation of waveforms which would allow extinction coefficients to be used to adjust for shadowing effects from upper canopy structures obscuring those below. Sensitivity of waveform structure to vegetation properties is furthermore being explored using the radiative transfer model FLIGHT (North, 1996; North *et al.*, submitted).

Large footprint LiDAR theory assumes that the final peak of multi-modal waveforms is that returned from the ground surface, whilst earlier peaks correspond to layers of vegetation or other features within the footprint. The centroid and variance from this final peak may be used to determine the mean surface elevation and breadth of the pulse due to slope and roughness. Additionally the maximum height within the footprint could be taken as the distance between the leading edge of the first return and the centroid of the last signal peak (Carabajal and Harding, 2001).

There are however, some circumstances in which the last peak will not represent the position of the ground: stepped surface heights such as a cliff edge; dense vegetation preventing penetrating to the ground; low lying vegetation may result in a combined return with actual ground level; a large building with multiple sloping roofs. A method is therefore needed to account for this potential discrepancy and the complications posed by the large footprint diameter and combined surface returned signal.

Uni-modal Waveforms

Interpretation of uni-modal waveforms where no overlying features are present is relatively simple, using the centroid to estimate the mean surface elevation and assuming the breadth of the pulse to be due to the effect of surface relief.

However, uni-modal waveforms originating from footprints containing overlying surface features are perhaps most difficult to interpret as the heights of the elements contributing to the waveform cannot be differentiated (Figure 3.6). The position of the centroid and the Gaussian variance are due to the combined influence of all intercepted components within the footprint, with the additional complication of the density and spatial arrangement of the overlying surface features being unknown.

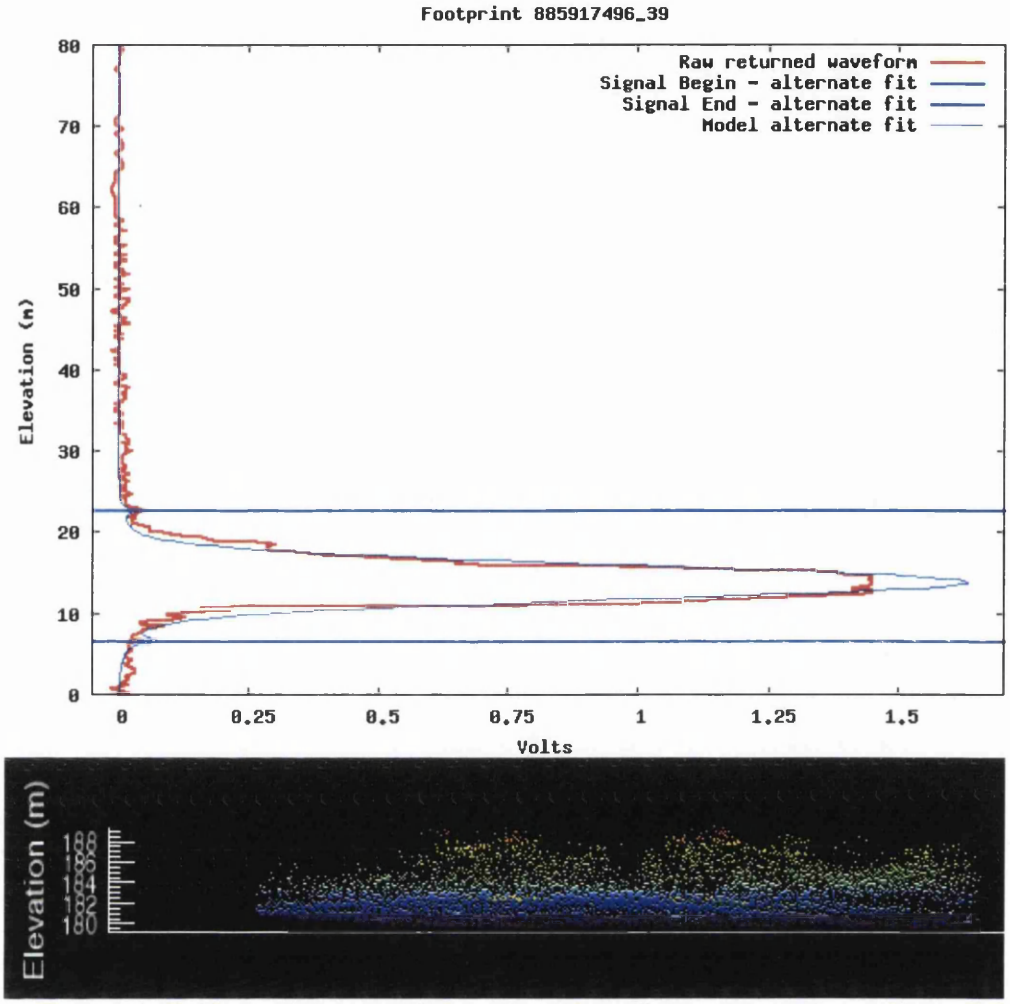


Figure 3.6 GLAS waveform example for a site with small stature vegetation and low relief combined within a single-peaked waveform as modelled by the GLAS alternate fit algorithm using Gaussian decomposition.

3.1.4 Data Pre-processing

A number of operations are performed to facilitate waveform understanding for further processing (Brenner *et al.*, 2003). There are four stages to the initial processing undertaken by NASA in order to produce products available for use by the scientific community:

I. Signal Detection

II. Waveform Digitisation

III. Gaussian Decomposition

IV. Elevation and Surface estimations

3.1.4.1 *Waveform Characterisation*

Signal Detection

For each transmitted laser pulse, the GLAS instrument will collect approximately 4,500,000 one ns samples. To reduce this to a manageable amount to enable it to be transmitted to Earth, of the entire laser echo returned to the spacecraft, only a limited Range Window of 544 data bins is recorded and used within data products.

On-board processing takes place using an algorithm which aims to maximise the chance of including the return from the Earth's surface. This uses an onboard SRTM-derived 90m resolution Digital Elevation Model (DEM) to indicate the region of the returned pulse where elevations relating to the earth's surface and overlying features are likely to be located in order to identify the maximum and minimum values of the Range Window.

The digitised waveform within this window is then passed through six filters which check for a viable signal and increase the probability of detecting echoes from sloped or rough terrain and reduce that of selecting those from clouds. Filter size starts with the smallest filter (4ns) which is increased progressively until a signal region is detected. The signal flag is set to zero (or one if no signal region is found).

The waveform is then smoothed using a Gaussian filter with a sigma equal to that of the filter. Signal start and end positions are then determined as the first and

last positions at which the waveform amplitude is greater than the mean noise value plus a multiple of the standard deviation of the noise (set at $4.5 \times \sigma$).

Finally, a noise threshold is determined for each filtered waveform. The surface echo is searched backwards in time to identify the last local maximum exceeding the threshold which is assumed to represent energy returned from the ground.

Waveform Digitisation

Using this reference point, the returned pulse from the surfaces is digitised within range bins. 544 range bins are designated for data retrieved over land. For laser operations L1A and L2A the waveform was recorded entirely within 1ns data bins (~15cm) allowing an elevation difference of 86.1 metres to be recorded. However on occasion, this resulted in the waveform being truncated where the vertical extent of the returned waveform exceeded the recorded waveform. Therefore, for subsequent data acquisitions, the uppermost 152 gates were recorded at 4ns resolution (91.2m) whilst the lower 392 gates remain sampled at 1ns resolution (58.8m). This produces a 1000ns waveform relating to a 150m Range Window.

Positions of indices within the waveform are provided as a negative number indicating the distance in metres from the lowest elevation within the Range Window. Further details are found in Brenner *et al.*, (2003); NSIDC, (2003); Harding and Carabajal, (2005); NSIDC, (2008).

Gaussian Decomposition

The transmitted pulse is Gaussian and, if the surface topography is flat with Lambertian reflectivity, the return is also expected to be Gaussian in form. Over

complex surfaces, the waveform is represented mathematically as a sum of Gaussians plus a bias (noise level). The modelled waveform can therefore be represented by the following equations:

$$w(t) = \varepsilon + \sum_{m=1}^{N_p} W_m \tag{3.1}$$

$$W_m = A_m e^{\frac{-(t-t_m)^2}{2\sigma_m^2}} \tag{3.2}$$

where

$w(t)$ is the amplitude of the waveform at time t

W_m is the contribution from the mth Gaussian

N_p is the number of Gaussians found in the waveform

A_m is the amplitude of the mth Gaussian

ε is the bias (noise level) of the waveform

t_m is the Gaussian position

σ^2_m is the $1/e$ half-width (standard deviation) of the mth Gaussian

Nonlinear least squares are used to calculate the model parameters $(\varepsilon, A_m, t_m \text{ and } \sigma_m)$ by fitting the theoretical model to the observed waveform. Parameters driving the algorithm for the land product are designed to achieve waveform fitting whilst preserving the peaks present. The results of this processing are then used to calculate L-1B GLA06 and L-2 GLA12-15 products which are presented in the following Chapter.

An overview of Gaussian decomposition of waveforms is provided below. This model fit to the waveform (the sum of six Gaussian peaks in the case of

waveforms for land) aims to simplify the complexities of the raw waveform whilst retaining the dominant trends in order to facilitate interpretation.

The second derivative of the smoothed waveform is used to identify the peaks; this is positive between peaks, negative at peaks and zero at the inflection points. Therefore when the second derivative changes from positive to negative, this indicates the first inflection point [T_1] identifying a Gaussian, the end point of which (second inflection point [T_2]) is marked by the second derivative reverting from negative to positive. The estimated Gaussian amplitude is the maximum smoothed waveform amplitude within [T_1, T_2].

For all but the maximum amplitude (A_{max}) Gaussian, the estimated Gaussian width is the smaller of $|T_1 - T_m|$ and $|T_2 - T_m|$. When the waveform is affected by saturation or forward scattering, the Gaussian width cannot be accurately estimated using inflection points and therefore the width of the maximum amplitude Gaussian is calculated using T_{1_80} and T_{2_80} plus T_{1_61} and T_{2_61} (the times before and after T_m at which W_m is 80% or 60.653% of A_m respectively). Definitions are as above.

Once all Gaussians have been identified, those with amplitudes less than the minimum peak amplitude (A_{min}) are removed and those which are closer than the determined minimum interval are combined with their closest neighbour. The Gaussians are then ranked by their area, the smallest of which are combined with their closest neighbour until the number of Gaussians does not exceed the maximum permitted (six in the case of the land product).

Elevation and Surface estimations

Interpretation of returns over land is complicated by the varied structures of the land surface and the small scales over which these can fluctuate. The effects of

these characteristics on the performance of laser altimeters is considered by Gardner, (1992) who identifies the centroid of the received pulse as being the most effective technique for estimating range when the pulse shape is unknown.

Harding *et al.*, (1994) explain this further. The centroid of the received pulse (T_S) is the mean round-trip time-of-flight range to surface features within the laser footprint, weighted by the reflectivity and area extent of the features. The pulse centroid is determined by integrating the received signal strength over a range gate of time duration T_G and the measured precision of T_S is a function of the time interval unit resolution (Δt):

$$T_S = \int_0^{T_G} \left(\frac{1}{T_G} \right) tP(t)dt \qquad (3.3)$$

where t is time and P is signal strength.

Due to the distinct requirements of the GLAS land product (i.e. using the description of the waveform and independent knowledge to permit interpretation regarding elevation, slope, roughness and heights of vegetation or other features), a 'land-specific range' is used. This is generally the centroid of the received waveform signal between the defined signal start and end points and is used to determine the final geolocated latitude and longitude and the footprint vertical location. This aims to provide a representative elevation of the signal, used as a point of reference for other waveform parameters.

Range offsets from this land-specific range can be used with the laser pointing vector to calculate the highest and lowest features detected within the footprint (when nadir-pointing, the range offsets will be equal to the elevation

offsets). The Land Range Offset parameter (*Ld_RngOff*) is used within this study to position the waveforms vertically in space.

The stages involved in pre-processing the GLAS waveform data in order to produce ICESat/GLAS data products are summarised below.

3.1.4.2 Overview of waveform pre-processing

I. Characterise the transmitted pulse and calculate the time for beginning the range calculation
II. Characterise the received waveform to determine if there is a signal and to determine the point on the waveform to be used to estimate the range and the preliminary footprint location on the Earth
III. Interrogate the database to determine the type(s) of surface at the footprint location
IV. Smooth the waveform and determine initial estimates for parameters
V. Fit the waveform:
    a. *Use zero-crossing of the second derivative to identify waveform peaks*
    b. *Merge peaks which are very close together (the initial number of peaks is a measure of waveform complexity and is stored in i_nPeaks*
    c. *Use the five peaks with the largest areas plus the peak furthest from the spacecraft as 'seeds' to determine initial Gaussian spacing, width and amplitude*
    d. *Carry out iterative least squares adjustment of the Gaussian positions, width and amplitude to minimise RMSE between Gaussian sum model and raw waveform*
    e. *For Gaussian amplitudes below a threshold (set very low), set amplitude to zero*
    f. *Report number of Gaussians (with non-zero amplitudes) in i_numPk*
VI. Calculate range to the mean surface and surface elevation distribution
VII. Calculate atmospheric delay and tidal values (Phillips *et al.*, 1999; Herring and Quinn, 2001)

VIII. Calculate a corrected range to the mean surface correcting for atmospheric delay and instrument effects
IX. Correct time for travel time
X. Calculate precise geolocation and mean surface elevation (Schutz, 2002)
XI. Apply the tides to the mean surface elevation
XII. Calculate region-specific parameters e.g. land, ice, ocean

(Brenner *et al.*, 2003 pages 36-37; pers. comms GLAS Science Team via NSIDC)

3.1.5 Data Limitations and Sources of Uncertainty

This section summarises the main known sources of uncertainty in GLAS waveform measurements and assesses the impacts on subsequent waveform interpretation and extraction of parameters which will be undertaken in this thesis. Error estimates derived from pre and post-launch validation exercises are discussed.

3.1.5.1 Instrument effects

Satellite LiDAR detector saturation has been previously known to cause truncation of returns and 'ringing' which is an artificial dip and peak after the returned signal (Carabajal *et al.*, 1999). To assist in overcoming these effects, waveform pre-processing selects either the centroid of the maximum amplitude Gaussian or the centroid of the returned signal to represent the elevation reference parameter for each waveform (to avoid the possibility of identifying the 'ringing' for this reference position). Additionally, fitting is carried out to the pulse leading edge to reduce the effect of truncated peaks. An amendment made to the original processing procedure to digitise the returned signal within a 150m Range Window further reduces the risk of waveform truncation.

Although prior to the launch of the ICESat mission, it had been possible to test the algorithms and make necessary modifications to the SLA-02 mission specification and adapt principles of large footprint airborne LiDAR, anticipated validation from the proposed VCL mission was not possible. Additionally, the GLAS instrument differs from those previously flown.

Harding *et al.*, (1994) underline how pointing errors would lead to a LiDAR-derived waveform being assigned to the nadir-oriented anticipated footprint rather than the actual illuminated location. A principal difficulty of geo-locating the relatively small laser footprints (in comparison with radar) relies on accurately determining the off-nadir angle; this is solved using star cameras and laser spot imaging. Bufton, (1989) identifies surface slope of the footprint area as further aggravating the knowledge of pointing angle and associated error (Figure 3.7). Further information regarding geolocation and laser pointing is discussed by other authors: Schutz, (2002); Fricker *et al.*, (2005); Harding and Carabajal, (2005); Luthcke *et al.*, (2005); Martin *et al.*, (2005); Schutz *et al.*, (2005). However it is anticipated that discrepancies are not excessive and therefore will not prevent adequate field validation and land surface assessment (Table 3.1). Detailed descriptions of determining and minimising uncertainty in precision orbit and spacecraft attitude are provided in Bae and Schutz, (2002); Rim and Schutz, (2002).

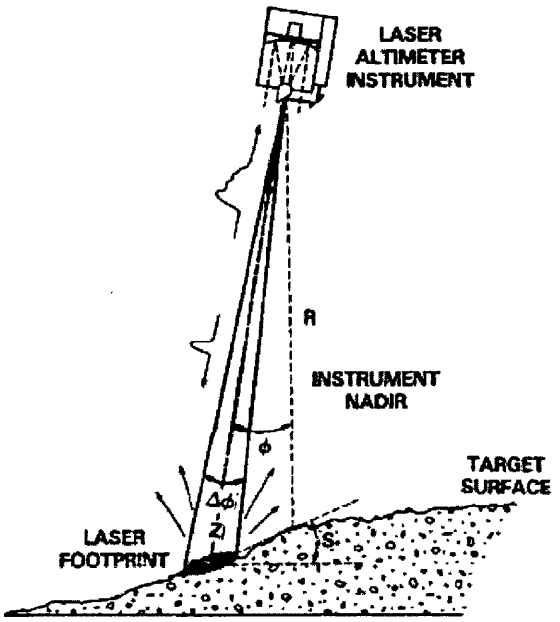


Figure 3.7. Effect of footprint surface slope broadening the pulse returned to the sensor and affecting footprint coverage on the ground and reflected energy. ø represents off-nadir pointing angle and *S* is slope (Bufton, 1989).

Potential sources of error and limitations of GLAS data are highlighted by Brenner *et al.*, (2003). Harding *et al.*, (1994) and Gardner, (1992) assess potential elevation accuracies from satellite laser altimetry considering the effects of slope and roughness. Their conclusions are summarised below.

The centroid of a narrow pulse can be more accurately determined than for a broad pulse and therefore, given the same total received energy, with greater pulse spreading due to surface complexity, the potential for range errors will increase. This is due to a lower signal to noise ratio (Harding *et al.*, 1994).

However, possible problematic waveforms caused by GLAS instrument effects can be identified as information regarding the quality of the data is provided in the form of flags, set to zero if no problem was present or to one if a problem was encountered in calculating any of the parameters for a product.

3.1.5.2 *Surface Properties*

All algorithms used to create GLAS products assume a Gaussian distribution of surface undulations. Deviations from this, plus sudden surface irregularities, will produce an uncertainty among slope, roughness and surface elevation calculations. For complex land surface returns, the issue of intricate waveform structure is addressed by fitting a simplified model to the waveform, determined as the sum of Gaussian peaks. However, interpretation of the fitted Gaussians (i.e. contribution of surface elevation, slope, roughness plus vegetation and other overlying features) relies on understanding the waveform shape (uni- or multi-modal) plus independent knowledge or assumptions about the land-cover type within the footprint.

The impacts on the waveform of assumptions relating to slope, roughness, surface elevation and reflectivity are discussed below.

Slope

Slope is assumed to be linear and planar, whilst the ground truth may be quite different. Significant error could be caused by a mound positioned at the centre of the footprint for example. Harding *et al.*, (1994) found that, for an individual pulse, predicted range error was predominantly dependent on slope and therefore suggests an average of subsequent measurements are taken. This however, would not allow for the detection of change and could be only meaningfully used in areas of stable elevation.

Harding and Carabajal, (2003) confirm the anticipated difficulty of vegetation and ground return mixing in areas of steep terrain. As previously mentioned, slope additionally contributes to the pointing uncertainty effect on the pulse centroid and

Harding and Carabajal, (2003) identify independent information regarding footprint location and ground slope as requirements for interpretation.

Roughness

As the laser beam has a near-Gaussian energy pattern, the centre of the footprint receives more energy; therefore the surface roughness calculation is most representative of this area.

If most variation is present towards the edge of a footprint, the roughness estimated from the waveform will tend to underestimate the ground truth. This is because pulse spreading is determined by the elevation profile within the footprint weighted by the normalised cross-section of the laser beam. As a result, roughness surrounding the footprint centre will be assigned smaller weighting and will therefore cause the pulse to spread less (Gardner, 1992).

Harding *et al.*, (1994) conclude that, at low slopes $\left(\sim 1^{\circ}\right)$ range errors are predominantly due to roughness. However as slope increases, elevation variation as a result of roughness is insignificant compared to elevation change due to sloped surfaces. Therefore smaller footprints produce more accurate results due to smaller surface variation within footprints. Gardner, (1992) also finds that range errors are generally less sensitive to surface roughness or reflectivity variations. The sensitivity of waveform shape to vegetation and ground surface characteristics is an area of current development (North *et al.*, submitted).

Surface elevation

Sampling of the same area at time intervals would allow observation of surface dynamics. This occurs at GLAS ground track crossover points (where ascending and descending passes cross) and for repeat orbits. However, at crossover

locations, interpolation could be needed as data sampled areas are obtained approximately 172m apart with a locational error of several metres due to pointing knowledge. Furthermore, even when the satellite path repeats a previous ground track, the illuminated section of the ground may be different as footprints will not necessarily fall upon the same point along the track. Brenner *et al.*, (2003) identify a research need to investigate repeatable results for detecting elevation changes.

Reflectivity and Scattering Events

A further potential error may be due to the assumption of Lambertian scattering (equal in all directions) of all components within a footprint, whilst in reality surfaces are likely to differ from this (Gardner, 1992; Brenner *et al.*, 2003).

Multiple scattering of photons within vegetation canopies will delay the pulse being returned to the sensor and thereby produce an artificially low trailing edge to the waveform. The effects of this and sensitivity to other vegetation parameters can be accounted for through radiative transfer modelling (North, 1996; North *et al.*, submitted).

3.1.5.3 Atmospheric Influence

The near infrared LiDAR wavelength causes the laser signal to be affected to a degree by atmospheric conditions producing signal attenuation; dense cloud may prevent or severely reduce energy penetration to the ground producing a highly distorted returned signal, whilst thinner clouds or aerosols produce forward scattering. This effect can be observed in Figure 3.8.

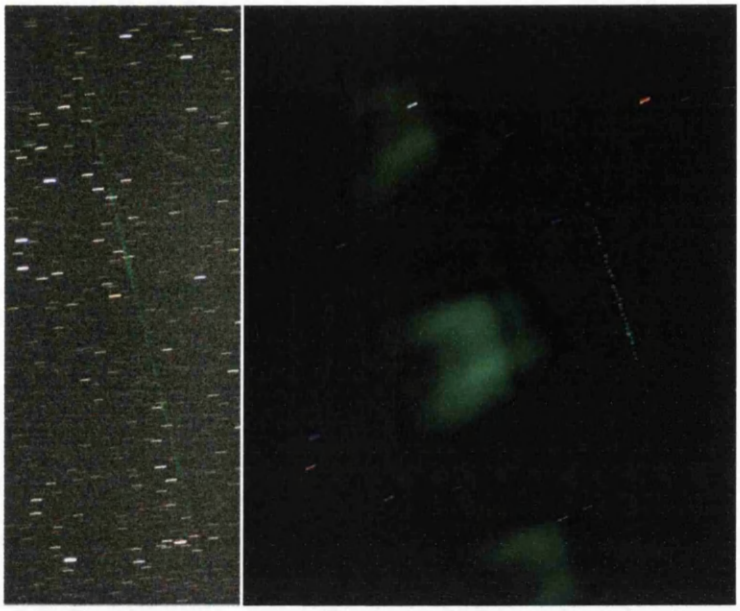


Figure 3.8. (left) Time lapse photograph illustrating ICESat pulses taken from the Mojave Desert, California. The blurred appearance of the stars is due to the Earth's motion during image exposure. (right) Scattering of green light photons (532nm) from thin cloud above Santa Rosa, New Mexico, seen as broad blocks of colour to the left of the laser pulses (NSIDC, 2003).

Forward scattering by clouds and atmospheric particles broadens the pulse and has been previously discussed. Corrections are applied to GLAS data to account for this (Herring and Quinn, 2001). However, these effects need to be recognised and reference made to quality flags to identify any potential error as a result.

3.2 Summary

This chapter has introduced the principles and limitations of ICESat/GLAS satellite LiDAR data and the pre-processing which is undertaken in order to create the GLAS Data Products which are used in this project. These products are presented in the following Chapter.

Satellite LiDAR datasets are challenging to interpret due to complex interactions with surface features and production of multiple waveform indices which require considered analysis. However, in contrast with estimates using optical

systems, GLAS data offer a unique opportunity to directly measure parameters based on the physical structure and properties of vegetation on scales which have previously been unfeasible. This data source provides a new opportunity to explore the proof of concept for analysis of vegetation using satellite LiDAR remote sensing.

Chapter 4. Data Sources and Methodology

This chapter introduces the sources of satellite and airborne LiDAR data used within this research and their formats and products as supplied to the end user. An overview of data processing tools is also provided.

The Forestry Commission and its Forest Research Agency have offered considerable support throughout the project. Their role in British forestry, forest organisational structure and data management tools are presented. The research study site is also managed by the Forestry Commission and a description is given of forest characteristics and management undertaken.

Finally, the field techniques which are used for ground-truth validation in this study are presented in detail. These measurements aim to incorporate both intercepted features represented within waveforms and biophysical parameters used within forestry.

4.1 Lidar Data Description and Products

4.1.1 Satellite LiDAR

The data source used within this study is the Geoscience Laser Altimeter System (GLAS), a full waveform LiDAR profiling system which is carried on the Ice, Cloud and land Elevation Satellite (ICESat), launched in January 2003.

4.1.1.1 ICESat/ GLAS Data Products

Fifteen ICESat/GLAS data products are publicly available from the National Snow and Ice Data Center (NSIDC) which distributes the data on behalf of NASA. Data products are divided into two Levels and those of relevance to this project are L-1A (GLA01), L-1B (GLA05-06) and L-2 data (GLA14). The product information which follows is summarised from information available from NSIDC, (2003).

Products contain the transit time of the pulse in Universal Co-ordinated Time referenced from noon on 1st January 2000 (J2000 UTC). One record consists of approximately one second of data, sampled at 40 shots second^{-1}. Each record is assigned a unique record index, based on the time of the Level-0 Header records after conversion to J2000 UTC. Thirty-nine delta times are provided which are added to the full UTC of the first transit time in each record to calculate the 40 laser shot times. Following accepted IDL convention, numbering of laser shots begins at zero. Therefore frame references consist of the unique record index (e.g. 885917496), the pulse number (0-39) plus J2000 UTC for each laser pulse. These allow cross reference between data products particularly since geolocational information is not

present within GLA01 (it contains a predicted orbit) and therefore spatial searching is not possible when ordering data for this product.

For GLA05, GLA06 and GLA14, orbits start and stop at the same demarcation position of ±50°. Each numbered track begins and ends at the ascending Equator crossing with Track 1 being closest to the Greenwich Meridian Line. The pass crossing the study site used within this research has track reference 1288 and was acquired at 08:28 on 22nd October 2005.

Data are available in the form of granules from the Search 'N Order Web Interface (SNOWI) or the Warehouse Inventory Search Tool (WIST); (NSIDC, 2003). If ordering data in this way, product GLA01 must be searched separately using acquisition times from higher products as reference criteria.

Data granules of each product vary in size and composition. GLA06 granules are approximately 7MB and contain 23 minutes of data (¼ orbit split at ±50°) with 56 granules being obtained per day. In contrast, GLA14 granules contain 14 orbits with a single data granule being obtained each day of laser operation. These can be 209MB in size.

For relatively small areas, a subsetter tool (NSIDC, 2003) can be used to restrict data searches for select products to the region of interest. For each laser operation, performing a search in this way returns a single file for each data product which combines all passes crossing the search area. The subsetter tool was mainly used within this project.

GLAS Product Summaries

GLA01 is the Global Altimetry Data product and was not originally intended for use by the general science community. It contains the raw waveforms and is also used to assess instrument health and data quality. Both received and transmitted pulses are provided in un-calibrated counts in nanoseconds (ns) and therefore must be converted to distance units in order to combine with parameters within higher products. The ordering of the received echo is in time-reversed order i.e. the value of the first sample is for that furthest from the spacecraft in time. The transmitted pulse is in time order.

GLA05 consists of Global Waveform-Based Range Corrections Data and contains the information necessary to characterise the waveform. This is used with GLA01 to create the L-1B (GLA06) and L2 (GLA14) elevation products. This product was also not originally anticipated for end users as the intention was that L-1B and L-2 products should contain all necessary information.

L-1B (GLA06) and L-2 regional products (GLA14) are at full 40 shots second^{-1} resolution. These data are geolocated to the centre of the footprint and **GLA06** is therefore used to identify ICESat/ GLAS footprint positions. GLA06 is the Global Elevation Data product and is used in conjunction with GLA05 to create the regional products GLA12-15.

Each L-2 regional product is written with an algorithm specific to each surface type (Ice Sheet; Sea Ice; Land; Ocean) in order to provide relevant surface elevation data i.e. different algorithms calculate the elevations in their respective products. Surface type masks define which data are written to each product and if any data within a record fall within a mask, the entire record is written to that product (therefore a Land product may contain Sea Ice data for instance).

GLA14 forms the Global Land Surface Altimetry product. Due to the complexities inherent in the returned waveform, a model fit to the returned signal is provided within this product, comprising the sum of six Gaussian peaks. This aims to retain the dominant features of the raw waveform in order to facilitate interpretation.

In addition to the above, quality flags are provided which indicate corrections and sources of concern regarding the data such as if the received return is saturated or if significant forward scattering is present which would misrepresent the surface characteristics and affect the accuracy of elevations.

Visualisation and Product Expansion Tools

As this data source cannot be read using conventional remote sensing software, tools are provided to enable users to work with the products (NSIDC, 2003). These are IDL-based and operate using both the full IDL licence and the freely available IDL Virtual Machine™ (ITTVIS, 2008). Specific use of tools to extract data for processing is discussed in subsequent chapters and an overview of their function is provided here.

The NSIDC GLAS Altimetry elevation extraction Tool (NGAT) allows information to be read from GLAS products GLA06 and GLA12-15. The following fields are exported in ASCII format: record reference; date; time of acquisition (Hours: Minutes: Seconds: fraction of seconds); latitude; longitude; elevation and elevation of the ICESat/GLAS ellipsoid above that of WGS-84 (the i_gdHt parameter is provided as a positive number with the ICESat elevations being greater than WGS-84). Co-ordinates and elevations have been converted to decimal degrees and to metres respectively. An important fact to note is that records are filtered, meaning that, where quality concerns are present, information is not extracted for all laser shots. However, as the shot reference is not output and time of day is reported

as opposed to J2000 UTC, co-ordinates must be used to relate information to parameters extracted using other GLAS tools.

IDL Readers are also available to expand each GLAS product and output variables as ASCII files. Extracted data are listed sequentially according to record and laser shot reference number and can then be queried using a programming language. Three sets of routines are available. The first of these extracts the raw waveform from product GLA01; the second expands elevation information from product GLA06 along with reference SRTM DEM data; and the final routine extracts all variables contained within each product.

Both the IDL Readers and NGAT require a control file to be edited which calls relevant program files and specifies input and output files plus contains user-defined conditions.

An alternative to these tools is the ICESat/GLAS Visualizer software which uses a graphical user interface and was designed within the NASA/Goddard Space Flight Center Science Computing Facility. Whilst data analysis is limited within this application, the Visualizer tool enables users to gain an immediate visual impression of the data and to export parameters of interest for external processing. Waveforms can be previewed and overlaid with indices (e.g. beginning or end of the detected signal). The trend of indices along a ground track can be displayed at record-level resolution providing an indication of elevation change along a pass (i.e. representing the first shot of each series of 40). Additionally, basic calculations using these indices can be displayed. The most significant waveform parameters can be exported in ASCII format at full resolution of 40 laser shots $second^{-1}$ although other parameters remain at only record-level resolution. For relatively small areas such as

that used within this project, this tool can readily provide a visual indication of trends.

4.1.1.2 Pre-Release ICESat Vegetation Product

The forthcoming ICESat Vegetation Product (IVP) aims to directly provide an estimation of mean canopy height and related parameters from ICESat/GLAS waveforms to address end-user needs for vegetation analysis. The algorithm for this product (Lefsky *et al.*, 2007) is under review and development, but it has been made available with the kind permission of the product developer, Professor Michael Lefsky, Colorado State University, USA.

4.1.2 Airborne LiDAR

The airborne LiDAR data which were used in this study were acquired using the Optech Airborne Laser Terrain Mapper 3033 (ALTM-3033), a discrete return laser scanner recording first and last pulses. The instrument was flown by the University of Cambridge, Unit for Landscape Modelling (ULM, year unknown) on behalf of the Natural Environment Research Council Airborne Research and Surveying Facility (ARSF, 2005). The data collection was commissioned by the Forestry Commission and the survey was flown using the Piper Chieftain Navajo (PA31) aircraft during August 2006.

Survey parameters are shown in Table 4.1. The instrument emits laser pulses at 1064nm wavelength with beam divergence of 0.2-1.0 mrads. Footprints of approximately 20cm diameter are produced with average point spacing of 45cm.

Table 4.1. Airborne LiDAR Survey Parameters

Laser Repetition Rate	33,000 Hz
Scan Frequency	29 Hz
Scanning half angle	15°
Mean flying altitude aslm	1050m
Flightpath overlap	65%

Resulting elevation data have an absolute RMS accuracy of better than ±15cm at 1200 metres flying altitude and horizontal placement of better than 53cm, calculated as 0.0005×altitude (ULM, year unknown).

Data are supplied in space-delimited ASCII format using the British National Grid co-ordinate system with field order as follows: Eastings, Northings, Elevation, Intensity (last pulse), Eastings, Northings, Elevation, Intensity (first pulse). Before distribution, the Unit for Landscape Modelling aggregated data into 2km by 2km tiles, irrespective of flightpath sequence.

Airborne LiDAR data were analysed using the IDL® based Queensland Remote Sensing Centre (QRSC) in-house LiDAR processing tools developed by John Armston, Department of Natural Resources and Water, Indooroopilly, Australia (Armston *et al.*, in preparation 2008). QRSC software allows all aspects of data classification, calculation of canopy metrics, terrain and vegetation statistics, plus output formats for visualisation and further analysis within most GIS software packages. Additionally Terrascan, a Bentley Microstation computer aided design plug-in, was used with the kind permission of Forest Research. Terrascan permits

greater user control over point classification. However analysis is not possible and classified data point clouds must therefore be exported for further processing.

4.2 Forestry Commission Data Sources

The Forestry Commission is the UK government department responsible for forestry. It is tasked with managing forest and non-forest resources (farmland, open mountain-tops, heathland, estuarine and riparian habitats) on behalf of the public. The management structure of this is outlined in Table 4.2 (Forestry_Commission, 2006; Forest_Research, year unknown).

Table 4.2. Forestry Commission Land Management Hierarchy.

GB-wide Forestry Commission-managed land
Forest Districts *GB divided into 30 districts*
Blocks *Discrete woodland areas/ design plans*
Compartments *Delineated by 'permanent' features e.g. roads/ rivers*
Sub-compartments *Discrete, relatively uniform areas within a compartment, determined by land-use or crop composition. Boundaries may vary over time due to crop felling, replanting, thinning, etc.*
Components *Up to nine distinct constituents of a sub-compartment e.g. dispersed species, two-storey crop and site attributes. Components are not spatially defined within sub-compartments.*

The distribution of Forest Districts and the location of Forest Research remote sensing 'super sites', where acquisition of remote sensing datasets is focused, are seen in Figure 4.1.

Figure 4.1. Forestry Commission management structure at a Forest District level. Remote sensing 'super sites' are indicated with red borders and are a focus for airborne remote sensing data acquisition and monitoring using field measurements (Source: Forest Research).

4.2.1 Forest Research

This project has been carried out in collaboration with Forest Research, the research agency of the Forestry Commission of Great Britain. Forest Research, and in particular Juan Suárez of the Northern Research Station, Roslin, Midlothian, have kindly provided access to data, processing software and expert advice regarding growth and process-based models, field techniques and airborne LiDAR data analysis.

4.2.2 Sub-compartment Database

The majority of publicly accessible woodland in Britain is managed by the Forestry Commission of Great Britain, a division of which, Forest Enterprise, maintains a sub-compartment database of inventory data for the Forestry Commission forest management units throughout Britain (Table 4.2). These sub-compartments are comprised of discrete, irregularly dispersed components with known species, physical conditions and management criteria (Forestry_Commission, 2006). Figure 4.2 shows an example of a 0.5km x 0.5km area of the Forest of Dean with the sub-compartment boundaries shown in red. The distribution and varying height of vegetation components within sub-compartments is visible in the aerial photography beneath.

Figure 4.2. Forestry Commission sub-compartment boundaries in red overlaid on an aerial photograph for a 0.5km x 0.5km area of the Forest of Dean. Sub-compartments may contain several components of different species, ages or canopy structure for example whose spatial distribution is not recorded. Components within sub-compartments are visible within the aerial photograph.

Whilst sub-compartments may have a mixed composition, their definition means that they should not be spilt by features such as roads, rivers or open space and should be relatively uniform in terms of relative mix of tree species and age classes, presence of canopy storeys, tree spatial distribution, yield class and habitat type (Forestry_Commission, 2006). Information contained in the sub-compartment database and associated component database files (.dbf) are listed in Tables 4.3 and 4.4.

Table 4.3. Component information fields (Forestry_Commission, 2006).

Component crop-level data
Land-use
Storey
Species
Origin
Propagation
Planting year
Yield class
Percentage area (of sub-compartment)
[Area]
Rotation
Mixture
Windthrow hazard class
Initial planted spacing
Stems per hectare
Stems per hectare date
Habitat code
Habitat condition

Table 4.4. Sub-compartment database fields (Forestry_Commission, 2006).

Sub-compartment site-level data
[Area]
Forest Park
Conservation Code
Forestry Commission Conservation Code
Ancient Woodland status
Soil type
Cultivation
Altitude
Terrain: condition, roughness, slope

Mixed composition sub-compartments occur in several ways: Sometimes felled areas are left to regenerate naturally, in which case Silver birch will often establish. Alternatively, when stands are ready for felling, they are offered out to tender as 'standing wood' for companies to bid for. When they are subsequently felled, the broadleaf species are often left standing as they are of little commercial value; in this situation the irregular distribution of components within sub-compartments is more incidental than deliberate. On occasion, the Forestry Commission will fell areas directly for sale as cut timber and in this case it may be more of an active decision to retain broadleaf components within stands. Furthermore, stable, broadleaf species may be retained along pathways to act as wind breaks and to protect exposed areas of young saplings or mature stands prior to felling. These factors can result in several species being dispersed among stands containing commercially planted crops (pers. comms. Judith Lack, Forest of Dean District Office).

4.2.3 Yield Models

Forestry Commission yield class models (Edwards and Christie, 1981) project a mean growth trend over time; this will vary between species and within species groups according to habitat conditions e.g. Figure 4.3. These models assume that tree density and stand composition at the time of planting remain constant and therefore do not take account of competitive suppression or dominance, nor of mortality due to natural or external factors e.g. wind-throw.

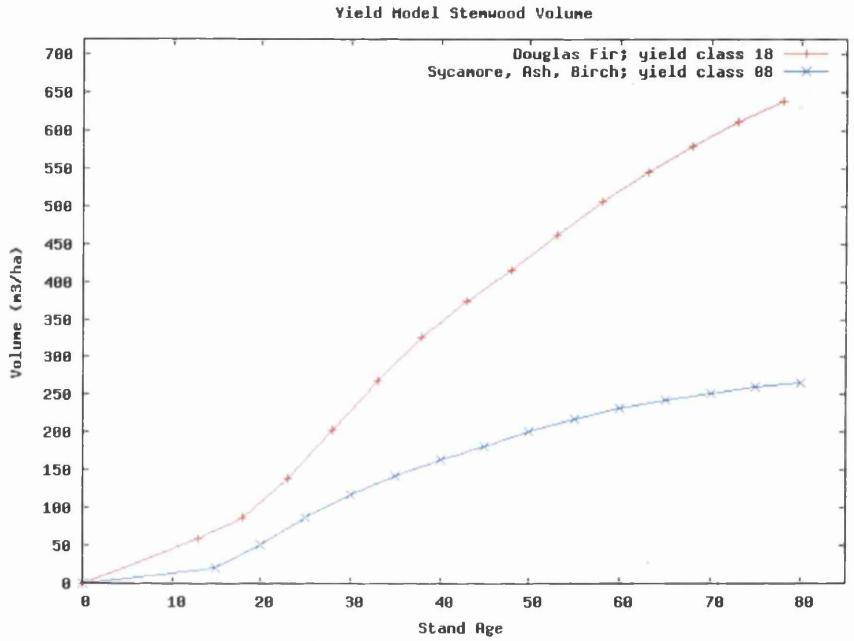


Figure 4.3. Non-dynamic yield models depict a mean trend by age group in order to predict biophysical parameters e.g. volume, top height, number of trees and basal area. Species, yield class, thinning regime and initial crop spacing are used as input parameters to select a model to apply.

The yield models were produced using plot-level data from thinning and spacing experiments collected since 1919. Each model was produced for a single thinning regime and therefore the actual use of a single management regime must be assumed in order to apply these models. Anticipated rate of growth for the given site conditions is determined by the yield class assigned to the crop (Edwards and Christie, 1981). At the time of their design, field data for old stands were not

available meaning that the trend of the model was extended beyond the validation data (Woodland Surveys, Forest Research, pers. comms.). Therefore greater potential discrepancies with actual growth patterns are to be expected in older stands.

Sub-compartment-level information is available as polygon shapefiles and an associated database file (.dbf) which can be linked with the related Component Area .dbf field within a Geographical Information System. ArcGIS 9.2 was used for this purpose. Species, yield class, thinning practice, initial spacing and planting year were extracted to refer to associated yield models whose names are formed as follows:

SP	06	IZ	14
Species code	*Yield class*	*Thinning regime*	*Spacing (decimetres)*

whereby thinning management practices and species codes are referred to as presented within Table 4.5 and Table 4.6 respectively.

Table 4.5. Forestry Commission thinning regime codes used to select yield models

Thinning Code	Thinning regime
IZ	Intermediate thinning no delay
IF	Intermediate thinning five years delay
IT	Intermediate thinning ten years delay
LZ	Line thinning no delay
LF	Line thinning five years delay
LT	Line thinning ten years delay
CZ	Crown thinning
NO	No thinning
T_x	User-defined thinning regime

Table 4.6. Forestry Commission species codes used to select yield models

Species code	Common name	Botanical name	Broadleaf/ conifer
AH	Ash	*Fraxinus excelsior*	B
AR	Alder	*Alnus spp*	B
AUP	Austrian pine	*Pinus nigra var nigra*	C
BCH	Bird cherry	*Prunus padus*	B
BE	Beech	*Fagus sylvatica*	B
BI	Birch	*Betula spp*	B
BIP	Bishop pine	*Pinus muricata*	C
CAR	Common alder	*Alnus gultinosa*	B
CLI	Common Lime	*Tilia europaea*	B
CP	Corsican pine	*Pinus nigra var maritima*	C
DF	Douglas fir	*Pseudotsuga menziesii*	C
EEM	English elm	*Ulmus procera*	B
EL	European larch	*Larix decidua*	C
EM	Elm	*Ulmus spp*	B
ESF	Silver fir	*Abies alba*	C
GAR	Grey alder	*Alnus incana*	B
GF	Grand fir	*Abies grandis*	C
HAZ	Hazel	*Corylus avellana*	B
HBM	Hornbeam	*Carpinus betulus*	B
HCH	Horse Chestnut	*Aesculus hippocastanum*	B
HL	Hybrid larch	*Larix leptolepis*	C
JCR	Japanese cedar	*Cryptomeria japonica*	C
JL	Japanese larch	*Larix kaempferi*	C
LC	Lawsons cypress	*Chamaecyparis lawsoniana*	C
LEC	Leyland cypress	*Cupressocyparis leylandii*	C
LI	Lime	*Tilia spp*	B
LLI	Large-leaved lime	*Tilia platyphyllos*	B
LP	Lodgepole pine	*Pinus contorta*	C
MAP	Maritime pine	*Pinus pinaster*	C
MB	Mixed broadleaves		B

MC	Mixed conifers		C
MCP	Macedonian pine	*Pinus peuce*	C
MOP	Mountain pine	*Pinus uncinata*	C
NF	Noble fir	*Abies procera*	C
NOM	Norway maple	*Acer platanoides*	B
NS	Norway spruce	*Picea abies*	C
OK	Oak	*Quercus spp*	B
OMS	Omorika spruce	*Picea omorika*	C
PDP	Ponderosa pine	*Pinus ponderosa*	C
PO	Poplar	*Populus spp*	B
RAN	Raoul	*Nothofagus procera*	B
RAP	Radiata pine	*Pinus radiata*	C
RAR	Red alder	*Alnus rubra*	B
RC	Western red cedar	*Thuja plicata*	C
ROK	Red oak	*Quercus borealis*	B
RON	Roble	*Nothofagus obliqua*	B
RSQ	Coast redwood	*Sequoia sempervirens*	C
SC	Sweet chestnut	*Castanea sativa*	B
SEM	Smooth-leaved elm	*Ulmus carpinifolia*	B
SLI	Small-leaved lime	*Tilia cordata*	B
SP	Scots pine	*Pinus sylvestris*	C
SS	Sitka spruce	*Picea sitchensis*	C
SY	Sycamore	*Acer pseudoplatanus*	B
VAR	Green alder	*Alnus viridis*	B
WCH	Wild cherry, Gean	*Prunus avium*	B
WEM	Wych elm	*Ulmus glabra*	B
WEP	Weymouth pine	*Pinus strobus*	C
WH	Western hemlock	*Tsuga heterophylla*	C
WSQ	Wellingtonia	*Sequoiadendron giganteum*	C
XB	Other broadleaves		B
XC	Other conifers		C
XF	Other firs (abies)	*Abies spp*	C
XP	Other pines	*Pinus spp*	C
XS	Other spruces	*Picea spp*	C

Yield class models therefore incorporate many aspects of forest management but are not available for all species or combination of factors. Where this situation was found, models for species with similar growth patterns or models most closely matching the required factors were used (priority was given to yield class, followed by thinning regime and spacing). For Lime, Common Alder, Sweet Chestnut and Mixed Broadleaf, the Sycamore/Ash/Birch model was used whilst European Larch was used for Mixed Coniferous stands as a conservative estimate previously applied to the Forest of Dean (Alan Walmsley, Forestry Commission, pers. comms.).

Yield class is determined by the maximum mean annual volume increment (m^3 ha^{-1}) for a stand and is given to the nearest even number. Thus a stand with maximum annual volume increment of 13.2 m^3 ha^{-1} during its growth cycle is assigned yield class 14. It should be noted that the time at which maximum growth is reached will vary considerably between species.

Currently the closest maximum increment curve to which a given stand is performing (and therefore yield class assigned) uses a known relationship between top height and cumulative volume production, dividing this by the stand age to derive mean annual increment. LiDAR estimates may provide an opportunity of improving this process either through confirmation of Top Height predictions (which can be determined from maximum canopy height) or volume estimates from LiDAR waveforms. A basis of this study is therefore to assess the potential of ICESat/GLAS for future validation or improvement of models of forest growth based on height estimates or as an indicator of vegetation distribution to complement conventional methods of forest inventory.

Using details within the Forestry Commission sub-compartment database and corresponding yield models (Edwards and Christie, 1981; Forestry_Commission,

2006), Top Height predictions were calculated for sub-compartments in which ICESat footprints are located. In several instances, footprints extended across adjacent sub-compartments. Use of Top Height predictions were explored for the principal component of the sub-compartment in which the footprint centre was positioned, for the principal component of all sub-compartments crossed and for all components within all sub-compartments encompassed by each footprint.

4.3 Research Study Site

4.3.1 The Forest of Dean

The study site used within this research is the Forest of Dean, Gloucestershire, UK, a highly heterogeneous, temperate forest within England and which borders south Wales (Figure 4.4). This has been identified by Forest Research as a remote sensing 'supersite' (Figure 4.1).

The Forest of Dean is classed as a semi-ancient forest and extends for an area of approximately 11,000 hectares (Forest_of_Dean_Partnership, 2006). The forest is unusual in terms of the United Kingdom, containing approximately 50% broadleaf and coniferous species, the majority of which are within planted, managed stands. The Forest of Dean also encompasses areas of non-intervention, classed as ancient woodland. These are predominantly oak stands which have been noted for their ecological importance and are protected as Sites of Special Scientific Interest (SSSI). Examples of these are Symmonds Yat, Speech House and Cannop Valley. These areas are fenced to prevent public access and no management is undertaken unless sites become dangerous (pers. comms. Judith Lack, Forest of Dean District Office).

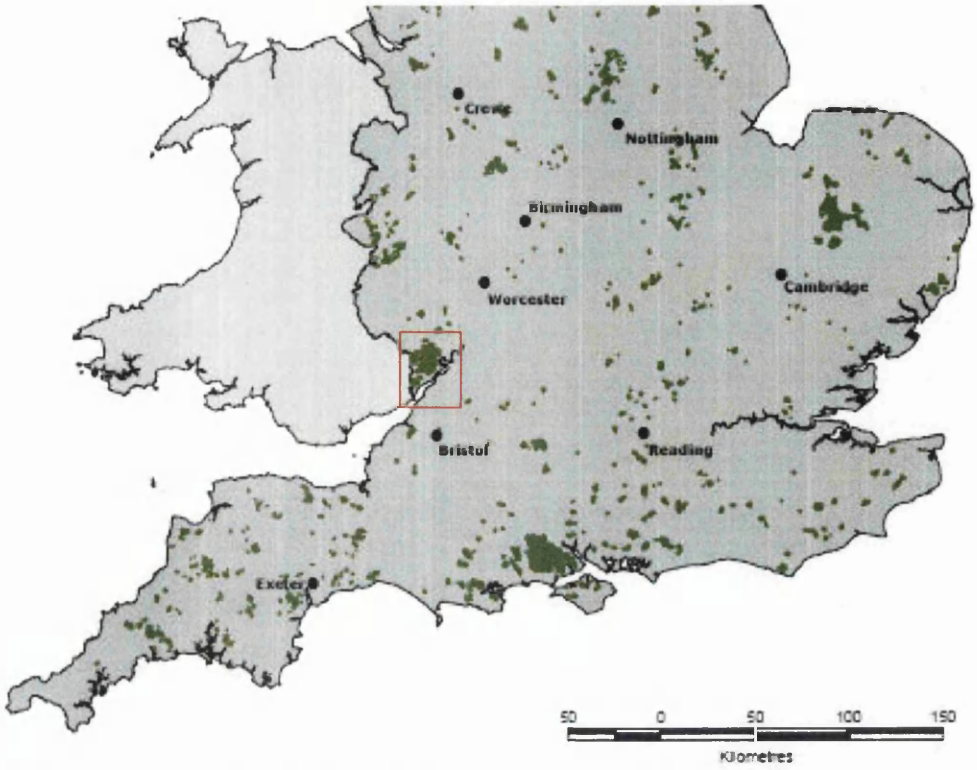


Figure 4.4. Map of Forestry Commission woodland in central and southern England with the Forest of Dean indicated (MAgiC, 2007).

The ICESat ground-track crossed the Forest between latitude 51.74° N and 51.88° N and longitude 2.54° W and 2.51° W. A section of the pass is seen in Figure 4.5. Of the stands sampled by ICESat/GLAS footprints, the most commonly occurring species were Norway Spruce (Picea abies), mixed broadleaf species, Oak (Quercus spp), Corsican Pine (Pinus nigra var maritima), Douglas Fir (Pseudotsuga menziesii), Scots Pine (Pinus sylvestris) and European Larch (Larix decidua). Vegetation was largely still in leaf at the time of GLAS data acquisition. Only footprints falling within the contents of the Forestry Commission sub-compartment database were used within this study and additionally, those footprints which partially or entirely traversed urban developments were excluded as both natural and unnatural surface features would contribute to waveform structure and could not be isolated from returned signals. This resulted in a total of 89 waveforms analysed.

Thus for each footprint location, the associated sub-compartment and constituent components permit reference to be made to relevant yield models (Edwards and Christie, 1981). This enables prediction of vegetation parameters for the given conditions.

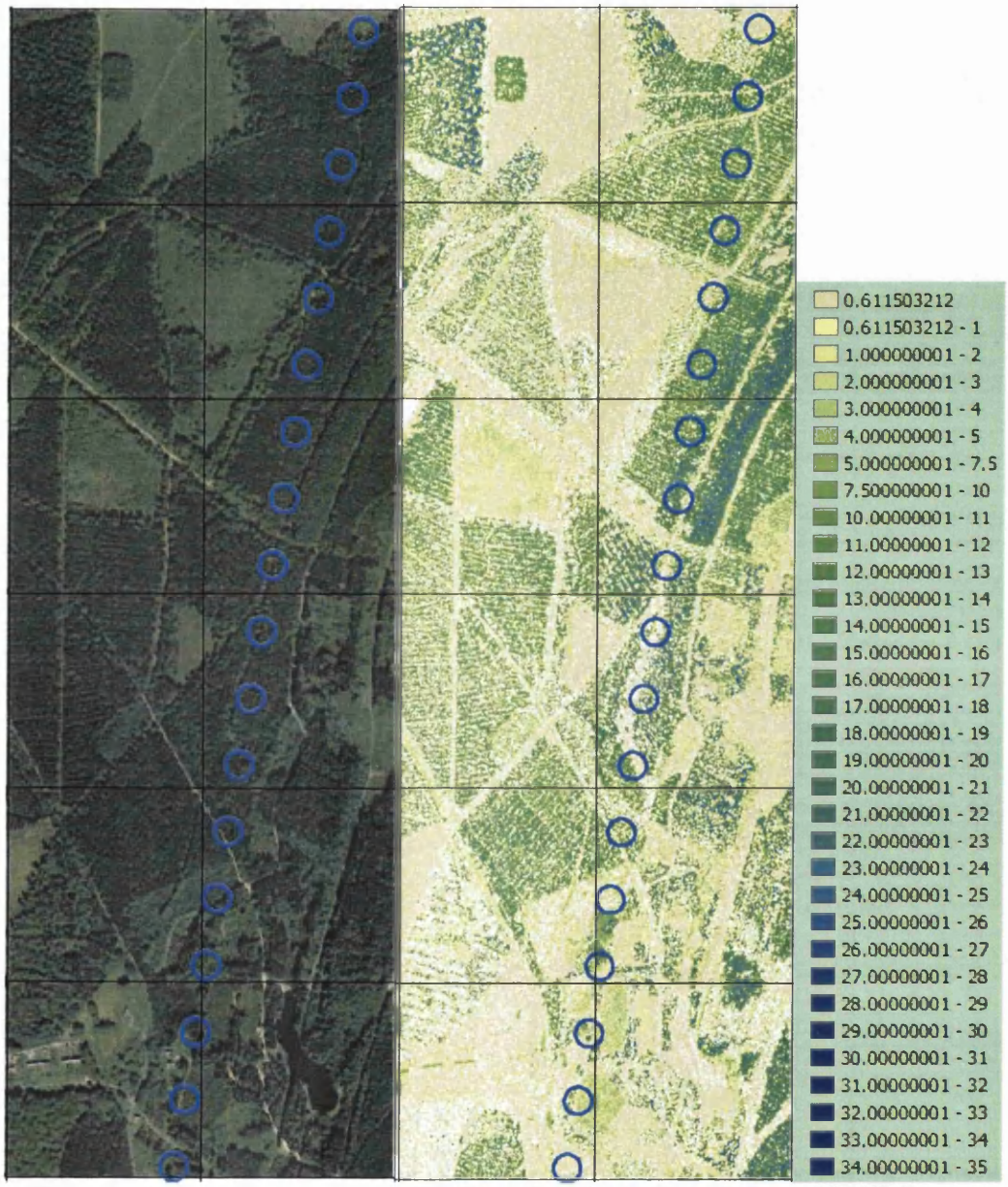


Figure 4.5. (left) Aerial photography (RGB) and (right) coincident airborne LiDAR Canopy Height Model (CHM) showing an area of 1x3 km within the Forest of Dean. Key illustrates height in metres for the canopy height model. Blue circles represent ICESat/GLAS footprints with centres distanced at 172m intervals. Minimum and maximum Eastings and Northings are respectively 363000, 209000 and 364,000, 212000.

4.3.2 Fieldwork

4.3.2.1 Objectives

Fieldwork was undertaken to allow an assessment to be made as to whether meaningful estimates of vegetation height could be extracted from large footprint LiDAR waveforms and to determine the accuracy of these estimates against measurements on the ground.

The principle underlying this was that conventional field methods of conducting stand-level assessments within forests rely on field surveys carried out for sample plots within stands to allow mean parameters to be estimated e.g. top height, an average of trees with largest DBH. This assumes that these measurements represent the mean stand conditions. Since GLAS footprints replicate this distribution of sample areas, should estimates of vegetation height derived from GLAS waveforms prove to give similar accuracy to field data, not only would this potentially allow a more frequent assessment of forest conditions to be made with repeat satellite passes, but would additionally provide the basis for developing methods of identifying indices within the region of waveforms returned from vegetation, which can be used for the estimation of further biophysical parameters.

The fieldwork was designed to measure the vegetation height and spatial variability for delineated footprint areas which coincide as closely as possible with the areas captured within GLAS waveforms. Particularly for such a heterogeneous site, this approach was taken to reduce error sources which would be introduced if using mean parameters estimated by stand-wide sampling.

To achieve this, adaptations were made to conventional field methods (Matthews and Mackie, 2006) in that measurements were confined to footprint boundaries yet the approach retains the accepted methodology of recording vegetation parameters of trees with largest diameter at breast height. This method may therefore be regarded as a modification of that used to calculate top height. In restricting field data to GLAS footprint areas, estimated vegetation height from waveforms could be directly evaluated against ground truth.

4.3.2.2 Field Measurements

Waveform-derived canopy height estimations were compared with field measurements of canopy height taken at nineteen footprint locations along the pass. A further four sites were visited and field observations were made although site access restrictions or felling operations prevented the fieldwork protocol from being followed. Field investigations were conducted in June 2006, creating an eight month discrepancy between satellite data acquisition and fieldwork. However, most of this period lies outside the growing season and only limited new growth could be observed.

Priority was given to sites classed by the Forestry Commission as principally evergreen, coniferous vegetation with high percentage cover within the first component of the sub-compartment in which the footprint centre is located. These were selected due to lower anticipated change in canopy cover during the eight month period October 2005 to June 2006 and aimed to permit more meaningful comparisons with yield model estimates as a result of less ambiguity from unknown distribution of components. However, by the heterogeneous nature of the site, almost

all ICESat/GLAS footprints sampled sub-compartments containing more than one component and several also crossed sub-compartment boundaries. In addition to the above, some sites with atypical waveform features were selected in order for this to be explored.

Where a discrepancy was identified between Forestry Commission classification and vegetated/non-vegetated status suggested by waveform estimates, field measurements at a further six sites were made during a second period of fieldwork during February 2007. This made field validation data available for a total of 25 footprint sites.

For areas of footprints where public access was prevented or where footprint areas crossed residential sites preventing comprehensive field measurements, site observations were made as well as limited field measurements. Accessibility via public rights of way was also a consideration for site selection for both health and safety reasons and due to access restrictions as a result of highly developed undergrowth. This may have produced a bias in field measurements as a disproportionate number of footprints areas measured may have crossed footpaths or clearings resulting in lower fractional cover in comparison with the study area mean. This in turn will be represented within returned waveform shape.

4.3.2.3 Fieldwork Protocol

At each site, eight height measurements (T1-8) were recorded of those trees with largest diameter at breast height (DBH), one within each 45° segment of 35m radius about the footprint centre, illustrated as below (Figure 4.6). In addition, nine

hemispherical photographs (A-I) were taken to record canopy cover at each site (Figure 4.7). This was carried out as follows, using the equipment listed in Table 4.7.

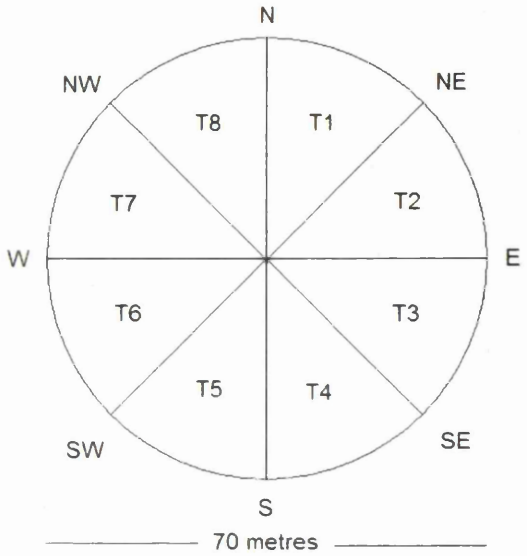


Figure 4.6. ICESat footprint tree height measurement protocol. Tree height, diameter at breast height (DBH), distance and bearing from the footprint centre were recorded of the tree with largest DBH within segments (T1 – T8).

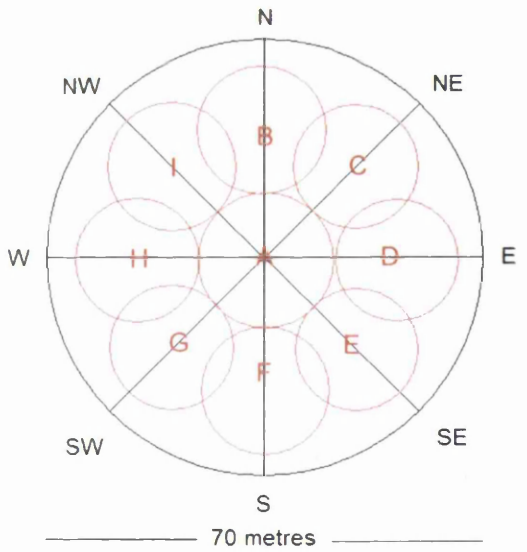


Figure 4.7. Hemispherical photograph reference positions at the footprint centre and at a distance of 20m from this location and estimated coverage, assuming a 20m diameter viewing area.

Table 4.7. Equipment used for field measurements.

Field equipment
Handheld GPS
'Chains' to delimit area
Trundle wheel
50 metre tapes
Two ranging poles
Clinometer
'Table top' camera tripod
Digital SLR camera with fish eye lens
Spirit level
Compass
Ordnance Survey 1:25,000 map

A handheld Global Positioning System (GPS) was used to locate the stated ICESat/GLAS footprint centre. Firstly, a hemispherical photograph was taken at this position before further equipment was set out. A low-standing 'table top' tripod was used meaning that the camera lens was positioned at approximately 35cm above the ground. This aimed to remove the effect of a dense grass or undergrowth layer which would adversely affect hemispherical photograph calculations, whilst retaining the full canopy profile which would be represented within waveforms. A bidirectional spirit level was used to ensure that the camera was as level as possible. The method aimed to take photographs during overcast conditions to reduce the intensity from the sun as fieldwork was undertaken throughout the day and so photographs could not be limited to low sun angles. The photograph number and reference location were then recorded.

A ranging pole was then placed at the footprint centre as the field site reference point. The first sector was delineated by placing 'chains' along 0° and 45° bearings from the footprint centre to a distance of 35 metres. Further hemispherical photographs were taken at a radius of 20 metres along these lines as described above. If this position fell at a tree or on unstable ground, the closest possible position was used.

Within the delineated sector, diameter at breast height - DBH (defined as 1.3m above the ground surface) of the broadest trees were systematically measured from the footprint centre to the field site boundary. A second ranging pole was placed beside the tree with the current largest diameter. When all possible contenders had been tested, the height of the tree with the largest DBH was measured as follows.

Tree height was measured using the tangent method of calculation. This is illustrated in Figure 4.8 where

$$Tree\_Height = H_1 + H_2 \tag{4.1}$$

$$\tan a = H_1 / D \text{ so } H_1 = \tan a \times D \tag{4.2}$$

$$\tan b = H_2 / D \text{ so } H_2 = \tan b \times D \tag{4.3}$$

$$\text{therefore } H_1 + H_2 = D(\tan a + \tan b) \tag{4.4}$$

The inclinometer used allowed direct height readings of H_1 *and* H_2 and using angles to the tree top and tree base from eye level at horizontal distances of either ten or fifteen metres from the tree.

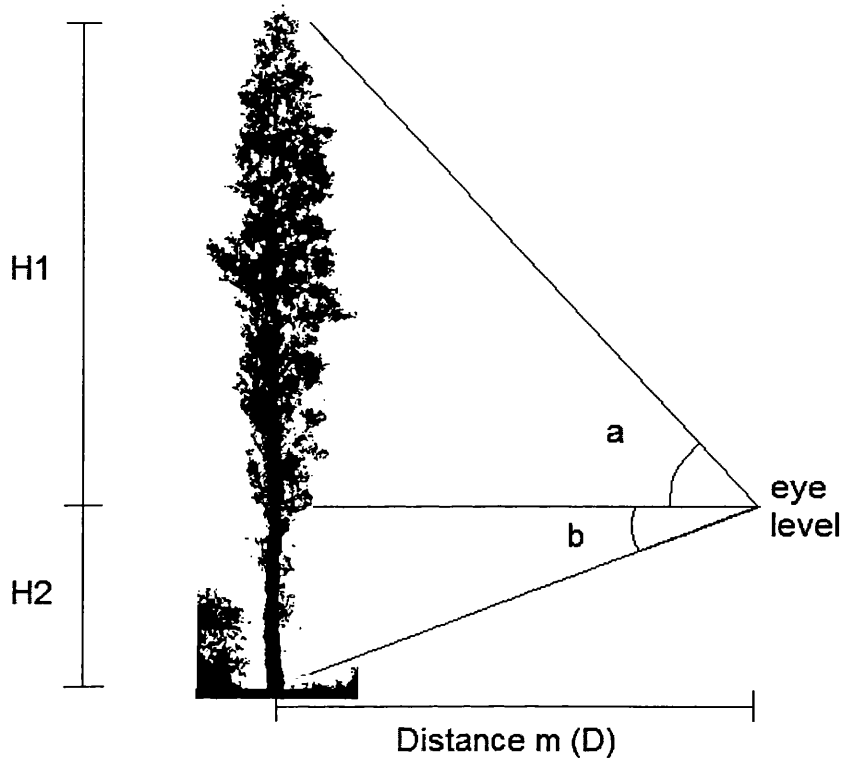


Figure 4.8. Calculation of tree height as the sum of H1 and H2 where these Opposite sides are calculated through trigonometry. Figure adapted from Skinner, (2002)

Where a slope was present, tree height measurements were taken along the slope contour where possible to allow calculation to be carried out as for level ground. To facilitate the identification of the tree crown top, where canopy density permitted, height measurements were carried out from a distance of 15m. H_1, H_2, distance and bearing of this tree from the footprint centre were recorded.

The first 'chain' was then rotated to a bearing of 90° from the footprint centre and the procedure was repeated. This was continued until 360° coverage had been completed.

Site observations were then made of species present, an approximation of percentage coverage of species (as a broad comparison with the sub-compartment

database records), single or multiple storey canopies, approximate height of any undergrowth, any litter or debris (which may indicate recent management that may explain any differences in waveform and field results). Where a slope was present, angle and approximate slope length were measured using ranging poles, inclinometer and 50 metre tape.

This protocol was repeated for each selected ICESat/GLAS footprint field site. Field measurements and Forestry Commission yield model and sub-compartment database calculations of Top Height are summarised in Table 4.8.

Table 4.8 Outline of field data collected and Top Height calculated from yield models for twenty-five ICESat/GLAS footprint locations at the Forest of Dean.

			Maximum height (m)		Canopy cover (%)
Footprint reference	Latitude	Longitude	Field measurements	Top Height	Hemispherical photographs
885917496_10	51.860584	-2.510985	20.25	15.78	91
885917496_11	51.859039	-2.511374	26.5	19.16	-
885917496_18	51.848198	-2.514125	3	0	-
885917496_19	51.846652	-2.514519	4	0	-
885917496_20	51.845108	-2.514911	6	0	-
885917496_22	51.842022	-2.51569	8	0	-
885917496_25	51.837399	-2.516834	6	10.9	-
885917496_26	51.835858	-2.51721	7.75	10.9	-
885917496_29	51.831216	-2.518329	17.25	18.58	89
885917496_36	51.820426	-2.520955	27.75	22	83
885917506_0	51.81427	-2.522491	26.5	25.1	89
885917506_6	51.804989	-2.52487	29	26.36	-
885917506_8	51.801901	-2.525664	23.5	20.7	90
885917506_9	51.80036	-2.52606	23.75	20.7	90
885917506_13	51.794195	-2.527646	28.25	22.86	91
885917506_14	51.792651	-2.528042	31	28.56	90
885917506_15	51.791106	-2.528435	31.5	28.56	89
885917506_16	51.789558	-2.528826	30.25	28.56	88
885917506_18	51.78646	-2.529599	26.25	22.66	89
885917506_19	51.784913	-2.529979	28	27.2	86
885917506_30	51.767937	-2.534066	28.25	27.34	84
885917506_32	51.764845	-2.534813	28	27.9	92
885917506_33	51.763297	-2.53519	28	26.3	78
885917516_4	51.746315	-2.539473	26	22.66	93
885917516_5	51.744763	-2.539875	24.75	23.9	90

4.4 Data Discussion and Error Assessment

The accuracy of inclinometer measurements accounting for both instrument error and distance measuring error stated within equipment documentation is better than ±1-1.5%. However Larsen *et al.*, (1987) note a tendency to underestimate tree height measurements using the tangent method, although they found 80% of measurements to be within 4.2% of actual height. With an average maximum canopy height of 26.6m found at the Forest of Dean, expected mean error may therefore be within 1.1m.

Assuming a circular area of radius 10m is captured by the hemispherical photographs, this sampling density is anticipated to provide adequate coverage of the footprint area (Figure 4.7).

Varying vegetation cover obscuring satellite reception resulted in GPS location accuracy ranging from 2m to 17m with mean accuracy of 10 metres. ICESat footprint location accuracy for the laser operation L3D used for this study, has yet to be released although is anticipated to be within several metres. This suggests a high degree of overlap between ICESat footprints and the areas sampled. Additionally, the major axis of laser 3D NIR elliptical footprint has resulted in length and standard deviation of 52.0 ± 1.1 metres (NSIDC, 2008) whilst earlier operations of laser 3 have produced ellipsoidal footprints of average size 47 x 61m. The field sample area of 70m diameter (following Carabajal and Harding, 2001) aims to compensate for location uncertainty and variation in footprint dimensions. The intention of 360° sampling was to record possible spatial variations in vegetation properties that would be represented within the returned waveform. The greatest of these field height

measurements for each site were used in comparison with waveform-derived estimates of maximum canopy height.

ICESat/GLAS coordinates and elevations are given with reference to the same ellipsoid used as TOPEX/Poseidon and Jason-1 which is approximately 70cm smaller than the WGS-84 ellipsoid. When comparing with co-ordinate systems used for other sources of data, this needs to be taken into account. Horizontal displacement of several centimetres is well within the uncertainty of footprint location mentioned above and is therefore ignored, however vertical differences are significantly larger and so the ICESat ellipsoid was converted as discussed in later Chapters to allow comparison with WGS-84 elevations (NSIDC, 2003).

As NASA continuously apply improvements and corrections to the pre-processing of GLAS data prior to their incorporation within ICESat products, new data releases are frequently announced. This research has made use of data release V026, since which, release 28 has been made available and re-processing by NASA for release 29 is currently in progress.

4.5 Summary

This chapter has introduced the study site, data sources and field measurements used within this investigation. The following chapters provide a description and evaluation of the research undertaken in the course of this project.

Chapter 5. Isolating the Waveform Canopy Return

This Chapter aims to ascertain whether estimates of vegetation height can be extracted from large footprint, satellite LiDAR data. Waveform height indices are compared with field-measured vegetation heights from coincident areas with a view to developing straightforward, repeatable methods which can be applied to large areas.

Two methods are developed. The first uses the limits of the waveform signal and requires a supplementary DTM dataset corresponding to the same area. The second method utilises parameters derived entirely from waveforms. Contributions to uncertainty within both of these methods are explored. Calculations within this research were carried out using the statistics package R, Version 2.3.1.

The results of these studies are further developed within research discussed in subsequent Chapters.

5.1 Vegetation Height Waveform Processing

Initial processing was undertaken before methods of estimating vegetation indices could be developed. The raw waveforms (GLA01) in the form of return time (ns) from the spacecraft to the intercepted surfaces were converted to one way

distance in metres so relative distances between features could be calculated. Following Carabajal and Harding, (2001), the alternate 'signal begin' and 'signal end' (GLA14), determined by exceeding a background noise threshold, were taken as estimates of the highest and lowest intercepted surfaces within a footprint. Using the terminology of Lefsky *et al.*, (2005), this distance is referred to as waveform extent (*WE*). For flat surfaces, vegetation typically produces a bimodal LiDAR waveform with a narrow abrupt peak indicating the ground return and a relatively broader, more complex return from the canopy (Figure 5.1). Thus, in keeping with Carabajal and Harding, (2001), the distance between 'signal begin' and a location within the waveform corresponding to the ground is assumed to be indicative of maximum canopy height.

Positions within waveforms are provided as negative offsets in metres with reference to the furthest gate from the spacecraft within the recorded range window (Section 3.1.4). In order to convert these to elevation, the following process must be followed:

$$Elevation_{parameter} = d\_elev + d\_lndRngOff - Offset_{parameter} - d\_gdHt \quad (5.1)$$

where *d_elev* is the waveform reference elevation; d_*lndRngOff* is the offset location of this within the waveform; $Offset_{parameter}$ is the offset of the parameter whose elevation is to be calculated; and d_*gdHt* is the geoid correction factor.

All methods were assessed against validation data collected in the field and discussed in Section 4.3.2.

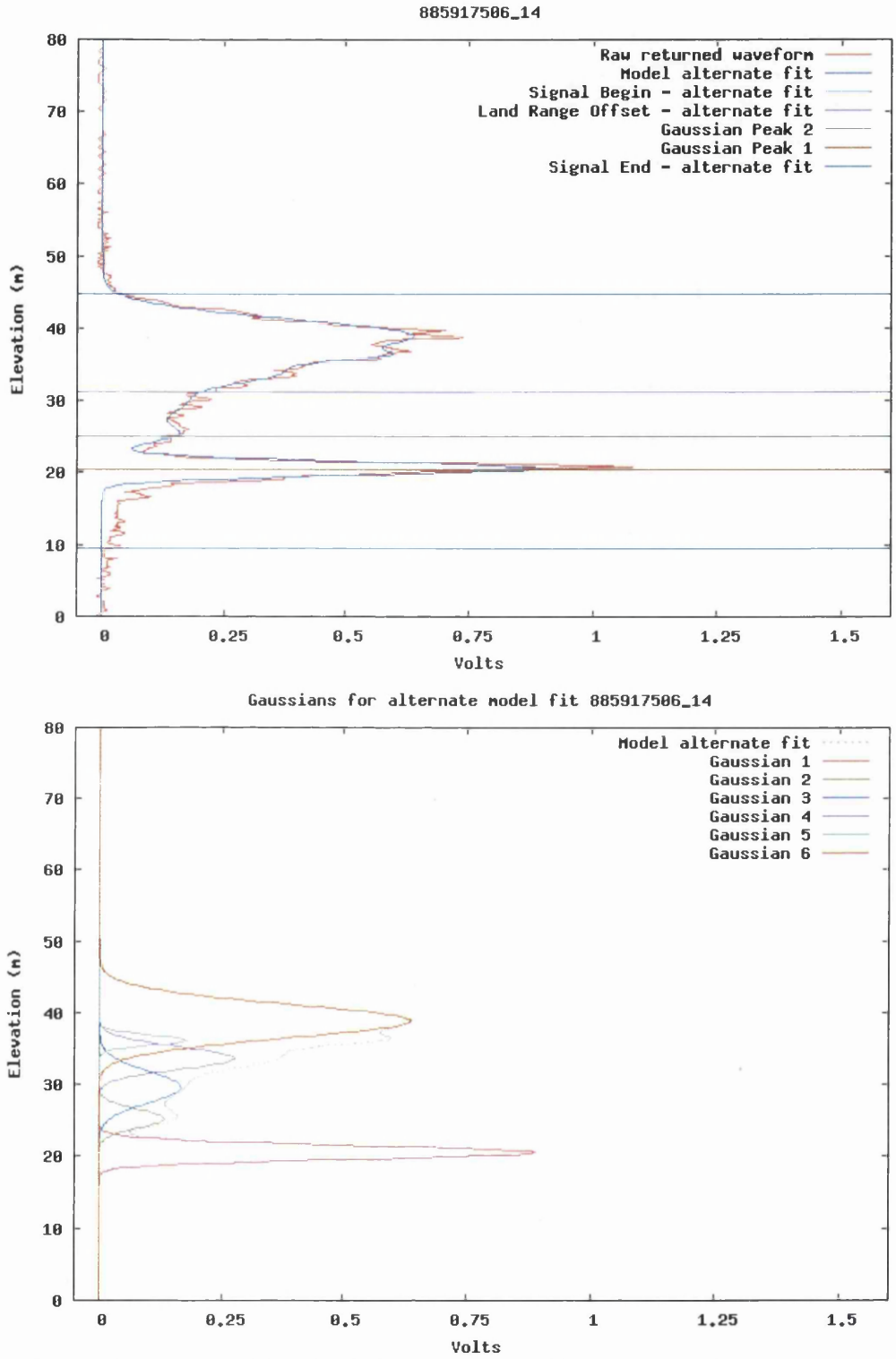


Figure 5.1 (above) Typical bimodal waveform, characteristic of a vegetated footprint. The raw returned waveform is shown with the smoother model fit. The figure shows Waveform Extent (signal begin to signal end) and waveform features as possible ground surface identifiers (Gaussian peak and land range offset positions) enabling estimations of maximum canopy height (signal begin to ground peak). (below) Model fit decomposition: the sum of six Gaussian peaks.

5.1.1 Waveform Extent/ Terrain Index Method

The first method uses the limits of the waveform (Waveform Extent), together with a calculation of terrain extent in order to estimate vegetation height.

To account for the contribution of topographic relief on waveform structure, a terrain index (*TI*) for each footprint location was calculated. This was formed by the difference in metres between the highest and lowest elevations contained within a 7x7 subset of the Ordnance Survey (OS) Land-Form PROFILE 10m resolution Digital Terrain Model - DTM (Ordnance_Survey, 2006) centred on the footprint co-ordinates. The matrix of elevations corresponds to the approximate ICESat/GLAS footprint area and therefore represents elevations contained within footprint boundaries.

This method is adapted from that of Lefsky *et al.*, (2005) and offers a means of removing the contribution of ground relief from the waveform leaving elevation due to overlying features.

Two approaches were explored. The first involves the subtraction of the terrain index from the waveform extent, thus removing the terrain extent from that of the returned signal (Equation 5.2). The second assumes a typical ground elevation can be calculated as a proportion of surface elevation difference. This uses a multiple linear regression to determine the relationship between waveform extent and the terrain index, calibrated with field-measured vegetation height (Equation 5.3).

$$GLAS_{ht} = WE - TI \qquad (5.2)$$

$$GLAS_{ht} = \alpha \times WE + \beta \times TI \qquad (5.3)$$

where $GLAS_{ht}$ represents maximum vegetation heights estimated using GLAS data, *WE* is Waveform Extent (m), *TI* is the Terrain Index (m) and α and β are coefficients applied to *WE* and *TI* respectively.

5.1.2 Gaussian Decomposition Method

The second method aims to derive estimates of vegetation height solely from returned waveforms using the structure of the returned signal to determine indices. The form of the returned waveform varies considerably depending on the characteristics of intercepted surfaces and their spatial arrangement (as laser energy diminishes towards the margins of the footprint). Thus the waveform represents returned energy from the ground surface and overlying features. In more open vegetation canopies, the majority of returns may be anticipated from the ground so forming the greatest amplitude peak within the returned waveform. However, for denser canopies, reduced laser penetration and energy attenuation through the canopy (Harding *et al.*, 2001; Parker *et al.*, 2001) would result in a less pronounced ground return in the waveform.

Therefore, given that the waveform maximum amplitude may be located within the canopy as opposed to a return from the ground, an alternative means of distinguishing the ground surface is required.

The possibility of developing an automated method of identifying the ground return signal from parameters provided within the GLAS products and the appropriateness of using this in order to estimate canopy height was explored. These parameters were also compared with height estimates from manual identification of

the ground return using visual inspection and local site knowledge. The suitability of approaches were then assessed using field validation data.

Gaussian decomposition provided within product GLA14 was used to achieve this. The ICESat processing software iteratively fits six Gaussian curves to the waveform as a means to summarise its complexity whilst retaining the dominant features of the original (Figure 5.1 - right). Several Gaussian fit parameters were compared as possible indicators of the ground return and these were evaluated with respect to their ability to estimate canopy height. These comprise the waveform reference elevation (*d_elev*), defined as the algorithm pick of the surface (NSIDC pers. comms.) and predominantly representing the centroid of the fitted waveform (Equation 5.4) and GLA14 Gaussian peaks 1 and 2, being of lowest elevations within the waveform (Equation 5.5). This is illustrated in Figure 5.1 which shows the ground return represented by the first fitted Gaussian peak in this instance.

$$GLAS_{ht} = SB - d\_elev \quad (5.4)$$

$$GLAS_{ht} = SB - Cnt_{GP1;2} \quad (5.5)$$

where *SB* is the elevation of the Signal Begin parameter, *d_elev* is that of the land elevation parameter and $Cnt_{GP1;2}$ is centroid elevation of Gaussian Peak 1 or 2, or a function of these. Maximum canopy height was thus estimated as the elevation difference between *SB* and the above parameters within waveforms.

5.2 Uncertainty Assessment

The work described in this section aimed to identify sources of error and to qualify their relative contribution.

Several factors are anticipated to contribute to error in waveform-derived height indices from large footprint LiDAR. These are discussed below. To quantify contributions to the small errors found in height estimates, multiple regression models were developed, to explain variation in error for both methods of estimated vegetation height from waveforms:

$$Error = \alpha Var_{canopy} + \beta species + \gamma Topography + \delta height + \varepsilon cover \quad (5.6)$$

Where *Error* is the difference between waveform estimates and field measurements, Var_{canopy} is upper canopy variability, *species* is species heterogeneity, *Topography* is accounted for using the Terrain Index, *height* is vegetation height and *cover* is canopy cover. Using field measurements, this aims to determine the coefficients α, β, γ, δ, ε, applied to the variables in order to predict error for the remaining footprints.

Uncertainty in estimates was calculated using the difference between waveform-derived estimates and field measurements, expressed as a percentage of field-measured vegetation height. Percentage error was used as an indicator of the significance of error; for example, a 2m discrepancy for a 30m tall tree being less critical than for a tree of 4m with regard to monitoring and inventory purposes (J. Suárez, pers. comms.).

Vegetation parameters derived from airborne LiDAR data were used to illustrate error dispersal and to calculate factors contributing to error. The processing of these data is presented in Chapter 7.

5.2.1 Field Measurement Uncertainty

Following the study by Larsen *et al.*, (1987), and taking account of both inclinometer and user error, field measurements are anticipated to be within ±4.2% of actual tree heights. Potential errors in tree heights measured in the field were therefore calculated according to this value.

5.2.2 Vegetation Height Variability

Height variability towards the upper part of the canopy is a potential source of error as the surface area of intercepted energy will vary with canopy surface roughness due to crown shape and individual tree heights comprising a stand. This will influence the energy returned and therefore the gradient of the start of the waveform signal.

Two indices of height variability at the uppermost canopy surfaces were explored as indicators of canopy roughness. The first follows the method of Lefsky *et al.*, (2007) to calculate the leading edge of the waveform. The leading edge is defined as the distance between the beginning of the signal and the highest position within the waveform at which the signal strength is half that of the maximum amplitude above background noise.

The second approach used the elevation difference between the beginning of the waveform signal and the greatest amplitude within the canopy return (domiant canopy height) as a proportion of GLAS-estimated maximum canopy height. The dominant canopy height was calculated as the centroid with maximum amplitude of

Gaussian peaks 2-6 if the first Gaussian is classified as the ground or of peaks 3-6 if the second Gaussian peak represents the ground.

Both methods aim to represent the gradient at the waveform leading edge and therefore indicate confidence in threshold detection.

5.2.3 Species Heterogeneity

Species heterogeneity, particularly for such a mixed composition forest, will affect waveform structure due to canopy shape, reflective properties and associated photon interactions (North *et al.*, submitted). Using the stand-level species mix from the Forest Enterprise sub-compartment database, the number of individual species comprising every stand covered by each footprint was used to take account of this.

5.2.4 Topography

A greater slope will broaden waveforms and increase the chance of vegetation and terrain signals being merged within the returned waveform. In order to incorporate this potential source of error, the terrain index (described in Section 5.1.1) was used to represent elevation difference of the ground surface within footprints.

5.2.5 Vegetation Stature

Vegetation stature has been found to influence the ability to estimate vegetation height from waveforms (Nelson, 2008). Estimates of vegetation height

from GLAS waveforms using both methods described within Sections 5.1.1 and 5.1.2 were used to investigate this.

5.2.6 Canopy Cover

Previous studies have also found canopy cover to affect estimates of vegetation height from large footprint LiDAR (Nelson, 2008).

This was tested using projected plant cover calculated from airborne LiDAR data for each footprint using return point counts above the interpolated ground surface. Canopy cover was estimated as the number of all vegetation points expressed as a fraction of total returns.

5.3 Results

5.3.1 Method using Waveform Extent and a Terrain Index

Using the difference between Waveform Extent (WE - elevation difference between signal begin and signal end) and Terrain Index (TI - difference between maximum and minimum DTM elevations within a 7x7 10m resolution subset) as an indicator of height of overlying features, produced R^2 of 0.89 and RMSE of 2.99m when compared with field measurements (n=19):

$$Field_{Ht} = 0.91*(WE\text{-}TI) + 4.86 \qquad (5.7)$$

coefficient significance $p<0.001$; intercept significance $p>0.01$. The Terrain Index varied over the site from 0.7m to 20m.

This offers itself as a simple method of estimating vegetation height from satellite LiDAR irrespective of the complexities of waveform structure. However a significant magnitude intercept is produced and this approach requires terrain data of appropriate resolution.

Following the method of Lefsky *et al.*, (2005), multiple regression was used to assess whether this could further improve the relationship between the Waveform Extent and Terrain Index with field measurements of maximum canopy height. The equation

$$Field_{Ht} = 1.0208*WE - 0.7310*TI \quad (5.8)$$

gave $R^2 = 0.90$, RMSE of 2.86m for n=19. Both coefficients are statistically significant in excess of 99.9% and the intercept was not significant and so could be removed forcing the trend through the origin as would be anticipated in the case of no vegetation cover. This method of regressing *WE* and *TI* against field measurements is subsequently referred to as R_{WT}.

5.3.2 Gaussian Decomposition

Vegetation height from waveforms was firstly calculated manually by visually identifying a ground return within the signal and estimating vegetation height as the difference between this position and the beginning of the signal. Estimates of vegetation height using the difference between a number of GLA14 waveform parameters and the Signal Begin parameter were then compared with these manual estimates. This aimed to develop an algorithm using waveform parameters available within product GLA14 which would enable a ground return to be estimated using waveform structure. The relationships are shown in Figure 5.2.

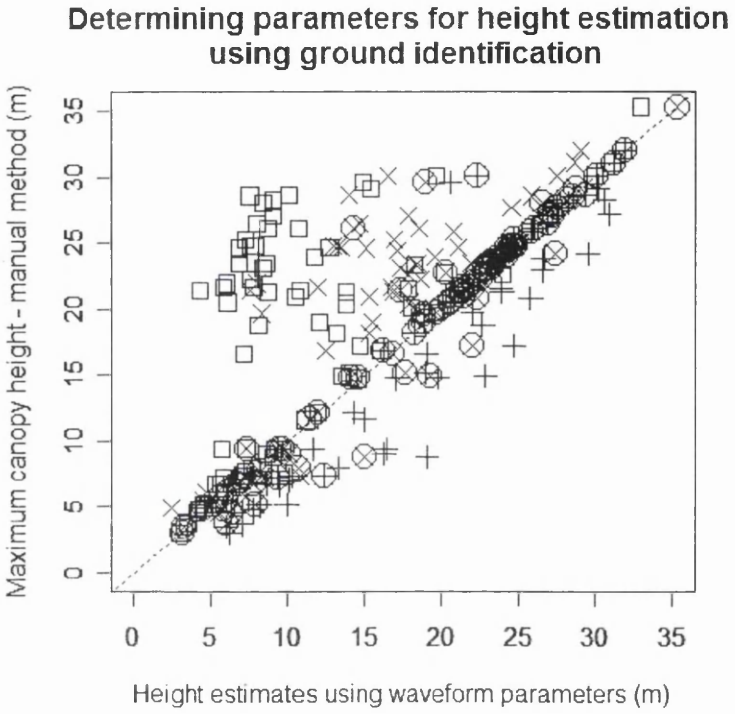


Figure 5.2 Relationships between maximum canopy height estimates using visual identification of the ground return (y axis) and estimates using waveform parameters where □ represents height estimates using GLA06 d_elev as the ground position; + uses Gaussian Peak 1; × is with Gaussian Peak 2 and ○ shows height estimates using the position of either Peak 1 or 2 determined by whichever has the greater amplitude.

The GLA06 land range offset position resulted in poorest correlation with maximum canopy height estimates using manual identification of the ground peak. Some points fall close to the 1:1 line and this is likely to be due to the fact that this parameter is predominantly located at the waveform centroid but is occasionally positioned at the echo peak which may correspond to the ground.

Use of Gaussian Peak 1 and 2 positions produced improved results with R^2 values of 0.90 and 0.70 respectively. However, it was recognised that a minor ground return from a surface below that of the majority of the footprint area may be unrepresentative and therefore use of Peak 1 may subsequently, on occasions, result in an over-estimation of canopy height.

A solution to this was explored and, compared with maximum height estimates using the manually identified ground surface, highest correlation of R^2 = 0.91 was achieved using the location within the waveform of either Gaussian Peak 1 or 2, whichever demonstrated the greater amplitude. This is hereafter abbreviated to GP_{MaxAmp}. The result of this is as follows:

$$Manual_{Ht} = 1.01 * GP_{MaxAmp} \quad (5.9)$$

RMSE=2.56m; coefficient significance $p<0.001$; intercept not significant.

The procedure was also repeated incorporating Peak 3, however this resulted in a reduced R^2 of 0.83 due to the probability of this Gaussian occurring within the canopy return signal rather than that corresponding to the ground surface.

Similar results were obtained when comparing waveform-derived estimates with field measurements of maximum canopy height. The GP_{MaxAmp} method produced the highest correlation with R^2 = 0.74 and RMSE 4.53m; coefficient significance $p<0.001$, intercept not significant:

$$Field_{Ht} = 1.05 * GP_{MaxAmp} \quad (5.10)$$

In this instance, additionally incorporating Peak 3 reduced R^2 to 0.59.

These results indicate the potential of estimating maximum vegetation height directly from returned satellite LiDAR signals using characteristics of the waveform structure.

5.3.3 Evaluation using Field Measurements

In order to identify the ground surface within waveforms, the Terrain Index procedure (R_{WT}) excludes a proportion of topographic limits from the Waveform

Extent whilst the method using ground identification (GP_{MaxAmp}) uses the elevation of the dominant ground signal. Both were found to produce statistically significant estimates of maximum canopy height.

A comparison of results for all approaches is detailed in Table 5.1 and Table 5.2. For this site, the GP_{MaxAmp} method improves on the concept often stated in the literature which assumes that the final Gaussian corresponds to the ground surface. The results of the R_{WT} method further improve on this, producing the highest correlation and lowest RMSE. Both methods prove the capability of estimating vegetation in a simple manner using GLAS data.

Table 5.1. A comparison of correlations produced by different methods of estimating maximum canopy height. Calculations compare estimates with nineteen coincident field measurements.

Method of vegetation height estimates	R^2	Intercept Level of significance	RMSE (m)
GLA06 *d_elev* ground identification	0.0828	19.49 $p < 0.001$	8.7579
GLA14 Pk1 ground identification	0.6242	Not statistically significant	5.5036
GLA14 Pk2 ground identification	0.4931	Not statistically significant	6.8615
GP_{MaxAmp}	0.7407	Not statistically significant	4.5264
Waveform Extent – Terrain Index	0.8927	4.86 $p < 0.01$	2.9949

Table 5.2 Regression equation results relating Waveform Extent and Terrain Index to field measurements using nineteen sites (R_{WT}). Both coefficients are statistically significant ($p < 0.001$)

R^2	Intercept	Coefficient a	Coefficient b	RMSE
0.9003	Not statistically significant	1.0208	-0.7310	2.8643

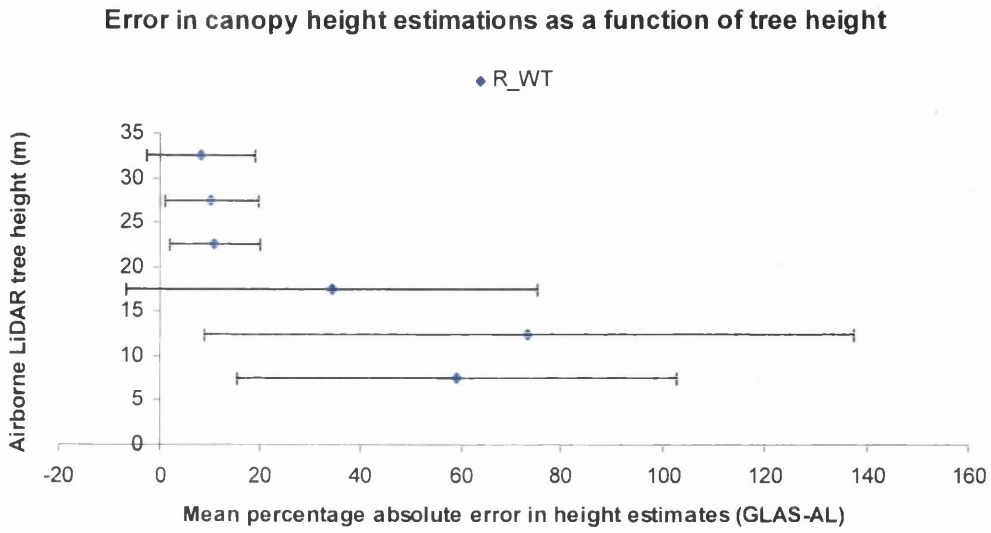
5.3.4 Uncertainty Analysis

Error calculated for waveform estimates by multiple regression indicate absolute error as a function of vegetation height variability, number of species within footprints, terrain index, vegetation height and canopy cover. Absolute error explained by equations 5.11 and 5.12 was 32% and 36% respectively for the two methods and so further factors are also expected to play a role. Coefficients are reported to two decimal places.

$$R_{WT}Error_{Abs} = -0.56Var_{canopy} + 0.08species + 0.10Topography - 0.01height + 1.17cover + 1.02 \quad (5.11)$$

$$GP_{MaxAmp}Error_{Abs} = 2.90Var_{canopy} + 0.55species + 0.15Topography - 0.01height + 3.47cover - 1.68 \quad (5.12)$$

Due to the primarily-managed nature of the Forest of Dean, numbers of instances of low canopy cover and tree height for example are too small to enable statistical analysis of effects; however their influence is nominally supported by data comparisons. The distribution of error in waveform vegetation height estimates as a function of the factors anticipated to play a role in uncertainty is illustrated within Figure 5.3, Figure 5.4 and Figure 5.5. In general, it can be seen that lower vegetation heights, greater slope and less dense canopy cover produce a larger error range.



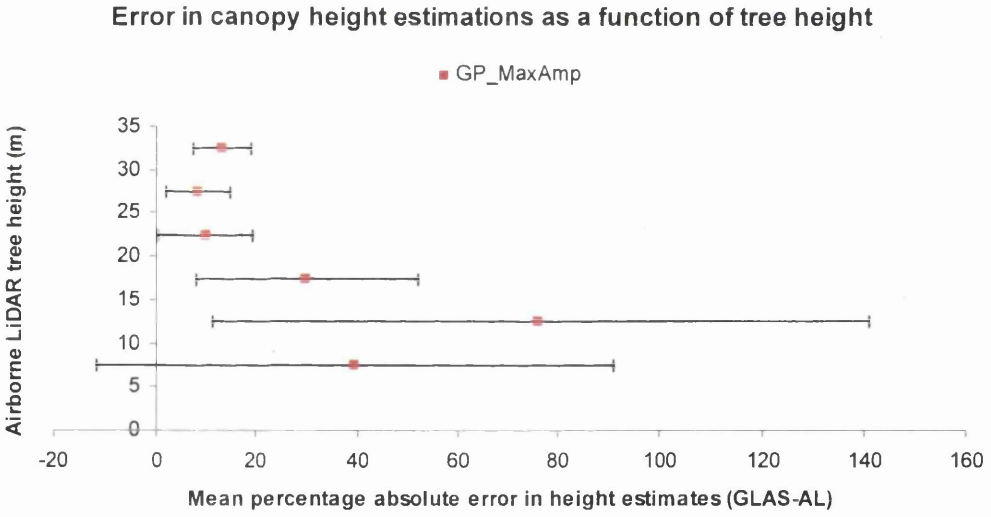
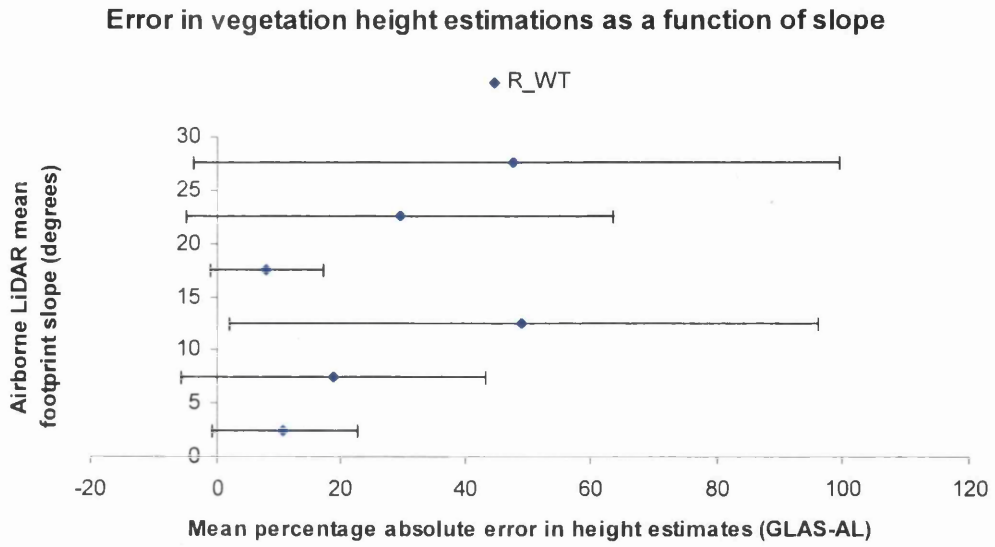


Figure 5.3. Illustration of error in waveform-derived estimates of vegetation height as a function of tree height. AL indicates airborne LiDAR-derived estimates. Error bars indicate standard deviation within five metre interval height bins.



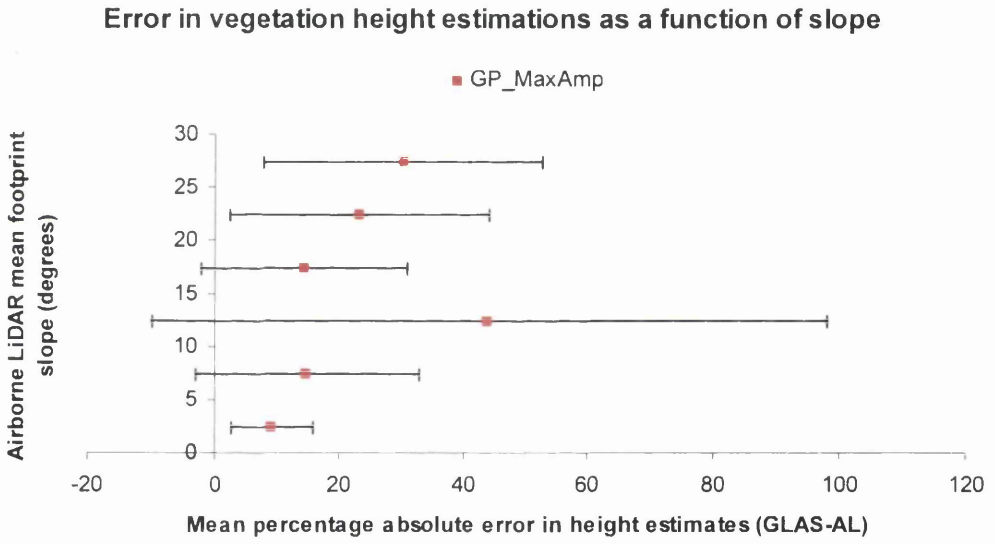
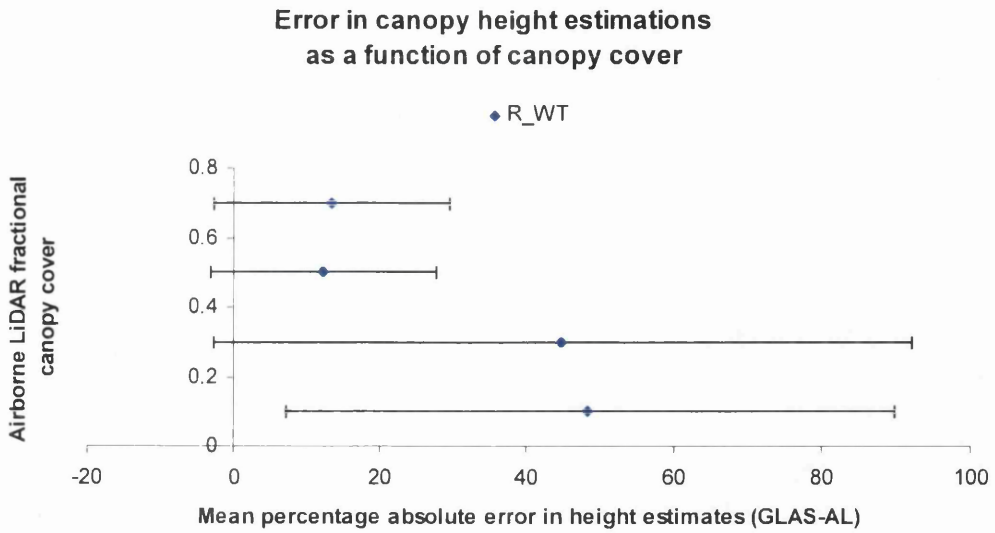


Figure 5.4. Illustration of error in waveform-derived estimates of vegetation height as a function of slope. AL indicates airborne LiDAR-derived estimates. Error bars indicate standard deviation within five degree interval slope data bins.



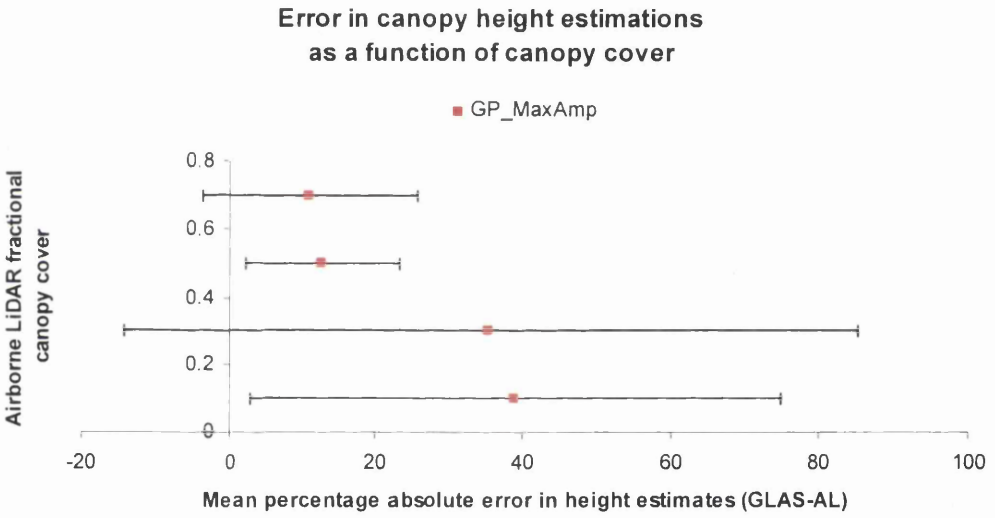


Figure 5.5. Illustration of error in waveform-derived estimates of vegetation height as a function of canopy cover. AL indicates airborne LiDAR-derived estimates. Error bars indicate standard deviation within 20% interval canopy cover data bins.

Sensitivity of waveform shape to structural and optical properties of vegetation and terrain are discussed in detail within North *et al.*, (submitted). This is illustrated within Figure 5.6 and Figure 5.7.

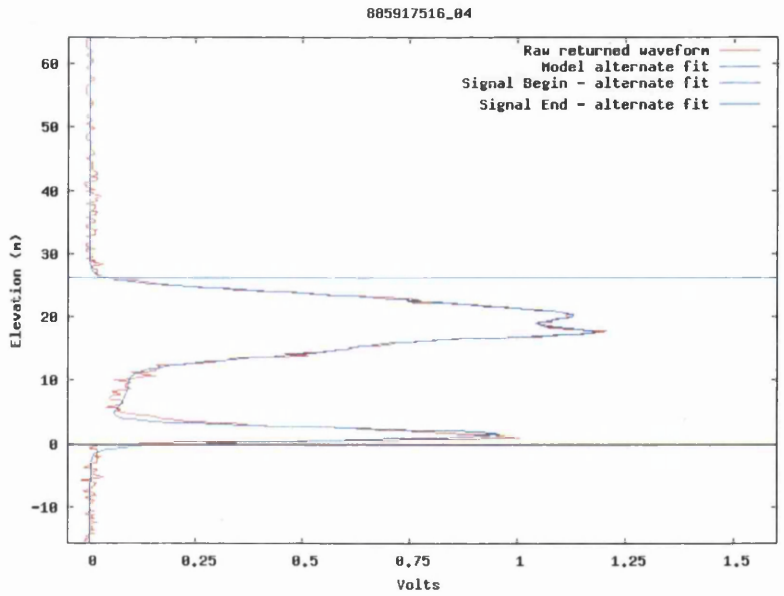


Figure 5.6. GLAS waveform returned for a relatively flat site indicating signal beginning and end positions, the raw waveform and the model fit to the waveform.Steep gradients at the beginning and end of the signal allow the signal begin and end positions to be located with more certainty.

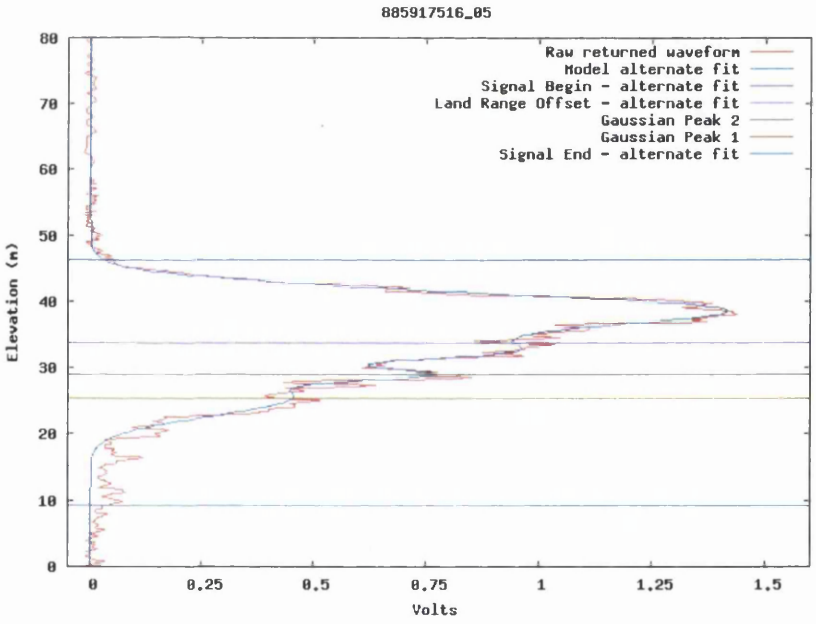


Figure 5.7. GLAS returned waveform for a vegetated footprint with terrain index of 18m. Signal beginning and end positions, the raw returned waveform and model fit to the waveform are indicated. Additionally, parameters explored as possible ground surface elevations are shown – Gaussian peaks 1 and 2 and the Land Range Offset (elevation of d_elev). Combined ground and vegetation signals mean a ground return cannot be visually determined.

Figure 5.8. shows the returned GLAS waveform for a site of low relief and relatively uniform vegetation height. The modelled effect of increasing slope (S),

carried out by Dr. P. North, Swansea University, shows the influence on the ability to estimate vegetation height. Simulated waveforms are for slopes of 0°, 10° and 20°.

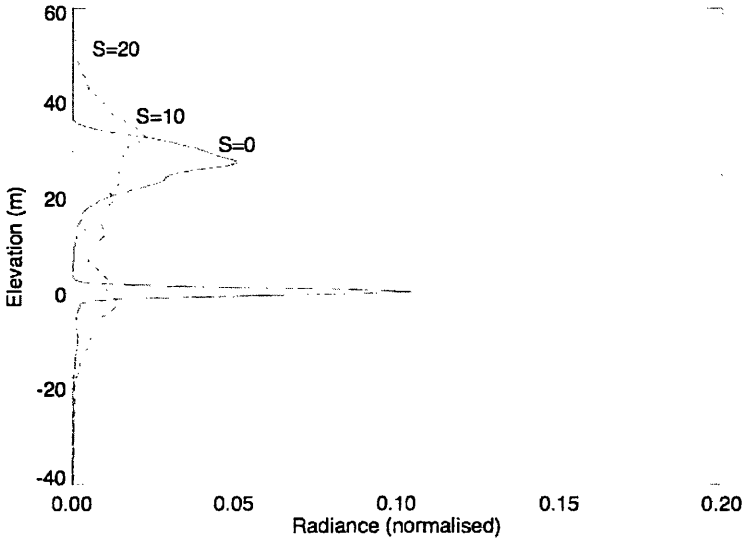


Figure 5.8. Simulated changes in waveform shape due to increasing slope (S) for 0, 10 and 20 degrees. Waveform amplitude is suppressed with increasing slope, limiting the presence of a distinct ground peak and reducing the gradient of leading and trailing edges to the waveform. The accurate identification of these three parameters is key to accurately estimating vegetation height Source: North *et al.*, (submitted)

5.3.5 Development of Results

The greatest divergence between estimates from waveforms and Forestry Commission Top Height predictions was seen for sub-compartments in which Forestry Commission records suggest no vegetation is present.

This may be explained if no formal planting has taken place in a felled or unmanaged area in which any invading species remain unaccounted for. Additionally, for flat reflective surfaces such as grassland, a 5ns emitted pulse would be expected to produce a 5ns return pulse to the sensor. An alternative explanation for apparent waveform elevations may be due to an unexpected effect of laser pulse width whereby the returned pulse width is greater than anticipated.

An investigation of this discrepancy using six footprints classified within the sub-compartment database as un-vegetated was carried out in February 2007. This revealed that the elevations suggested within the waveforms were in fact representing above-surface features present within the footprint areas, such as buildings, isolated trees and shrubs. For these footprints, deviations from conditions suggested within the sub-compartment database were not due to error within GLAS waveforms. Observed maximum heights for these locations were between approximately three and eight metres and mean offsets were produced of -1.07m and +1.32m for maximum height estimates using R_{WT} and GP_{MaxAmp} methods respectively.

Incorporating these additional field validation measurements resulted in the following revised results (Table 5.3) which confirm the ability to estimate vegetation height from GLAS waveforms by either identifying the ground surface using waveform structure or through use of an independent DTM. As previously, the R_{WT} method showed the best correlation and lowest RMSE. This suggests that, where a suitable DTM is available, this may offer the most reliable method of estimating vegetation height. Since the OS 10m DTM has country-wide coverage, this relatively simple method could be applied nationally. Equally, the GP_{MaxAmp} method has demonstrated that, for similar conditions and where such a DTM is not available, there is sufficient information within the waveform structure to permit vegetation height to be estimated solely from GLAS waveforms.

Table 5.3 Revised results of R_{WT} and GP_{MaxAmp} methods. Coefficients are statistically significant ($p < 0.001$). NS indicates, 'Not Significant'.

Method	R^2	Intercept	Coefficient a	Coefficient b	RMSE (m)
R_{WT}	0.9235	NS	0.96209	-0.52849	2.81
GP_{MaxAmp}	0.8426	NS	1.06	-	3.83

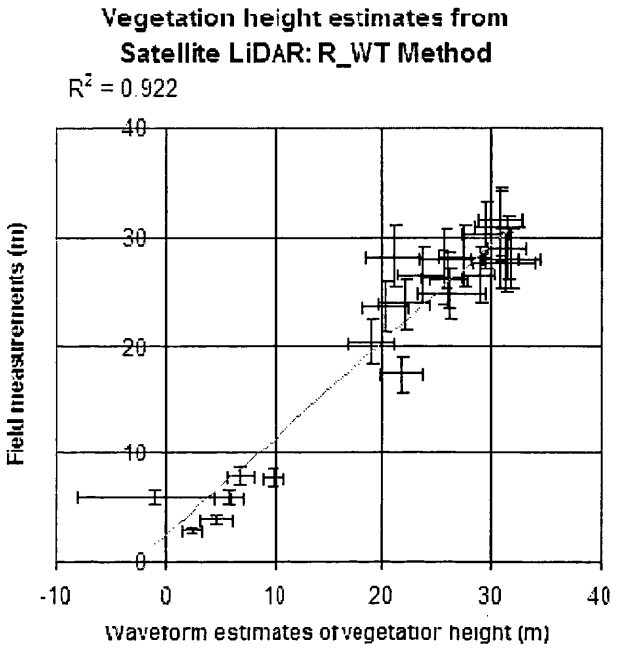


Figure 5.9. Potential error for field measurements (combined instrument and user error) and waveform estimates of vegetation height using the waveform extent and terrain index method (R_{WT}). Waveform-derived estimate error is calculated as a function of slope, vegetation heterogeneity, upper canopy variability, vegetation stature and canopy cover,

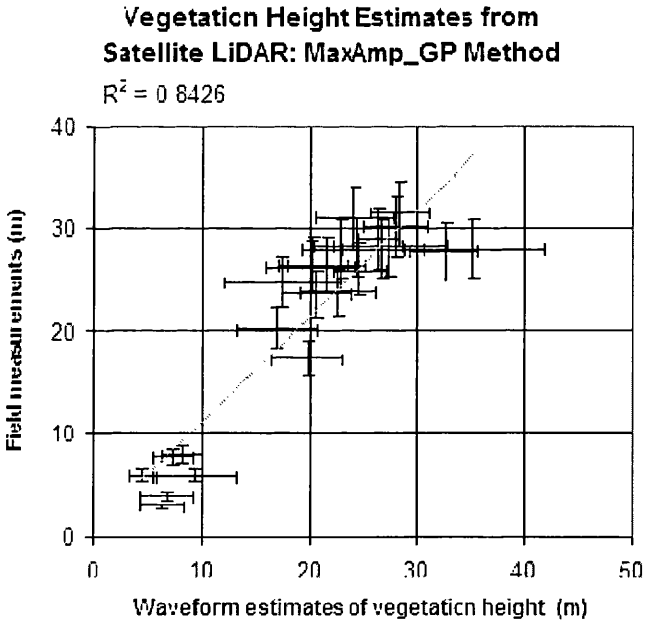


Figure 5.10. Potential error for field measurements (combined instrument and user error) and waveform estimates of vegetation height using the Gaussian decomposition method (GP_{MaxAmp}). Waveform-derived estimate error is calculated as a function of slope, vegetation heterogeneity, upper canopy variability, vegetation stature and canopy cover,

Representation of uncertainty within the revised results (Section 5.3.4) is shown through error bars in Figure 5.9 and Figure 5.10.

An overview of differences between waveform-derived estimates and field measurements is provided in Table 5.4. The Table clearly demonstrates the benefits of using the GP_{MaxAmp} method over either the first or the second Gaussian (*GP1* and *GP2*). This method marginally underestimates vegetation height which, as previously discussed, is typical of both small footprint and large footprint LiDAR systems. As waveform structure is used to identify the ground surface, this is likely to be due to misallocation of the start of the waveform signal. The R_{WT} method is shown to produce the smallest mean error and range of error. This disregards waveform structure, enabling vegetation height to be estimated as with discrete first and last return LiDAR plus a supplementary DTM. The overestimation is likely to be the result of the ground elevation being assigned below the true surface due to error of the end of the waveform signal. This is analysed further in the Discussion section.

Both methods demonstrate that, for forest-wide assessment, mean estimates of vegetation height obtained from waveforms compared with field measurements may be expected to be within approximately half a metre.

Table 5.4 Error range for GLAS estimates of vegetation height in comparison with ground truth data (GLAS – field measurements). Parameter *d_elev* is the algorithm pick of the surface, *GP1* and *GP2* are Gaussian peaks 1 and 2 respectively. The GP_{MaxAmp} uses amplitude of Gaussian 1 or 2 to identify the ground surface and the R_{WT} method uses regression of the waveform extent and a terrain index. NB method R_{WT} is calibrated using field measurements.

Method	Mean (m)	Minimum (m)	Maximum (m)
d_elev	-7.88	-23.96	5.97
GP1	1.00	-6.76	12.76
GP2	-2.48	-15.00	7.35
GP_{MaxAmp}	-0.55	-7.37	7.35
R_{WT}	0.41	-6.73	4.88

5.3.6 Comparison of Results

The methods of estimating maximum canopy height from waveforms described in this Chapter have been found to produce good relationships with field measurements. Figure 5.11 illustrates vegetation height differences along the ICESat/GLAS pass crossing the Forest of Dean. All estimates can be seen to similarly replicate height variation between footprints throughout the pass.

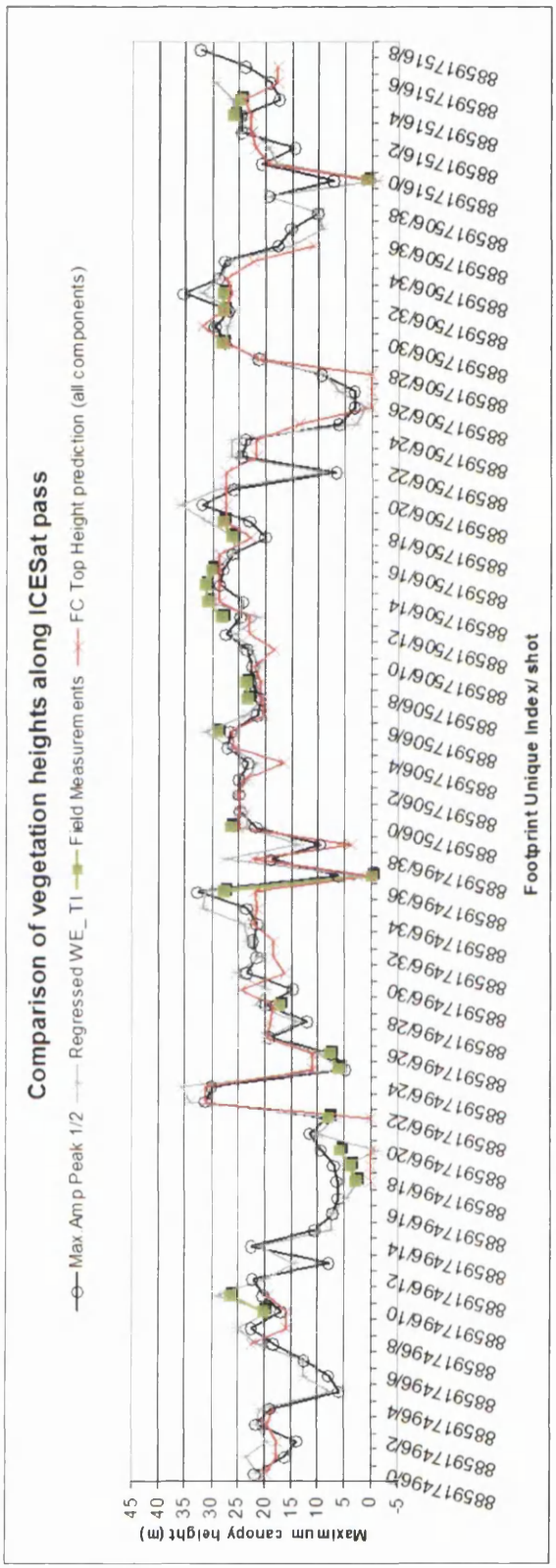


Figure 5.11 Relative canopy height estimates and field measurements for the Forest of Dean footprint locations (ICESat pass 22nd October 2005). *X* axis Footprint Unique Index and Shot reading from left to right represent the ICESat pass orientated NNE to SSW.

The ability of satellite LiDAR to retrieve data for such a complex and diverse area further indicates the potential of this technique for both carbon accounting and forest management.

5.4 Discussion

5.4.1 Method Intercomparison

Data from the Geoscience Laser Altimeter System aboard the Ice Cloud and land Elevation Satellite offer an unprecedented opportunity for canopy height retrieval at a regional to global scale. In addition, the data provide useful information for forest stand level assessment at coincident locations. In this Chapter, height indices from LiDAR waveforms were explored as a means of extracting canopy height; these were examined with reference to a mixed temperate forest in Gloucestershire, UK, containing planted stands with mean age of 51 years and mean maximum height of 26.6m.

Results have shown a mean underestimation of vegetation height from GLAS waveform structure (GP_{MaxAmp}) of 0.55m (R^2 = 0.84; RMSE = 3.83m). A likely explanation may be the misplacing of the beginning of the signal either due to spatial distribution of vegetation (e.g. Figure 5.12) as laser energy diminishes towards the margins of footprints or as a result of a low intercepted foliage surface area failing to trigger the waveform amplitude threshold immediately.

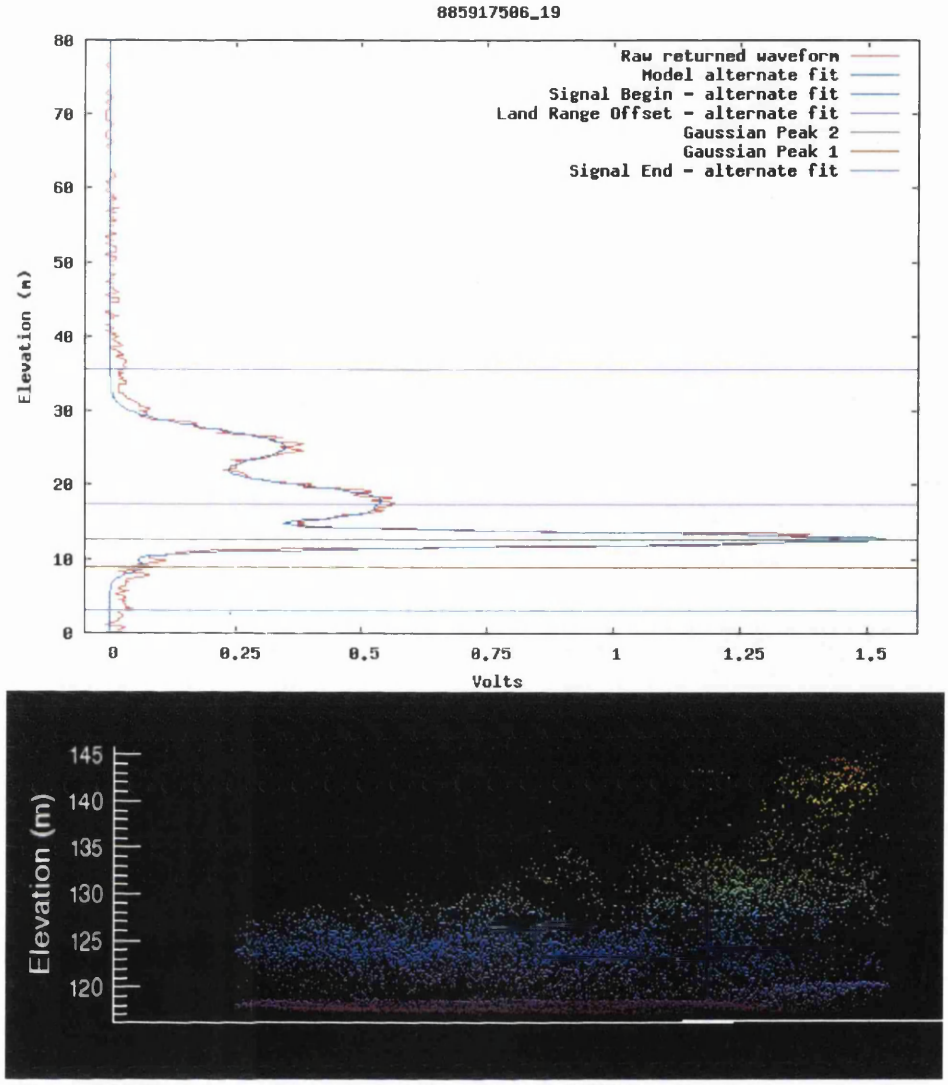


Figure 5.12 Waveform and corresponding airborne LiDAR point cloud covering three management components with the following predicted Top Heights: Common Alder 10.9m (observed heights were approximately 6-8m) and Scots Pine (16.5m and 27.2m). The beginning of the signal appears to partially miss taller vegetation towards footprint boundaries. Colours of the airborne LiDAR point cloud indicate elevation differences.

A predisposition of broad footprint LiDAR to under-predict vegetation heights is supported by Lefsky *et al.*, (1999a) who note an underestimation in upper canopy variability, whilst Harding *et al.*, (1998) and Ni-Meister *et al.*, (2001) also discuss the effects of this on determining signal begin and end positions. A tendency to underestimate canopy height has also been noted for small footprint airborne LiDAR (Chapter 2).

Within-footprint slope magnitude is anticipated to most influence height estimations which identify the ground surface using the shape of the returned waveform. Where relatively dense vegetation is present upon steep slopes, there is a greater likelihood for the returned signal from ground and canopy surfaces to be combined within the waveform. This may increase the possibility of wrongly identifying the Gaussian peak relating to the ground surface. For the Forest of Dean, where the within-footprint terrain index ranged from 0.7m to 20m, with a mean difference of 6.9m, this rarely prevented the presence of a distinct ground peak. However, the method of estimating vegetation height using waveform extent (elevation limits of intercepted features) plus an independently sourced DTM (R_{WT}), could avoid this potential difficulty.

Nevertheless, the GP_{MaxAmp} method avoids the necessity of using a supplementary dataset and so is suitable where topographic data may not be available. A further benefit of this method is that is aims to overcome possible errors in ground surface identification if a prominent 'tail' is present within waveforms by using the Gaussian amplitude to determine the most probable ground elevation.

In their study of Northern China, Sun *et al.*, 2008 found correlations between field measurements and vegetation height estimated from GLAS with $R^2 = 0.57$ and residual standard error of 4.46m for N=84. Their field method measures vegetation within footprint areas on the ground however the approach differs from that of this research. The Sun approach differs additionally in estimating vegetation height from GLAS as the elevation difference between the beginning of the waveform signal and the centroid of the final Gaussian peak. The results presented in this Chapter improve on this approach by aiming to assign a more representative ground elevation using dominant amplitude of the lowest two Gaussians.

Conversely the R_{WT} method presented in this Chapter has shown a minor positive bias of 0.41m (R^2 = 0.92; RMSE = 2.81m). This is likely due to the trailing edge of the waveform being extended through multiple scattering and resulting in the estimated ground elevation using a terrain index being below that of the footprint mean.

The accuracy of this method is dependent on the resolution of the available DTM and how well this may be expected to represent the ground elevation range within footprints of ~64m diameter. For example, Ordnance Survey provide a 10m resolution DTM nationwide for Britain whilst for the USA, the SRTM DEM is at 30m. However beyond this, publicly available SRTM data are only at 90m accuracy which, for the purposes of identifying terrain within LiDAR footprints of ~64m, may be insufficient.

The R_{WT} method does not rely on a clear ground signal being distinguished as it only requires the waveform limits to be determined (the equivalent of discrete first and last return LiDAR). However, accuracy of these indices is most sensitive to local conditions as vegetation structure will influence the ability to assign the beginning of the waveform to the elevation of the highest vegetation and multiple scattering will extend the waveform tail.

Results of the R_{WT} method compare favourably with those of Lefsky *et al.*, (2005) in which maximum R^2 value achieved was 0.68 and lowest RMSE of 4.85m (the most comparable site to this study was composed of temperate, coniferous forest of Douglas Fir and Western Hemlock of between near zero to 65m maximum canopy height and had mean slope greater than 18%. Results for this study area in Oregon, USA produced R^2 of 0.64 and RMSE 12.66m for N=24). It should be noted that it is not possible to draw direct conclusions by comparing results at the Forest of Dean

and findings in the above paper due to differences in slope and vegetation height. Field measurements undertaken in the above study are coincident with GLAS footprints however the method differs from that undertaken in this research. Additionally, the finer resolution DEM, described in this paper, provides a more detailed representation of terrain variation than the 3x3 subset, 30m SRTM DEM for Oregon. The inclusion of lower vegetation height observations and waveform estimates also had the effect of improving correlation in the case of the Forest of Dean.

Of the 89 footprints considered, on four occasions, the difference between Waveform Extent and Terrain Index resulted in a negative value. This ranged from -0.4 to -6.2m, the latter of which occurred at a site used for field measurements and represents a hilltop with low vegetation growth but devoid of tree cover. A number of sources of error are possible: algorithm misidentification of signal begin and signal end, error in footprint location, discrepancy between the DTM subset used to calculate Terrain Index and the ellipsoidal footprint dimensions or the effect of diminishing laser energy distribution towards the edge of the footprint resulting in less effect of extremes of slope in this region of footprints.

It should be noted that field measurements may not correspond to the height of the tallest tree as many species extend their height disproportionately before broadening their trunk. Waveforms can be seen to reveal distinct height profiles of illuminated stands. Where footprints covered more than one sub-compartment with distinct mean vegetation heights, this vertical complexity could be observed as a stepped canopy return within the waveform (Figure 5.12).

No differentiation has been made between mixed and pure stands as only six footprints covered a single, pure sub-compartment. These footprints contained

species of similar heights and therefore conclusions for homogeneous vegetation cover could not be determined for this site.

Work undertaken at Monks Wood National Nature Reserve, Cambridgeshire, UK in 2000 (Patenaude *et al.*, 2004), using airborne LiDAR data resulted in tree canopy height mean offset of -2.12m and error range between -5.23 and +0.11m. In the study presented here, ICESat/GLAS estimates resulted in a low mean offset (0.41m and -0.55m), but they produced a greater error range (-6.73m to +4.88m and -7.37 to +7.35m for maximum height estimates using R_{WT} and GP_{MaxAmp} methods respectively). Errors may be potentially greatest for low stature objects situated within an indentation below the mean footprint ground surface.

The RMSE of ~2.8-3.8m found for vegetation height during this research, suggests that estimates from GLAS waveforms may be insensitive to short term vegetation growth. However, Sun *et al.*, (2008) have demonstrated the mean of vegetation heights from near coincident passes, separated by 60m, to be 17.75m for 19th June 2005 and 18.22m for 24th June 2006. This implies that mean calculations over large areas may permit vegetation change to be detected. However a relatively low R^2 of 0.46 and RMSE of 4.87m produced for vegetation height estimates of the two dates raises some concern and suggests the need for further work to explore this.

5.4.2 Quantification and Assessment of Uncertainty

The combined contribution of terrain elevation difference, species heterogeneity, vegetation surface roughness, vegetation height and canopy cover explained approximately a third of the variance in error between field-measured

canopy height and waveform-derived estimates. The significance of each is expected to vary considerably according to local conditions.

Areas of little vegetation cover may be subject to greater error in estimations from large footprint LiDAR due to the relatively small intercepted area for low fractional cover causing a delay in triggering the threshold for determining the beginning of the signal. Additionally, for low stature vegetation, there is a greater probability of signals from the ground and vegetation being combined even with relatively low relief.

Canopy surface roughness and slope are both addressed by Lefsky *et al.*, (2007) in their estimation of mean canopy height. In the study described in this Chapter, use of the leading edge of the waveform helped explain error in height estimates less well than the elevation of the dominant canopy layer. However, as noted by the authors above, the leading edge parameter demonstrates an indirect relationship with the correction factor applied to the waveform to compensate for height variability.

Where vegetation height is relatively consistent and spatially dispersed, the gradient and characteristics of the start of the signal will reflect the terrain in addition to canopy roughness (Lefsky *et al.*, 2007). Where vegetation of uniform height follows the contours of the terrain, in fact it may be necessary to estimate tree height as the difference between the beginning of the signal and the *highest* ground surface as opposed to the centre of the ground peak (P. North, pers. comms.). However, often vegetation distribution is related to topography (vegetation obtaining shelter and stability within valleys for example or clearing occurring on more level ground) and so the representation within the waveform is complicated further.

The simulated waveforms from FLIGHT (North *et al.*, submitted) illustrate the considerable effect of varying slope whilst keeping all other input values constant. The influence on both methods of estimating vegetation height can be inferred. Gaussian decomposition of the waveform to identify the ground surface becomes more problematic with increasing slope and may lead to a Gaussian being mistakenly allocated to this classification. This is because variation in waveform amplitude is less pronounced and therefore the ground return becomes less defined. Similarly, the waveform is broadened by increased slope and, although the use of the waveform extent and a terrain index utilises an independent DTM to compensate for this, if vegetation follows the ground surface, the signal amplitude will be suppressed, causing greater potential error in identifying the beginning of the signal. Additionally, multiple scattering will cause the assigned ground elevation using this method to be positioned below that of the actual ground surface.

Nevertheless, uncertainty produced within waveform-derived estimates of maximum canopy height remains relatively small and has not prevented valid estimations.

5.5 Conclusion

This Chapter is closely based on the contents of Rosette *et al.*, (2008c) and Rosette *et al.*, (submitted). The study has explored the use of the ICESat/GLAS satellite LiDAR for tree height retrieval over a semi-ancient, managed, mixed temperate forest of varied relief. Use of a Terrain Index to adjust the Waveform Extent provided the least dispersed estimates of canopy height when compared with field height measurements at footprint locations (R_{WT}: $R^2 = 0.92$, RMSE = 2.81m). In

the absence of a DTM, maximum canopy height estimates using an automated approach to ground identification based on iterative fitting of Gaussian peaks to the waveform (GP_{MaxAmp}) explained 84% of variance (RMSE = 3.83m) when compared with field measurements. The results suggest that maximum canopy height estimates from ICESat/GLAS can provide a reliable indicator of actual canopy height for a mixed temperate forest.

Contributions to the error in waveform estimates of vegetation height estimates using large footprint Satellite LiDAR are formed by complex interactions between many factors with the physical and optical properties of the intercepted features. These have been shown to include vegetation stature, upper canopy surface roughness, canopy cover, slope and species heterogeneity. Greater understanding is needed of the effects of topography and canopy properties on waveform composition. Radiative transfer modelling by Dr. P. North using FLIGHT has been shown to simulate this. For direct interpretation of waveform structure for footprints containing diverse features, local knowledge may be helpful due to the complex representation of intercepted surfaces combined within the waveform. Despite this complexity, estimates of vegetation height closely relate with those of validation field measurements, thereby supporting the opportunity for assimilation within forest growth models, vegetation monitoring and quantification of large areas to be achieved using Satellite LiDAR.

This Chapter aimed to determine whether the signal representing vegetation could be identified within waveforms returned from the broad dimensions of GLAS footprints. The study has presented two straightforward, repeatable methods for reliably extracting maximum vegetation height estimates for a mixed temperate forest.

This in itself may be of interest for complementing field assessment of stands, for spatial representation of vegetation distribution for vegetation monitoring or as observed inputs to Dynamic Vegetation Model components of General Circulation Models. The following Chapter further applies these methods to develop potential approaches for estimating vegetation top height and stemwood volume.

Chapter 6. Forest Parameter Estimation

Carbon accounting requires prior knowledge of dispersal and, as an effective carbon sink, understanding distribution of vegetation volume and quantifying and monitoring changes. For the purposes of forest management, quantifying timber volume is of importance for commercial viability and assessing stand processes such as regeneration (Forest_Research, 2006; UNFCCC, 2007).

Additionally, top height is an important parameter used to assess stand-level properties. The ability to successfully estimate this will provide meaningful information for forestry practitioners, managers and the operational research community. This parameter is also incorporated within yield models (Edwards and Christie, 1981; Forestry_Commission, 2006) and provides an input for process-based models of stand growth and risk factors.

Methods of estimating top height and stemwood volume have therefore been developed and explored in this Chapter. The objectives of this study are to evaluate the use of ICESat/GLAS data for sampling-based forest inventory and to develop uncomplicated, replicable methods for estimating vegetation top height and stemwood volume that can be applied to regional and national scales. In particular, an examination is undertaken of how existing methods for top height and stemwood

volume estimation based on yield models compare with estimates derived from satellite LiDAR remote sensing.

The study described here uses the methods presented in the previous Chapter to identify the region of the waveform attributable to vegetation. Several possible approaches of estimating top height and stemwood volume from GLAS waveforms are subsequently explored and compared with predictions from Forestry Commission yield models. Maximum canopy height and functions of this, height of cumulative energy percentiles, dominant canopy height and area under the waveform were considered as potential estimators. Top height was calculated for all GLAS footprints coincident with stands within the sub-compartment database. Stand-level stemwood volume was estimated for a single species and for the mixed composition stands found at the Forest of Dean.

6.1 Biophysical Parameters

Forestry Commission yield models are empirically-derived and estimate growth over time accounting for habitat conditions and consequently differing growth characteristics between and within species groups. Initial spacing of individuals, species, yield class (defined as an annual increment in m^3/ha/year) and management (e.g. thinning regime) are used for each model to estimate vegetation parameters which comprise top height, individual tree volume, volume per hectare and mean diameter at breast height by age.

6.1.1 Top Height

Top Height is an important vegetation parameter used within forestry and estimated within yield models. Where diameter at breast height is measured at 1.3m above ground level, Top Height is defined as

> the average height of a number of 'top height trees' in a stand where a 'top height tree' is the tree of largest breast height diameter in a 0.01 ha sample plot. (Edwards and Christie, 1981)

Using the Forestry Commission sub-compartment database and corresponding yield models (Edwards and Christie, 1981; Forestry_Commission, 2006), Top Height predictions were calculated for sub-compartments in which ICESat footprints are located. The majority of ICESat/GLAS footprints contained multiple management components (Chapter 4) and in several instances, extended across adjacent sub-compartments.

Top Height predictions were calculated as an indicator of vegetation distribution throughout the satellite pass. The following approaches were explored:

I. using the largest management component of the sub-compartment in which the footprint centre was positioned

II. for the largest-sized management component within all sub-compartments covered by the footprint

III. using all components within all sub-compartments encompassed by each footprint.

As components are not regularly dispersed, this aimed to examine the probability of a component species being present within a footprint. Top height

calculated from yield models was assumed to be representative of the ground truth. Heights of waveform percentiles of cumulative energy were explored as possible estimators of this.

6.1.2 Vegetation Stemwood Volume

Forestry Commission yield models were used to calculate stemwood volume coincident with ICESat footprints. Stemwood volume is defined by Edwards and Christie, (1981), as living tree over-bark volume (m^3/ha) which, for coniferous species, includes main stem timber of 7cm diameter or greater. This was estimated for sub-compartments sampled by each ICESat/GLAS footprint making reference to the sub-compartment database and relevant yield models. Two measures of stemwood volume are used in this study:

6.1.2.1 Single Species Stemwood Volume

Only six GLAS footprints crossing the Forest of Dean sample sub-compartments which contain a single species and, furthermore, several footprints cross more than one sub-compartment. In the absence of sufficient pure stands at the study site to permit reliable analysis, waveform indices were compared with yield model stemwood volume calculated for the tallest species within each footprint as determined from the sub-compartment database. This approach was based on the principle that the returned signal from the tallest species could be identified within the waveform (i.e. R_{WT} and GP_{MaxAmp} estimates), and aims to indicate the potential for stemwood volume estimation within pure stands. Footprints were then

distinguished according to whether the tallest species was broadleaf or coniferous in order to determine whether this would result in an improved correlation using waveform parameters.

6.1.2.2 *Mixed Composition Stemwood Volume*

The second measure represents the mixed composition of stands and uses a weighted stemwood volume calculation, taking account of the percentage distribution of species within all components of each sub-compartment covered by footprints. Footprints were then discriminated according to whether broadleaf or coniferous species formed the greatest percentage cover and relationships with waveform indices were calculated.

6.2 Method

Top Height is a significant parameter in forestry used in the assessment of stand-level properties and as driving data for process-based models. Similarly, knowledge of tree volume is a useful means of assessing the economic potential of stands as well as quantifying the distribution of woodland which is necessary for national forest inventory. Therefore methods of estimating these parameters using remote sensing techniques can offer a useful opportunity to complement conventional field methods and inform model development.

Conventional field methods routinely gather stand-level data, central among which is the calculation of top height (Edwards and Christie, 1981), estimated as the mean height of trees with the largest diameter at breast height (and therefore a

function of maximum canopy height). Tree volume is a function of both height and DBH (Edwards and Christie, 1981; Matthews and Mackie, 2006), the latter determining basal area and so stability of the stem. For a given species, a positive relationship is expected with a greater diameter permitting the stem to be extended proportionally whilst maintaining stem stability. Since DBH is additionally a function of height, vegetation height and functions of this are anticipated to show a relationship with stand-level volume. This theoretical basis has been demonstrated through LiDAR-derived estimates of volume or related parameters (biomass or carbon) using height metrics (e.g. Lefsky *et al.*, 1999b; Hyyppä *et al.*, 2001; Nelson *et al.*, 2004) or the square of these (Lefsky *et al.*, 2002; Lefsky *et al.*, 2005).

Therefore, in order to estimate forest parameters, a number of indices using height metrics and waveform area, which may contain information relating to top height and stemwood volume, were defined and extracted from GLAS waveforms. Regression relationships were then explored for these using estimations of top height and stemwood volume calculated from yield models and information contained within the Forestry Commission sub-compartment database.

Yield models intrinsically contain a degree of error in their predictions as forest growth is influenced by many factors throughout their lifespan which may differ from initial model input conditions. Therefore, an assessment is made to indicate potential uncertainty within stemwood volume estimates derived from yield models. This is achieved through the sum of individual tree volume calculations using parameters obtained from airborne LiDAR crown delineation (discussed fully in the following Chapter).

Finally the Chapter concludes with an evaluation of results and a discussion of the implications of findings.

6.2.1 Waveform-Derived Indices

The research within this chapter aims to identify indices from waveforms which may produce consistent relationships with forest parameters that are useful for inventory and management purposes. The intention is to derive methods applicable for easily-obtainable forest-type classes which could therefore be applied to large areas.

The following waveform parameters were explored as potential indicators of top height and/or stemwood volume:

I. Maximum canopy height estimates (R_{WT} and GP_{MaxAmp})

II. Heights of cumulative energy percentiles

III. Dominant canopy height

IV. Waveform area

6.2.1.1 Maximum Canopy Height Estimates

Use of maximum canopy height estimations (R_{WT} and GP_{MaxAmp}) presented previously were examined. The square of these values were also considered for stemwood volume estimates as an improved relationship had been previously achieved by Lefsky *et al.*, (2005). NB the aforementioned authors used squared height in the absence of calculated canopy structure indices applied in airborne studies (i.e. energy quartiles) and so although the use of a squared height lacks a theoretical foundation as to why a relationship with volume might be expected, testing this parameter allows the results to be compared.

6.2.1.2 Heights of Cumulative Energy Percentiles

R_{WT} and GP_{MaxAmp} methods were used to locate the portion of the waveform representing the vegetation return. Within these regions of the waveform, percentiles of cumulative energy (adapted from Harding *et al.*, 2001) were calculated and the heights at which percentiles occurred were compared as potential estimators of top height and stemwood volume.

6.2.1.3 Dominant Canopy Height

Dominant canopy height was earlier defined as the height above the ground surface to the position of the maximum amplitude within the canopy return. This was considered to offer possible information relating to volume as it may be more representative of mean stand conditions than maximum canopy height which, by its nature, is formed by one or few trees.

This approach used Gaussian decomposition to identify ground and canopy return maxima and estimated the dominant canopy height as the difference between these.

6.2.1.4 Waveform Area

An alternative means of attributing waveform structure to volume of intercepted vegetation was investigated using the area under the canopy return signal. Waveform amplitude is, in part, attributable to the intercepted surface area which is expected to be positively related to volume.

Two ways of removing the ground surface contribution from the waveform were explored, based on the two methods of estimating maximum canopy height (R_{WT} and GP_{MaxAmp}).

The first approach calculated the area beneath the waveform between the alternate fit signal begin position to the upper limit of the Terrain Index elevation.

The second approach assumes the canopy return area to be the sum of areas under Gaussian peaks 2-6 if peak 1 had been identified as the ground peak or the total of areas under Gaussian peaks 3-6 if the ground position was designated as the centroid of peak 2.

6.2.2 Analysis of Yield Model Uncertainty

An assessment was made of potential uncertainty within yield model stemwood volume predictions using estimates of stand-level volume calculated using the following:

I. A data subset within the Forest Research Environmental Database (FRED) kindly made available by the Forestry Commission.

II. Canopy delineation from airborne LiDAR data (presented in the following Chapter), which provided estimates of individual tree crown areas, heights and locations.

III. Field measurements taken within ICESat/GLAS footprint areas (Section 4.3.2)

The uncertainty analysis involved the following steps:

I. Extract regression relationships from FRED in order to estimate DBH from crown width and tree height.

II. Identify GLAS footprints with known species distributions.

III. Calculate crown width for each tree from crown area (identified using airborne LiDAR delineation).

IV. Apply the regression equations (I) to estimate DBH for each individual tree.

V. Use trigonometry to calculate radius at the base of each tree.

VI. Determine individual tree volume as a right-circular cone using tree height and radius at each tree base.

VII. Estimate GLAS footprint volume as the sum of individual tree volumes within footprint boundaries. Extrapolate this value in order to estimate stand-level stemwood volume (m^3/ha).

Estimation of Diameter at Breast Height

The Forest Research Environmental Database (FRED) contains in excess of 15,000 entries from field data taken throughout Britain. Data are at an individual tree level and comprise key vegetation indices including species, co-ordinates, tree height, DBH and crown width.

Using a method also presented by Suárez *et al.*, (2008a), allometric relationships were developed to estimate DBH as a function of vegetation height and crown width using ground truth data within an area of 200 x 200km, centred on the Forest of Dean. This area is anticipated to have relatively comparable climatic conditions and therefore similar growth responses.

Six footprints were selected to assess yield model uncertainty, based on the ability to identify species, performance of the canopy delineation algorithm and for which field measurements had been taken. This is discussed in further detail in the following chapter.

For the species within these footprints, the regression models that were developed estimated DBH for Douglas Fir with $R^2 = 0.92$, RMSE = 0.05m; Norway Spruce $R^2 = 0.85$, RMSE = 0.06m and Oak $R^2 = 0.88$, RMSE = 0.05m (Figure 6.1, Figure 6.2 and Figure 6.3). This suggests that these provide reasonable relationships to estimate DBH from tree height and canopy width.

These estimates of DBH were validated against field measurements. An underestimation of DBH for Oak was found (ground truth ~0.6m). Therefore a correction was applied by removing the intercept of -0.10m.

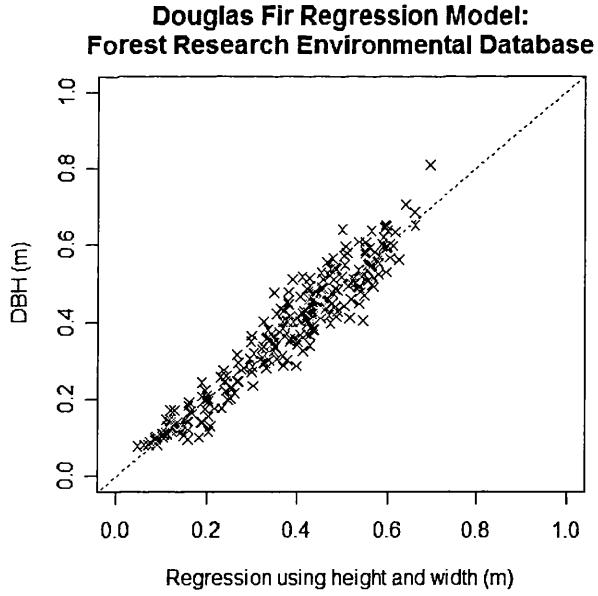


Figure 6.1 Regression model for Douglas Fir, estimating DBH from tree height and canopy width. The model was developed using individual tree-level field measurements contained within the Forest Research Environmental Database for an area of 200x200km, centred on the Forest of Dean.

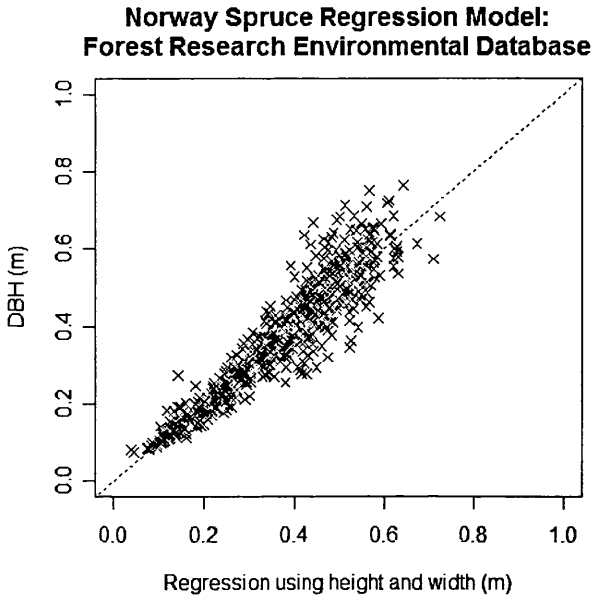


Figure 6.2 Regression model for Norway Spruce, estimating DBH from tree height and canopy width. The model was developed using individual tree-level field measurements contained within the Forest Research Environmental Database for an area of 200x200km, centred on the Forest of Dean.

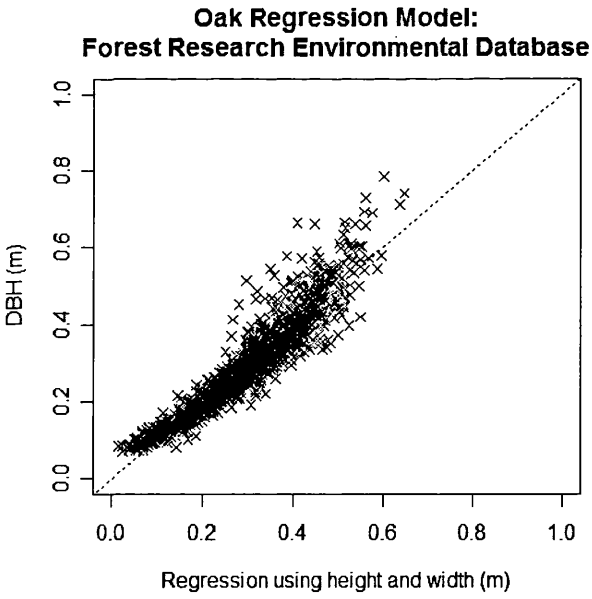


Figure 6.3 Regression model for Oak, estimating DBH from tree height and canopy width. The model was developed using individual tree-level field measurements contained within the Forest Research Environmental Database for an area of 200x200km, centred on the Forest of Dean.

Airborne LiDAR tree delineation, (Chapter 7), was used to represent ground truth of individual tree dimensions. This provided estimates of tree location, height and canopy area ($Area_{canopy}$) from which canopy width ($Width_{canopy}$) was estimated, modelling crowns as being circular (Equation 6.1).

$$Width_{canopy} = 2\sqrt{(\text{Area}_{\text{canopy}}/\pi)} \quad (6.1)$$

A correction was made to compensate for a 5% negative bias observed in airborne LiDAR estimates of individual tree height. Using regression equations derived from FRED for each species, DBH was estimated for trees identified within the footprint boundaries. Species within footprints could not be discriminated within the scope of this study and therefore priority in calculations was firstly given to footprints for which field measurements had been taken (allowing the delineation to be validated) and then those containing a single species, with large percentage cover by one species or crossing sub-compartments containing single species.

Calculation of Tree Base Radius

Using trigonometry, the radius at the base of each tree was calculated using estimated DBH with the following stages:

$$TanA = \frac{DBH/2}{h - 1.3} \quad (6.2)$$

$$r_{base} = hTanA \quad (6.3)$$

Where A is the angle formed at the top of a right-angled triangle in a vertical plane, with opposite side at 1.3m (height of DBH) and adjacent side passing through the centre of the tree stem; DBH is diameter at breast height estimated using regression equations, h is the adjusted estimated tree height from airborne LiDAR data and r_{base} is the radius at the base of the tree (Figure 6.4).

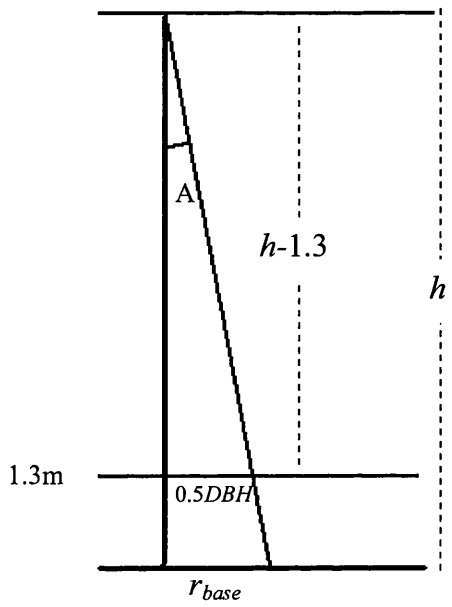


Figure 6.4 Calculation of the radius of the tree base from trigonometry using Diameter at Breast Height (DBH 1.3m above the ground surface) and tree height.

Stemwood Volume Estimation

Stemwood volume for each individual tree ($SWvol_{ind}$) was estimated as a right-circular solid cone and therefore assumes a regular taper along the stem:

$$SWvol_{ind} = \frac{1}{3} \pi r_{base}^2 h \qquad (6.4)$$

This procedure was repeated for all delineated trees within the footprint boundaries, applying a 5m buffer within ArcGIS in order to include trees whose outer canopies crossed the footprint border. The sum of these estimates indicates the footprint-level stemwood volume and, assuming this sample area is representative of the stand mean, this sum (0.38ha) was multiplied to estimate the stemwood volume per hectare ($SWvol_{stand}$ m^3/ha):

$$SWvol_{stand} = \sum SWvol_{ind} * 2.598 \qquad (6.5)$$

These stand-level estimates were subsequently compared with stemwood volume calculated from yield models.

6.2.2.1 Yield Model Uncertainty Assessment Results

For all six footprints, stemwood volume calculated using airborne LiDAR canopy delineation produced a lower estimation with respect to yield model estimates. This ranged from 14-22% for footprint areas positioned within the centres of stands. Those footprints crossing footpaths produced discrepancies from yield model estimates of 44-55%.

The validity of this assessment is dependent on the accuracy of the airborne LiDAR canopy delineation, however the results suggest that yield model predictions used to assess waveform-derived stemwood volume estimates may prove to be an over-estimation.

6.3 Results: Biophysical Parameter Estimates

6.3.1 Top Height

When comparing Forestry Commission yield model estimates of top height with field measurements, highest correlation of $R^2 = 0.96$, RMSE of 1.90m, was found using the greatest of predicted Top Heights (*TH*) for all components contained within all sub-compartments encompassed by each footprint (equation 6.6). This suggests that the species that comprises a component is generally well dispersed within subcompartments.

$Field_{Ht} = 1.09*TH$ (6.6)

coefficient significance $p<0.001$, intercept not significant.

This relationship is illustrated in Figure 6.5. This approach was therefore used as a reasonable indicator of top height ground truth throughout the pass.

Top Height Estimation using Waveform Canopy Metrics

The above calculation of Top Height was therefore used to assess the ability to extract estimates of the same using GLAS waveform indices.

Greatest correlation was found for percentiles towards the uppermost part of the canopy (Table 6.1) although little difference in strength of the relationship and RMSE is seen among the highest percentiles.

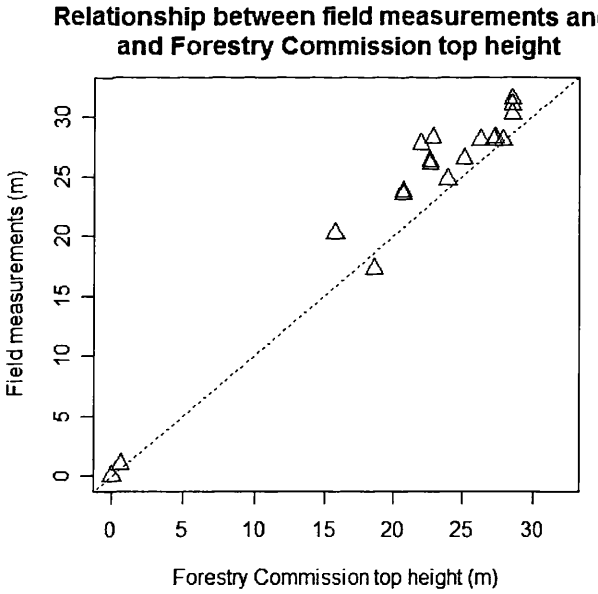


Figure 6.5 Relationship between Forestry Commission Top Height predictions from yield models and field measurements. $Field_{Ht} = 1.09*TH$; $R^2 = 0.96$, RMSE = 1.90m, coefficient significance $p<0.001$, intercept not statistically significant.

A degree of error is expected among yield model estimates of top height particularly due to the heterogeneous nature of the forest but nevertheless, waveform-derived indices succeeded in estimating this parameter. This supports the possibility for direct observation-based estimates to be applied in support of forest management.

Table 6.1 Waveform estimates of Top Height. NS indicates not significant, coefficients are significant p<0.001

Method	Percentile	R^2	Intercept	Coefficient	RMSE (m)
	Max. height	0.76	NS	0.89	3.9
R_{WT}	99th	0.75	NS	0.95	4.01
	98th	0.75	NS	0.98	4.06
	Max. height	0.73	NS	0.95	4.4
GP_{MaxAmp}	99th	0.73	NS	1.01	4.5
	98th	0.72	NS	1.04	4.5

The R_{WT} method produced the highest correlation for estimating Top Height. Due to the association between Top Height and maximum height, both R_{WT} and GP_{MaxAmp} methods found that maximum canopy height provided the best estimates of Top Height as opposed to lower height percentiles of cumulative energy (R_{WT} R^2 = 0.76, RMSE 3.9m; GP_{MaxAmp} R^2 = 0.73, RMSE 4.4m). These correlations are seen in Figure 6.6. This is in contrast with the typical findings using airborne LiDAR for which percentiles in the high 90s tend to produce the most robust estimates (e.g. Doce *et al.*, 2008).

Discrepancies can be observed among the lowest Top Height estimates. As discussed in the Chapter 5, for some footprints, heights detected within waveforms are formed by buildings or unmanaged vegetation which is not accounted for within the sub-compartment database.

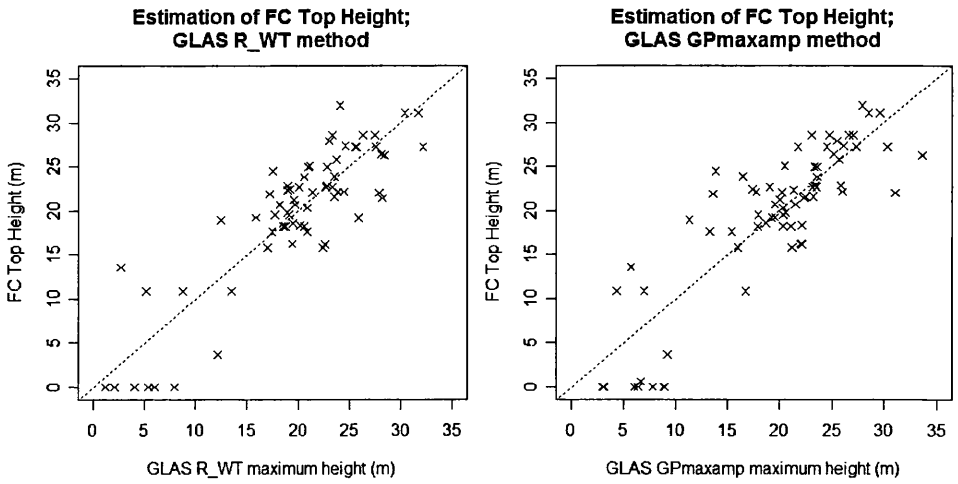


Figure 6.6 Estimation of Forestry Commission Top Height using maximum canopy height from GLAS waveforms. R_{WT} method: $R^2 = 0.76$, RMSE 3.9m; GP_{MaxAmp} method: $R^2 = 0.73$, RMSE 4.4m; coefficient significance $p<0.001$, intercept not statistically significant.

6.3.2 Stemwood Volume

6.3.2.1 Single Species Stemwood Volume

R_{WT} Method

A summary of the results of the regression analysis using waveform indices and yield model stemwood volume estimates are shown in Table 6.2. Groups include common footprints classified as un-vegetated. NB when calculating waveform indices for estimating stemwood volume, the number of footprints used differs from the GP_{MaxAmp} method due to discrepancies between Waveform Extent and estimated maximum height (i.e. two anomalies in which estimated vegetation height was negative were excluded from the analysis).

Table 6.2 Tallest species stemwood volume estimates using the R_{WT} method. Coefficients and intercepts are statistically significant ($p < 0.001$) except where stated. NS indicates not significant

Waveform parameter	Tallest species	R^2	Intercept	Coefficient	RMSE (m^3/ha)	Number
Maximum canopy height (R_{WT})	Conifers	0.63	NS	13.8	90.3	53
	Broadleaf	0.62	NS	8.27	70.2	17
	Combined	0.56	NS	12.8	100.2	67
$(R_{WT})^2$	Conifers	0.57	94.7 p>0.001	0.37	96.4	53
	Broadleaf	0.54	NS	0.31	82.5	17
	Combined	0.51	83.4 p>0.001	0.36	105.2	67
Height of 95th percentile cumulative energy	Conifers	0.59	NS	16.2	93.9	53
	Broadleaf	0.64	NS	9.59	68.7	17
	Combined	0.52	NS	15.0	103.7	67

This exercise has simulated the situation in pure stands by using yield model predictions of stemwood volume for the tallest intercepted tree species and has demonstrated the ability to estimate this volume from waveforms by extracting information relating to the maximum canopy height. This illustrates the potential for estimates of volume for homogeneous forests to be made using ICESat/GLAS waveforms.

Estimated maximum canopy height (R_{WT}) produced the best relationship with stemwood volume for the tallest species within all footprints with R^2 of 0.56 and RMSE of 100.2m^3/ha. Top height (as a function of maximum height) is an input for model estimation of volume and therefore explains the correlation found for this parameter. Differentiating between broadleaf and coniferous species improved the

correlations to R^2 of 0.64, RMSE of 68.7 m^3/ha and R^2 of 0.63, RMSE of 90.3 m^3/ha respectively. These relationships are illustrated in Figure 6.7.

Neither heights of cumulative energy percentiles nor the square of maximum canopy height improved upon these estimates of stemwood volume. Using other waveform parameters (waveform area, dominant canopy height or multiple regressions combining variables) did not produce statistically significant results or showed weak correlations (Table 6.6).

In fact, a non-linear relationship such as a quadratic fit may produce an improved correlation. However the use of linear regression analysis using GLAS is supported within the literature (Nelson, 2008; Nelson *et al.*, 2008a). Furthermore, the two clusters of high and low volumes make it difficult to evaluate non-linearity, particularly considering the uncertainty among yield model estimates of the lower volumes.

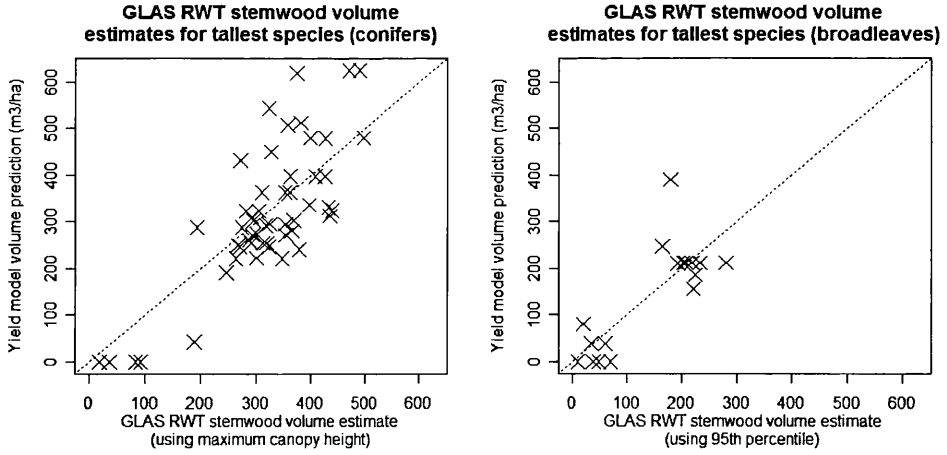


Figure 6.7 Relationship between stemwood volume predictions from yield models and volume estimates using (left) Maximum Canopy Height R_{WT}, for coniferous species, $R^2 = 0.63$, RMSE 90.3 m^3/ha and (right) 95th percentile for broadleaf species, $R^2 = 0.64$, RMSE 68.7 m^3/ha.

GP_{MaxAmp} Method

Regression analysis was also carried out using GP_{MaxAmp} method waveform indices against yield model stemwood volume estimates for the tallest species within each footprint. Results are outlined in Table 6.3.

Estimated maximum canopy height produced the best relationship with R^2 of 0.59 (RMSE 100.8 m^3/ha). Differentiating between coniferous and broadleaf species () did not significantly improve the estimation of stemwood volume for conifers where using the height of the 99th percentile of cumulative energy produced R^2 of 0.59 (RMSE 98.3 m^3/ha). Considering broadleaf species in isolation however, resulted in a substantial improvement in correlation of $R^2 = 0.75$ and RMSE of 59.1 m^3/ha using height of the 98th percentile of cumulative energy.

This provides further proof of concept for the capability of estimating stemwood volume for homogeneous forests using indices from GLAS waveforms.

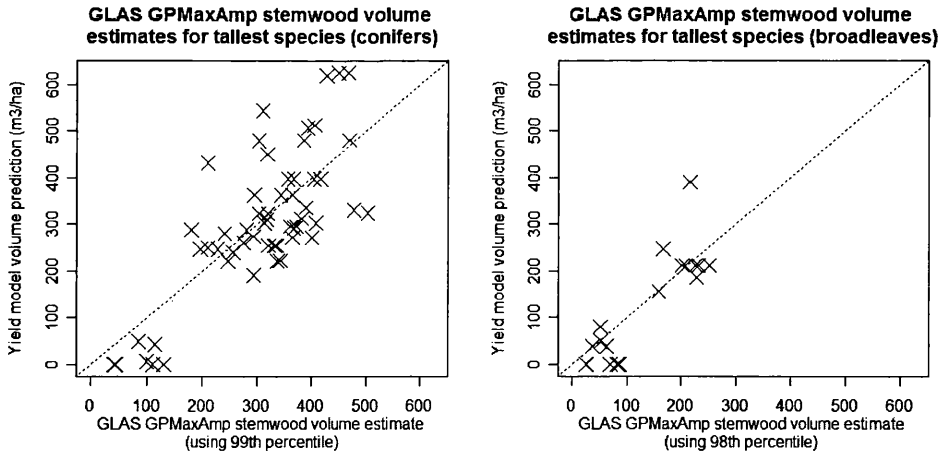


Figure 6.8 (left) Stemwood volume estimates from height of 99th percentile of cumulative energy for footprints where the tallest species is coniferous; $R^2 = 0.59$, RMSE 98.3 m^3/ha (right) Stemwood volume estimates using height of 98th percentile of cumulative energy for footprints where tallest trees are broadleaves; $R^2 = 0.75$ and RMSE 59.1 m^3/ha

Table 6.3 Tallest species stemwood volume estimates using the GP_{MaxAmp} method. Coefficients and intercepts are statistically significant ($p < 0.001$) except where stated. Groups include common footprints classified as un-vegetated.

Waveform parameter	Tallest species	R^2	Intercept	Coefficient	RMSE (m^3/ha)	Number
Maximum canopy height (GP_{MaxAmp})	Conifers	0.59	NS	15.55	99.0	55
	Broadleaf	0.75	NS	9.39	61.0	19
	Combined	0.59	NS	13.76	100.8	69
$(GP_{MaxAmp})^2$	Conifers	0.51	108.19 P<0.001	0.39	109.5	55
	Broadleaf	0.71	NS	0.42	63.2	19
	Combined	0.53	84.64 p>0.001	0.40	107.4	69
99th percentile	Conifers	0.59	NS	15.48	98.3	55
98th percentile	Broadleaf	0.75	NS	10.33	59.1	19
99th percentile	Combined	0.58	NS	14.63	100.5	69

Other waveform indices or multiple regression analysis incorporating waveform area or dominant canopy height did not improve results further.

6.3.2.2 Mixed Composition Stemwood Volume

R_{WT} Method

Using the same waveform parameters, regression analysis was repeated for the mixed stand weighted stemwood volume estimates. Results are summarised in Table 6.4 below. As previously, groups include common footprints classified as un-

vegetated and those with no dominant vegetation type. Two waveforms with anomalous 'negative' height estimates were excluded.

Table 6.4 Mixed stand stemwood volume estimates using the R_{WT} method. Coefficients and intercepts are statistically significant ($p < 0.001$) except where stated. NS indicates not significant.

Waveform parameter	Predominant species	R^2	Intercept	Coefficient	RMSE (m^3/ha)	Number
Maximum canopy height (R_{WT})	Conifer	0.49	NS	9.83	99.8	56
	Broadleaf	0.46	NS	5.69	76.4	23
	Combined	0.36	NS	8.90	107.5	67
$(R_{WT})^2$	Conifer	0.46	NS	0.36	103	56
	Broadleaf	0.43	NS	0.21	78.5	23
	Combined	0.33	51.9 p>0.05	0.25	110.2	67
Height of 95th percentile cumulative energy	Conifer	0.55	NS	11.7	93.8	56
	Broadleaf	0.46	NS	6.82	76.3	23
	Combined	0.42	NS	10.6	103.1	67

Correlation was found to be lower for estimates of mixed stand stemwood volume than for the single species estimates above. However, inaccuracies within yield model predictions are compounded by uncertain species distribution and footprint location. The height of the 95th percentile of cumulative energy produced the best relationship with R^2 of 0.42, RMSE 103.1 m^3/ha for all footprints sampled.

Footprints were then distinguished according to whether the predominant vegetation cover consisted of broadleaf or coniferous species. The 95th percentile achieved respectively R^2 of 0.46, RMSE 76.3 m^3/ha for broadleaf species and R^2 of 0.55, RMSE 93.8 m^3/ha for conifers. These relationships are shown in Figure 6.9.

The results of other waveform parameter analyses are outlined in Table 6.4 and Table 6.6.

This supports the principle that waveform structure can provide additional information regarding biophysical parameters representing the composition of the species contained within intercepted stands. This is of importance for the assessment of natural forests where monocultures are less likely to predominate.

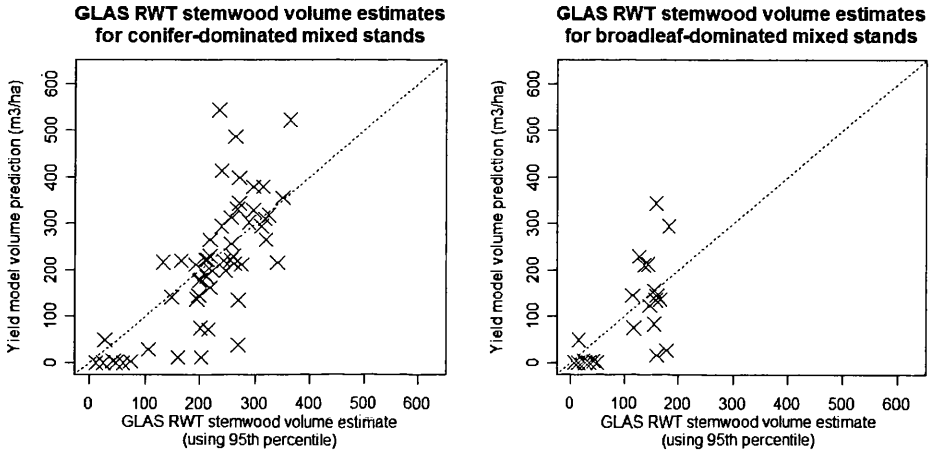


Figure 6.9 Relationship between weighted stemwood volume predictions and volume estimates using height of the 95th percentile of cumulative energy for footprints with greatest percentage cover by: (a) coniferous species, R^2 of 0.55, RMSE 93.8 m^3/ha and (b) broadleaf species, R^2 = 0.46, RMSE 76.3 m^3/ha.

GP_{MaxAmp} Method

Weighted stemwood volume estimates accounting for the mixed species composition of stands were then used to regress waveform-derived indices using the GP_{MaxAmp} method. Table 6.5 outlines the results.

Positions within the canopy return of the waveform for the GP_{MaxAmp} method also offer closer correlations with mixed stand composition calculations of volume than those provided by maximum canopy height estimates. This furthermore supports the opportunity of widely applying satellite LiDAR estimates of stemwood volume.

Table 6.5 Mixed stand stemwood volume estimates using the GP_{MaxAmp} method. Groups include common footprints classified as un-vegetated and those with no dominant vegetation type. Coefficients and intercepts are statistically significant ($p < 0.001$) except where stated. NS indicates not significant.

Waveform parameter	Tallest species	R^2	Intercept	Coefficient	RMSE (m^3/ha)	Number
Maximum canopy height (GP_{MaxAmp})	Conifers	0.63	-84.46 p>0.05	14.34	86.6	58
	Broadleaf	0.46	NS	5.97	76.6	24
	Combined	0.46	NS	9.65	102.1	69
$(GP_{MaxAmp})^2$	Conifers	0.62	NS	0.43	87.9	58
	Broadleaf	0.37	NS	0.22	84.1	24
	Combined	0.41	42.93 P>0.05	0.31	105.3	69
95^{th} percentile	Conifers	0.66	-63.22 P>0.01	15.76	82.5	58
98^{th} percentile	Broadleaf	0.47	NS	6.66	75.6	24
95^{th} percentile	Combined	0.5	NS	11.41	97.8	69

Greatest correlation was seen for all mixed stand weighted stemwood volume estimates using the height of the 95^{th} percentile of cumulative energy. This produced R^2 of 0.50 and RMSE of 97.8 m^3/ha. Height of the 95^{th} percentile of cumulative energy also produced the best estimates when only considering coniferous species (R^2 of 0.66 and RMSE of 82.5 m^3/ha). However, considering broadleaf species separately produced a poorer correlation with R^2 of 0.47 and RMSE of 75.6 m^3/ha for height of the 98^{th} percentile of cumulative energy. These relationships are seen in Figure 6.10.

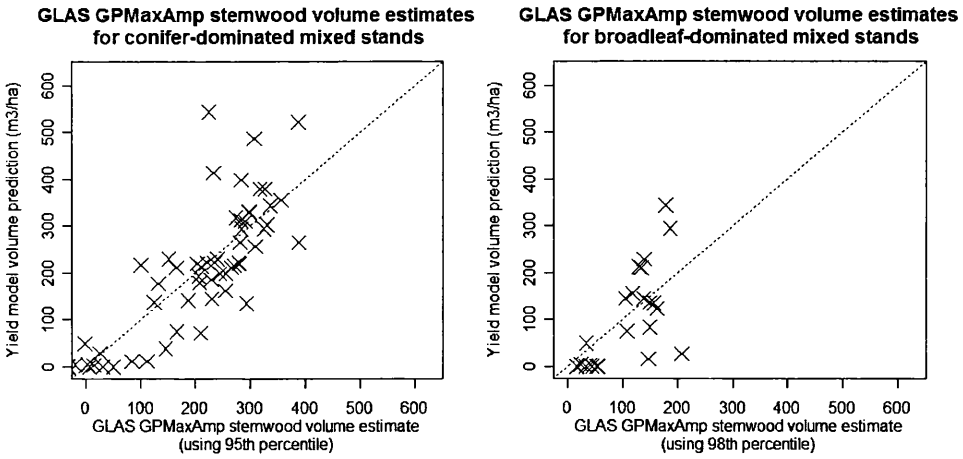


Figure 6.10 Relationship between mixed stand stemwood volume estimates and volume estimated by (left) height of 95th percentile of cumulative energy for footprints dominated by conifers; R^2 = 0.66, RMSE 82.5 m^3/ha. (right) height of 98th percentile of cumulative energy for footprints dominated by broadleaf trees; R^2 = 0.47 RMSE 75.6 m^3/ha.

Further regression analysis relationships

Relationships using multiple regression with height indices and area under the canopy return of the waveform or dominant canopy height did not significantly improve correlations and coefficients were not statistically significant. These parameters alone were not statistically significant when estimating either stemwood volume approach using both R_{WT} or GP_{MaxAmp} methods of identifying the waveform canopy return. For reference, the results of these correlations are summarised in Table 6.6. The lack of correlation using these indices may represent the heterogeneous nature of the Forest of Dean study site and therefore an inconsistent contribution to waveform structure.

Table 6.6 Additional correlation ranges. Parameters shown were not found to be statistically significant or produced weak correlations. Ranges stated are for both stemwood volume methods.

Waveform parameter	R^2 range	RMSE range m^3/ha
Dominant canopy height	0.05 – 0.34	100.2 – 143.4
Waveform canopy area (Gaussian peaks)	0.08 – 0.19	98.9 – 149.7
Waveform canopy area (to upper terrain index limit)	0.05 – 0.24	93.4 – 140.1

6.4 Discussion: Biophysical Parameter Estimation

Approaches of estimating vegetation top height and stemwood volume using large footprint LiDAR may provide a future means of assessing actual against predicted growth using yield models, without the necessity of more costly and time-consuming field measurement validation. This may permit timber volume for stands to be quantified more efficiently to assist with carbon accounting and could additionally have implications for timber income predictions. Currently, to achieve accurate quantification, field measurements are taken prior to felling operations as actual volume may differ significantly from yield model predictions.

6.4.1 Top Height

Cumulative energy percentile heights were used to estimate top height estimated from yield models. Using both R_{WT} and GP_{MaxAmp} methods of identifying the vegetation return from waveforms, the estimated maximum canopy height provided the best estimate of top height (R^2 = 0.76, RMSE 3.9m and R^2 = 0.73,

RMSE 4.4m respectively). The greater variability among top height compared with estimates of field measurements may be at least in part due to deviations between yield model predictions and ground truth. Error is known within yield models and is anticipated to be particularly evident among older stands such as those found within the Forest of Dean.

A field measurement validation of top height could quantify this error. For each GLAS footprint-sized area, this would require the mean to be calculated of 'top height trees' within 38 10x10m plots within each stand. For mixed composition stands such as at the Forest of Dean, this method would need to be modified (Matthews and Mackie, 2006).

Attempts to combine ICESat/GLAS estimates with Forestry Commission top height data did not result in a significant improvement of predictions of field measurements for this study of the Forest of Dean. Forestry Commission yield models are not dynamic and therefore have some limitations as they depict a mean trend devoid of any natural disturbance. For old stands this assumption is unrealistic which may be the case for the Forest of Dean sub-compartments sampled by GLAS, as mean stand age of principal components is 51 years.

The measurements also differ from the method employed by the Forestry Commission in which Top Height is an average of 'top height trees' for a stand. The Forestry Commission method could not be replicated precisely due to the heterogeneity of the sites sampled, the necessity of limiting measurements to the footprint area and the requirement for future work of representing canopy height variation within footprints.

Pure stands may produce smaller errors as laser penetration would be expected to differ between species. As an indicator of this, a comparison between

GLAS maximum canopy height estimates using R_{WT} with Forestry Commission Top Height predictions, showed mean offset for pure stands was reduced from 1.89m to 1.52m. NB Top Height is a function of maximum canopy height and therefore is not directly comparable.

6.4.2 Stemwood Volume

Using the heights of the percentiles of cumulative energy plus waveform parameters such as dominant canopy height and area under the waveform allowed different canopy return indices to be considered with respect to their ability to estimate stemwood volume as predicted by independent calculations derived from yield models. It is anticipated that, whilst higher waveform positions are largely the result of returns from the tallest species within footprints, returns from lower canopy elevations might better represent the mixed species composition within stands. The UK National Inventory of Woodland and Trees (NIWT) provides class distinctions for broadleaf species and conifers, therefore indicating that methodologies to estimate volume using such broad classes could be applied regionally or nationally.

Due to expected deviation in relationships between conifers and broadleaf species, these groups were first analysed together as a sample forest population before being considered in isolation. Nelson *et al.*, (2004) found estimates of volume and biomass using their Portable Airborne Laser System (PALS) to be significantly more accurate for conifers than for their hardwood grouping. This study found conifers to be better estimated than broadleaf species for mixed composition volume calculations but conversely was able to estimate volume of the tallest broadleaf species with greater correlations using the GP_{MaxAmp} method.

RMSE is lower for broadleaf-dominated footprints. However, mean values of stemwood volume for vegetated stands are 327.7 m^3/ha and 174.7 m^3/ha for coniferous and broadleaf species respectively and therefore a lower RMSE would be anticipated for the latter as a result of this alone. The results of this study suggest that differentiation may be necessary between conifers and broadleaves in order to use generalised relationships for broader scale stemwood volume estimates.

Maximum canopy height is, by its nature, determined by little foliage of one or few trees of the tallest trees within the stand. Higher elevations within the canopy return of the waveform therefore demonstrated a potential means of estimating stemwood volume for pure stands.

Particularly for mixed stands, positions within the waveform and characteristics of the canopy return are anticipated to better represent the combined contribution of the canopy elements for the diverse species present within footprints. GP_{MaxAmp} mixed stand estimates show marginally greater correlations at higher percentiles of cumulative energy for stands predominantly containing broadleaf species than for conifers, possibly due to canopy structure and leaf area affecting laser penetration.

Area under the waveform canopy return was not a statistically significant estimator of stemwood volume and multiple regressions using this failed to significantly improve estimates of stemwood volume. A contributory factor could be the considerable variation in reflectivity that may be expected between species. Therefore, for such a species-diverse forest, the principal reason for differing waveform amplitude may be reflectivity as opposed to intercepted surface area (anticipated to be related to volume). The validity of the methods using waveform canopy area are dependent on the degree to which reflectance between species differs

at the measured wavelength (1064 nm) in addition to the effects of canopy profile on multiple scattering events and their representation within waveforms.

However the probable explanation for waveform area lacking relationships with volume is due to the fact that returned signal strength is mainly determined by atmospheric transmission and therefore area under the waveform can vary greatly as a result of cloud cover which can fluctuate at local scales. Waveform area therefore needs to be normalised to permit this to be assessed.

The results suggest that, for such highly mixed stands, if a relationship does exist with dominant canopy height, the broad distinction between broadleaf and coniferous coverage is insufficient or that values are inconsistent due to variation in species structure, reflectivity or atmospheric transmittance.

Dispersion among smaller volume values is observed and may be attributable to unmanaged re-growth or the effect of the 5 ns emitted pulse width producing an artificial minimum elevation difference even for flat surfaces. It is recognised that differences in relationships between maximum canopy height or percentiles of cumulative energy and stemwood volume within both coniferous species and broadleaf species groups are likely to contribute to a proportion of the variation observed. Investigation of a more homogenous site may reveal improved correlation and a possible significant contribution of other waveform parameters.

Forestry sub-compartments that were listed by the Forestry Commission but classified as unpopulated (i.e. zero anticipated stemwood volume) were included in comparisons and the incorporation of these lower values will have improved correlations. Additionally, the study refers to relationships using waveform data acquired whilst vegetation was predominantly still in leaf. Correlation may be anticipated to vary with seasonal differences in LAI as noted by Sun *et al.*, (2008).

6.4.2.1 Method Inter-Comparison

The R_{WT} method requires a two-stage process, deriving maximum canopy height from a multiple regression using the Waveform Extent (distance between Signal Begin and Signal End) plus a terrain index (using a DTM centred on the footprint co-ordinates) and calibrating against field measurements of within-footprint tree height. These estimates are dependent on the accuracy and availability of field measurements and the DTM in addition to how representative the selected footprints are of the complete pass and the stand mean.

Overall, results using the GP_{MaxAmp} method produce better correlations than the R_{WT} method. However, this may be partly a result of two footprints being excluded in the R_{WT} analysis due to height errors produced by this method. Stemwood volume estimates using the GP_{MaxAmp} method for the tallest species within footprints are considerably better for broadleaf species than for conifers (a possible effect of upper canopy shape), whilst for mixed stand estimates, greater correlation is seen for stands with greatest cover formed by conifers than by broadleaf species.

However, the Gaussian decomposition method offers improvements for stemwood volume estimates for the tallest broadleaf trees (from R^2 of 0.64, RMSE 68.7 m^3/ha using R_{WT} to R^2 of 0.75, RMSE 59.1 m^3/ha using GP_{MaxAmp}) and for mixed stands dominated by conifers (from R^2 of 0.55, RMSE 93.8 m^3/ha (R_{WT}) to R^2 of 0.66, RMSE 82.5 m^3/ha (GP_{MaxAmp})). Methods which do not necessitate an additional source of information may simplify the process of waveform interpretation, potentially allowing broader application where supplementary data are not available.

The stemwood volume results presented in this Chapter are comparable with those obtained by other authors in similar studies. Lefsky *et al.*, 2005 applied their method of using Waveform Extent and a Terrain Index to estimate above ground biomass in Santarem, Brazil. Using the square of maximum estimated height, they found R^2 of 0.73, RMSE 58.3 Mg/ha for N=19.

Sun *et al.*, 2008 identified the vegetation return as the difference between the beginning of the waveform signal and the centroid of the lowest Gaussian and applied this to estimate biomass in Northern China. For their sites combining evergreen conifers and deciduous species plus data from late Autumn and early Summer GLAS campaigns (N=84), they found correlations with R^2 of 0.68 and residual standard error of 29.35 Mt/ha. A higher correlation (R^2 of 0.78 and residual standard error of 30.58 Mt/ha) was found when considering data from late Autumn alone (leaf off conditions). This cannot be attributed solely to date of acquisition as this data subset (N=44) contains considerably less deciduous species. When deciduous trees were considered alone for data obtained in early Summer, the relationship with field estimated biomass produced R^2 of 0.59 and residual standard error of 24.56 Mt/ha for N=40.

6.4.2.2 Yield Model Accuracy

Limitations of estimating top height and stemwood volume using yield models to assess the potential of using waveform-derived parameters are recognised. Forestry Commission yield models are not dynamic and therefore do not take account of changes in growth or stand composition due to competition, damage affliction or mortality. Forest_Research, (2009), citing a Forestry Commission

internal report, state that the yield models developed for use in Britain (Edwards and Christie, 1981) "are based on an extensive permanent sample plot network maintained across the UK and are considered to be robust. However, a recent study has highlighted possible deficiencies in the predictions of volume development made by the models, particularly for the latter stages of a rotation typical in the UK at present (Matthews, 2003)." Estimates for coniferous stands for example, have been found by the Forestry Commission to overestimate actual volume. The canopy delineation yield model uncertainty analysis carried out within this study suggests that this may be the case for the Forest of Dean.

Stemwood volume estimates used in this study include some common stands which were contained within the sub-compartment database but not listed as planted. These zero volume values may have improved the relationships and may go some way to explaining the spread among lower waveform estimates: initial observations at footprints locations have revealed the presence of unmanaged trees, shrubs or buildings in some cases which are contributing to waveforms.

An assumption is also made in the calculation of stemwood volume for mixed stands, that components are regularly distributed within sub-compartments rather than individuals forming clusters or being dispersed along a linear feature such as a footpath.

Deviations from actual stand volumes are anticipated due to errors inherent in yield model predictions as individual stands may not perform in accordance with expectations. If stand performance and management treatments differ from the yield model assigned or if habitat anomalies are present, long-term forecast production for an individual stand may vary from actual production by 20% (Edwards and Christie,

1981). This is also supported by the 15-22% discrepancy found between canopy delineation and yield model stemwood volume estimates found during this study.

However, updates are made to the sub-compartment database annually and two-three stand performance assessments based on field measurements are made during the lifetime of each crop to revise projected growth if necessary. Therefore the yield models demonstrate a mean trend representing the best available estimates of current conditions. Recent work conducted by the Forestry Commission at a stand level using yield model estimates calibrated with field measurements, produced vegetation height accuracy of 98% whilst, in the course of this study, a comparison of greatest field height measurements within footprints with corresponding yield model estimates of Top Height for the tallest species revealed R^2 of 0.94. Since Top Height is used within yield models they may therefore be anticipated to provide a reasonable indication of the vegetation present.

Error Analysis

Discrepancies within yield models are expected to vary between species and with stand age. Within-footprint diameter distribution measurements may provide a more accurate assessment of actual volume present and yield model error. Uncertainty assessment was undertaken in order to validate yield model estimates of stemwood volume and quantify any discrepancies. An alternative estimation of stand-level stemwood volume was therefore calculated using airborne LiDAR canopy delineation as a surrogate for tree-level field data.

This was used to estimate diameter distributions within footprints, from which stand-level volume estimates were calculated in order to compare with yield model predictions. This method is not without its own error: Calculations were limited to sites in which field validation of delineation results was possible and, as

species identification was beyond the means of this study, stands were selected where any different species could be easily distinguished or where they comprised a small percentage of total area. Small differences in drawn sub-compartment boundaries will also affect the accuracy of this approach. A lower performance of the delineation algorithm was noted for shorter trees and therefore sites used to assess yield model estimates took account of this. These criteria limited the number of footprints used for the analysis. However this may partially account for lower volume estimates if shorter trees were not identified by the algorithm or were obscured by higher canopies. Additionally, field-measured footprints often contain a lower percentage cover due the necessity of access via rights of way. For these sites, calculations were found to significantly underestimate the stand mean.

Although the regression equations using FRED take no account of soil type, aspect, drainage, etc., the large number of field measurements selected from the database aim to indicate a mean trend for an appropriate area surrounding the study site.

Deviations may be expected from a regular decline in diameter along the tree stem, which would influence the accuracy of volume estimates. Therefore for more refined individual tree volume estimates, species-specific taper functions could be applied (Matthews and Mackie, 2006). The method is also highly sensitive to tree height using airborne LiDAR data. This was found to underestimate field-measured heights and therefore a calibration was applied. However a small inaccuracy in estimated height may have a great influence on the calculated volume.

Despite these sources of error, the exercise supports Forestry Commission suggestions of an overestimation in stemwood volume by yield models. This

assessment would need to be extended to determine whether there is a systematic trend and also the accuracy of volume estimates using canopy delineation.

Nevertheless, the Forestry Commission database and yield models are widely used in forest management and have provided the best available indication of vegetation volume distribution for the diverse stands throughout the Forest of Dean. They have therefore formed useful points of reference against which to explore methods of estimating stemwood volume from GLAS waveforms.

6.5 Conclusion

The contents of this Chapter are based on work presented within two papers: Rosette *et al.*, (2008a); Rosette *et al.*, (2008b). Despite known challenges within yield models and the Forest Enterprise sub-compartment database, GLAS waveform estimates of top height succeeded in explaining 76% and 73% of variance in model predictions for the R_{WT} and GP_{MaxAmp} methods respectively. The RMSE of 3.9/4.4m may not be dissimilar to possible error in field measurements compounded by the numerous plot-level measurements needed in order to calculate top height. Further field work would be required to confirm this.

LiDAR stemwood volume estimates with accuracy in excess of 60-70% are anticipated to be welcomed by forestry practitioners to contribute to national forest inventory – NIWT (Forestry_Commission, 2003; Forest_Research, 2006). The results of this study suggest that for homogeneous sites, this can be achieved. Best results produced $R^2 = 0.75$, RMSE = 59.1m^3ha^{-1} for broadleaf species and $R^2 = 0.63$, RMSE = 90.3m^3ha^{-1} for conifers.

However, to consistently reach this threshold for areas with more complex species composition and to reduce error to an acceptable level would require greater refinement. Highest correlations found were R^2 = 0.47, RMSE = 75.6m^3ha^{-1} for stands dominated by broadleaf species and R^2 = 0.66, RMSE = 82.5m^3ha^{-1} for greatest coverage by conifers.

Estimates of stemwood volume were found to be similar to recent studies of related parameters by other authors using the GLAS sensor. The work of Sun *et al.*, 2008 suggests that date of data acquisition in relation to leaf on/off conditions may play a role in quantifying vegetation distribution. This provides a sound foundation from which to further develop stand assessment techniques to complement forest management and provide inputs to process-based models (e.g. Gardiner *et al.*, 2004).

The results suggest the potential for satellite LiDAR estimates of top height and volume to be extended to regional and national scales and that drawing a similar broad distinction between broadleaf or coniferous species for volume estimates may assist the quantification of vegetation distribution for the requirements of forest inventory. The use of such broadly-defined classes is consistent with the definitions of interpretive forest types used within NIWT (Forestry_Commission, 2003; Broadmeadow and Matthews, 2004; Forest_Research, 2006). Furthermore, the fact that relationships were improved using such broad classes suggests the potential for these to be further refined by developing more specific regression models using individual species or stratification by species groups. Future approaches might also involve exploring the use of multiple waveform indices. The capability to spatially map the distribution of vegetation volume has been shown using sampling-based LiDAR profiling. This could assist in addressing needs of vegetation distribution

uncertainty, locating deforestation and assessing Land Use and Land Use Change in Forestry (IPCC, 2003; FAO *et al.*, 2008; UNEP, 2009).

Using satellite LiDAR waveforms, this Chapter has developed means of determining indices of biophysical parameters of interest to carbon accounting and vegetation monitoring, model requirements and for forest management and assessment. The following Chapter evaluates the GLAS dataset with respect to airborne LiDAR data which are able to represent a finer spatial resolution of intercepted surfaces and whose value is acknowledged for vegetation application purposes and analysis.

Chapter 7. Airborne and Satellite LiDAR Comparison

This Chapter provides a comparison of remote sensing technologies using small footprint, discrete return, airborne laser scanning and large footprint, full waveform, satellite LiDAR profiling.

Capabilities of the two systems are explored with regard to their estimation of stand-level parameters suitable for management and forest inventory purposes. Airborne LiDAR is already a well-established field of commercial and operational forestry applications which has developed beyond the purely research and development stage. Therefore, in particular, the performance of satellite LiDAR data is evaluated against airborne platforms to establish whether challenges posed by the large footprint dimensions, signal convolution and atmospheric transmittance, etc. result in degradation of the ability to derive useful biophysical parameters.

In addition, the facilities offered by airborne LiDAR for the derivation of individual tree-level properties and their application for forest management and process-based modelling are explored and evaluated.

7.1 Method

The above aims are met in this Chapter as follows:

Firstly the motivation for exploring satellite LiDAR profiling for forest parameter retrieval is justified based on previous work using airborne laser scanning. An overview of the two systems is provided and processing of the airborne LiDAR data is presented. Subsets were created of airborne LiDAR data for a circular area of 35m radius surrounding GLAS footprint co-ordinates. Estimates within these coincident areas could then be compared. Subsets were also taken of similar areas offset from published footprint centres in order to investigate possible misplacing of footprints and the effect of this.

Analysis was undertaken for GLAS footprints for which there was airborne LiDAR coverage (N=59). As previously, this excluded footprints which intercepted urban areas as solid-surface features will contribute to waveform structure differently than porous vegetation plus two sites where elevation was known to have changed between GLAS, airborne LiDAR and field data acquisition. These comprised an area where a quarry was present and a second site where construction of a recreational facility had taken place.

Canopy height metrics and detected elevation limits using both systems were compared. Estimates of ground surface elevation, Top Height and Stemwood Volume using GLAS waveforms are evaluated with respect to those obtained from airborne laser scanning. A method is developed to estimate canopy cover from satellite LiDAR and is evaluated using airborne LiDAR estimates of the same, validated using hemispherical photography.

Further processing of airborne LiDAR data then enabled the delineation of individual tree crowns and the extraction of parameters which would be beyond the capabilities of sampling-based techniques e.g. using satellite LiDAR profiling or field plot measurements. This delineation was evaluated using field measurements and has been applied in the validation of yield model estimates of stemwood volume presented in the previous Chapter.

This Chapter concludes with a discussion of the opportunities and limitations offered by both LiDAR systems.

7.2 Stand Level Analysis

Conventional methods of assessing stand-level parameters within forestry in the UK require field measurements within sample plots of 10x10m to determine stand mean data (Matthews and Mackie, 2006). Airborne LiDAR has been shown to provide useful data for the management, assessment and quantification of forest stands, particularly at local scales and has also demonstrated that such small plot areas may be insufficient to accurately represent stands.

A study within Forest Research investigated the use of airborne LiDAR data 'windows' of differing sizes in order to identify optimal sampling areas for the estimation of stand-level parameters (Doce *et al.*, 2008). The 99th percentile of a normalised canopy height model (CHM), discussed in Section 7.3, was used as an approximation of top height. The study demonstrated that window sizes of 30x30m and above produced stand-level estimations of top height with approximately 95% accuracy (Figure 7.1). In fact this suggests that the conventional field plot size may

pose the risk of unrepresentative stand assessment if the number and locations of sample areas do not adequately measure the full variation present.

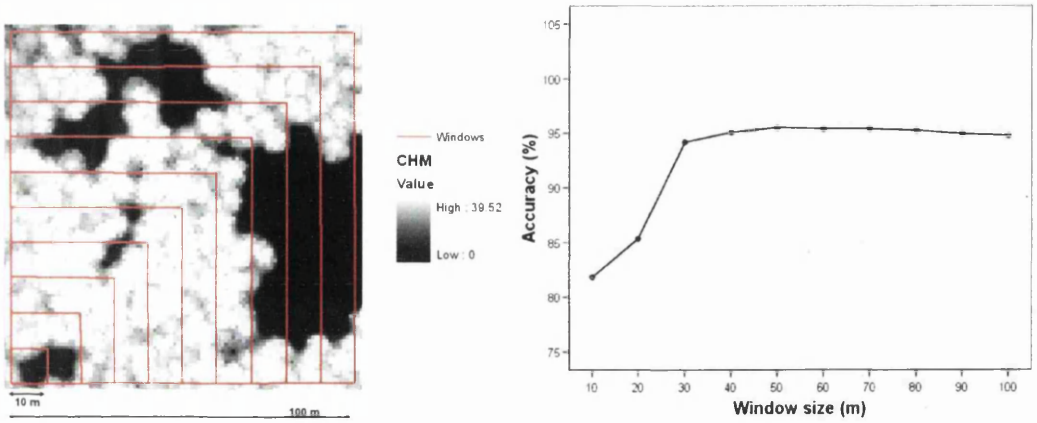


Figure 7.1 (left) Effect of the size of airborne LiDAR subset 'windows' on (right) the estimation accuracy of top height (Doce *et al.*, 2008). This demonstrates that LiDAR subsets of equivalent dimensions to GLAS footprints can provide accurate estimations of stand level parameters.

The results therefore support the use of LiDAR subsets of comparable size to ICESat/GLAS footprints for stand-level assessment. Despite challenges of such systems, this study therefore considers whether large footprint satellite LiDAR data, may nevertheless provide comparable results to LiDAR data from airborne platforms for the estimation of forest parameters.

7.2.1 Data processing

For reference, specifications of the two LiDAR systems are outlined in Table 7.1. Both systems emit pulses at the same wavelength eliminating this as a source of discrepancies. The airborne LiDAR footprint spacing and diameter allows near continuous spatial coverage thereby effectively sampling the GLAS footprint areas.

Table 7.1. Specification overview of the airborne and satellite LiDAR systems

Platform	Airborne	Satellite
Instrument	Optech ALTM	GLAS
LiDAR system	Discrete return	Full waveform
Spatial coverage	Laser scanning	LiDAR profiling
Footprint diameter	~0.2m	~64m
Footprint spacing	0.45m	172m
Wavelength	1064nm	1064nm

7.2.1.1 Airborne Lidar Analysis

The Optech Airborne Laser Terrain Mapper-3033 is a discrete return LiDAR system recording first and last echoes only for each emitted laser pulse. These echoes represent the highest and lowest elevations at which the energy returned to the sensor exceeds a designated threshold.

The footprint diameter of approximately twenty centimetres allows the penetration of gaps between foliage and woody biomass and therefore, whilst some first echoes may be returned from the uppermost canopy surface, others fall within the tree crown or may reach the ground. Conversely, last return echoes may be intercepted within the canopy before reaching the ground surface. Therefore filtering of last returns is necessary to remove points at elevations above the ground surface in order for the topography to be accurately identified and represented for further analysis. The data were therefore processed as follows using the Queensland Remote Sensing Centre (QRSC) IDL® based in-house LiDAR processing tools with the kind permission of John Armston, Department of Natural Resources and Water, Indooroopilly, Australia.

The raw, unprocessed airborne LiDAR data are supplied as ASCII files, formatted as eight fields representing Easting, Northing, Elevation and Intensity for last and first returns respectively (Table 7.2) enabling processing to be undertaken according to echo type.

Using the open source GNU AWK programming language (Free_Software_Foundation, 2006), data in the vicinity of ICESat footprints were subset in order to allow their import into the QRSC LiDAR software tool. The QRSC tool enables ASCII format data to be imported and converted into industry-standard binary LAS format (http://www.lasformat.org/).

Table 7.2 Sample of raw data acquired by the airborne LiDAR flight campaign. Int. represents intensity. Data are provided in ASCII format.

Last Return				First Return			
Eastings	Northings	Elevation	Int.	Eastings	Northings	Elevation	Int.
364568.97	216000.29	166.25	59	364568.79	215999.96	172.53	5
364570.17	216000.23	165.69	23	364569.89	215999.73	174.41	9
364571.17	216000.14	165.44	23	364570.8	215999.54	176.81	1
364572.01	216000.07	165.47	12	364571.64	215999.49	176.22	12
364579.35	216000.23	173.09	38	364579.35	216000.23	173.11	38

Due to sensor time limitations in distinguishing and registering returns, elevation differences between first and last returns of less than five metres were disregarded as effects of noise and therefore considered to be a single return. Subsets of coincident airborne LiDAR data were then created within a radius of 35m about the geo-located ICESat/GLAS footprint locations (allowing for footprint position uncertainty and eccentricity) in order to compare the surfaces detected by the two LiDAR systems and field measurements.

Topography Estimation

An elevation filter was then applied, using the QRSC LiDAR tools, to eliminate non-ground points from the last returns. This method is based on the algorithm of Zhang *et al.*, (2003) in order to identify and remove non-ground LiDAR returns from 'porous' surfaces like vegetation and solid surface structures such as buildings. Initial identification of 'seed' ground points is undertaken using the lowest detected elevation within cells whose size is determined by average LiDAR point spacing. An elevation difference threshold (using slope present at the study site) is utilised to preserve gradually-increasing topographic features whilst identifying abrupt elevation changes assumed to be associated with buildings and vegetation. This factor aims to prevent excessive filtering of ground points and artificial lowering of the estimated ground surface. A series of iterations is performed to apply a progressively larger search area 'window size' to the airborne LiDAR data to enable detection of non-ground objects of varying sizes. This removes returns from trees and objects which are less than the window size whilst retaining those above it (maximum window size is set as maximum tree crown diameter for the study area).

The software requires the following input parameters for this process: Initial disk radius (m) is usually set to equal the average return spacing which was calculated for each footprint area; mean return point spacing was 0.45m with a range of 0.41m to 0.51m. Expected maximum crown radius was entered as seven metres for all footprints to take account of wide crown dimensions of broadleaf species observed during fieldwork. Within-footprint slope was estimated using elevation difference within the coincident 10m resolution OS Land Form Profile DTM described previously and expressed as a percentage with respect to the subset area diameter of 70m. Relative elevation accuracy was set to the system specification

accuracy of 0.15m. The classification of ground/vegetation points resulting from processing using the above input parameters was then displayed within the 3D viewer (QRSC LiDAR tools adapted from Streutkers, 2008) in order to visually determine whether non-ground returns had been satisfactorily filtered.

Since points were regularly distributed with little variation across the study area, ground class surface models for each footprint area were created using linear interpolation with Delaunay triangulation (creating a Triangulated Irregular Network - TIN).

Table 7.3 QRSC LiDAR Tools input parameters

QRSC Parameter	Value
Initial disk radius (m)	0.45
Mean return point spacing (m)	0.45
Expected maximum crown radius (m)	7
Maximum slope (%)	28.6
Relative elevation accuracy (m)	0.15

Stand-level Vegetation Analysis

Return count height percentiles and maximum canopy height above this interpolated ground surface were calculated using 15cm height bins. This is comparable with the 1ns resolution within GLAS waveforms.

Some approaches to processing airborne LiDAR data consider first and last return echoes separately. Within this study, all returns (first, last and single) were used for data processing with the aim of better estimating the continuous profile of returned energy from throughout the canopy seen within full waveforms.

Figure 7.2 presents a 0.5m resolution digital terrain model (DTM) produced following filtering of non-ground last returns.

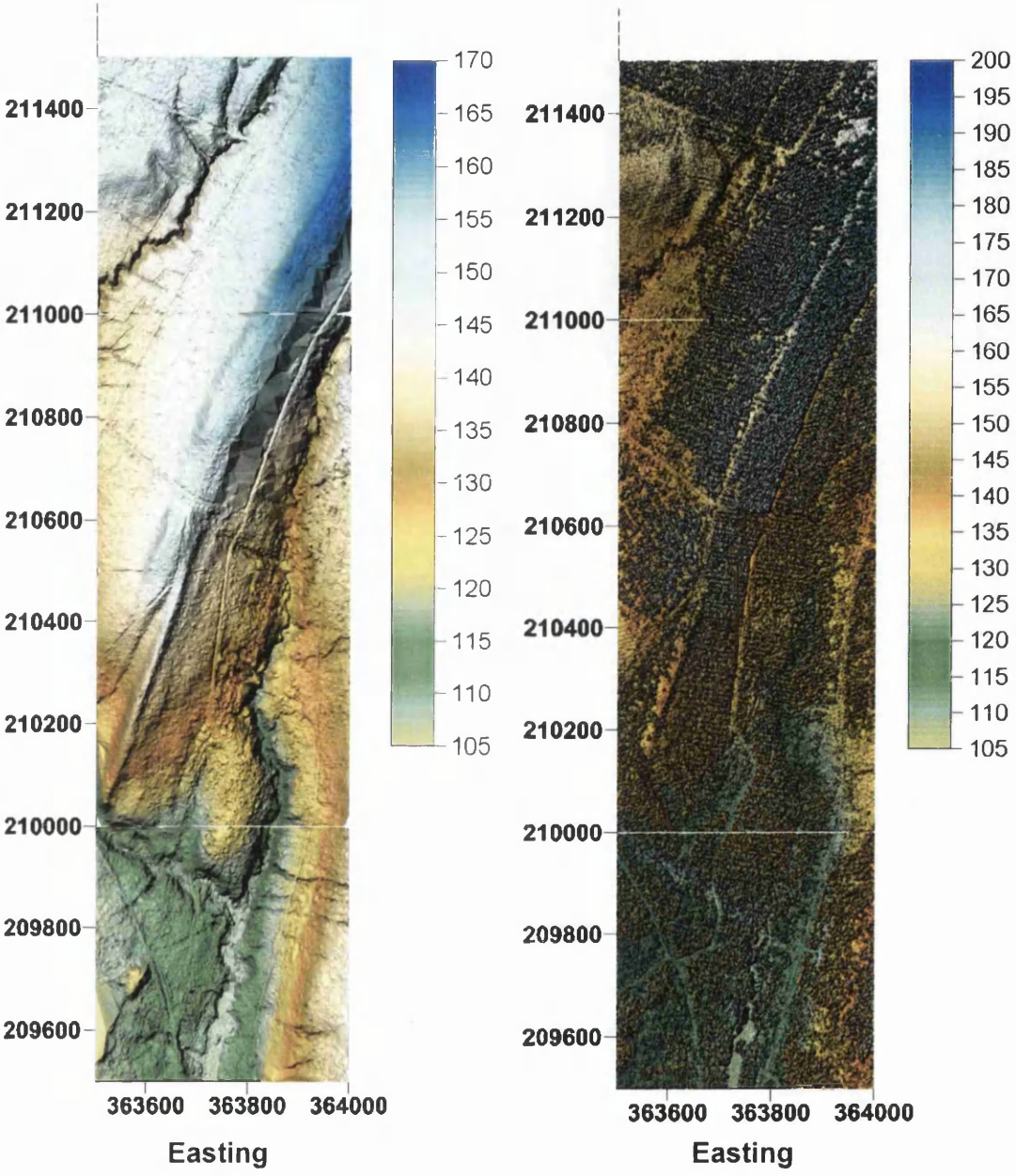


Figure 7.2. (left) 0.5m resolution DTM with shaded relief illuminated from the northwest showing a 0.5km x 2km area of the Forest of Dean and (right) coincident DSM derived from airborne LiDAR data. Colour scales represent elevation in metres.

The image to the right is a digital surface model (DSM) of first returns, thereby representing the spatial distribution of the upper canopy surface. The images show a region of 0.5km by 2km covering a section of the study area within the Forest

of Dean. Both models were created with TIN interpolation using Golden Software Surfer 8.

Differences between the highest intercepted canopy surface and lowest ground elevation within airborne LiDAR subsets were calculated in order to compare these with ICESat/GLAS waveform extent. This was carried out with the aim of assessing whether the effect of footprint size, energy distribution or atmospheric transmittance would result in significant deviations between detected surface limits.

Using the airborne LiDAR ground class, mean slope within footprints was calculated in order to assess the extent to which any differences observed between estimates from the two systems or field measurements may be a function of slope.

Projected plant cover (Armston *et al.*, in preparation 2008) was calculated for each footprint using return point counts above a linearly-interpolated ground surface as described above. A 0.5m height threshold was used to exclude the effects of low cover by ferns, brambles or grass to prevent artificial estimates of cover but to include energy distribution throughout the canopy in order to be comparable as far as possible with the GLAS waveform energy profile. As previously, 0.15m height bins were used for consistency with waveform resolution. Using these criteria, canopy cover was estimated as the number of all canopy points expressed as a fraction of total returns. During this process, a ground/canopy reflectivity ratio was calculated using the mean intensities of the ground class from last returns and canopy class using first returns.

7.2.1.2 ICESat/GLAS Data

Vegetation Height Metrics

The methods described in previous Chapters were used to identify the region of the waveform estimated to be returned from the vegetation (R_{WT} and GP_{MaxAmp}) and to calculate percentiles of cumulative energy for this area. Heights of these percentiles were subsequently used in comparison with airborne LiDAR return count height percentiles. Both methods of estimating vegetation height from GLAS were evaluated.

Waveform Extent (the difference between the beginning and end of the waveform signal) was also compared with elevation limits within coincident GLAS footprint subsets of airborne LiDAR data.

Assessment of Sensitivity to Geolocation Accuracy

Horizontal geolocation accuracy mean and standard deviation have varied between laser operations: L1A 4.6±9.3m; L2C 37.7±53.4m; L3A 0.0±2.7m; L3B 17.4±22.8m. Position accuracy for other laser operation periods, including L3D used in the study, has yet to be determined by NASA.

Therefore an exercise was carried out to assess the effect of possible footprint offsets for L3D using subsets of airborne LiDAR data. Taking the stated footprint centre co-ordinates, subsets of airborne LiDAR data were created for circular areas of 70m diameter, at a distance of 20m and at 45° intervals about this position. This distance was selected as a mid-way point between minimum and maximum reported location uncertainty for GLAS laser 3 campaigns.

Maximum canopy height and percentiles were estimated from the airborne LiDAR data subsets for each offset direction. The resulting range of height estimates were incorporated to assess the potential error in ICESat/GLAS-derived vegetation height due to geolocation uncertainty. Although there may be some drift during the course of a laser operation (approximately one month), the three second duration needed to cross the Forest of Dean suggests that the assumption is reasonable that any offset would be systematic.

Vegetation Canopy Cover

The concept of Gaussian decomposition to identify the vegetation return was also applied to estimations of canopy cover using the returned waveforms: if Gaussian Peak 1 was previously found to have greater amplitude than Gaussian 2, peak 1 was assumed to represent the ground return and the sum of areas under Gaussian Peaks 2-6 to indicate the returned energy from intercepted canopy surfaces. Similarly, if Peak 2 had the greater amplitude, both Gaussian Peaks 1 and 2 were estimated to represent the returned terrain signal whilst the area under Gaussian Peaks 3-6 gives that returned from the overlying vegetation.

Given that waveform amplitude is a function of both surface area of intercepted features plus returned energy from those surfaces, the effect of differences in reflectivity between canopy elements and ground surfaces needs to be accounted for. Therefore a method is needed to isolate and remove this influence, normalising the waveform structure for intercepted surface area alone. Possibilities to achieve this are offered by optical data or airborne LiDAR remote sensing.

Aerial photography was obtained at the same time as the airborne LiDAR data and therefore presented the opportunity of using this to apply a correction for reflectivity differences. However, several effects introduce error:

I. Mixed pixel values combine the reflectivity from different surfaces.

II. Pure ground pixels are expected to be either entirely in sun or shade.

III. Vegetation pixel reflectivity is likely to be from the canopy surface rather than within the crown.

IV. Differences in wavelength.

V. Colour infrared imagery did not extend across all areas covered by ICESat and airborne LiDAR. Therefore an assumption would need to be made for reflectivity beyond this.

Therefore the decision was made to correct the waveform amplitude of vegetation and ground components directly using airborne LiDAR intensity (Figure 7.3).

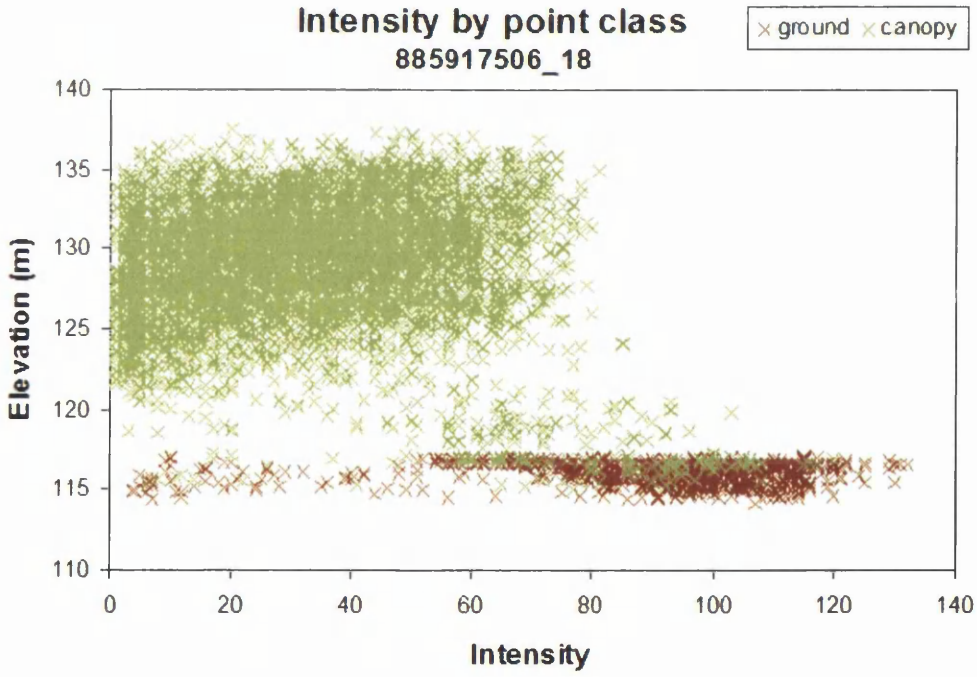


Figure 7.3. Example of intensities of ground and vegetation classes using a subset of airborne LiDAR data for GLAS footprint 558917506_18. For each footprint this was used to remove the effect of differences in intensity for ground and vegetation from GLAS waveform amplitude

This approach offers the following advantages:

I. Differentiating between ground and vegetation airborne LiDAR point classes allows these to be considered entirely separately.

II. Using the airborne LiDAR classification previously discussed, ground point intensities are identified irrespective of whether they are located beneath the forest canopy or in clearings.

III. Likewise, vegetation class intensities upon the canopy surface and within the crown are both included.

IV. The airborne LiDAR intensity record is expected to correspond most closely with the effects of energy differences within waveforms due to the properties of intercepted features.

V. Both systems emit pulses at the same wavelength.

Therefore the adjustment to modify the area under the waveform canopy signal was carried out using the ground/canopy intensity ratio from classified airborne LiDAR points. Please refer to Figure 7.3 for an example of how intensities of ground and vegetation classes differ at 1064nm. Discussions relating to this ratio are also provided by other authors (Harding *et al.*, 2001; Ni-Meister *et al.*, 2001). A comparison was made between reflectance ratios calculated using first returns and all returns to assess the sensitivity to echo type. R^2 of 0.96 was found and therefore the assumption was made that echo type would not significantly alter results.

Waveform amplitude from the canopy (area beneath Gaussians 2-6 or 3-6 as above) was multiplied by this ratio to remove the influence of reflectivity (intensity) differences from the returned waveform amplitude. This was designed to leave the waveform amplitude remaining which is representative of the intercepted surface

profile. It is based on the assumption that the ratio is determined by the reflectivity of intercepted surfaces alone. Footprint canopy cover was then estimated by expressing the modified canopy area as a fraction of total modified waveform area.

Ground Surface Identification within Waveforms

Elevations of the estimated ground positions within waveforms were calculated in order to explore the degree to which a representative ground surface could be identified using large footprint LiDAR and to assess the ability of each ICESat/ GLAS method (R_{WT} and GP_{MaxAmp}) to estimate ground elevation with respect to airborne LiDAR and Ordnance Survey Land Form Profile 10m DTM mean elevations. Waveform ground surface elevations were calculated as follows:

$$Ground = d\_elev + d\_ld\_RngOff - d\_SigBegOff - R_{WT} - d\_gdHt \quad (7.1)$$

$$Ground = d\_elev + d\_ld\_RngOff - d\_SigBegOff - GP_{MaxAmp} - d\_gdHt \quad (7.2)$$

where *d_elev* is the reference elevation of the ellipsoid which is generally the centroid of the standard fit waveform though on occasion is the waveform maximum; the land range offset, *d_ld_RngOff*, indicates the offset position within the waveform of *d_elev*; *d_SigBegOff* provides the offset of the beginning of the waveform signal; R_{WT} and GP_{MaxAmp} are estimated maximum vegetation heights; *d_gdHt* is the height of the WGS-84 geoid below that of the ICESat ellipsoid.

All waveform parameters used are from product GLA14 as original units converted to metres. Waveform offset positions are provided as a negative number with reference to the final bin furthest from the spacecraft recorded in each 150m 'window' and indicate the distance from this position in metres.

7.2.1.3 *Hemispherical Photography*

As described in detail within Section 4.3.2, fifteen ICESat footprints were sampled using hemispherical photography with the aim of assessing the mean footprint fractional cover. For each footprint, nine photographs were taken: at the footprint centre and, using a radius of 20m from this point, at 45 degree intervals from the North position. Hemiview 2.1 processing software was then used to estimate fractional cover for each photograph.

The calculation boundaries were manually placed around the circular limits of each photograph and a threshold was visually applied to create a classification of sky/non sky areas (Figure 7.4). The validity of the binary classification is dependent upon having high contrast photographs in which canopy elements are consistently darker than the sky and yet avoiding glare from the sun through foliage.

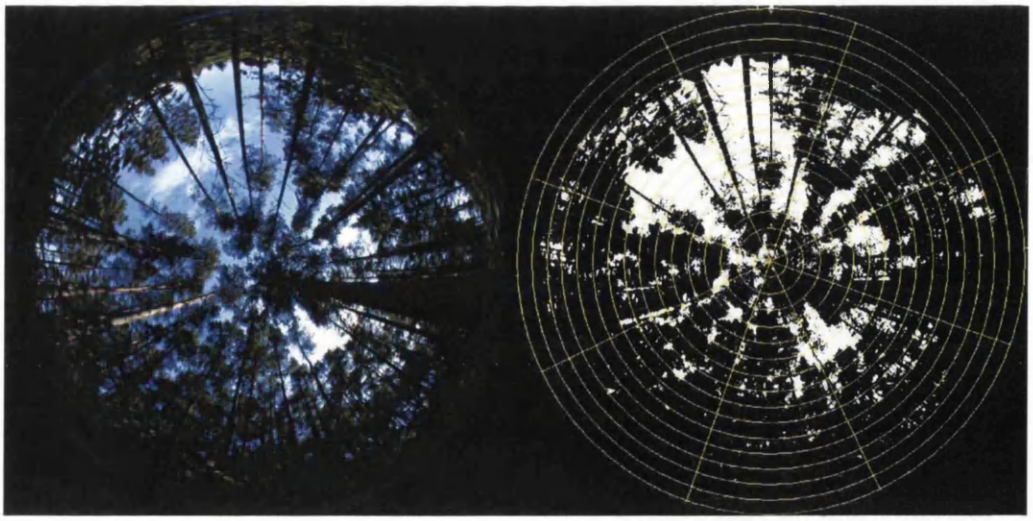


Figure 7.4 (left) Example of a hemispherical photograph taken for footprint 885917496_36. (right) A sky/non-sky threshold is applied visually and the resulting classification is divided into sky sectors (yellow lines). These sectors are weighted for the proportion of sky represented, enabling canopy cover to be estimated.

In practice, processing hemispherical photographs involves a compromise between overestimating large canopy clearings near the zenith and underestimating small canopy gaps near the horizon.

The angular area of each sky sector is calculated for its associated bounding zenith and azimuth angles (θ, α). The software enables gap fraction to be calculated, based on the percentage of visible sky, weighted for the proportion of the hemisphere which each sector of the photograph represents:

$$\mathrm{VisSky}_{\theta,\alpha} = SkyGap_{\vartheta,\alpha} \times SolidAng_{\vartheta} \times PixValid_{\vartheta,\alpha} \Big/ \sum_{(\vartheta,\alpha)} SolidAng_{\vartheta} \times PixValid_{\vartheta,\alpha} \qquad (7.3)$$

where:

- $\mathrm{VisSky}_{\theta,\alpha}$ is the proportion of visible sky in a given sky sector, relative to the entire hemisphere of sky directions, excluding areas ignored using threshold allocation
- $\mathrm{SkyGap}_{\theta,\alpha}$ is the gap fraction for the sky sector
- $\mathrm{SolidAngle}_{\theta}$ is the angular area of the sky sector on a hemisphere of unit radius
- $\mathrm{PixValid}_{\theta,\alpha}$ is the proportion of the sector which is not ignored

(HemiView2.1, 1999)

Canopy cover was then estimated from the hemispherical photographs as 1-[proportion of visible sky]. The mean of the nine calculations for each footprint was used to assess the validity of airborne LiDAR canopy cover estimates.

7.2.2 Results

7.2.2.1 Height Percentiles

An analysis of the relationship between airborne LiDAR and field measurements of maximum canopy height produced R^2 of 0.83 and RMSE 4.2m for N=25; the intercept was not statistically significant (Figure 7.5). This is below that achieved using GLAS waveform estimates but is due to an outlier.

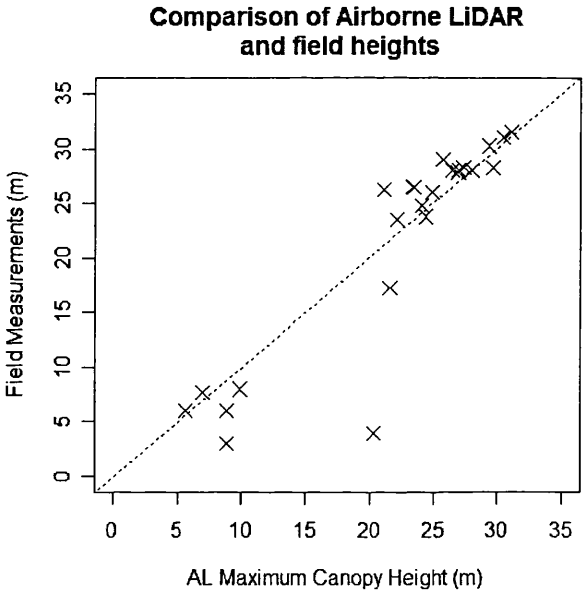


Figure 7.5. Relationship between estimated maximum canopy height from airborne LiDAR (AL) and field measurements. R^2 = 0.83, RMSE 4.2m, intercept not significant.

The canopy metrics calculated for the airborne and satellite LiDAR data showed strong associations, particularly with the GP_{MaxAmp} method of estimating canopy height (Table 7.4). Maximum canopy height estimations from satellite LiDAR explained 68% of variance from airborne LiDAR estimates, RMSE 4.4m (seen in Figure 7.7). For the upper percentiles, strong correlations were found which decreased with greater depth through the canopy below the 95th percentile.

Of particular interest was the relationship with the 98th and 99th airborne LiDAR height percentiles which often correspond with Top Height estimations within forestry. These percentiles were estimated with R^2 of 0.76 (RMSE 3.4m) and R^2 of 0.75 (RMSE 3.5m) respectively.

This demonstrates that, despite the large differences in scale, satellite LiDAR waveforms and coincident airborne LiDAR point clouds can be seen to similarly represent canopy profiles, particularly among the upper region of the canopy. Airborne LiDAR is becoming more widely used and accepted in operational forestry. The correspondence found between results from satellite and airborne platforms suggests the prospects for future practical applications for satellite LiDAR remote sensing.

Table 7.4. Comparison of airborne and satellite LiDAR percentiles using areas covered by stated GLAS footprint co-ordinates. N=59

Position	GP_{MaxAmp} R^2 (RMSE)	R_{WT} R^2 (RMSE)
Maximum height	0.68 (4.4 m)	0.61 (4.9m)
99th percentile	0.75 (3.5 m)	0.70 (3.9m)
98th percentile	0.76 (3.4 m)	0.71 (3.8m)
95th percentile	0.75 (3.5 m)	0.69 (3.9m)
90th percentile	0.67 (3.8 m)	0.63 (4.1m)
85th percentile	0.65 (3.8 m)	0.59 (4.2m)
80th percentile	0.63 (3.9 m)	0.56 (4.3m)
75th percentile	0.61 (3.9 m)	0.50 (4.4m)
50th percentile	0.54 (3.8 m)	0.34 (4.7m)

This is further supported by the similar estimations of Top Height produced using both systems. For this diverse site, the 98th airborne LiDAR percentile estimated Top Height with R^2 of 0.73; RMSE 4.5m (Figure 7.6). Whilst the satellite

LiDAR maximum canopy height using the GP_{MaxAmp} method explained 77% of variance from yield model Top Height estimates (Chapter 6).

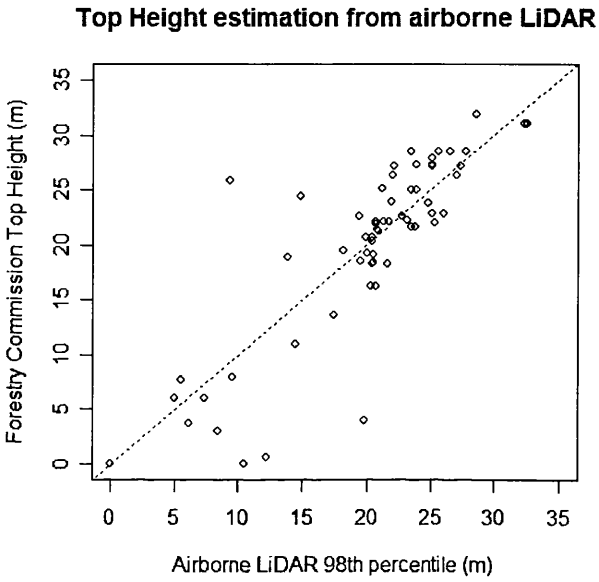


Figure 7.6. Estimation of Top Height using Airborne LiDAR 98th height percentile. $R^2 = 0.73$; RMSE 4.5m, intercept not significant, coefficient significant $p<0.001$. N=59

Although there is high correspondence between airborne and satellite LiDAR vegetation height estimates, the dispersal which is seen may be a result of high species heterogeneity within the Forest of Dean, meaning that a small discrepancy in footprint location or shape could produce significant deviations.

An analysis was undertaken of correlations with offset airborne LiDAR point clouds. Given the high spatial variability in species distribution within some stands, offsetting the footprint area about the stated co-ordinates of footprints revealed potential height differences of up to -18m and +20m (Figure 7.7).

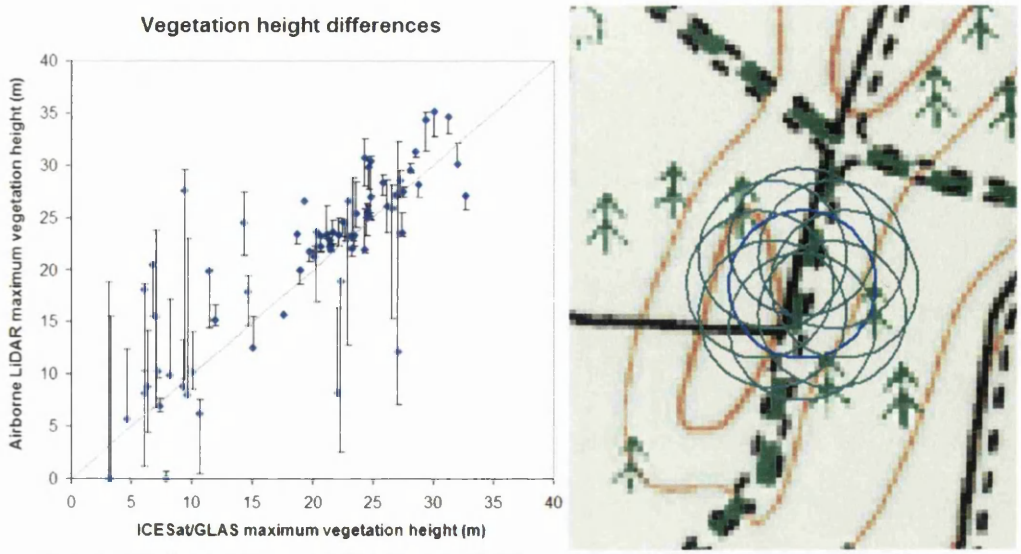


Figure 7.7. (left) Airborne LiDAR and GLAS maximum height estimates with error bars representing canopy height differences within 20m of stated footprint co-ordinates; N=59. (right) Illustration of subset areas, offset from the footprint centre, for which local heterogeneity was determined.

The optimum fit of the nine trials suggests that the GLAS pass used for this study may in fact be positioned approximately 20m to the northeast of published locations (revised maximum canopy estimates: $R^2 = 0.79$, RMSE 4.0m, Figure 7.8; improved from $R^2 = 0.68$, RMSE 4.4 m).

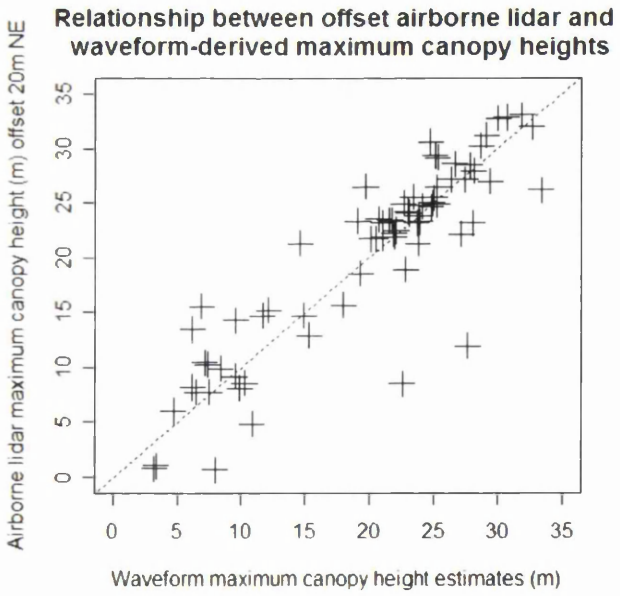


Figure 7.8. Comparison of maximum canopy height estimates from ICESat/GLAS and subsets of airborne LiDAR data, 20m to the northeast of 59 stated GLAS footprint co-ordinates.

7.2.2.2 *Ground Surface Detection*

The difference between the uppermost and lowest detected surfaces from GLAS (Waveform Extent) demonstrated a close correlation with coincident airborne LiDAR elevation limits, producing R^2 of 0.71 and RMSE of 5.0m with no significant intercept (Figure 7.9).

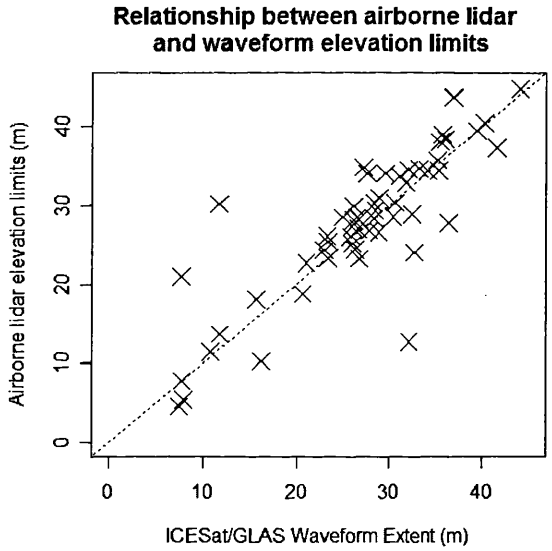


Figure 7.9. Comparison of detected surface limits between airborne LiDAR (difference between lowest ground elevation and highest canopy surface) and satellite LiDAR (Waveform Extent) for 59 GLAS footprint areas.

Figure 7.10 shows estimations of within-footprint mean ground elevation using Ordnance Survey Land-Form Profile 10m resolution DTM, mean ground class elevation using airborne LiDAR and ICESat/GLAS estimated ground surface using both GP_{MaxAmp} and R_{WT} methods. Estimates can be seen to correspond closely throughout the pass with the differences found in Table 7.5.

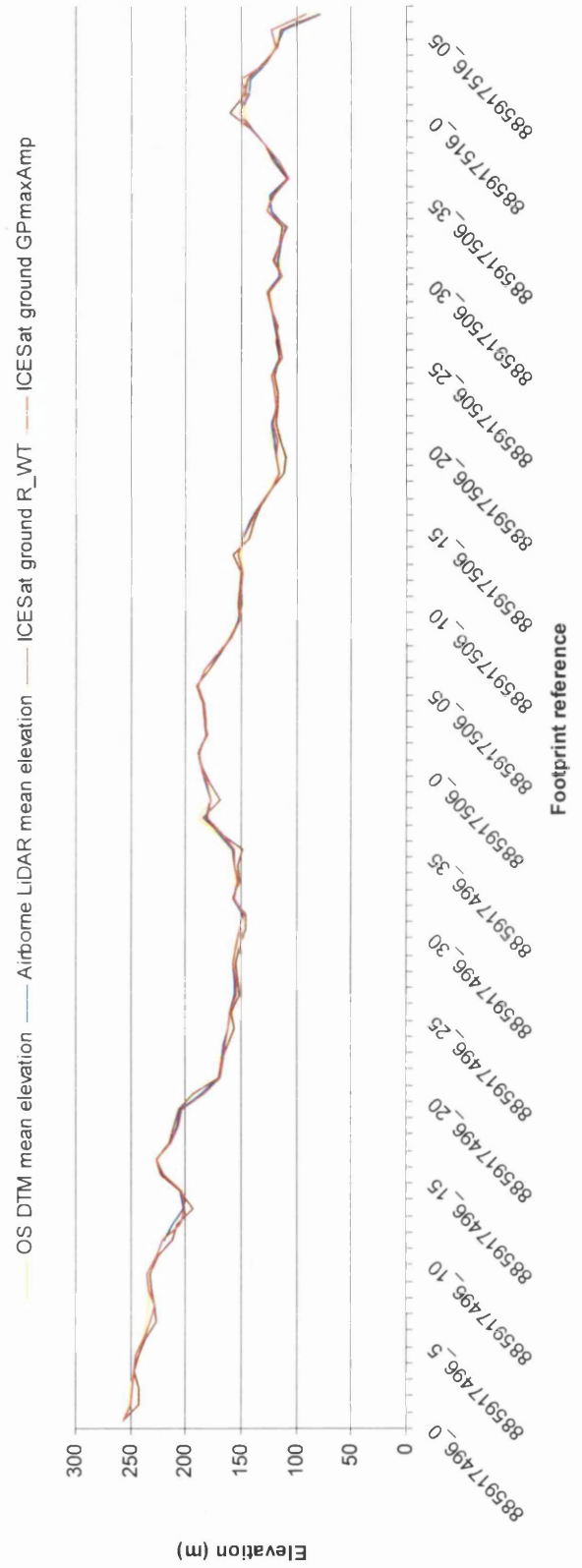


Figure 7.10. Comparison of ground identification using mean elevation of airborne LiDAR ground class points, satellite LiDAR estimation of the ground from R_{WT} and GP_{MaxAmp} methods and a reference dataset: 10m resolution Ordnance Survey DTM.

Along with the allocation of the beginning of the waveform signal, the ability to identify the elevation of the ground surface within waveforms is a key factor in estimating vegetation height. Assessment of this is therefore important for identifying sources of error within estimates.

The use of waveform structure to identify a ground elevation (GP_{MaxAmp}) has resulted in the closest estimate to both the reference OS DTM and the fine spatial resolution estimate obtained from the airborne LiDAR dataset. In contrast, the R_{WT} method relies on the representation of the independent DTM and the accuracy of the signal end position but nevertheless has succeeded in estimating the ground surface with mean bias of approximately one metre.

The distribution of error found with LiDAR data in relation to the OS DTM reference dataset is illustrated in Figure 7.11 and statistical relationships are presented in Table 7.5. The Gaussian decomposition method (GP_{MaxAmp}) underestimated mean ground elevations of airborne LiDAR by 0.32m and the Ordnance Survey DTM (OS) by 0.10 m. By comparison, the R_{WT} method shows a greater range of error however also produces a relatively small mean bias from OS data of -0.83m (-1.05m with AL mean elevation).

Table 7.5. Comparison of estimated ground surfaces using Ordnance Survey and LiDAR data.

Comparison (m)	AL-OS	R_{WT} -OS	GP_{AMP} -OS	R_{WT}-AL	GP_{AMP}-AL
Mean offset	0.22	-0.83	-0.10	-1.05	-0.32
Max. difference	3.05	12.07	9.05	9.02	11.47
Min. difference	-2.95	-10.43	-6.98	-9.64	-7.36
Standard Deviation	1.28	3.92	2.21	3.66	2.26

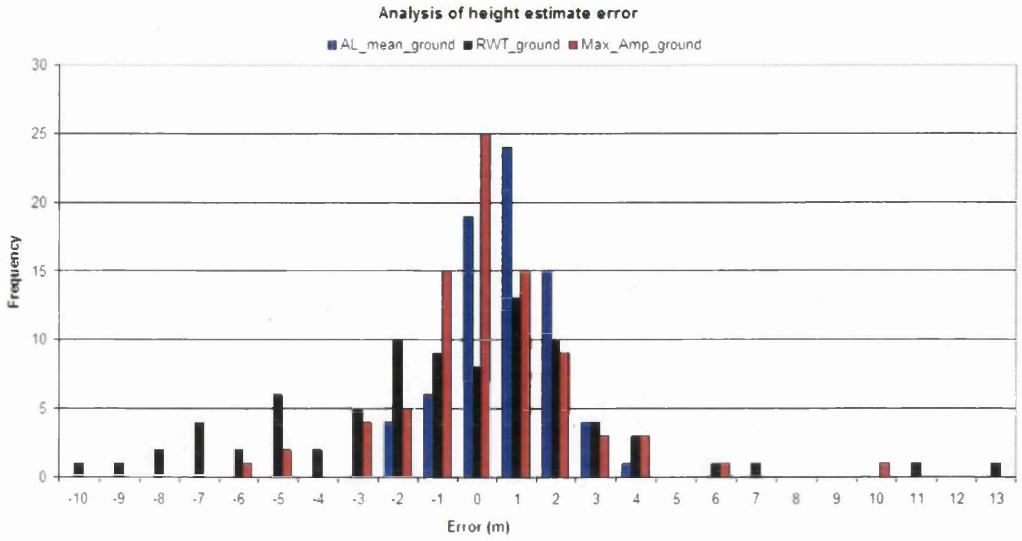


Figure 7.11 Distribution of error in estimates of the ground surface using GLAS R_{WT} and GP_{MaxAmp} methods and airborne LiDAR mean ground surface with respect to mean OS DTM elevation within ICESat/GLAS footprints.

When compared with airborne LiDAR ground surface, mean slope calculated from the airborne LiDAR ground class explained 39% and 0.5% of the percentage error using GP_{MaxAmp} and R_{WT} estimates of the ground surface respectively (Equation 7.4). The ground elevation difference within footprints using airborne LiDAR data explained 38% and 1% of the percentage error using GP_{MaxAmp} and R_{WT} methods respectively. This suggests that the R_{WT} method has succeeded in removing the effect of terrain as an error source.

$$AbsError_{(GLAS_{grd} - AL_{grd})} = \alpha \times AL_{slope} \tag{7.4}$$

where $GLAS_{grd}$ is the estimated elevation of the ground within GLAS waveforms (Section 2.2.1); and AL_{grd} and AL_{slope} are the mean elevation and the calculated slope respectively for GLAS footprint subsets using airborne LiDAR ground class points.

7.2.2.3 *Stemwood Volume*

Single Species Stemwood Volume

For all vegetated footprints across the study site, estimations of stemwood volume for the tallest species within footprints using the 98th airborne LiDAR percentile produced R^2 of 0.60 and 103.4 m^3ha^{-1} RMSE with no significant intercept.

Differentiating between stands with greatest coverage provided by broadleaf or coniferous species improved the relationships as follows. The 75th percentile showed the best relationship when estimating stemwood volume for broadleaf species with R^2 of 0.80, RMSE of 55.95 m^3ha^{-1}. The 99th percentile best estimated conifer stemwood volume with R^2 of 0.63, RMSE of 98.55 m^3ha^{-1}. Intercepts were not statistically significant. These correlations are seen in Figure 7.12.

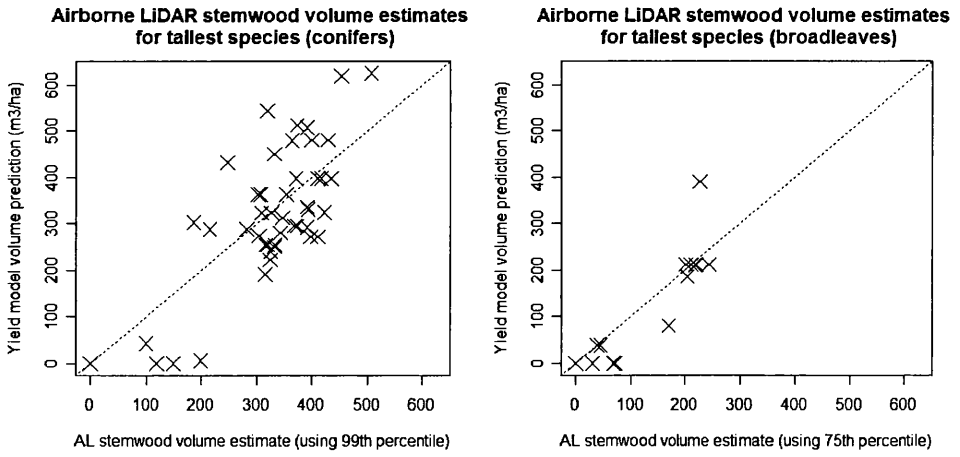


Figure 7.12. Estimation of stemwood volume for the tallest species within stands using airborne LiDAR. (left) relationship for conifers $R^2 = 0.63$, RMSE 98.55 m^3ha^{-1} (right) correlation for broadleaf species; $R^2 = 0.80$, RMSE 55.95 m^3ha^{-1}

Mixed Composition Stemwood Volume

Stemwood volume estimates for all vegetated footprints taking account of the mixed composition of stands produced R^2 of 0.46, RMSE of 102.9 m^3ha^{-1}. Whilst considering broadleaves and conifers separately produced R^2 of 0.53, RMSE of 79.93

m^3ha^{-1} for broadleaf species and R^2 of 0.59, RMSE of 92.22 m^3ha^{-1} for conifers. In all cases the intercept was not statistically significant. Correlations for broadleaves and conifers are seen in Figure 7.13.

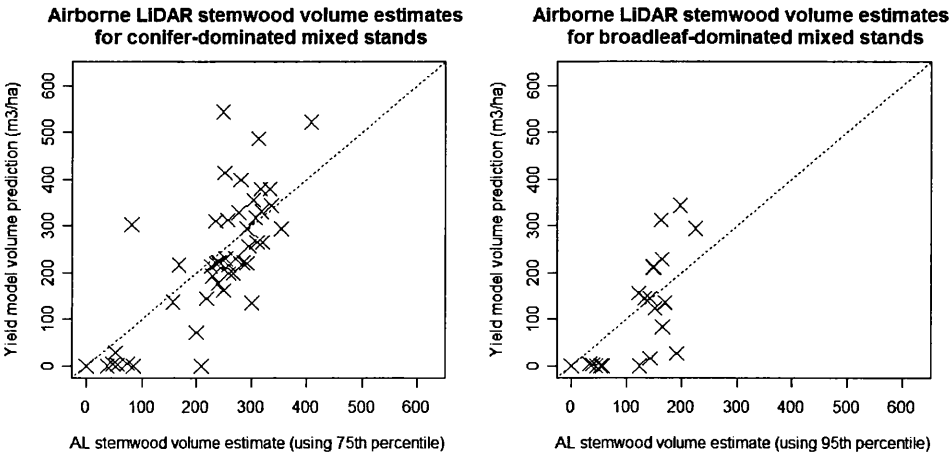


Figure 7.13. Estimation of stemwood volume for the mixed composition of stands using airborne LiDAR. (left) relationship for conifers $R^2 = 0.59$, RMSE 92.22 m^3ha^{-1} (right) correlation for broadleaf species; $R^2 = 0.53$, RMSE 79.93 m^3ha^{-1}

For reference, Table 7.6 shows a comparison of stemwood volume estimates for both LiDAR systems. Greatest similarity in analysis trends is seen between estimates using airborne LiDAR data and the GLAS GP_{MaxAmp} method. This higher correspondence could be due to fact of the latter method using waveform structure to isolate the canopy return which more similarly replicates the airborne LiDAR method of calculating height percentiles above the classified ground surface.

The ability of GLAS waveform indices to estimate stemwood volume similarly to airborne LiDAR data supports the potential to apply large footprint LiDAR estimation of biophysical parameters more widely.

Table 7.6. Comparison of stemwood volume estimate correlations using airborne and satellite LiDAR

	Species type	Tallest species R^2 (RMSE m^3/ha)	Mixed stand R^2 (RMSE m^3/ha)
Airborne LiDAR	Broadleaves	0.8 (55.95)	0.53 (79.93)
	Conifers	0.63 (98.55)	0.59 (92.22)
	Combined	0.6 (103.4)	0.46 (102.9)
GLAS R_{WT}	Broadleaves	0.64 (68.7)	0.46 (76.3)
	Conifers	0.63 (90.3)	0.55 (93.8)
	Combined	0.56 (100.2)	0.42 (103.1)
GLAS GP_{maxamp}	Broadleaves	0.75 (59.1)	0.47 (75.6)
	Conifers	0.59 (98.3)	0.66 (82.5)
	Combined	0.59 (100.8)	0.5 (97.8)

An example of mixed stemwood volume estimates along the satellite pass crossing the Forest of Dean using both LiDAR systems and yield model calculations is seen in Figure 7.14. Gaps in estimations are found where ICESat/GLAS footprints cross urban constructions or do not fall within Forestry Commission sub-compartments. Airborne LiDAR data are not available for the entire extent of the Forest of Dean.

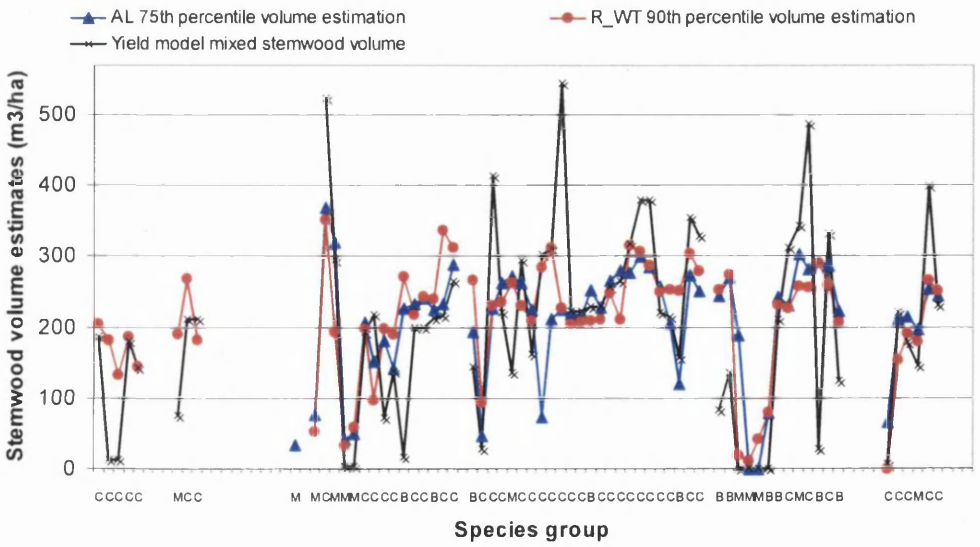


Figure 7.14 Mixed composition stand stemwood volume estimates using airborne LiDAR data, GLAS R_{WT} method and calculations from Forestry Commission yield models. X-axis categories indicate the dominant species group for each GLAS footprint: C=conifer, B=broadleaf and M=mixed stands of approximately equal composition.

7.2.2.4 *Canopy cover*

Canopy cover provides an indication of stand regeneration potential, stage in stand succession (greater cover is often seen with younger stands once established, which decreases with competition and mortality) and suitable conditions to promote biodiversity (light penetration permitting multi-storey or sub-canopy vegetation growth).

Despite the small data range, hemispherical photography estimates of canopy cover produced a strong linear relationship with airborne LiDAR cover calculations (R^2=0.77; RMSE=0.02) thereby supporting their use as an indication of ground truth across the entire area (Figure 7.15).

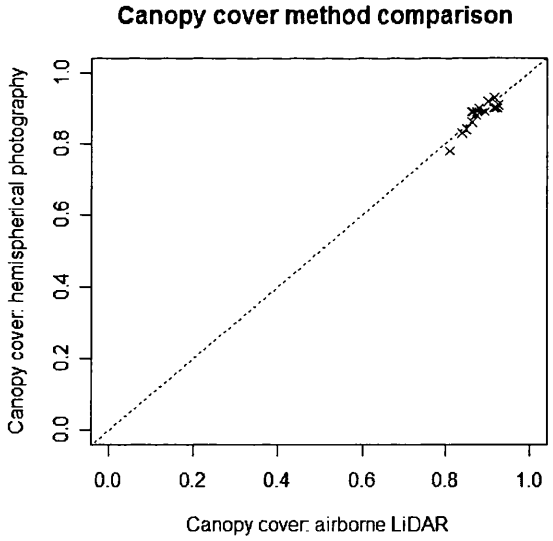


Figure 7.15. Estimates of canopy cover from hemispherical photograph and airborne LiDAR vegetation class points as a percentage of total returns. A strong correlation is found for field-sampled footprints using hemispherical photography despite the small data range, thereby supporting the use of airborne LiDAR estimates.

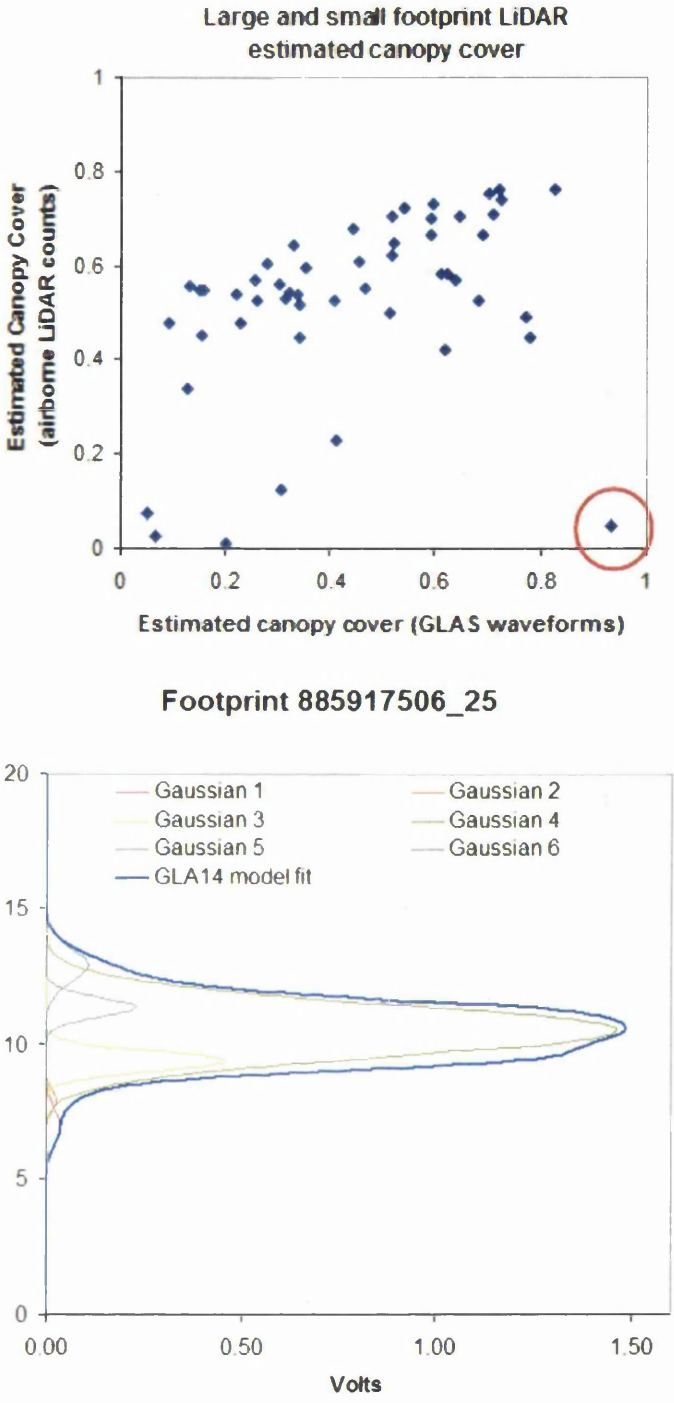


Figure 7.16. (above) Direct comparison of airborne and satellite LiDAR estimates of canopy cover. (below) Gaussian decomposition of the anomalous waveform for the encircled outlier.

Figure 7.16 shows ICESat/GLAS and airborne LiDAR system canopy cover relationships. The example raises an interesting situation in which the waveform (below) representing the extreme outlier indicated, is broadened by a low stature,

lone tree on a gentle slope meaning the Gaussian fit has placed the greatest amplitude for the 3rd peak causing the ground surface to be misclassified as canopy.

Multiple linear regression using waveform-derived cover and vegetation height addresses the error source though produces R^2 of just 0.32 (RMSE 0.16):

$$Cover = \alpha Area_{Canopy} + \beta R_{WT} \quad (7.5)$$

Where *Cover* is the estimate of fractional cover calculated from airborne LiDAR data; $Area_{Canopy}$ is the modified area beneath the canopy return as a percentage of total modified area beneath the GLAS waveform; R_{WT} is maximum vegetation height estimated from GLAS; α and β are coefficients applied to the variables.

Distinguishing greatest coverage by conifers or broadleaf species improved relationships to R^2 0.41 (0.16 RMSE) and R^2 0.63 (0.11 RMSE) respectively (Figure 7.17). The use of multiple regressions combining dominant canopy height or percentile associated with volume were explored, however these did not further improve estimates of canopy cover.

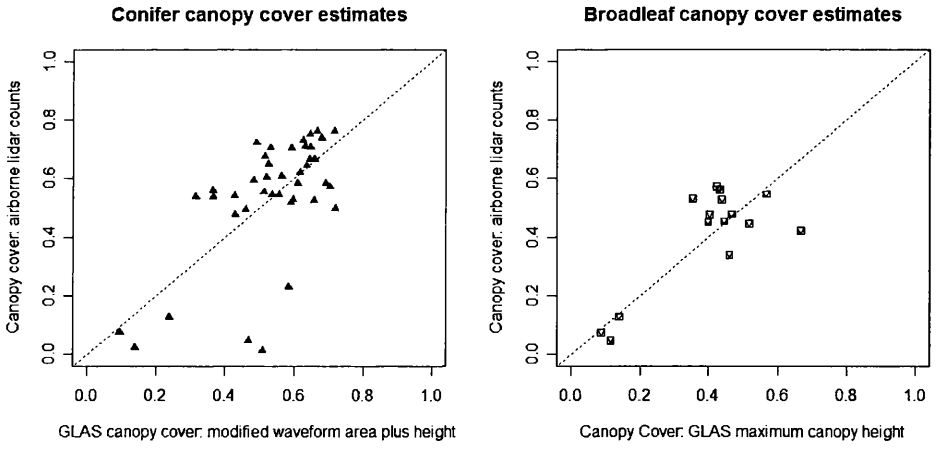


Figure 7.17. Canopy cover estimates for broadleaf and conifer-dominated stands using multiple regression for GLAS waveform area (modified for reflectance differences and to express canopy area as proportion of total area) and estimated vegetation height in comparison with canopy cover estimated from airborne LiDAR data.

7.3 Crown Delineation and Canopy Level Analysis

Irrespective of along-track footprint frequency and ground-track concentration, individual tree level vegetation analysis is beyond the capabilities of large footprint LiDAR applications as the aggregation of returned energy from the intercepted surfaces within the footprint area cannot be spatially differentiated. Small footprint LiDAR therefore exceeds the potential of large footprint systems in this respect as the distribution of intercepted surfaces and estimated parameters can be spatially represented and analysed at finer scales than those possible with the former.

Individual tree dimensions and volume are important factors for carbon accounting and forest management. Additionally, topographic relief and shielding from adjacent vegetation need to be considered when modelling susceptibility to wind. Quantifying vegetation distribution and providing inputs to inform process modelling will not only identify and allow appropriate management of site-specific risks but also permit optimum and targeted felling of stands or individuals. Airborne LiDAR data enables the locations, heights and canopy dimensions of individual trees to be mapped and therefore addresses many of the data requirements for assessing vulnerability to wind damage and uprooting during storm events (Suárez *et al.*, 2008b).

In this Section, a method of delineating canopies based on object-orientated analysis is explored which not only demonstrates the capabilities and applications of airborne LiDAR data, but also provides a means of assessing yield model stemwood volume estimates from the previous Chapter.

7.3.1 Data Processing

Following the method of Suárez *et al.*, (2008a), a ground return class was determined from last return echoes using Terrascan 007.008 Software, an extension application for Bentley Microstation V8 2004. This ground class and all first return data were subsequently exported as LAS files within 0.5x0.5km grids covering the Forest of Dean and converted into regular 0.5m resolution raster geotiffs using Delaunay triangulation with linear interpolation within the QRSC LiDAR processing software. A canopy height model (CHM) was calculated as the difference between the digital terrain model (DTM) created from the ground class and the digital surface model (DSM) from the first return data using ArcGIS 9.2. An example area of this is seen in Figure 7.18 (right).

Definiens Developer 7.0 was used to delineate individual tree canopies solely from the LiDAR-derived 0.5m resolution CHM. Conditions relating to the study site which may influence the performance of algorithms within this software are outlined below.

When delineating canopies, interlocking crowns can pose difficulties in distinguishing foliage between adjacent trees. This may be particularly so for broadleaf species where crown structure may not produce such distinct local elevation minima between crowns as seen with conifers. This may therefore be a challenging factor for the Forest of Dean. Additionally, the nature of sensors observing from above can obscure suppressed vegetation meaning that dominant and co-dominant trees are best represented at the expense of others. Furthermore, delineation algorithms need to restrict crown shape to prevent possible expansion into intra-crown spaces producing elongated or irregularly-shaped crowns.

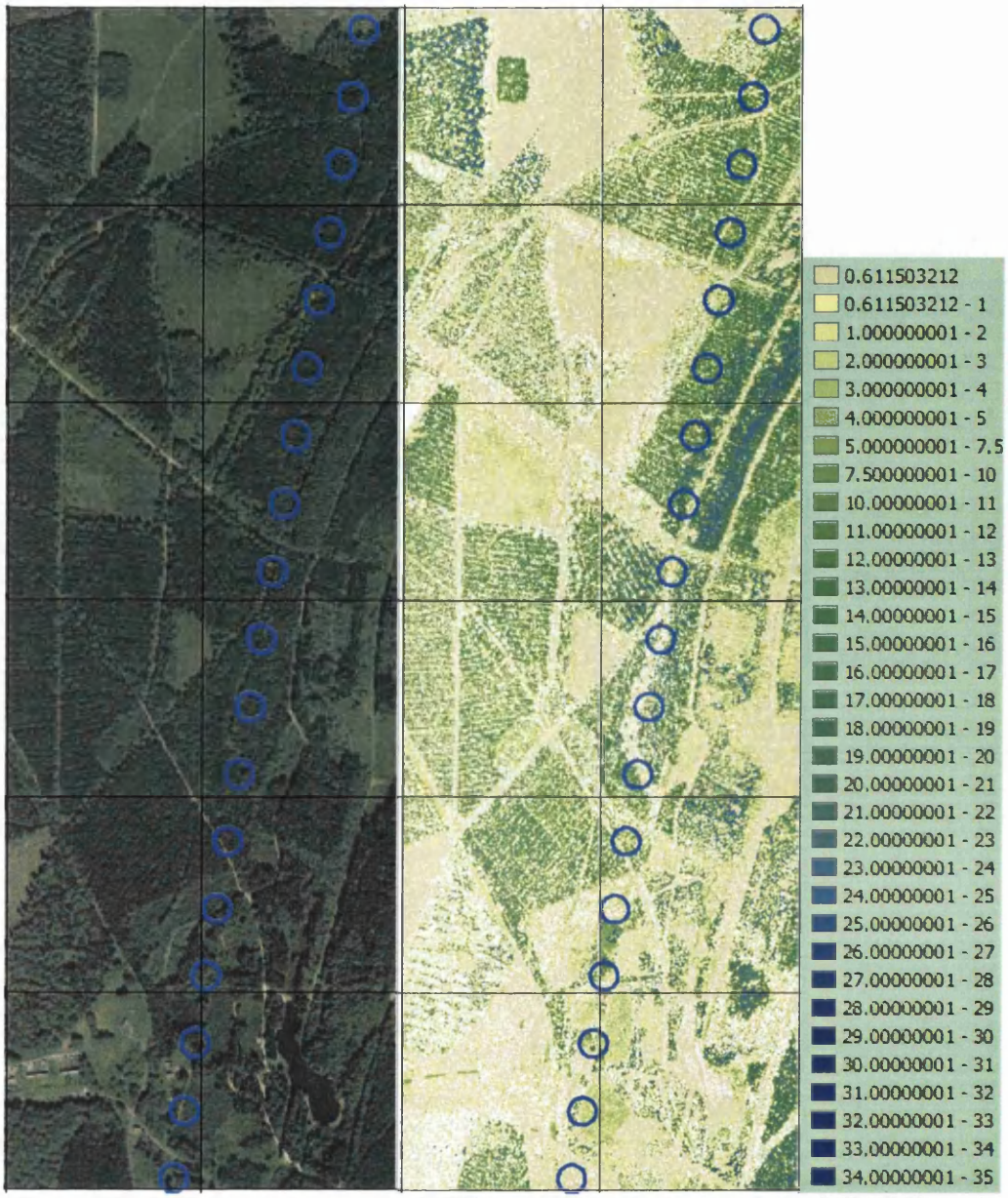


Figure 7.18. (left) Aerial photography (RGB) and (right) coincident airborne LiDAR Canopy Height Model (CHM) showing an area of 1x3 km within the Forest of Dean. Key illustrates height in metres. Minimum and maximum bounding co-ordinates: 363000, 209000 and 364,000, 212000.

For some sites with distinct and widely differing vegetation properties, it may be necessary to apply segmentation, e.g. using stand boundaries, in order to adjust algorithm parameters to suit vegetation characteristics. The method used within this

study aims to account for different crown properties by vegetation height, allowing the algorithm to assign broader crowns to taller trees and to distinguish individual regularly-shaped crowns within closed canopies.

Process of Tree Crown Delineation

The algorithm used to achieve this is based on site-specific refinements made to a Definiens ruleset which was developed by Juan Suárez, Forest Research, Northern Research Station.

This ruleset performs a series of routines to identify tree tops and crown boundaries using the LiDAR canopy height model. Modifications were made to this in order to provide optimum classification for target GLAS footprint areas for which field measurements had also been taken.

Firstly the CHM was smoothed using a Gaussian blur of 5 in order to reduce the effect of clumping and height variability within crowns. Local maxima were then identified, located and classified as tree tops. Areas of ground or understorey vegetation below a 2m threshold, plus canopy 'edges' were identified using local minima and used as boundaries to prevent further canopy extension.

The algorithm was subsequently run to extend tree tops radially until they either met an adjacent crown or designated boundaries. A mask was applied to limit irregularly shaped polygons producing a more natural shape (e.g. Figure 7.19).

Tree top locations were then saved as point shapefiles and polygons representing individual canopies were also exported with associated maximum crown height, crown area, maximum and minimum radii, width and polygon centroid co-ordinates.

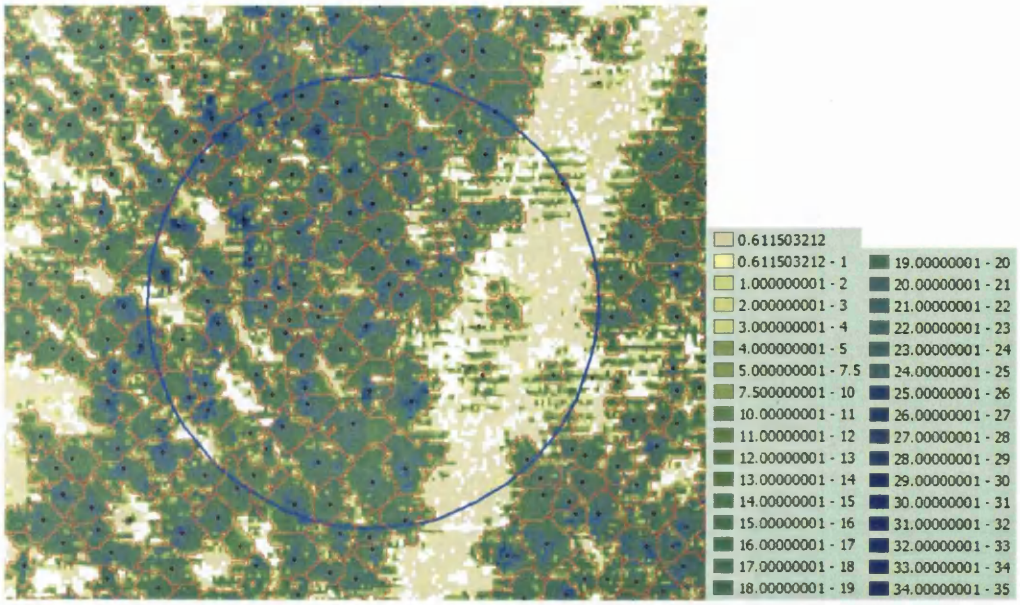


Figure 7.19. An example of canopy delineation of vegetation within footprint 885917506_13 (blue circle – diameter 70m). Colour key illustrates height in metres for the canopy height model. Calculated tree crown boundaries are outlined in red; blue points represent the highest elevation (tree top) within each tree crown.

7.3.2 Crown Delineation Analysis and Results

The validity of the above delineation was tested using field data. In order to be able to calculate individual tree volume (Chapter 6), six footprints were selected for which field measurements had been made and which contained a high percentage of a single species or species with known spatial distribution (Table 7.7). This enabled a species-specific regression equation to be applied for the calculation of DBH using individual tree height and crown dimensions.

Using the eight field-measured tree heights recorded within each footprint area, heights and locations of these were compared with those extracted from the crown delineation for coincident areas. NB footprint 885917506_14 showed a discrepancy between the area measured (located using a standard hand-held GPS) and the actual GLAS footprint area identified using footprint co-ordinates. For this site, the delineation was therefore verified using the area measured.

Table 7.7 Species composition of stands used to verify airborne LiDAR delineation. Some areas of sub-compartments are unplanted and therefore the sum of components need not equal 100%. DF represents Douglas Fir, LI is Lime, OK is Oak and NS indicates Norway Spruce.

GLAS Footprint	Species: Sub-compartment 1		Species: Sub-compartment 2	
885917506_13	96% DF	2% LI	-	-
885917506_14	90% DF	10% OK	80% DF	-
885917506_15	90% DF	10% OK	-	-
885917506_16	90% DF	10% OK	-	-
885917516_04	85% NS	-	-	-
885917516_05	100% DF	-	100% OK	-

The result of this analysis found that positions of trees were located to within three metres and delineation of the canopy height model underestimated individual tree height by a mean value of 5% (range = -22% to 22.5%). Tree heights for delineated trees were therefore adjusted to compensate for this mean 5% height underestimation. The accuracy of this investigation was comparable with that found by Suárez *et al.*, (2005a) and so estimates using this method were accepted as a useful surrogate for field measurements for the stands listed in Table 7.7.

The algorithm was found to perform better for trees above approximately 12m in height than for shorter trees. Below this height, tree tops were identified but the algorithm failed to adequately delineate crowns. Closed canopies did not prevent individual trees above this height from being identified and located. Use of canopy delineation to provide surrogate ground truth validation data would require segmentation of analysis by vegetation height, in order to produce reliable canopy representation for all tree heights. Additionally, the use of differential GPS for field measurements would assist with accurately placing field plots and locating trees within this.

7.4 Discussion

7.4.1 Stand-Level Parameters and Forestry Applications

This study has explored the relationships between canopy metrics derived from small footprint, discrete return airborne laser scanning and large footprint full waveform LiDAR profiling from a spaceborne platform. Previous studies using airborne LiDAR have shown data subsets of comparable size to ICESat/GLAS footprints to provide a good representation of stand characteristics. This suggests that, where GLAS footprints are coincident with forestry management units, stand characteristics could be estimated from returned waveforms.

Vegetation Height Estimation

Airborne LiDAR data estimated field measurements of maximum canopy height with $R^2 = 0.83$, RMSE = 4.2m. Disregarding the effect of an outlier, airborne LiDAR estimates of maximum vegetation height corresponded closely to field measurements (R^2 0.94, RMSE 2.4m). A similar degree of accuracy was found between estimates of field height derived from both airborne and satellite LiDAR systems.

Greatest correlation in height percentiles between satellite and airborne LiDAR is observed in the uppermost canopy, particularly between the 95th and 99th percentiles (Table 7.4). This may be due to the increased influence of changes of intensity for shaded foliage and leaf angle distribution with greater depth through the canopy (waveform amplitude being a function of both surface area and intensity relationships by elevation), or the effects of crown structure and its interaction with airborne laser scanning angle. The higher correlation between airborne LiDAR

vegetation height estimates and the GLAS method using Gaussian decomposition may be due to both similarly underestimating vegetation height.

Ground Surface Identification

This study has shown the ability to identify a representative ground surface from satellite LiDAR waveforms. For the Forest of Dean, the method using Gaussian decomposition to estimate ground elevation within the waveform ground peak (GP_{MaxAmp}) produced the smallest mean error in comparison with both airborne LiDAR and Ordnance Survey Land Form Profile DTM mean ground elevations. Slope was identified as a contributory factor for the minor negative offset using Gaussian decomposition whereas this had been successfully addressed using the Waveform Extent/Terrain Index method (R_{WT}). Use of a terrain index may similarly reduce the effect of slope on percentage error for the former method however would defeat the object of deriving estimates entirely from waveforms. A means to account for this may be using the width of the ground peak return as an indication of slope. A further explanation may be offered by the fact that the model fit is produced by the sum of Gaussian peaks and therefore the centroid of the Gaussian Peak with greatest amplitude may not always represent the most common ground elevation. Use of the largest amplitude inflexion point within the ground return may address this small error.

Assuming the OS DTM to be an accurate indication of ground truth, for both GLAS methods, a small negative bias is seen in the estimation of the ground surface. This amounts to a mean error of less than one metre. For the R_{WT} method, this may be a result of the waveform 'tail' extending below the true lowest ground surface. The airborne data differs from the satellite system in marginally overestimating the

ground surface, possibly as a result of detection of ground-cover vegetation although error range is significantly smaller (within approximately 3m) than using the large footprint LiDAR.

The results suggest that, for situations such as the Forest of Dean in which dense canopy cover or extreme slope do not prevent a representative ground surface from being detected, Gaussian decomposition may offer the most appropriate means of estimating ground elevation. Furthermore, GLAS estimations of ground elevation have shown high consistency across different laser operations. Sun *et al.*, (2008) compared 260 waveforms, covering a distance of approximately 4.5km, captured during the laser 2A campaign (September-November 2003) and laser 3F operation (May-June 2006). Footprint pairs were positioned at a mean distance of 82.6m apart (range 75.4m-89.8m). The relationship between ground surface elevations of these near repeat passes produced $R^2 = 0.997$ and RMSE 4.1m. This strongly supports the ability to consistently estimate ground elevation beneath vegetation from GLAS waveforms.

The strong relationship between elevation limits found by airborne and satellite LiDAR suggests that issues of varying atmospheric transmittance, footprint size or energy distribution have not prevented intercepted surfaces from being similarly detected.

Site Complexity

On a single footprint basis, direct interpretation of waveform structure can be challenging without local site knowledge. Figure 7.20 shows two footprints with complex terrain and vegetation distribution. The canopy height and ground surface estimates for these sites are found in Table 7.8.

This demonstrates how the broad ICESat/GLAS footprint may be problematic for sites such as this whereby the canopy signal may be broadened and signals from ground and vegetation surfaces may be combined. In the case of 885917506_30, the distribution of terrain and above-surface features has not deterred a representative estimation of vegetation height. The example of 885917506_33, however, demonstrates the advantage of airborne LiDAR to estimate vegetation height above a moving ground surface. For this footprint, both satellite LiDAR methods overestimate vegetation height due to the necessity of attempting to select a single ground elevation which is representative of complex ground topography and the difference from the highest intercepted surface (Signal Beginning). Nevertheless, when considering a series of estimates across a forest as shown in this study or potentially for significantly larger areas, overall errors can be small.

Table 7.8. Example estimates for footprints with complex spatial distribution of features shown in Figure 7.20. WF indicates waveform, AL represents airborne LiDAR and OS is Ordnance Survey.

Footprint	885917506_30	885917506_33
Field maximum height (m)	28.25	28
WF R_{WT} height (m)	27.64	31.82
WF GP_{MaxAmp} (m)	27.52	35.35
AL max height (m)	27.46	28.14
Mean OS DTM (m)	114.1	112.0
WF R_{WT} ground (m)	112.9	111.8
WF GP_{MaxAmp} ground (m)	113.1	108.3
AL mean ground elevation (m)	115.7	109.4

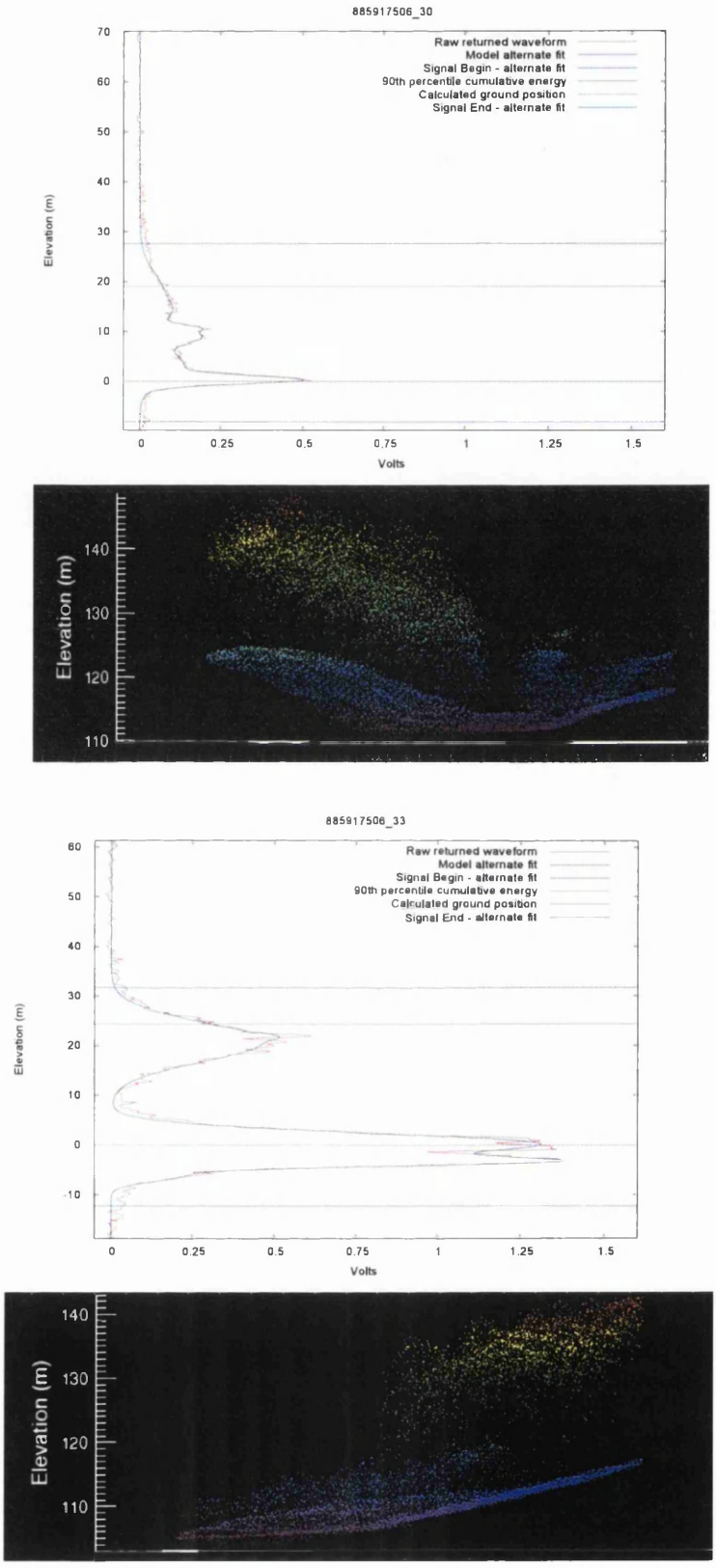


Figure 7.20. GLAS waveforms and coincident airborne LiDAR point clouds (70m horizontal extent) for two challenging sites. In both cases the spatial distribution of features and similar elevations of ground and vegetation may influence the ability to estimate vegetation height.

Biophysical Parameter Estimation

For forestry purposes, the ability to estimate top height (98^{th}/99^{th} percentile) may indicate the broader applicability of large footprint LiDAR profiling (Figure 7.21). Furthermore, other authors have demonstrated how these principles may be extended to biophysical parameter estimates for significantly larger areas (Nelson *et al.*, 2004; Nelson *et al.*, 2008a).

Figure 7.21. Representation of LiDAR profiling using airborne LiDAR subsets of GLAS footprints. Footprint centres are distributed at 172m intervals. Regular sampling in this way effectively samples forest variability in a similar way to conventional field techniques.

Stemwood volume estimates using satellite LiDAR are similarly estimated using airborne LiDAR, particularly with regard to the GP_{MaxAmp} method where use of the waveform structure replicates the airborne LiDAR classification closer. Although this suggests that estimates using satellite waveform indices has not degraded estimates of volume, a considerable RMSE is found in all cases which may be the result of errors within yield model calculations or an indication that further refinements are needed to estimate stemwood volume from LiDAR data.

Canopy cover is a challenging parameter to estimate using both systems. Figure 7.22 shows an example of a cross section of multiple echo, high density (35 points/m^2) airborne LiDAR data for an area of Glen Affric, Scotland, showing laser penetration throughout the canopy. For areas such as the Forest of Dean, with less dense footprint coverage, the recorded points may most represent the upper canopy surfaces. This may affect estimates of canopy cover, although use of canopy point counts of all echo types as a proportion of total points aims to account for this.

Figure 7.22. An example of high density airborne LiDAR data for Glen Affric, Scotland (illustration produced using Bentley MicroStation and Terrascan software). This cross section shows laser energy penetration throughout the canopy. Lower point densities are likely to represent lower portions of the canopy profile less well.

Estimates of canopy cover from hemispherical photography are based on the subjective judgement of the threshold applied to classify sky/non-sky sectors of the photograph. The sensitivity of this threshold to resulting estimates of gap fraction was explored and was not found to cause significant differences in estimations if the threshold used was not extreme. The use of the mean of nine calculations for each footprint also serves to reduce the effect of any misjudgement.

Correlations of $R^2 = 0.41$ (RMSE 16%) and $R^2 = 0.63$ (RMSE 11%) are observed between satellite and airborne LiDAR estimates of canopy cover for conifer and broadleaf-dominated stands. The factors which may explain the dispersal seen amongst estimates are varied. This may be due simply in part to seasonal variability for the ten month discrepancy in data capture between the LiDAR systems. Additionally, spatial heterogeneity and the nature of sampling using laser scanning and the continuous signal obtained from the full waveform can further explain differences seen. The latter may provide a possible explanation of the greater gap fraction observed from the GLAS waveforms with respect to the airborne LiDAR counts.

A further explanation may be found in the complex interactions between canopy structure, leaf orientation or surface reflectivity and the returned intensity of airborne LiDAR points which was used to modify the amplitude of the GLAS waveforms. Intensity is an uncalibrated and unitless record which, as yet, requires greater understanding. As well as the influence of reflectivity, it is been shown to vary due to laser scanning incidence angle, range between emitter and target, echo type, signal pre-processing, beam divergence and atmospheric attenuation. Vegetation attributes such as leaf area, leaf angle inclinations, species and tree density can also affect intensity (Boyd and Hill, 2007; Kaasalainen *et al.*, 2007; Ørka *et al.*, 2007). For the purposes of this study, airborne LiDAR point class intensities are considered to be entirely the result of interception by the classified surface. However, sub-footprint gaps may also control the ratio in addition to the reflective properties of the surfaces (J. Armston, Pers.Comms.) hence the relationship may be affected by canopy volume, density, clumping and foliage size.

Despite these issues, the correction to waveform amplitude is applied using intensity records from airborne LiDAR, a related system which additionally emitted pulses at the same wavelength. A ratio of simultaneously-acquired, class-related intensities is used rather than attempting to infer properties from absolute values. The airborne LiDAR data are coincident with the GLAS footprints and so canopy structure and interactions with surface features are consistent as far as possible. Likewise, when footprints are differentiated between dominance by broadleaf or coniferous species, clearer relationships are found, supporting this approach to estimating canopy cover.

Limitations to airborne LiDAR systems' ability to distinguish small distances between first and last echoes may hinder the assessment of low-lying vegetation,

although system improvements and the wide use of small footprint, relatively high density specifications may compensate for this (e.g. Hill, 2007). However, the profile throughout the canopy provided by full waveform LiDAR overcomes this. In addition, full waveforms more closely match radiative transfer theory to LiDAR. This enables biophysical parameters to be directly estimated from remote sensing techniques using interactions with the surfaces under investigation (Maltamo *et al.*, 2007; Reitberger *et al.*, 2007; Wagner *et al.*, 2007).

7.4.2 Tree-Level Parameters and Forestry Applications

There are applications for which small footprint LiDAR exceeds the applications of the aggregated surface information derived from large footprint waveforms. The identification of a dynamic topographic surface and spatial distribution of vegetation returns allows individual trees to be located. These can then be related to the ground surface underlying the canopy and adjacent vegetation in order to generate model inputs to assess competition indices and wind risk or slope stability. This offers a unique means of producing detailed DTMs beneath forest layers which can be used to determine access routes or appropriate management approaches for forestry applications, to identify hydrological systems or areas of subsidence. The spatially represented point cloud allows forestry practitioners or other interested parties to 'virtually' return to a field site as frequently as necessary to assess conditions at the time of data capture. This format and presentation can also be intuitively interpreted by those not familiar with LiDAR data and principles.

The potential applications of tree-level parameters are multifold. This study has demonstrated how field validation data can be used to develop allometric

relationships using tree dimensions such as height, crown width and DBH. Suárez *et al.*, (2008b) demonstrate one such application using mean canopy width and tree heights to provide inputs to the forest wind risk model, ForestGALES (Geographical Analysis of the Losses and Effects of Storms in Forestry). This is a process-based model which enables risk of wind damage to be assessed for different management scenarios and with changing conditions due to stand growth (Gardiner *et al.*, 2004; Forestry_Commission, 2008; Suárez *et al.*, 2008b). Given local site characteristics, the model therefore identifies the wind speed at which windthrow will occur. Reconstruction of the location and dimensions of individual trees permits the risk of windthrow to be mapped for specific trees and therefore provides a more meaningful analysis of risk than stand mean data (Suárez *et al.*, 2008a).

In addition, canopy delineation methods can be applied to estimate stand-level mean parameters e.g. fractional cover, crown dimensions, vegetation height etc. which may provide input parameters such as for the radiative transfer model FLIGHT (North, 1996) in order to link to forest light regime and photosynthesis.

This study has shown the ability to determine individual tree-level parameters from airborne LiDAR remote sensing. These estimates were verified using coincident field data. The delineation has therefore been applied as a substitute to comprehensive field measurements in order to calculate stand volume presented in the previous Chapter. This illustrates the potential offered by small footprint airborne LiDAR data which enable the detailed reconstruction of a forest.

7.5 Airborne and Satellite LiDAR Conclusion

As previously discussed, challenges can arise from the broad GLAS footprints as returned signals from the vegetation upon sloped ground surfaces can be combined within waveforms. Additionally, issues of atmospheric transmittance may distort the returned waveform, either by way of breadth or amplitude, which may influence the detection of the beginning of the waveform signal returned from the uppermost intercepted surface.

However, the near-global sampling of ICESat/GLAS LiDAR has the potential to estimate important forest parameters for regional-national scales. The sampling pattern of ICESat also offers the opportunity for inventory or validation estimates supplementary to field measurements at a spatial and temporal frequency that may not be economical or feasible using conventional field techniques or frequent acquisitions of airborne LiDAR.

The research discussed in this chapter contains work presented in Rosette *et al.*, (accepted for publication) and Rosette *et al.*, (submitted). The crown delineation and canopy level analysis represent the further application of techniques used in the preparation of Suárez *et al.*, (2008a).

This Chapter has compared the estimation of vegetation properties derived from coincident small footprint, discrete return airborne LiDAR scanning data with those extracted from large footprint, full waveform LiDAR profiling using GLAS. Airborne LiDAR return point count height percentiles were compared against satellite LiDAR heights of cumulative energy percentiles. Maximum canopy height estimates produced a relationship with R^2 of 0.68 (RMSE 4.4m). A comparison of 98^{th} and 99^{th} percentiles (comparable with top height estimates in forestry) explained

76% and 75% of variance with RMSE of 3.4m and 3.5m respectively. Satellite LiDAR estimates of field measurements and Top Height or stemwood volume derived from yield models were comparable with those achieved using airborne LiDAR.

Detection of surface elevation limits corresponded well between the two systems, producing R^2 of 0.71 and 5m RMSE. ICESat/GLAS estimates of ground elevation succeeded in producing a offset of -0.32m when compared with airborne LiDAR mean ground class elevation and -0.10m from mean surface using a coincident Ordnance Survey Profile 10m DTM.

Waveform-derived canopy cover explained just 32% of estimates obtained from airborne LiDAR data. This improved to 41% and 63% when differentiating between greatest coverage by coniferous or broadleaf species respectively.

The results presented suggest that the broad ICESat/GLAS footprints can provide estimates of mixed vegetation biophysical parameters which are comparable to those obtained from relatively high density airborne LiDAR data.

Airborne LiDAR estimates at a stand level have been shown to produce comparable or improved results to conventional or field measurements at other sites e.g. Maltamo *et al.*, (2007) demonstrate that, for Finland, in addition to mean plot level tree heights, area-level basal area estimations allowed optimum felling time to be identified whilst tree-level diameter distributions are often important model inputs. Due to the near vertical perspective of LiDAR incidence scanning angle, both basal area and diameter parameters are estimated using indirect relationships with height percentiles and the same might be possible from satellite systems for stand level analysis. Therefore field measurements may be used to complement satellite LiDAR waveform indices in order to develop suitable regression models with

indirect factors such as DBH, basal area and crown properties for model input where precise locations of individual trees are not necessary.

The methods presented in this Chapter have been tested for a highly diverse, mixed forest for which a small difference in location may potentially produce larger errors than more homogeneous stands. However, the high degree of footprint overlap with anticipated maximum offsets and use of field methods to take account of possible location differences appear to address discrepancies. The high consistency between GLAS and airborne LiDAR retrieval, in particular for percentiles of estimated canopy height and detected surfaces, suggests that the broad footprint dimensions produced by ICESat/GLAS can nevertheless produce comparable results to the use of relatively high density airborne LiDAR data.

Where repeat orbit satellite LiDAR profiling stands apart is in its potential to comparatively estimate vegetation parameters at a stand level or above for large areas which would be unfeasible or too costly using conventional field techniques or repetitive airborne LiDAR surveys. This further supports the application of satellite laser altimetry for vegetation applications at scales ranging from forest management requirements of assessing mean stand conditions to national forest inventory; global estimates for model assimilation or to reduce uncertainty in vegetation distribution.

Chapter 8. Conclusions

This Chapter provides a summary of the research undertaken during this project and the relevance of the key findings to the fields of carbon accounting, environmental modelling and forestry. Advantages and limitations of Satellite LiDAR data and research methods are considered and opportunities offered by alternative small footprint LiDAR technologies are discussed. Directions for future research are suggested with respect to proposed missions, prospects for extending applications and development of methods. The thesis concludes with an outline of the principal aspects of the project results relating to satellite LiDAR data analysis of vegetation.

8.1 Research Summary and Original Contributions

This research has explored the potential offered by satellite LiDAR data for carbon accounting and vegetation monitoring, model applications and forest management operations.

Chapters 1 and 2 have provided an introduction and given the scientific context of the research. An evaluation was presented of the state of the art of forest parameter retrieval using LiDAR remote sensing. **Chapter 3** described the

pioneering GLAS instrument and pre-processing of the data used during the course of this research.

New material has been presented in **Chapters 4-7**. The principal original contributions are discussed below:

Chapter 4 defined the methods developed to address the project research questions and presented the data sources used. Yield models, data and forest management systems maintained by the Forestry Commission were described, along with characteristics of the project study site. These sources have been used in support of the work throughout the project with the kind permission of Forest Research. Field methods used to gather data for sites coincident with ICESat/GLAS footprints were discussed in detail. These results were subsequently used to validate waveform estimates.

Chapter 5 investigated potential means of extracting maximum canopy height from large footprint satellite LiDAR waveforms. Two methods were identified as offering reliable estimates of the region of the waveform signal returned from vegetation. The first requires the use of a supplementary DTM dataset to account for ground relief within footprints. Multiple regression using waveform limits plus a coincident terrain index were used to form vegetation height estimates. Field data are required to calibrate the regression equation. This adaptation of the Lefsky *et al.*, (2005) method, produced improved results and succeeded in removing the contribution of slope from uncertainty in estimates.

The second method derives estimates entirely from the returned signal, using waveform structure to identify a representative ground surface and estimating vegetation height as the difference between this elevation and the beginning of the waveform signal. The accuracy of this method is dependent upon the strength of the

signal returned from the ground. This new approach demonstrated the possibility of estimating maximum canopy height exclusively from returned waveforms although resulted in a slightly lower correlation than the method above. It is concluded that, for the Forest of Dean, the broad footprint dimensions rarely prevented a ground return within waveforms.

The two methods defined in Chapter 5 were applied and further developed in **Chapter 6** in order to estimate Top Height and Stemwood Volume. Estimates were evaluated and compared with those from the Forestry Commission sub-compartment database and associated yield models. An exercise was undertaken using airborne LiDAR canopy delineation (Chapter 7) to calculate potential error within yield model estimates. This demonstrated that the assumption of stationarity within yield model predictions may overestimate stemwood volume by up to 22%.

Height percentiles using airborne LiDAR data have previously been found by other authors to offer useful estimators of forest parameters in the UK (e.g. Patenaude *et al.*, 2004; Suárez *et al.*, 2005b). A method was therefore devised using GLAS waveforms to extract equivalent heights above the allocated ground surface of cumulative energy percentiles. These were used as waveform vegetation height indices for the estimation of both Top Height and Stemwood Volume.

This approach enabled Top Height to be extracted from large footprint waveforms. This parameter is significant in terms of forestry as an input to model predictions for stand-level assessment of growth, wind exposure and timber potential. The ability to estimate this variable for large areas therefore offers prospects for regional or national UK forest monitoring. Additionally, analysis using waveform percentiles demonstrated the ability for estimates of stemwood volume to

be derived from GLAS data. This supports the potential of their future application for large area estimation of biomass distribution and carbon accounting.

In **Chapter 7** methods were developed to compare vegetation parameter estimates from large footprint, full waveform satellite LiDAR and discrete return, airborne laser scanning data. Subsets of airborne LiDAR data were created, coincident with GLAS footprints. This aimed to consider whether the challenges posed by possible combined signals from ground and vegetation would impede the estimation of vegetation indices. Maximum canopy height estimates and detected surface elevation limits of both airborne and satellite systems were evaluated. Airborne LiDAR height percentiles were used to estimate Top Height and Stemwood Volume in order to assess any deterioration in capabilities using large footprint data. A high degree of correspondence was found in all cases, substantiating the wider use of large footprint systems for forestry purposes.

Subsets of airborne LiDAR data were also used to indicate possible offsets in ICESat/GLAS footprint locations. This exercise concluded that actual geo-location of GLAS co-ordinates for the laser campaign crossing the Forest of Dean are likely to be offset by approximately 20m to the northeast of published locations. For highly mixed sites, uncertainty in footprint location may result in significant discrepancies between field measurements and waveform estimates due to this fact alone.

Using coincident elevations from within the Ordnance Survey 10m resolution DTM as a validation dataset, an assessment was made of the ability to identify a representative ground surface within GLAS waveforms and from airborne LiDAR ground class data. The airborne data differ from estimates using the satellite system in marginally overestimating the ground surface (mean difference = 0.22m, standard deviation = 1.28m), possibly as a result of detection of ground-cover vegetation. The

results suggest that, for situations such as at the Forest of Dean in which dense canopy cover or extreme slope do not prevent a representative ground surface from being detected, Gaussian decomposition of GLAS waveforms offers an appropriate means of estimating ground elevation (mean difference = -0.10m, standard deviation = 2.21m).

An approach to calculating canopy cover from GLAS waveforms was developed and evaluated against estimates using airborne LiDAR data. The latter were verified using hemispherical photography. The agreement found provides the grounds for further exploration of regeneration potential or calculations of leaf area index using large footprint laser profiling.

Finally, the investigation of the capabilities of airborne LiDAR data was extended to the extraction of individual tree-level parameters. The delineation method of Suárez *et al.*, (2008a) was adjusted for the Forest of Dean in order to identify tree locations and heights (validated using field measurements) plus crown dimensions. Such capacity far exceeds that of large footprint sampling for the reconstruction of forests and parameterisation of process-based growth, quality and wind-risk models.

8.2 Implications of LiDAR Remote Sensing

The monitoring of biophysical parameters at scales required to understand global environmental processes are only available through remote sensing. Passive optical data provide a two dimensional view of surface features enabling indirect indications of vegetation quantity using reflectance properties of objects e.g. NDVI and fAPAR.

LiDAR remote sensing, however, offers the possibility of systematic sampling of the Earth's surface, providing a vertical aspect to land cover classes using physical interactions with the target surfaces and thus producing a vertical profile of intercepted surfaces.

8.2.1 Satellite LiDAR

Large footprint satellite LiDAR data offer a unique opportunity of assessing elevation profiles of both topographic and above-surface features at global scales. The limitations and advantages of such systems, illustrated by ICESat/GLAS, have been explored throughout this thesis and are outlined below. Satellite LiDAR data uniquely provide remote, geo-referenced elevation data to a high degree of accuracy on a global scale. Returned waveforms are produced using the physical interactions between photons and the structural and reflectance characteristics of the intercepted features. This has enabled canopy profile estimations to be extracted from waveforms which can be applied to derive vegetation height and canopy cover parameters and a non-destructive means of estimating vegetation volume. Such estimates are not directly obtainable at these scales using other means.

The innovative GLAS system provides an unprecedented opportunity to derive global topographic and above-surface elevation data for an extended period of time. The system is not optimally configured for vegetation analysis, as a dedicated LiDAR system would likely produce greater density of footprints with diameter on a par with tree crown width. Nevertheless, ICESat/GLAS has provided data which have enabled this research to be undertaken to assess the potential of satellite LiDAR in this respect and in support of a sensor designed for this purpose.

ICESat/GLAS data are freely available from the NSIDC. The seasonal repetition of data acquisition offers a unique means of quantifying vegetation change on a relatively frequent basis without incurring excessive costs associated with airborne LiDAR campaigns. Equally, this degree of field survey sampling would be cost and time-intensive and would not be viable to repeat for large areas with such frequency.

Some characteristics of large footprint LiDAR systems by their nature may impede forest parameter estimations. Broad footprint diameters increase the chance for signals from the ground and vegetation to be combined within the returned waveform. For low vegetation, this may be the case even with minor surface relief. In such situations, accuracy of determining the waveform canopy return is likely to be reduced.

Due to the premature failure of laser 1, GLAS campaign specifications were modified to extend the life of the remaining lasers. For the UK, this has reduced longitudinal ground track density to approximately 36km. This sparse coverage is further exacerbated as data capture is dependent on atmospheric conditions, since scattering from cloud will affect data quality of returned waveforms. This may interrupt continuity along the footprint profile and obstruct the assessment of vegetation change.

There is some uncertainty with respect to footprint locations which has varied between GLAS laser campaigns. Additionally, imprecise footprint repetition along ground tracks and varying footprint eccentricity between laser campaigns would require interpolation between waveform estimates to be made or assumptions regarding stand-level representation of these. The accuracy of this will depend on forest homogeneity.

8.2.2 Airborne LiDAR Systems

Airborne LiDAR data can fulfil and surpass the capabilities of conventional field measurements and data provided by large footprint LiDAR profiling. These latter approaches are currently limited to analysis of mean stand-level parameters using sampling techniques.

Airborne laser scanning can provide high resolution elevation data to accurately reproduce a scene to a degree and extent which would usually be unfeasible using conventional field methods. This technique has the advantage of enabling a site to be 'revisited' by practitioners and re-processed for different purposes e.g. producing a DTM for access, slope stability assessment and archaeological identification or classifying vegetation returns to analyse for forestry applications. Additionally, the point cloud visualisation techniques provide an intuitive, recognisable scene which is possible for non-specialists to understand and navigate.

Uniquely, canopy analysis permits delineation of individual trees, enabling their height, volume and canopy dimensions to be estimated. In this way, a forest may be accurately reconstructed and trees placed in the context of topography, infrastructure and proximity to neighbours. Furthermore, these permit the simulation of competition and growth indices and risk factors such as wind to be calculated through inputs to process-based models.

8.3 Future Prospects of Satellite LiDAR

Three sensors are currently proposed for future satellite LiDAR remote sensing: ICESat II is anticipated for launch in approximately 2015, following the original specifications of the first ICESat mission. A debate is currently in progress regarding the emitted pulse duration of ICESat II. A lengthening of the pulse may produce waveforms which are less sensitive to vegetation than is currently the case and this may have implications for the continuity of biophysical parameter estimation using comparable systems.

However, if alternative system designs such as DESDynI and LIST are realised, these aim to offer more pertinent data for the estimation of vegetation parameters. DESDynI proposes to combine multi-beam, 25m footprint LiDAR profiling and L-band RaDAR data. The LiDAR element of this system is deemed more appropriate for vegetation analysis than ICESat/GLAS as footprint diameters aim to maximise the probability of intercepting the canopy top whilst lessening challenges posed by combined ground and vegetation signals within waveforms. This dual sensor platform would provide the opportunity to explore the different absorption properties of intercepted surfaces at microwave and near-infrared frequencies.

The LIST proposal would produce continuous global LiDAR coverage using swath mapping with 5m diameter footprints. This density of elevation profile data and medium footprint size may offer surface evaluation more on a par with airborne techniques and reduce uncertainties associated with assumptions of sampling through LiDAR profiling.

Both DESDynI and LIST are intended for launch before 2020 although system specifications are subject to change. The considerable investment in the development of global vegetation LiDAR systems demonstrates the potential for this, recognised by NASA.

8.3.1 Implications for Carbon Accounting and Monitoring

This research has shown large footprint LiDAR profiling to contribute to reducing the uncertainty in vegetation distribution. This is achieved by systematically sampling the Earth's surface. Height variability is recorded within waveforms which directly represents intercepted surfaces within the illuminated areas. The analysis of waveform energy indices has additionally illustrated the ability for information to be extracted relating to vegetation profile plus canopy properties such as fractional cover. This has demonstrated the contribution offered by satellite LiDAR for improving knowledge of the properties of forest carbon sinks. These capabilities therefore contribute to key research questions identified by the UK Natural Environment Research Council (NERC).

The RMSE for waveform-derived height estimates of approximately 2.8-3.8m may signify that estimates are not sensitive enough to permit year on year growth observation, however longer term periodic quantification and monitoring of changes in carbon stock are feasible. If this is to be applied, Nelson, (2008) notes the importance of consistency in the use of regression models, as considerable discrepancies in estimates may result from using different models alone.

Additionally, this project has demonstrated how large footprint LiDAR offers a suitable means of detecting more significant changes in vegetation height along the

satellite ground track. This may suggest LiDAR profiling as a future additional remote sensing technique which could contribute to global forest resources assessment (FAO, 2007) for the Reduction of Emissions from Deforestation and Forest Degradation in developing countries – REDD (FAO *et al.*, 2008; UNEP, 2009) and assist with reporting on Land Use and Land Use Change in Forestry (IPCC, 2003).

The specifications of future mission proposals will further improve the contributions to carbon accounting and monitoring which may be offered by satellite LiDAR remote sensing.

8.3.2 Model Contributions

Satellite LiDAR data offer a means of providing a third dimension to land cover maps which has previously only been inferred indirectly using optical sensor reflectance properties. Biophysical parameter estimates using footprint-based sampling, such as demonstrated through this project, have the potential to be extrapolated using cartographic classifications of vegetation type and coverage. This approach provides a means of enhancing global landcover maps by attributing spatial variability of biophysical parameters to generalised plant functional types offering more representative input characteristics for Dynamic Vegetation Models (e.g. Sellers *et al.*, 1996a; Sitch *et al.*, 2003; JULES, 2008).

The ability to apply methods of estimating vegetation height from GLAS globally has been shown by Los *et al.*, (submitted) who illustrate correlations with NDVI and use estimates to model the role of land use change on drought severity. The spatial variation in GLAS-derived vegetation height RMSE is likely to vary

considerably, however the work provides proof of concept of the global modelling applications for biophysical parameters derived from this dataset.

The use of radiative transfer modelling has also been used during the course of this research by North *et al.*, (submitted) to explore the sensitivity of GLAS waveforms to optical and structural properties of vegetation. Further application in this field aims to assist the understanding and interpretation of large footprint LiDAR waveforms and the uncertainty among estimates.

8.3.3 Forestry Applications

For forest operations and management, conventional methods rely upon aggregated statistical data gathered within numerous field plots to produce mean stand-level parameters. ICESat/GLAS footprint distribution may be regarded as effectively replicating this sampling technique which is used both at stand-level and at national scale for forest inventory.

Where footprints are coincident with stands, this work has shown large footprint LiDAR profiling to offer equivalent stand-mean information for vegetation top height and volume predicted from independently-derived yield models plus a first step to investigating canopy cover. These estimates were additionally found to be comparable with those from airborne LiDAR data, the use of which is supported by the work of many other authors for local scale studies (e.g. Lefsky *et al.*, 1999b; Hyyppä *et al.*, 2001; Parker *et al.*, 2001; Næsset, 2002; Lefsky *et al.*, 2005).

The sparse sampling provided by the ICESat mission currently does not offer a reliable substitute for intensive and expensive stand-level field surveys for forest management. Nevertheless, this research has shown satellite waveform estimates to

complement field-acquired data, indicating the potential of future proposed sensors providing greater footprint density, such as LIST.

Additionally, methods developed during this research of estimating mean stand-level vegetation biophysical parameters have the potential to provide input parameters for data assimilation within processed-based models of stand conditions. The ability to estimate top height has been demonstrated within this thesis whilst further potential is suggested by airborne LiDAR results which indirectly estimate DBH from LiDAR percentiles. This may permit observed parameters to be applied nationally to offer inputs to models such as to assess wind-throw risk (Gardiner *et al.*, 2004; Suárez *et al.*, 2008b). Estimates may furthermore assist with production forecasting as the ability to determine top height may allow the yield class assigned to forest stands to be validated.

Large footprint LiDAR profiling offers a means of complementing national forest inventories by supplying data sampling of forests at a density which may not be cost or time-effective using field surveys. Additionally, the seasonal repetition of ground tracks may allow more frequent updating of inventories than would be possible by others means. Fundamentally, LiDAR profiling offers a means of reducing uncertainty for private woodland within the UK national forest inventory for which access is often not permitted nor data available for monitoring purposes (Forestry_Commission, 2003; Broadmeadow and Matthews, 2004; Forestry_Commission, 2007).

The benefits of LiDAR data for forest inventory are beginning to be implemented in several countries. In particular, Norway, Finland and Ehime Prefecture Japan have begun to investigate discrete return airborne LiDAR systems for this purpose (e.g. Ene *et al.*, 2007; Maltamo *et al.*, 2007; Tsuzuki *et al.*, 2008).

However, flight campaigns of such large extent are costly. The freely-available satellite LiDAR data may offer a suitable option for nation-wide vegetation quantification.

8.4 Conclusion

In the absence of spaceborne missions designed specifically for vegetation analysis, the innovative GLAS sensor has provided a valuable opportunity to assess the potential of full waveform, satellite LiDAR profiling for vegetation applications.

This thesis has developed and evaluated methods for identifying the signal returned by vegetation and biophysical parameter retrieval from GLAS waveforms. Vegetation heights estimated from GLAS waveforms corresponded well with field measurements taken at coincident areas at the Forest of Dean, Gloucestershire, UK. A relationship of $R^2 = 0.92$, RMSE = 2.81m was found.

Heights of cumulative energy percentiles within the waveform canopy return were used as estimators of key forestry parameters calculated from Forestry Commission yield models, in particular Top Height and Stemwood Volume. Top Height estimates produced $R^2 = 0.76$, RMSE = 3.9m and may permit adjustment or verification of yield model predictions. Stemwood volume estimates were improved through differentiation using broad vegetation classes which reflect interpretive forest type classes used within the National Forest Inventory in Britain. The stemwood volume estimates distinguished trees solely according to whether stands contained conifer or broadleaf species. Highest correlations of stemwood volume estimates for the tallest species within stands showed $R^2 = 0.75$, RMSE = 59.1m^3ha^{-1} for broadleaves and $R^2 = 0.63$, RMSE = 90.3m^3ha^{-1} for conifers. Taking account of

mixed species composition produced $R^2 = 0.47$, RMSE = 75.6m^3ha^{-1} and $R^2 = 0.66$, RMSE = 82.5m^3ha^{-1} for stands containing predominantly broadleaf and coniferous species respectively. Further stratification of analysis using more refined species-type segmentation may allow regression models with still higher correlations to be developed.

Estimates of canopy cover from waveforms may offer a means of determining regeneration potential of forest stands. Multiple linear regression using waveform-derived canopy cover estimates and vegetation height produced R^2 0.63, 11% RMSE for broadleaf-dominated stands and R^2 0.41, 16% RMSE for those containing mostly conifers.

These results have demonstrated the unique potential for national or global scale biophysical parameter estimation using full waveform, satellite LiDAR profiling. Such data have filled a gap in previous research capabilities through provision of a third dimension to remotely sensed data. Prospects for applications in carbon accounting, model improvement, forest management and inventory are supported by the strong relationships found between estimates developed using GLAS waveforms with coincident field measurements, yield model predictions and comparisons with airborne laser scanning data. More advanced satellite LiDAR systems designed specifically for vegetation applications will further exploit the potential for this field of research.

References

ABSHIRE, J., SUN, X., RIRIS, H., SIROTA, J., MCGARRY, J., 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**: L21S02.

AIRBORNE_LASER_MAPPING, date unknown, A Reference Source on an Emerging Technology. [online] available from http://www.airbornelasermapping.com accessed

ARMSTON, J.D., DENHAM, R., DANAHER, T., MOFFIET, T. and SCARTH, P., in preparation 2008, Prediction and Validation of Foliage Projective Cover from LandSat for Queensland, Australia. *Photogrammetric Engineering and Remote Sensing*

ARSF, 2005. Airborne Research and Survey Facility, Available online from http://arsf.nerc.ac.uk/instruments/altm.asp. Last accessed October 2007

AXELSSON, P., 1999, Processing of Laser Scanning Data - Algorithms and Applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**: 138-147.

BAE, S. and SCHUTZ, B., 2002. Precision Attitude Determination (PAD); Algorithm Theoretical Basis Document Version 2.2. NASA Goddard Space Flight Center.

BALTSAVIAS, E., 2007. Introduction to airborne laser scanning, Seminar presentation: The world in 3D: Applications of LiDAR Remote Sensing for the management of natural resources and infrastructure (translation), Cuenca, Spain; June 2007.

BALTSAVIAS, E.P., 1999, Airborne Laser Scanning - Basic Relations and Formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**: 199-214.

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**: 115-122.

BOYD, D.S. and HILL, R.A., 2007, Validation of Airborne Lidar Intensity Values from a Forested Landscape Using HYMAP data: Preliminary Analyses. In: P.

Rönnholm, H. Hyyppä and J. Hyyppä (Editors), Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Espoo, Finland

BREIDENBACH, J., KUBLIN, E., MCGAUGHEY, R., ANDERSON, H.-E. and REUTEBUCH, S., 2008, Mixed-effects models for estimating stand volume by means of small footprint airborne laser scanner data. *Photogrammetric Journal of Finland*, **21**(1)

BRENNER, A., ZWALLY, H., BENTLEY, C., CSATHO', B., HARDING, D., HOFTON, M., MINSTER, J.-B., ROBERTS, L., SABA, J., THOMAS, R. and YI, D., 2003. Derivation of Range and Range Distributions from Laser Pulse Waveform Analysis for Surface Elevations, Roughness, Slope and Vegetation Heights, Algorithm Theoretical Basis Document Version 4.1. NASA Goddard Space Flight Center.

BROADMEADOW, M. and MATTHEWS, R., 2003. Forests, Carbon and Climate Change: the UK Contribution. Information Note. Forest Research, Edinburgh, UK.

BROADMEADOW, M. and MATTHEWS, R., 2004. Survey Methods for Kyoto Protocol Monitoring and Verification of UK Forest Carbon Stocks, Forest Research, Alice Holt, Surrey, UK [online] available from: http://www.nbu.ac.uk/ukcarbon/docs/DEFRA_Report_2004_Section8.pdf accessed May 2009

BUFTON, J., 1989, Laser Altimetry Measurements from Aircraft and Spacecraft. *Proceedings of the IEEE*, **77**(3): 463-477.

CARABAJAL, C. and HARDING, D., 2001, Evaluation of Geoscience Laser Altimeter System (GLAS) Waveforms for Vegetated Landscapes using Airborne Laser Scanning Data. *International Archives of Photogrammetry and Remote Sensing*, **XXXIV**(October 3/W4): 125-128.

CARABAJAL, C., HARDING, D.H., LUTHCKE, S.B., FONG, W., ROWTON, S.C. and FRAWLEY, J.J., 1999. SLA-02 Data Products Distribution. NASA.

CHARLTON, M.E., LARGE, A.R.G. and FULLER, I.C., 2003, Application of Airborne Lidar in River Environments: The River Coquet, Northumberland, UK. *Earth Surface Processes and Landforms*, **28**: 299-306.

CHEN, X., VIERLING, L., ROWELL, E. and THOMAS, D.F., 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**: 14-26.

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**: 121-138.

CROW, P., BENHAM, S., DEVEREUX, B.J. and AMABLE, G.S., 2007, Woodland Vegetation and its implications for archaeological survey using LiDAR. *Forestry*, **80**(3): 241-252.

DENNING_RESEARCH_GROUP, 2005. Simple Biosphere Model (SiB); Biocycle_Website.[online] available from http://biocycle.atmos.colostate.edu/html/s

DESSLER, A.E., PALM, S.P. and SPINHIRNE, J.D., 2006, Tropical Cloud-top Height Distributions Revealed by the Ice, Cloud, and Land Elevation Satellite (ICESat)/Geoscience Laser Altimeter System (GLAS). *Journal of Geophysical Research*, **111**(D12215): doi:10.1029/2005JD006705.

DOCE, D.D., SUÁREZ, J.C. and PATENAUDE, G., 2008. Assessing stand and data variability using airborne laser scanner. In: L. Bernard (Editor), Lecture Notes in Geoinformation and Cartography, AGILE 2008, Girona, Spain.

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.D., 2002, Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, **79**: 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**: 147-159.

DUBAYAH, R., 2008. The use of large footprint waveform lidar for landscape characterization: past experience and future prospects. In: R.A. Hill, J. Rosette and J. Suárez (Editors), Proceedings of Silvilaser 2008: *8th International conference on LiDAR applications in forest assessment and inventory*. ISBN: 978-0-85538-774-7, Edinburgh, UK.

EARTH_SCIENCE_OFFICE, 2005. Sparcle Tutorial, [online] available from: http://www.ghcc.msfc.nasa.gov/sparcle/sparcle_tutorial.html Date accessed: January 2005

EARTH_SCIENCE_OFFICE, date unknown. Sparcle Tutorial, [online] available from: http://www.ghcc.msfc.nasa.gov/sparcle/sparcle_tutorial.html Date accessed: January 2005

EDWARDS, P.N. and CHRISTIE, J.M., 1981. Yield Models for Forest Management; Booklet 48. The Forestry Commission, Edinburgh.

ENE, L., NÆSSET, E. and GOBAKKEN, T., 2007, Simulating Sampling Efficiency in Airborne Based Laser Scanning Forest Inventory. In: P. Rönnholm, H. Hyyppä and J. Hyyppä (Editors), Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Espoo, Finland

FANG, J.-Y., WANG, G.G., LIU, G.-H. and XU, S.-L., 1998, Forest Biomass of China; An estimate based on Biomass-Volume Relationship. *Ecological Applications*, **8**(4): 1084-1091.

FAO, 2007. Global Forest Resources Assessment 2010; Options and Recommendations for a Global Remote Sensing Survey of Forests. Forestry Department, Food and Agriculture Organization of the United Nations, [online] available from: ftp://ftp.fao.org/docrep/fao/010/ai074e/ai074e00.pdf accessed May 2009.

FAO, UNDP and UNEP, 2008. United Nations Collaborative Programme on Reducing Emissions from Deforestation in Developing Countries. [online] available from: http://www.undp.org/mdtf/UN-REDD/docs/Annex-A-Framework-Document.pdf accessed May 2009.

FEDERATION_OF_AMERICAN_SCIENTISTS, year unknown, Altimetry. Available online at: www.fas.org Accessed 4th January 2004

FOREST_OF_DEAN_PARTNERSHIP, 2006. Royal Forest of Dean; History, Available online at: http://www.visitforestofdean.co.uk/history.htm (accessed February 2006)

FOREST_RESEARCH, 2006. UK Forest Carbon Inventory, Available online at: http://www.forestresearch.gov.uk/website/forestresearch.nsf/ByUnique/INFD-62XH5R (accessed June 2006)

FOREST_RESEARCH, 2009. The CARBINE accounting model. [online] available from http://www.forestresearch.gov.uk/fr/INFD-633DXB accessed May 2009.

FOREST_RESEARCH, year unknown. Knowledge, Innovation, Sustainability. Forest Research; Research Agency of the Forestry Commission, Edinburgh.

FORESTRY_COMMISSION, 2003. National Inventory of Woodland and Trees - Great Britain. Forestry Commission, Edinburgh, UK, ISBN: 0 85538 602 9.

FORESTRY_COMMISSION, 2006. Operational Guidance Booklet 6, Survey Handbook. The Forestry Commission, Edinburgh.

FORESTRY_COMMISSION, 2007. Forestry Facts and Figure 2007; A summary of statistics about woodland and forestry. Economics and Statistics, Forestry Commission, Edinburgh, UK, ISBN: 978-0-85538-742-6.

FRAWLEY, GARVIN, HARDING, BLAIR and BUFTON, 2005. Shuttle Laser Altimeter II: First Science Results. [online] available from http://dtam.gsfc.nasa.gov accessed 14th January 2005

FREE_SOFTWARE_FOUNDATION, 2006. Gawk: Effective AWK Programming. [online] available at: http://www.gnu.org/software/gawk/manual/Date accessed October 2005

GARDINER, B.A., SUÁREZ, J.C., ACHIM, A., HALE, S. and NICHOLL, B., 2004. ForestGALES 2.0; A PC-based wind risk model for British Forests. Forestry Commission Publications. ISBN 0855386320, Edinburgh, pp. 60.

GARDNER, C., 1992, Ranging Performance of Satellite Laser Altimeters. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(5): 1061-1072.

GAULTON, R. and MALTHUS, T.J., 2008. LiDAR mapping of canopy gaps in continuous cover forests; a comparison of canopy height model and point cloud based techniques. In: R.A. Hill, J. Rosette and J. Suárez (Editors), Proceedings of *Silvilaser 2008: 8th International conference on LiDAR applications in forest assessment and inventory*, Edinburgh, UK, pp. 407-416. ISBN: 978-0-85538-774-7.

GEOLAS_CONSULTING, date unknown, Imaging laser Altimetry. [online] available from: http://www.geolas.com accessed January 2005

GOODWIN, N.R., COOPS, N.C., BATER, C. and GERGEL, S.E., 2007, Assessment of sub-canopy structure in a complex coniferous forest. *IAPRS*, **XXXVI** (3 / W52)

GOODWIN, N.R., COOPS, N.C. and CULVENOR, D.S., 2006, Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, **103**(2006): 140-152.

HARDING, D., BUFTON, J. and FRAWLEY, J., 1994, Satellite Laser Altimetry of Terrestrial Topography: Vertical Accuracy as a Function of Surface Slope, Roughness and Cloud Cover. *IEEE Transactions on Geoscience and Remote Sensing*, **32**(2, March): 329-339.

HARDING, D. and CARABAJAL, C., 2003, ICESat Observations of Forest Canopy Height, American Geophysical Union Fall meeting

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

HARDING, D.J., BLAIR, J.B., RABINE, D.R. and STILL, K., 1998. SLICER: Scanning Lidar Imager of Canopies by Echo Recovery Instrument and Product Description. NASA; June 1998.

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**: 283-297.

HEMIVIEW2.1, 1999. HemiView Online Help, © 1999 Delta-T Devices Ltd. [online] available from: www.delta-t.co.uk accessed April 2008

HERRING, T. and QUINN, K., 2001. Atmospheric Delay Correction to GLAS Laser Altimeter Ranges; Algorithm Theoretical Basis Document Version 2.1. NASA Goddard Space Flight Center.

HESE, S., LUCHT, W., SCHMULLIUS, C., BARNSLEY, M., DUBAYAH, R., KNORR, D., NEUMANN, K., RIEDEL, T. and SCHROTER, K., 2005, Global biomass mapping for an improved understanding of the CO2 balance - the Earth observation mission Carbon-3D. *Remote Sensing of Environment*, **94**: 94-104.

HILL, R. and SMITH, G.M., 2005, Land cover heterogeneity in Great Britain as identified in Land Cover Map 2000. *International Journal of Remote Sensing*, **26**(24): 5467-5473.

HILL, R. and THOMSON, A.G., 2005, Mapping woodland species composition and struture using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, **26**(17): 3763-3779.

HILL, R.A., 2007, Going Undercover: Mapping Woodland Understorey from Leaf-On and Leaf-Off Lidar Data. In: P. Rönnholm, H. Hyyppä and J. Hyyppä (Editors), Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Espoo, Finland

HINSLEY, S.A., HILL, R., 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**: 1399-1406.

HINSLEY, S.A., HILL, R., GAVEAU, D.L.A. and BELLAMY, P.E., 2002, Quantifying woodland structure and habitat quality for birds using airborne laser scanning. *Functional Ecology*, **16**(6): 851-857.

HOFTON, M.A. and BLAIR, J.B., 2002, Laser Altimeter Pulse Correlation: a Method for Detecting Surface Topographic Change. *Journal of Geodynamics*, **34**: 477-489.

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] available from: http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_contents.html accessed May 2009.

ITTVIS, 2008. ITT Visual Information Solutions IDL Virtual Machine, [online]: http://www.ittvis.com/ProductServices/IDL/VirtualMachine.aspx Accessed July 2008

JANSMA, P., MATTIOTI, G. and MATIAS, A., 2001, Slicer Laser Altimetry in the Eastern Caribbean. *Surveys in Geophysics*, **22**: 561-579.

JENKINS, J., CHOJNACKY, D., HEATH, L. and BIRDSEY, R., 2003, National-Scale Biomass Estimators for United States Tree Species. *Forest Science*, **49**(1): 12-35.

JULES, 2008, Joint UK Land Environment Simulator; Science in JULES - Vegetation, [online] available from http://www.jchmr.org/jules/science/vegetation.html accessed July 2006

KICHAK, R., 2003. Chairman's Executive Summary. Independent GLAS Anomaly Review Board, [online] available from: http://icesat.gsfc.nasa.gov/docs/IGARB.pdf accessed July 2008.

KOTCHENOVA, S., SHABANOV, N., KNYAZIKHIN, Y., DAVIS, A., DUBAYAH, R. and MYNENI, R., 2003, Modeling lidar waveforms with time-dependent stochastic radiative transfer theory for remote estimations of forest structure. *Journal of Geophysical Research*, **108**(D15 ACL 12): 1-13.

LARSEN, D., HANN, D. and STEARNS-SMITH, S., 1987, Accuracy and precision of the Tangent Method of Measuring Tree Height. *Western Journal of Applied Forestry*, **2**(1): 26-28.

LEFSKY, M., COHEN, W., HARDING, D., PARKER, G., ACKER, S. and GOWER, S., 2002, Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography*, **11**: 393-399.

LEFSKY, M., HARDING, D., KELLER, M., COHEN, W., CARABAJAL, C., DEL BOM ESPIRITO-SANTO, F., HUNTER, M., DE OLIVEIRA JR., R. and DE CAMARGO, P., 2005, Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, **32** L22S02.

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., HARDING, D., 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**: 83-98.

LEFSKY, M.A., KELLER, M., PANG, Y., DE CAMARGO, P. and HUNTER, M.O., 2007, Revised method for forest canopy height estimation from the Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, **1**: 1-18.

LIM, K., TREITZ, P., BALDWIN, K., MORRISON, I. and GREEN, J., 2003a, Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, **29**(5): 658-678.

LIM, K., TREITZ, P., WULDER, M., ST-ONGE, B. and FLOOD, M., 2003b, LiDAR remote sensing of forest structure. *Progress in Physical Geography*, **27**(1): 88-106.

LOS, S.O., ROSETTE, J.A. and NORTH, P.R.J., submitted, Observational evidence for linkages between increased drought severity and land-cover change. *Geophysical Research Letters*

LUTHCKE, S., CARABAJAL, C. and ROWLANDS, D., 2002a, 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**: 447-475.

LUTHCKE, S.B., EKHOLM, S. and B., B.J., 2002b, Introduction. *Journal of Geodynamics*, **34**: 343-345.

MAGIC, 2007. GIS Digital Boundary Datasets; Forestry Commission Woodland, [online] available from: http://www.magic.gov.uk/staticmaps/national.htm accessed January 2007

MALTAMO, M., PACKALÉN, P., PEUHKURINEN, J., SUVANTO, A., PESONEN, A. and HYYPPÄ, J., 2007, Experiences and Possibilities of ALS Based Forest Inventory in Finland. In: P. Rönnholm, H. Hyyppä and J. Hyyppä (Editors), Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Espoo, Finland

MATTHEWS, R. and MACKIE, E., 2006, Forest Mensuration; A handbook for practitioners. Forestry Commission, Edinburgh, UK. ISBN: 0 85538 621 5.

MOLERO, F. and JACQUE, F., 1999, The laser as a tool in environmental problems. *Optical Materials*, **13**: 167--173.

MORSDORF, F., MEIER, E., KOTZ, B., ITTEN, K.I., DOBBERTIN, M. and ALLGOWER, 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**: 353-362.

MORSDORF, F., NICHOL, C., MALTHUS, T., PATENAUDE, G. and WOODHOUSE, I., 2008, Modelling multi-spectral LIDAR vegetation backscatter – assessing structural and physiological information content. In: R.A. Hill, J. Rosette and J. Suárez (Editors), Proceedings of *Silvilaser 2008: 8th International*

conference on LiDAR applications in forest assessment and inventory, Edinburgh, UK. ISBN: 978-0-85538-774-7, pp. 257-265.

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 November)

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**: 88-99.

NELSON, R., 2008, Model Effects on GLAS-based Regional Estimates of Forest Biomass and Carbon. In: R.A. Hill, J. Rosette and J. Suárez (Editors), Proceedings of *Silvilaser 2008: 8th International conference on LiDAR applications in forest assessment and inventory*, Edinburgh, UK, pp. 207-215.

NELSON, R., NÆSSET, E., GOBAKKEN, T., STÅHL, G. and GREGOIRE, T., 2008a, Regional Forest Inventory using an Airborne Profiling LiDAR. *Journal of Forest Planning*, **13**(Special Issue 'Silvilaser'): 287-294.

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**: 500-511.

NELSON, R.F., NÆSSET, E., GOBAKKEN, T., STÅHL, G. and GREGOIRE., T., 2008b, Regional Forest Inventory Using an Airborne Profiling LiDAR. *Journal of Forest Planning*, **13**: 287-294.

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., 1996, Three-Dimensional Forest Light Interaction Model Using a Monte Carlo Method. *IEEE Transations on Geoscience and Remote Sensing*, **34**(4 July): 946-956.

NORTH, P.R.J., 2002, Esimation of fAPAR, LAI and vegetation fractional cover from ATSR-2 imagery. *Remote Sensing of Environment*, **80**: 114-121.

NORTH, P.R.J., ROSETTE, J.A., SUÁREZ, J. and LOS, S.O., submitted, A Monte Carlo radiative transfer model of satellite waveform lidar. *International Journal of Remote Sensing* (Special Issue SilviLaser 2008)

NSIDC, 2003. ICESat/ GLAS Data at NSIDC. National Snow and Ice Data Center, [online] available from http://nsidc.org/daac/icesat/.

NSIDC, 2008. ICESat Laser Operation Period Attributes.[online] available from: http://nsidc.org/data/icesat/data.html accessed June 2008

ORDNANCE_SURVEY, 2006. OS Land-Form PROFILE DTM 1:10,000, © Crown copyright/ database right 2006. An Ordnance Survey/EDINA supplied service

ORDNANCE_SURVEY, date unknown. Ordnance survey glossary; Mapping Terminology and Acronyms, [online] available from: www.ordnancesurvey.co.uk/oswebsite/aboutus/report/misc/glossary.html accessed January 2005

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**: 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**: 368-380.

PHILLIPS, H.A., RIDGEWAY, J.R., MINSTER, J.-B., YI, D. and BENTLEY, C., 1999. Tidal Corrections; Algorithm Theoretical Basis Document Version 2.0. NASA Goddard Space Flight Center.

RIANO, D., VALLADARES, F., CONDES, S. and CHUVIECO, E., 2004, Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology*, **124**: 269-275.

RIM, H.J. and SCHUTZ, B.E., 2002. Precision Orbit Determination (POD); Algorithm Theoretical Basis Document Version 2.2. NASA Goddard Space Flight Center.

ROSETTE, J., NORTH, P. and SUÁREZ, J., 2008a, Stemwood Volume Estimates for a Mixed Temperate Forest using Satellite LiDAR. *Journal of Forest Planning*, **13**(Special Issue 'Silvilaser'): 205-214.

ROSETTE, J.A., NORTH, P.R.J. and SUÁREZ, J.C., 2008b, Satellite LiDAR Estimation of Stemwood Volume; A Method using Waveform Decomposition. *Photogrammetric Journal of Finland*, **21**(1): 76-85.

ROSETTE, J.A., NORTH, P.R.J., SUÁREZ, J.C. and ARMSTON, J.D., accepted for publication, A Comparison of Biophysical Parameter Retrieval for Forestry using Airborne and Satellite LiDAR. *International Journal of Remote Sensing*, **Special Issue ForestSat 2007**

ROSETTE, J.A., NORTH, P.R.J., SUÁREZ, J.C. and LOS, S.O., submitted, Uncertainty within Satellite LiDAR Estimations of Vegetation and Topography. *International Journal of Remote Sensing*, **Silvilaser 2008 Special Issue**

ROSETTE, J.A.B., NORTH, P.R.J. and SUÁREZ, J.C., 2008c, Vegetation Height Estimates for a Mixed Temperate Forest using Satellite Laser Altimetry. *International Journal of Remote Sensing*, **29**(5): 1475-1493.

SCHUTZ, B., 2001. GLAS Altimeter Post-Launch Calibration/Validation Plan v1. Center for Space Research University of Texas at Austin.

SCHUTZ, B., ZWALLY, H., SHUMAN, C., HANCOCK, D. and DIMARZIO, J., 2005, Overview of the ICESat Mission. *Geophysical Research Letters*, **32**: L21S01.

SCHUTZ, B.E., 2002. Laser Footprint Location (Geolocation) and Surface Profiles, Algorithm Theoretical Basis Document Version 3.0. NASA Goddard Space Flight Center.

SELLERS, P.J., LOS, S.O., TUCKER, C.J., JUSTICE, C.O., DAZLICH, D.A., COLLATZ, G.J. and RANDALL, D.A., 1996a, A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part II: The Generation of Global Fields of Terrestrial Biophysical Parameters from Satellite Data. *Journal of Climate*, **9**: 706-737.

SELLERS, P.J., RANDALL, D.A., COLLATZ, G.J., BERRY, J.A.F., C.B., DAZLICH, D.A., ZHANG, C., COLLELO, G.D. and BOUNOUA, L., 1996b, A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs; Part I: Model Formulation. *Journal of Climate*, **9**: 676-705.

SHIRAH, G. and KEKESI, A., 2003. ICESat Lithograph, NASA/Goddard Space Flight Center Scientific Visualization Studio. [online] available from: svs.gsfc.nasa.gov accessed May 2005

SHUMAN, C.A., ZWALLY, H.J., SCHUTZ, B.E., BRENNER, A.C., DIMARZIO, J.P., SUCHDEO, V.P. and FRICKER, H., 2005, ICESat Antarctic Elevation Data: Preliminary Precision and Accuracy Assessment. *Geophysical Research Letters*, **33**: L07501.

SITCH, S., SMITH, B., PRENTICE, I.C., ARNETH, A., BONDEAU, A., CRAMER, W., KAPLAN, J.O., LEVIS, S., LUCHT, W., SYKES, M.T., THONICKE, K. and VENEVSKY, S., 2003, Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dydnamic global vegetation model. *Global Change Biology*, **9**: 161-185.

SKINNER, L., 2002, Estimating forest parameters from synthetic aperture radar: a case study of Thetford Forest, Swansea University.

SPINHIRNE, J., HART, W. and WELTON, J., 2003. Atmospheric and Surface Reflectance Measurements by the Geoscience Laser Altimeter System, American Geophysical Union Fall meeting.

STREUTKERS, D., 2008. ENVI LiDAR Tools, Available online: http://geology.isu.edu/BCAL/tools/EnviTools/

SUÁREZ, J., ONTIVEROS, C., SMITH, S. and SNAPE, S., 2005a, Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. *Computers and Geosciences*, **31**(2): 253-262.

SUÁREZ, J., ROSETTE, J., NICHOLL, B. and GARDINER, B., 2008a, A practical application of airborne LiDAR for forestry management in Scotland. In: R.A. Hill, J. Rosette and J. Suárez (Editors), Proceedings of *Silvilaser 2008: 8th International conference on LiDAR applications in forest assessment and inventory*, Edinburgh, UK. ISBN: 978-0-85538-774-7, pp. 581-585.

SUÁREZ, J.C., GARCIA, R., GARDINER, B. and PATENAUDE, G., 2008b, The Estimation of Wind Risk in Forests Stands using Airborne Laser Scanning (ALS). *Journal of Forest Planning*, **13**: 165-185.

SUÁREZ, J.C., SMITH, S., BULL, G., MALTHUS, T.J., DONOGHUE, D. and KNOX, D., 2005b, The use of remote sensing techniques in operational forestry. *Quarterly Journal of Forestry* **99**(1): 31-42.

SUN, G., RANSON, K.J., MASEK, J., GUO, Z., PANG, Y., FU, A. and WANG, D., 2008, Estimation of Tree Height and Forest Biomass from GLAS Data. *Journal of Forest Planning*, **13**(Special Issue 'Silvilaser'): 157-164.

TAKAHASHI, T., AWAYA, Y., HIRATA, Y., FURUYA, N., SAKAI, T. and SAKAI, A., 2008, Effects of Flight Altitude on LiDAR-Derived Tree Heights in Mountainous Forests with Poor Laser Penetration Rates. *Photogrammetric Journal of Finland*, **21**(1)

TODD, K.W., 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.

TSUZUKI, H., NELSON, R. and SWEDA, T., 2008, Estimating Timber Stock Volume of Ehime Prefecture, Japan using Airborne Laser Profiling. *Journal of Forest Planning*, **13**(Special Issue 'Silvilaser')

ULM, year unknown. Unit for Landscape Modelling LiDAR Survey, [online] available from: http://www.uflm.cam.ac.uk/lidar.htm accessed October 2007

UNEP, 2009. Reducing Emissions from Deforestation and Degradation. [online] available from: http://www.unep.org/climatechange/UNEPsWork/REDD/tabid/245/language/en-US/Default.aspx accessed May 2009.

UNFCCC, 2007. Kyoto Protocol. [online] available from : http://unfccc.int/kyoto_protocol/items/2830.php accessed April 2007.

WEHR, A. and LOHR, U., 1999, Airborne Laser Scanning - An Introduction and Overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**: 68-82.

WEHR, A., LOHR, U. and BALTSAVIAS, E., 1999, Editorial: Theme Issue on Airborne Laser Scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**: 61-63.

YANG, Y., MARSHAK, A., CHIU, J.C., WISCOMBE, W.J., PALM, S.P., DAVIS, A.B., SPANGENBERG, D.A., NGUYEN, L., SPINHIRNE, J.D. and MINNIS, P., 2008, Retrievals of Cloud Optical Depth from the Geoscience Laser Altimeter System (GLAS) by Calibration of Solar Background Signal. *Journal of the Atmospheric Sciences*, **doi: 10.1175/2008JAS2744.1.**

ZHANG, K., CHEN, S., WHITMAN, D., SHYU, M., YAN, J. and ZHANG, C., 2003, A Progressive Morphological Filter for Removing Nonground Measurements from Airborne Lidar Data. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(4)

ZWALLY, H., SCHUTZ, B., ABDALATI, W., ABSHIRE, J., BENTLEY, C., BRENNER, A., BUFTON, J., DEZIO, J., HANCOCK, D., HARDING, D., HERRING, T., MINSTER, B., QUINN, K., PALM, S., SPINHIRNE, J. and THOMAS, R., 2002, ICESat's Laser Measurements of Polar Ice, Atmosphere, Ocean and Land. *Journal of Geodynamics*, **34**: 405-445.