

# Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data

**Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben,  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op dinsdag 8 juni 2010 om 12.30 uur  
door

**Hieu van Duong**

Master of Engineering, Space Technology and Remote Sensing (STARS),  
Asian Institute of Technology, Thailand

geboren te Ho Chi Minh, Vietnam

Dit proefschrift is goedgekeurd door de promotoren:  
Prof. dr.ir. P.J.G. Teunissen  
Prof. dr. ir. M.G. Vosselman

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. M.G. Vosselman	University of Twente, promotor
Prof. dr. ir. P.J.G. Teunissen	Technische Universiteit Delft, promotor
Dr. R.C. Lindenberg	Technische Universiteit Delft, copromotor
Univ.Prof. Dipl.Ing. Dr.techn. N. Pfeifer	Vienna University of Technology
Prof. Dr. M. Menenti	Technische Universiteit Delft
Dr.Ing. B. Jutzi	Universitaet Karlsruhe
Dr. M.A. Lefsky	Colorado State University
Prof. Dr. D.G. Simons	Technische Universiteit Delft

Copyright © 2010 by Hieu van Duong.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

ISBN 978-90-9025478-4

Typeset by the author with the L<sup>A</sup>T<sub>E</sub>X Documentation System.  
Author email: duongvanhieu@yahoo.com or hieu.duong@colostate.edu

*In loving memory of my parents, I wish they could know my success!  
Gui den Ba va Mo yeu quy cua con!*



# Acknowledgements

The work presented in this thesis was done in collaboration with and with great support of many people. It is my pleasure to acknowledge their substantial contributions.

First of all, I would like to thank my daily supervisor Dr. Roderik Lindenberg. My PhD would not be done if I did not have great support from him. I have learned a lot from him in setting up and organizing articles, research plans and even on my Dutch life.

I also would like to thank Prof. Norbert Pfeifer for his great and substantial ideas for my research. I would like to thank his parents that gave me the opportunity to stay in their lovely house during my 3-month visit to Vienna to study airborne full waveform data.

I never forget to thank Prof. George Vosselman for his sharp and top comments during my research. I have gained a lot of knowledge from his remarks that definitely contributed to the quality of my work.

I would like to express my appreciation to Prof. Peter Teunissen who gave me the chance to join a high rank university, Delft University of Technology, and provided me with a nice working environment, with friendly groups and a friendly university. This was a chance to change my life too.

Moreover, I would like to thank Prof. Massimo Menenti for his qualified remarks that helped me to make my thesis well-structured and organized.

I would like to send my sincere thanks to Lidwien for many administrative support during my study, my colleagues for interesting discussions and my friends for helping me through many years.

Finally, I deeply would like to thank my 'hidden' beloved for their unforgettable support during my study and my life.



# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research context	1
1.1.1 Discrete airborne laser scanning systems	1
1.1.2 Multiple return echoes	1
1.1.3 Full waveform systems	2
1.1.4 ICESat large footprint full waveforms	5
1.1.5 Research questions	8
1.2 Scope of research	9
1.3 Research methodology	9
1.3.1 Characteristics of ICESat full waveform data	9
1.3.2 Waveform processing	11
1.3.3 Validation studies	13
1.3.4 ICESat applications	13
1.4 Outline of the thesis	14
<b>2 Sensors and instruments</b>	<b>15</b>
2.1 Introduction	15
2.2 Principal of operation of a laser scanning system	15
2.3 Development of full waveform scanning systems	18
2.3.1 Scanning Lidar Imager of Canopies by Echo Recovery (SLICER)	20
2.3.2 Laser Vegetation Imaging Sensor (LVIS)	20

2.3.3	A typical commercial airborne full waveform scanning system (LMS-Q560)	23
2.4	ICESat/GLAS spaceborne laser scanning system	25
2.4.1	ICESat/GLAS Science objectives	25
2.4.2	Overview on instruments/missions	26
2.4.3	ICESat track coverage	28
2.4.4	ICESat data products and processing releases	29
2.5	Conclusion	33
<b>3</b>	<b>Full waveform processing</b>	<b>35</b>
3.1	Introduction	35
3.2	Interaction of the laser pulse with the surface	35
3.3	Gaussian decomposition	39
3.3.1	Mathematical model of transmitted pulse and return full waveform	39
3.3.2	Count to volt conversion	40
3.3.3	Waveform normalisation	40
3.3.4	Real waveform signal	42
3.3.5	Smoothing the waveform	43
3.3.6	Initial parameter estimation	45
3.3.7	Fitting algorithm	45
3.3.8	Quality of fit	50
3.4	Waveform deconvolution	52
3.5	Waveform simulation	55
3.6	List of waveform parameters used	58
3.7	Conclusion	61
<b>4</b>	<b>Validation of large footprint full waveform processing</b>	<b>63</b>
4.1	Introduction	63
4.2	Related work on ICESat data calibration and validation	64
4.2.1	Definition of calibration and validation	64
4.2.2	Error sources	64
4.2.3	Current calibration/validation approaches	67
4.3	Case study 1: ICESat vs. AHN over The Netherlands	72
4.3.1	Background	72
4.3.2	Study Area and Data	73
4.3.3	Methodology	77
4.3.4	Results and comparison	87
4.3.5	Conclusions and further research	94
4.4	Case study 2: overlapping European ICESat footprints	96
4.4.1	Background	96
4.4.2	Study area and input data	97
4.4.3	Methodology	98
4.4.4	Results and discussion	100

4.4.5	Case study: coinciding waveforms . . . . .	102
4.4.6	Conclusion and further research . . . . .	104
4.5	Conclusion . . . . .	104
<b>5</b>	<b>Applications of large footprint full waveform data</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.2	Literature review . . . . .	108
5.2.1	Elevation applications . . . . .	108
5.2.2	Waveform applications . . . . .	111
5.3	Case study 1: Single & two epoch over forested areas . . . . .	113
5.3.1	Background . . . . .	113
5.3.2	Input forest data . . . . .	115
5.3.3	Methodology . . . . .	119
5.3.4	Results and discussion . . . . .	124
5.3.5	Recognised problems . . . . .	128
5.3.6	Conclusions and further research . . . . .	129
5.4	Case study 2: ICESat data for land cover classification . . . . .	131
5.4.1	Background . . . . .	131
5.4.2	Study area and data input . . . . .	132
5.4.3	Methodology . . . . .	134
5.4.4	Classification results and discussion . . . . .	137
5.4.5	Conclusion and further research . . . . .	141
5.5	Conclusion . . . . .	142
<b>6</b>	<b>Conclusion and future work</b>	<b>145</b>
6.1	Data handling . . . . .	145
6.2	Full waveform processing techniques . . . . .	146
6.3	Cloud detection and saturation . . . . .	147
6.4	Validation . . . . .	147
6.5	New applications . . . . .	148
6.6	Upcoming missions . . . . .	150
<b>A</b>	<b>Ordering ICESat data</b>	<b>151</b>
<b>B</b>	<b>Waveform parameters are displayed along ICESat tracks</b>	<b>153</b>
<b>C</b>	<b>Classification results</b>	<b>157</b>
<b>D</b>	<b>ICESat data over Europe</b>	<b>159</b>
<b>E</b>	<b>Look-up table: Count to Voltage</b>	<b>161</b>
	<b>Bibliography</b>	<b>163</b>
	<b>List of Abbreviations</b>	<b>175</b>

<b>Summary</b>	<b>179</b>
<b>Samenvatting</b>	<b>185</b>
<b>List of Publications</b>	<b>191</b>
<b>Curriculum Vitae</b>	<b>193</b>



# List of Figures

1.1	Picture showing (a) image data (Google Earth), (b) high resolution airborne laser data (AHN) and (c) large footprint full waveform data (ICESat). The airborne laser data in Figure 1.1(b) is decomposed into terrain points (grey), vegetation (green) and building (remaining colors). Figure 1.1(c) shows a typical ICESat waveform (dashed line) displayed with a selection of AHN points within the ICESat footprint (magenta ellipse) in Figure 1.1(b). . . . .	2
1.2	Multiple return echos. (a) Laser signal intercepting at different height levels. (b) Return signal as a function of time. Breakdown of the return signal into different echos: (c) Only first and last echo and (d) multiple return echos. Figure adapted from <a href="#">Lefsky et al. (2002)</a> . . . . .	3
1.3	Schematic overview of discrete and full waveform laser scanning systems. . . . .	4
1.4	Top left: Validation of an ICESat vertical accuracy over The Netherlands. Bottom: Validation of a large data set of ICESat waveform pairs over Europe. Top right: Application of forest canopy changes by repeated tracks occurred at the black line marked by a small rectangle. Moreover, application of land cover classification. . . . .	6
1.5	ICESat full waveforms over bare earth (a) and urban areas (b). . .	10
1.6	Two waveforms with clearly separation between terrain surface and feature heights like forest (a) and buildings (b). Feature heights can be separated easily. Two more complicated waveforms that mixed features: (c) trees and buildings and (d) tree and slope surface. . .	10

1.7	ICESat track located nearby Haarlem city, The Netherlands. The transmitted pulse is shown in panel (a). Panel (b) shows ICESat waveforms in its vertical direction. The horizontally connected lines present along track elevation profiles. The black line is potentially a digital terrain profile and the blue line is possibly a digital surface profile. Below the waveforms the following waveform parameters are displayed: (c) waveform extent/width; (d) total energy; (e) number of Gaussian components/modes; and (f) waveform begin. The land cover classes (g) are colored as vegetation (green), urban areas (red) and bare land (white). At the bottom, ICESat derived land cover classes together with CLC200 class information is superimposed on Landsat7 image data. . . . .	12
2.1	Laser altimeter subsystem components. Extracted from <a href="#">Bufton (1989)</a> . . . . .	16
2.2	Laser pulse waveforms for range measurement and waveform digitization. Extracted from <a href="#">Bufton (1989)</a> . . . . .	17
2.3	All ground tracks from the two SLA missions. SLA-01 operating with a 28.5 degree orbit inclination and SLA-02 with a 57.9 degree inclination. Extracted from <a href="http://denali.gsfc.nasa.gov/lapf_web/">http://denali.gsfc.nasa.gov/lapf_web/</a> . . . . .	19
2.4	SLICER 5-beam Swath Geometry. Extracted from <a href="#">Harding et al. (2000)</a> . . . . .	21
2.5	Relative footprint sizes of laser systems. A typical airborne footprint size of less than 1 m is displayed by the black dot. . . . .	22
2.6	A flight configuration of the LVIS airborne laser altimeter. A 1-km wide swath was generated using forty 25-m wide footprints from 8 km above the ground. Smaller footprint sizes were possible, as were overlapping footprints or less dense sampling schemes. Extracted from <a href="#">Blair et al. (1999)</a> . . . . .	22
2.7	Riegl LMS-Q560 system. . . . .	23
2.8	Principal of operation of ICESat. Extracted from <a href="#">Zwally et al. (2002)</a> . . . . .	26
2.9	GLAS instrument showing on the left the 3 laser boxes (yellow) on the optical bench, the 1-m diameter telescope with heat pipe (red) and side radiators, and on the right the GLAS star tracker (pink), electronics boxes, and the small telescope (grey) of the stellar reference system. Figures extracted from <a href="#">Zwally et al. (2002)</a> . . . . .	27
2.10	ICESat footprints are visualized. The 3D image (top left) is the power distribution within one footprint. The footprint shape and orientation (bottom left) are provided in some ICESat products like GLA06 and GLA14. Figures extracted from <a href="#">Abshire et al. (2005)</a> ; <a href="#">Lefsky (2009)</a> . . . . .	28
2.11	Track spacing (between ascending or descending track pairs) as a function of latitude. Figure extracted from ( <a href="#">Abdalati et al., 2007</a> ). . . . .	29

2.12	All ICESat tracks of data campaign L2a are visualized over the world, north and south polar regions. Figures are taken from <a href="http://nsidc.org/data/icesat">http://nsidc.org/data/icesat</a> .	
3.1	(a) An example of a transmitted pulse of the ICESat laser altimetry system. (b) A bell shape or Gaussian function used to describe the transmitted pulse with peak position $t_x$ , width $\sigma_x$ and amplitude $A_x$ .	36
3.2	Different scenarios of waveform visualization. Waveforms are simulated from the elevation points (black dots).	38
3.3	Waveform data in counts (left) and in voltage (right).	41
3.4	A normalised waveform (in grey) is displayed together with its cumulative distribution curve (in black).	41
3.5	The real waveform signal is within the two vertical lines that are defined based on a threshold value (horizontal line).	42
3.6	Example of a noisy waveform that will be discarded (a) and a rather noisy waveform that is still accepted for further processing (b).	43
3.7	A waveform (grey line) smoothed by a Gaussian kernel (dotted line). The black line represents the smoothed waveform. The Gaussian kernel is located at the 342th bin to compute a smoothed value of the waveform at this position.	44
3.8	Waveform displayed in grey. Initial Gaussian decomposition into single modes is displayed in black.	45
3.9	Gaussian decomposition is applied in a least square estimation sense.	48
3.10	Fitting results: (a) The fitted ICESat waveform (black solid) is displayed together with the raw ICESat waveform (grey thick line) and the four Gaussian components (grey thin line). (b) The ICESat transmitted pulse is displayed in grey and the fitted Gaussian mode in black. (c) Airborne transmitted pulse and (d) Airborne return waveform (grey) with three Gaussian modes (black). The airborne data set is provided by Institute of Photogrammetry and Remote Sensing (IPF), Vienna University of Technology.	49
3.11	The noise level $\sigma_{N_i}$ of a waveform is computed where the deviations occur in the range between 1 to 150 nanoseconds. The waveform level $\sigma_{WF_i}$ is determined by the difference (dashed line) between the raw waveform and the fitted line.	51
3.12	Quality of fit in terms of the noise standard deviation: the solid curve (local) and dashed curve (global).	51
3.13	(a) ICESat waveform, (b) ICESat transmitted pulse and (c) Deconvolution result.	53
3.14	Results of a Gaussian decomposition (left) and deconvolution (right) are displayed in black line with the original ICESat waveform in grey.	54

3.15 (a) Selected raw AHN data within ICESat footprint. (b) AHN data projected to YZ plane. (c) Histogram of raw AHN data. (d) Gaussian distribution of power of a transmitted pulse. (e) Power distribution projected to YX plane. The top color bar is the elevation scale for (a)–(b), and the middle one for (d)–(e). (f) The weighted histogram is derived by applying the Gaussian distribution of (d)–(e) to AHN points (a)–(b). (g) Emitted pulse as system response of the GLAS system. (h) Simulated waveform (thick black curve) displayed with the histogram (f) in grey and the convolved result (dotted curve). (i) ICESat waveform (dashed line) displayed with simulated waveform (solid line). Note that waveforms are scaled for the purpose of visualization. . . . . 56

3.16 AHN simulated waveform (solid curve) and ICESat waveform (dashed curve) displayed with AHN points (grey dots) over: (a) bare land, (b) forest and (c) google image of (b) displayed with the ICESat footprint. The footprint partly covers the building (top) and high trees (left). . . . . 58

4.1 Three components contribute to the error budget of an individual laser shot. . . . . 66

4.2 Study area: Six ICESat ground tracks (magenta) displayed with the actual height model of The Netherlands (AHN). The upward arrows indicate ascending tracks and the downward arrows descending tracks. ICESat reference track numbers are given in blue. . . . . 74

4.3 Procedure of coordinate system conversion. . . . . 79

4.4 Principal of ICESat geolocation and surface elevation determination. 80

4.5 A waveform (solid curve) is georeferenced by matching the waveform centroid (horizontal dotted line) to a GLA14 elevation point (cross). The GLA01-derived elevation is the centroid of the last peak (circle). Waveform start and waveform end are defined by threshold crossing locations. . . . . 81

4.6 AHN points (grey dots) and its mean height (horizontal solid line), lowest segment points (black dots) and its mean height (horizontal dashed line) are displayed with AHN simulated waveform (solid curve) and ICESat waveform (dashed curve). (a) City with roofs and trees. (b) Canal with vegetation on the embankment. (c) Staired surface. . . . . 82

4.7 Definitions of feature heights are visualized: (a) forest height,  $H_{LM-WB}$ , (b) building height with a flat roof,  $H_{LM-FM}$ , (c) building height with a tilted roof,  $H_{LM-FMH}$ , and (d) building height in which the tree is higher than the building,  $H_{LM-WB}$ . Moreover, the ICESat waveform (dashed line) and AHN simulated waveform (solid line) are displayed together with the AHN points (black dots). . . . . 83

4.8	Procedure of incident angle determination of the ICESat laser pulse. Details of this procedure are found in <a href="#">Molijn (2009)</a> . . . . .	85
4.9	(a) Coverage of the ICESat footprint by AHN points. (b) Cloud presence resulting in a height difference $\geq 10$ m. (c) Saturated waveform with a saturation flag value of 3. The range of saturation, between 39 m to 45 m, is clearly visible. . . . .	86
4.10	Examples of outliers in the ICESat–AHN comparison. The ICESat waveform is represented by the dashed black curve, the AHN simulated waveform by the solid black curve, the mean elevation of the ICESat last mode by the horizontal dashed line, the mean elevation of the AHN lowest segment by the solid black line. (a) the ICESat last mode at $-0.2$ m is not extracted because of fitting errors. (b) the AHN lowest segment does not give a good corresponding elevation compared to the ICESat last mode. (c) the AHN lowest segment is not well defined for a slant surface. . . . .	88
4.11	Histograms of height differences obtained after applying conditions C1–C4 over: (a) broadleaf, (b) mixed, (c) needleleaf, (d) urban and (e) bare land. Median and standard deviation values have been estimated by the Median/MAD robust statistics ( <a href="#">Muller, 2000</a> ). . . . .	89
4.12	Water waveforms (top panel) with number of AHN points (top right) within ICESat footprints displayed together with the Google Earth image (middle panel) and the AHN elevation image (bottom panel). In Figure 4.12(a)–(d), the horizontal black line represents the last mode height and the horizontal red line is the mean height of AHN ground points. The green dots are the AHN points located inside the ICESat footprint and the grey dots are the AHN points outside the ICESat footprint. . . . .	91
4.13	Histograms of feature heights extracted from AHN and ICESat data for (a) forest, (b) buildings, and (c) low vegetation. The histograms of the AHN feature heights are in grey, the ICESat histograms are bounded by a solid black line. For both AHN and ICESat, the mean value and standard deviation are given in the corresponding text boxes. Moreover, the histograms of $H_{LM-WB}$ feature height differences between AHN and ICESat data are shown in figures (d)–(f) for different terrain classes. In each plot the median/MAD value is indicated. . . . .	93
4.14	Big changes on the earth surface. The AHN waveform (solid line) is simulated from the AHN points (black dots) that does not contain any data points from above the terrain surface, Figure 4.14(a). However, the ICESat waveform (dashed line) properly includes other objects, e.g. new houses. Similarly, Figure 4.14(b), very few vegetation points in the AHN data correspond to a weak peak in the AHN simulated waveform, however there seems to occur, e.g. new trees in the ICESat waveform. . . . .	94

4.15	ICESat campaigns, right, and ICESat tracks over Europe during the Winter 2006 campaign (L3e), overlaid on CLC2000 land cover data, left. . . . .	97
4.16	Crossover pair, left, and two repeated track pairs, right. . . . .	99
4.17	A flowchart of finding crossover and repeated pairs. . . . .	100
4.18	Histogram of waveform extent differences of epoch 1. . . . .	102
4.19	3 case studies with 2m distance difference between footprint centers: waveform pairs (top) and footprint pairs over images of Google Earth (bottom). . . . .	103
5.1	Digital elevation models over Antarctica (left, 500m resolution) and Greenland (right, 1km resolution). Figures downloaded from <a href="http://nsidc.org/data/icesat/data.html">http://nsidc.org/data/icesat/data.html</a> . . . . .	109
5.2	Maps of slope in degree (left) and roughness in meter (right) over the Greenland ice sheet. Figures are extracted from <a href="#">Yi et al. (2005)</a> . . . . .	109
5.3	The definitions of freeboard ( <a href="#">Zwally et al., 2008</a> ) and grounding zone ( <a href="#">Fricker and Padman, 2006</a> ) are illustrated in (a) and (b) respectively. . . . .	110
5.4	Biophysical parameters are defined and extracted from the waveform shape over low-relief surfaces. Figure extracted from <a href="#">Harding and Carabajal (2005)</a> . . . . .	111
5.5	Biophysical parameters are defined and extracted from the waveform shape over high-relief surfaces. Figure extracted from <a href="#">Lefsky et al. (2007)</a> . . . . .	113
5.6	ICESat groundtracks from February (blue) and September 2003 (red) overlaid on the CORINE land cover image. . . . .	116
5.7	Left: histogram of inter pair distance and its mean (vertical black line). Right: visualisation of footprint locations in winter (thick black) and summer (thin black) and the 100m CLC2000 map with bare land (dot pattern), needleleaf (horizontal line pattern) and broadleaf (slant line pattern). . . . .	118
5.8	Two intensity histograms displayed together: February, 2003 (dashed) and September, 2003 (grey) with its mean intensity. The mean intensity of the summer data is 2 times the mean intensity of the winter data. . . . .	119
5.9	Two fitted waveforms (black solid) are displayed together with the raw waveform (grey starred line) and the Gaussian components (grey solid) for February, 2003 (left) and September, 2003 (right). . . . .	120
5.10	The original summer waveform is kept fixed (grey solid). The original winter waveform (dashed dot line) is shifted by: (a) complete waveform, (b) last mode, and (c) modified last mode. The shifted waveform is displayed in a black line. . . . .	121

5.11	Waveform parameters: maximum canopy height ( $dEcho = 32.15$ m (214.3 ns)), canopy return, ground return, height of median energy ( $HOME = 20.60$ m (137.3 ns)). The $rGround$ for this waveform is 0.16. . . . .	122
5.12	Histograms of canopy intensity differences ( $\Delta I$ ) with $-5$ m $< \Delta dEcho < +5$ m in broadleaf (left), mixed (middle) and needleleaf (right). . . . .	126
5.13	Along track topographic profile (grey) with the three forest types given by the CLC2000 (top). Remaining and removed broadleaf waveform pairs (2 <sup>nd</sup> panel). The third panel shows the slopes between the paired footprint centres. . . . .	130
5.14	Study area: ICESat ground tracks (blue) and CLC2000 land cover map (100m resolution). A close up of the black rectangle is shown in top left corner box, with elliptical footprints of 95 m x 52 m. The track passes from the North Sea (in white) onto land (in color). . .	133
5.15	Waveform parameters derived from a typical waveform: (a) number of modes, (b) Total return energy, (c) waveform begin and (d) waveform extent. The crossed-line represents the normalized waveform and the solid thin line represents the fitted waveform. . . . .	134
5.16	Classification scheme. . . . .	136
5.17	Top Left: High vegetation. Top Right: Urban. Bottom Left: Bare land or low vegetation. Bottom Right: Water. . . . .	137
5.18	Right: land cover classification results displayed with two ICESat ground tracks and Landsat-7 image. Left: typical ICESat waveforms (black) and its Gaussian components (yellow) are displayed over the Landsat-7 image considering three case studies. The white arrows indicate the waveform location. . . . .	140
B.1	Over the Waddenzee mud flat region. . . . .	154
B.2	Over an area near the Haarlem city. . . . .	155
C.1	(a) A correct classification result of bare land. However, there are still some misclassification results: (b) Urban class (Haarlem city) is classified as high vegetation; (c) Water classified to bare land at between Noordoost polder and Flevoland regions; (d) Water (with planted trees) classified to high vegetation at Noordoost polder. . .	158
D.1	ICESat data tracks over Europe. . . . .	160





# List of Tables

2.1	Preliminary Error Budget for SLA. . . . .	19
2.2	Specifications of typical full waveform scanning systems. . . . .	24
2.3	Standard GLAS data products. . . . .	31
2.4	GLAS data campaigns. . . . .	32
3.1	Parameters of a single waveform and their physical explanation. . .	59
3.2	Parameters of a single waveform and their physical explanation. . .	60
3.3	Parameters of repeated waveform pairs and their physical explanation.	61
4.1	Single shot error budget for ICESat elevation measurements. Note that pointing error assumes 1.5 arcsec pointing error on 1° surface slope (Schutz, 2001). . . . .	65
4.2	Differences in elevation between GLAS data and other data. . . . .	69
4.3	Number of ICESat waveforms used with energy and footprint infor- mation. . . . .	76
4.4	Number of ICESat waveforms used: Forest (Broadleaf, mixed and needleleaf), Urban, Bare land, and Water. . . . .	77
4.5	Height difference and its standard deviation in parenthesis between ICESat last mode and AHN lowest segment. . . . .	87
4.6	Feature height difference between ICESat and AHN and its stan- dard deviation in parenthesis. . . . .	92
4.7	ICESat epochs, laser name, orbit repeat period, acquisition date, release, and number of points . . . . .	98
4.8	ICESat epochs, total number of pairs, individual footprint diam- eters, distance between footprint centers and number of close by waveforms. . . . .	100
4.9	Differences between repeated waveforms . . . . .	101

5.1 Number of waveforms used . . . . . 117

5.2 Median and standard deviation values of the shift comparison are  
 computed using Median/MAD robust statistics (Muller, 2000). . . 125

5.3 Forest parameters derived from waveform with  $-5m < \Delta dEcho < +5m$  127

5.4 Confusion matrix and classification results . . . . . 128

5.5 Confusion matrix . . . . . 138

5.6 Classification results . . . . . 139

A.1 Four methods to order ICESat data. . . . . 151

E.1 Look-up table: Count to Voltage for ICESat full waveform data. . . 162

# Introduction

## 1.1 Research context

### 1.1.1 Discrete airborne laser scanning systems

Airborne laser scanning data have been used in many applications like generation of digital terrain models, archaeology, forest inspection and 3D city modeling. An airborne laser scanning data set can be characterized as a point cloud consisting of three dimensional points, where additionally for each point an intensity value may be available. Such kind of data is acquired by airborne laser scanning systems. In principal, a laser scanning system transmits a single pulse and records a signal return from the ground surface to compute a travel time of the pulse resulting in a range distance. Together with positioning information acquired from the Global Positioning System (GPS) and information on the attitude of the system, georeferenced coordinates of a three dimensional point are obtained. Figure 1.1(a) shows a photo of an urban area with buildings, trees and ground surface, while Figure 1.1(b) shows a picture from the same area that actually consists of a set of three dimensional points acquired from an airborne laser scanning system.

### 1.1.2 Multiple return echoes

A transmitted laser pulse has a certain width. When the pulse arrives at the surface, for example over an area containing trees, see Figure 1.2(a), a part of the energy of the transmitted pulse intercepts the canopy while another part continues down to interact with the ground. The return signal can therefore be expressed

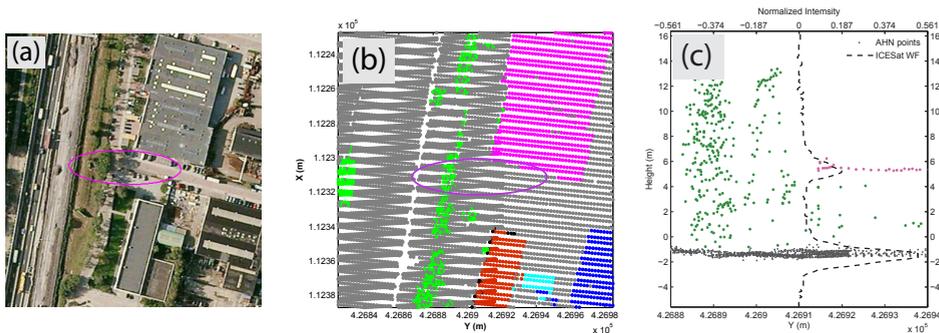


Figure 1.1: Picture showing (a) image data (Google Earth), (b) high resolution airborne laser data (AHN) and (c) large footprint full waveform data (ICESat). The airborne laser data in Figure 1.1(b) is decomposed into terrain points (grey), vegetation (green) and building (remaining colors). Figure 1.1(c) shows a typical ICESat waveform (dashed line) displayed with a selection of AHN points within the ICESat footprint (magenta ellipse) in Figure 1.1(b).

as the return power as a function of time, as shown in Figure 1.2(b).

In the earliest systems, the laser return signal was detected once, for example when it passed a background noise level for the first time. Systems that measured both the first and the last time the return signal was larger than the noise level, resulted in so-called first and last return echo data, see Figure 1.2(c). In current commercial laser scanning systems, multiple return echoes can be recorded. The first return presents the highest point of interception (top of tree or house) and the last return the lowest surface (ground). The remaining returns in between the last and the first echo give additional information on for example the canopy structure, Figure 1.2(d). However, these systems still have the limitation that the full return signal as a function of time cannot be recorded and therefore possible information on the vertical structure of objects hit by the laser pulse is lost.

### 1.1.3 Full waveform systems

Fortunately, nowadays, laser scanning systems are available that are able to capture the full return signal. Such system samples the return signal at a certain temporal resolution. As a consequence the full return signal can be reconstructed, see Figure 1.3(d)–(e). Given the full shape of the return signal, structural information on the vertical distribution of objects between the height levels corresponding to the first and the last echo becomes available.

Laser scanning systems have been mounted on helicopters, small airplanes, large airplanes and satellites. The difference in scanning height together with the dif-

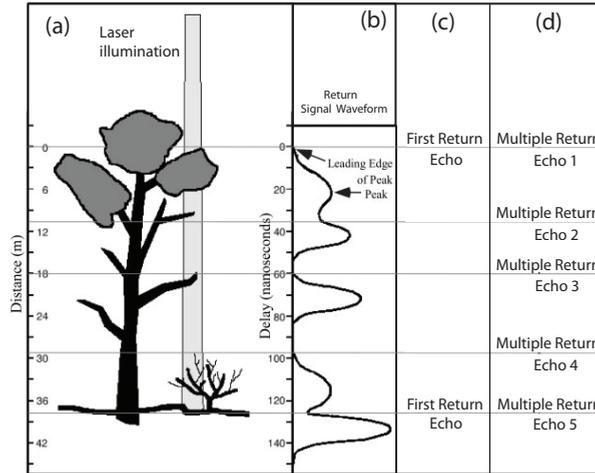


Figure 1.2: Multiple return echos. (a) Laser signal intercepting at different height levels. (b) Return signal as a function of time. Breakdown of the return signal into different echos: (c) Only first and last echo and (d) multiple return echos. Figure adapted from [Lefsky et al. \(2002\)](#).

ferences in the laser systems itself results in large differences in the size of the laser footprints, i.e. the area illuminated by one laser pulse on the ground, see Figure 1.3. If the footprint is assumed to be a circle, the footprint diameter on the ground surface is obtained at first approximation by multiplying the platform altitude to the beam divergence. For example, the footprint diameter for the Riegl airborne system, operating at about 1–2 km height is between 15–30 cm, while for the spaceborne GLAS system operating at about 600 km the illuminated footprint diameter is in the order of 70 m.

The full waveform signal encodes the complete response of the interaction between emitted laser pulse and the illuminated target and therefore potentially carries more information on the target than a return signal consisting of a few single elevations. Given the full waveform signals, the main issue is how to extract the information on the vertical structure from these signals. Development of new methods to deal with such full waveforms is one of the important goals of this thesis. This thesis will mainly focus on the analysis of large footprint full waveform signals as obtained by the Geoscience Laser Altimeter System (GLAS) mounted on the Ice, Cloud and land Elevation Satellite system (ICESat).

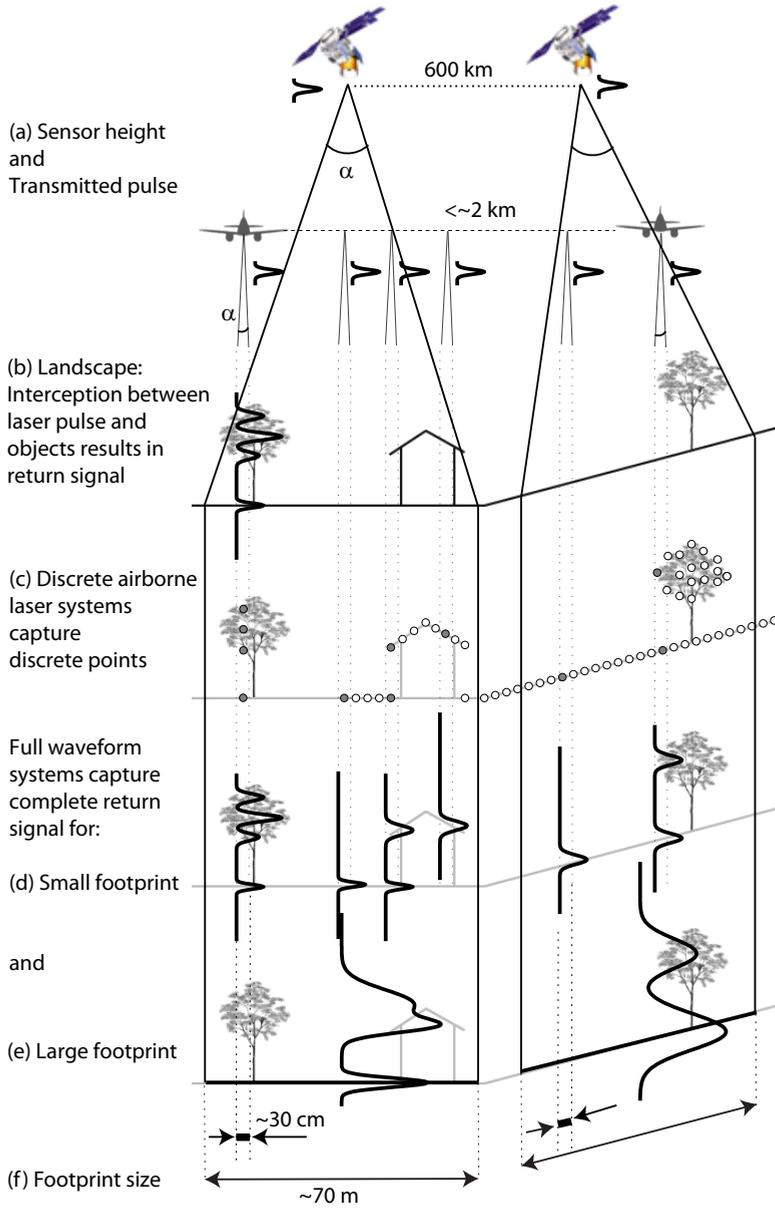


Figure 1.3: Schematic overview of discrete and full waveform laser scanning systems.

### 1.1.4 ICESat large footprint full waveforms

ICESat was launched in January 2003 to observe the polar regions, the atmosphere and also to measure land topography profiles and canopy heights (Zwally et al., 2002) and is orbiting at an altitude of 600 km. GLAS carries three different laser altimeters but uses one laser altimeter at a time to transmit a laser pulse of 10 nanoseconds pulse duration and to consecutively record the return pulse as reflected from the approximately 70m-diameter footprint on the ground. GLAS systematically samples the energy returned from the surface to obtain its waveforms (Harding and Carabajal, 2005). Such waveforms are digitized at a vertical resolution of 15 cm (NSIDC, 2005). GLAS acquires elevation profiles of the entire earth along tracks. Along each track every 175 m a waveform is recorded. Initially it was planned that ICESat would operate in long campaigns of approximately 6 months. Due to problems with the lasers, the campaigns were rescheduled and shortened to a period of about 4–6 weeks per campaign. The current operational mode of the system acquires a maximum of three campaigns a year. Currently, i.e. June 2009, 18 campaigns were successfully finished and only 14 campaigns (up to March 2008) are released. The first campaign ran from 2003-02-20 to 2003-03-21. The last campaign covers a period of four weeks in June 2009.

Two major advantages of the ICESat mission are first that it captures tracks world wide and second, that these tracks are also approximately repeated. The latter potentially allows to assess along track elevation changes. Major disadvantages of the ICESat orbital configuration are however that major gaps exist between consecutive ascending or descending tracks, and that repeated tracks are not always overlapping well. ICESat operates at a inclination of 94 degree, thus there is a gap of 4 degree at both North and South pole. Close to the poles, ICESat data are significantly more dense compared to lower latitude locations, as for example The Netherlands. Within one campaign, the distance between consecutive ICESat ground tracks over The Netherlands is approximately 50 km. Over The Netherlands, ICESat data is sampled along a total of 15 different ascending and descending tracks. Some of these tracks are visible in Figure 1.4, top left.

Because the laser pulse is transmitted from space, the return signal and its positioning in a georeferenced coordinate system will be affected by four main components: platform, instrument, atmosphere and ground surface.

Platform effects are mainly caused by the orbit positioning systems. The satellite operates at an altitude of approximately 600 km above the ground surface, however the satellite height (Luthcke et al., 2005) varies with respect to latitude. Additionally, jitter induced by the solar panels, and temperature variations in the different platform components are likely to affect the ICESat range measurements.

The Geoscience and Laser Altimeter system consists of three lasers. Shortly after its operational start in 2003, the first laser died. Since then the remaining two

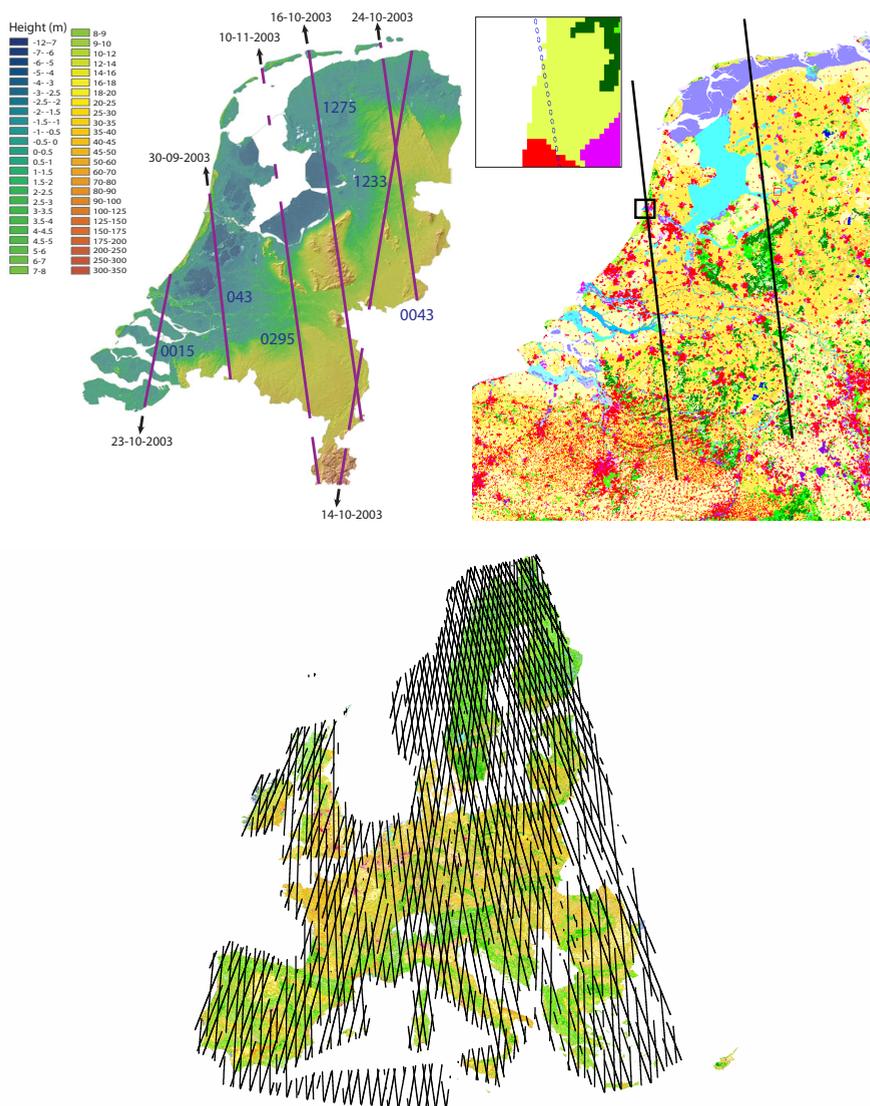


Figure 1.4: Top left: Validation of an ICESat vertical accuracy over The Netherlands. Bottom: Validation of a large data set of ICESat waveform pairs over Europe. Top right: Application of forest canopy changes by repeated tracks occurred at the black line marked by a small rectangle. Moreover, application of land cover classification.

lasers were used to obtain waveforms, until 2008, when also the third laser died. As a consequence of the performance of the lasers, the emitted energy was strongly varying during operation. Another effect that affects the accuracy of waveform derived parameters is the laser pointing angle. Because of the large distance between laser and Earth, even a small error in determining the pointing angle of the operational laser will result in a relative large error in the positioning of the laser footprint on the Earth surface.

In contrast to for example radar radiation, laser light is not able to penetrate thick clouds. As a consequence, over a country as The Netherlands that is often covered by clouds, more than half of the emitted laser pulses will not or poorly return to the satellite. The presence of clouds is in principle indicated in the GLAS data products, but it turns out in practice that the cloud detection methodology is also affected by other factors, like the emitted laser power.

Surface conditions are also affecting the ICESat measurements. Slope and roughness of the ground surface both result in a widening of the waveforms and their effect is difficult to distinguish. Within an approximately 70 meter footprint the ground surface is mostly not perfectly flat. In many cases, ground surfaces within a 70 meter footprint are neither perfectly planar nor rough in a perfect homogeneous way. Instead, the surface typically displays a mixture of slope and roughness effects.

Forest features may cause a range underestimation. When a transmitted pulse hits a very dense forest, the return signal may not penetrate to the ground surface and only reflects from the forest canopy. In another case the signal may go through the canopy and reflect back from the ground, however, the return signal may be further weakened by the canopy obstruction or understorey beneath the canopy. As a consequence, the energy reflected from the ground may not be detected at all. Therefore, these measured ranges are not suited for determining the terrain elevation.

In addition, characteristics of the ground surface, like its reflectivity will result in varying levels of return energy. Pure water almost completely absorbs the energy of the transmitted 1064 nm laser pulse, while bare ground results in a high return energy.

The shape of large footprint size waveforms is more complicated compared to the shape of small-footprint waveforms. The reason for this is that a large footprint potentially contains a mix of many objects, i.e. buildings, trees and roads, all with possibly different reflectivity properties. Figure 1.1(c) shows a typical ICESat full waveform resulting from illuminating a footprint containing high trees, buildings and ground surface.

### 1.1.5 Research questions

Although the above shows that there are many potential challenges in working with spaceborne full waveform laser data, a big advantage of the ICESat system is that full waveform data is available all around the Earth, and that the repetition of approximately coinciding ICESat tracks is opening promising research activities.

In order to assess and minimize the impact of the different error components as sketched above, the associated processing flow, starting from readily available data products, should be validated over an area where high quality validation data is available. For this purpose The Netherlands is chosen as main region of interest, while in some cases the region of interest is enlarged to the European continent.

Therefore, within this context, the following research questions will be considered in this thesis:

- How can ground surface elevations be derived in an optimal way from ICESat products? And what horizontal and vertical accuracies can be obtained over a variety of land surface classes?
- What applications of large footprint full waveform parameters exist and can be further developed?
- What are possible applications of analysing waveforms obtained from nearly overlapping footprint locations, acquired at either nearly the same time or in different seasons like in summer and winter.

To answer these questions, the following research topics will be investigated in this thesis:

(1) Two cases are studied as validation. First, quality and vertical accuracy of ICESat derived elevations are assessed by comparison to accurate airborne laser data over The Netherlands, Figure 1.4 (top left). Second, an error analysis method is proposed by constructing a large data set of ICESat waveform pairs at overlapping footprints and with acquisition time differences of maximally a few weeks over the European continent, Figure 1.4 (bottom).

(2) Besides existing applications of waveform parameters, also two new applications are developed. First, an analysis of a set of overlapping waveform pairs over The Netherlands acquired at different seasons (winter and summer) is studied for assessing forest change with respect to different forest species: broadleaf, needle-leaf and mixed forest, Figure 1.4 (top right). Second, full waveform data and derived waveform parameters are applied to classify the footprint locations of the waveforms into the land cover classes high vegetation/forest, urban areas, bare land/low vegetation and water, Figure 1.4 (top right).

## 1.2 Scope of research

In this thesis a decomposition method will be described and partly developed for processing both large and small footprint full waveform data. The method is tested on ICESat large footprint data and Riegl small footprint data. Alternative tools to assess waveform parameters will also be described, notably waveform deconvolution and waveform simulation. The interaction between the emitted pulse and the vertical distribution of terrain features is used to study waveform parameters. An overview of full waveform parameters, both newly defined parameters and parameters described in previous research is given. Note that waveform parameters are defined not only for single waveform data but also for overlapping waveform pairs.

The aim of the research described in this thesis is not to correct errors caused by platform, instrument, atmosphere and ground surface. Instead it will be studied how to minimize the effect of errors as much as possible. For this purpose, filtering constraints will be developed to remove affected signals as much as possible. The final aim is to develop procedures that are efficient enough to enable the analysis of GLAS waveforms at the individual shot level.

To support these research purposes, study areas will be selected where reliable reference data are available. The auxiliary data used in this research are the CORINE land cover data base covering the European continent, and dense airborne laser scanning data from the national Dutch elevation product AHN (Actueel Hoogtebestand Nederland). Two different ICESat products are extensively studied and evaluated in this research: GLA01 - the GLAS/ICESat L1A Global Altimetry Data full waveform product and GLA14 - the GLAS/ICESat L2 Global Land Surface Altimetry Data elevation product.

## 1.3 Research methodology

In this section, methodology is described step by step together with specific objectives. For each step, a visualization of typical examples is included. Moreover, results of specific case studies are stated. Four steps are described: characteristics of ICESat full waveform data, waveform processing, two validation case studies and two new applications.

### 1.3.1 Characteristics of ICESat full waveform data

Figure 1.5(a) shows a full waveform over a simple surface, in this case flat bare land, and Figure 1.5(b) shows the waveform response resulting from a complex

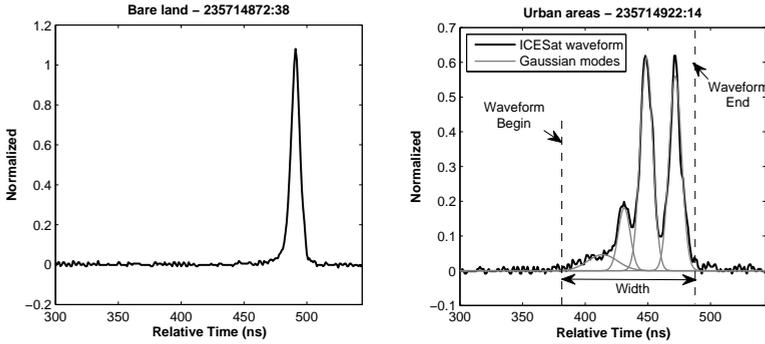


Figure 1.5: ICESat full waveforms over bare earth (a) and urban areas (b).

surface, in this case an urban area. The raw ICESat waveforms are displayed by black lines. The unit of the x-axis is nanoseconds and the unit of the y-axis is voltage. The waveform acquired over bare land has a single peak. In contrast, the waveform over urban areas has at least three significant peaks. In order to get a better of understanding of full waveform data, they are visualized together with accurate airborne laser data in Figure 1.6.

Figure 1.6 shows different raw ICESat waveforms, all with two dominant peaks that are acquired over forest and urban areas. These waveforms (solid black line) are displayed together with traditional airborne laser data, decomposed in points belonging to the terrain surface (grey dots), vegetation (green dots) or building roofs (red dots). From a comparison between the terrain points and the lower peak

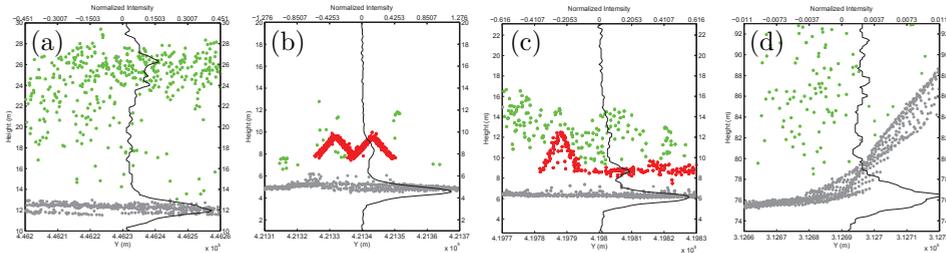


Figure 1.6: Two waveforms with clearly separation between terrain surface and feature heights like forest (a) and buildings (b). Feature heights can be separated easily. Two more complicated waveforms that mixed features: (c) trees and buildings and (d) tree and slope surface.

of the ICESat waveform, it is clear that the peak location along the vertical axis is well presenting the mean elevation of the terrain. The width of this peak contains information on surface slope and terrain roughness. Flat and smooth surfaces in (b) and (c) correspond to narrow last peaks. In (a) the surface is rougher while in (d) the surface is partly sloped. As a result, the ground peaks have widened. In addition, the top peak corresponds to the canopy or roof structure. The distance between the top and lower peak possibly provides the feature height of e.g. forests and buildings, Figure 1.6(a)–(b), respectively.

However, over complex surfaces, large footprint waveforms will have complicated shapes, Figure 1.6(c)–(d), and thus an extraction of tree or building heights is more difficult. The waveform shape is a result of a mixture of different reflecting objects. In Figure 1.6(c) both trees, buildings and terrain can be distinguished. The footprint in Figure 1.6(d) is sampling a forest over both flat and sloped terrain.

### 1.3.2 Waveform processing

In this thesis, a method that extracts waveform parameters that is able to at least partly solve problems of instrument and atmosphere will be described. Full waveform data are assumed to be a sum of a number of superimposed bell functions or Gaussian components. Each Gaussian component may describe properties of one particular reflecting object within a footprint. An analysis of such a mode can provide the user with attributes like elevation, surface slope or roughness. Therefore an approach based on linear and nonlinear least square estimation to decompose the raw full waveform data into single Gaussian components is implemented. This method is also able to process small footprint waveform data. With this method, input parameters like maximum number of Gaussian components and distance between Gaussian components can be adjusted to match different circumstances for later research applications.

In Figure 1.5, two typical ICESat waveforms over bare land and urban areas are visualized together with waveform parameters that can be derived from Gaussian decomposition. The fitted line and the Gaussian components are the black thick line and thin grey line respectively. Bare land waveform typically has one Gaussian component. Urban waveform conducts more something like four Gaussian components. Moreover, the width of the urban waveform over urban is in general larger than for bare land waveforms. Next to the decomposition result, other waveform parameters encoding useful information include the total return energy, the signal begin and end of the waveform. The waveform parameters along an ICESat track are spatially shown in Figure 1.7.

Another approach used to decompose waveform data is so-called deconvolution. A waveform is assumed to be a convolution of a transmitted pulse and a surface response function. Both the return waveform and the transmitted pulse are

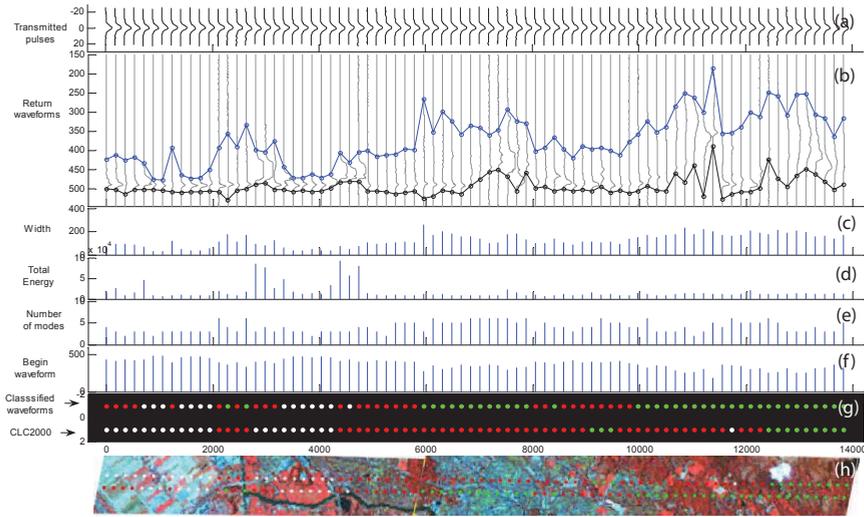


Figure 1.7: ICESat track located nearby Haarlem city, The Netherlands. The transmitted pulse is shown in panel (a). Panel (b) shows ICESat waveforms in its vertical direction. The horizontally connected lines present along track elevation profiles. The black line is potentially a digital terrain profile and the blue line is possibly a digital surface profile. Below the waveforms the following waveform parameters are displayed: (c) waveform extent/width; (d) total energy; (e) number of Gaussian components/modes; and (f) waveform begin. The land cover classes (g) are colored as vegetation (green), urban areas (red) and bare land (white). At the bottom, ICESat derived land cover classes together with CLC200 class information is superimposed on Landsat7 image data.

recorded, the surface response can be obtained by deconvolving the transmitted pulse out of the waveform.

To avoid and discard unwanted effects of notably instruments and atmosphere, filtering constraints are defined. ICESat quality flags including e.g. atmosphere conditions during acquisition and sensor behavior are studied and if necessary applied.

A tool applied in this thesis is waveform simulation by using airborne laser data. For this purpose, small footprint airborne laser data is used as a representation of the vertical distribution of the terrain. A full waveform signal can be considered a convolution of a surface response function and a system response. Elevation points are assumed to be reflecting points or scatterers. The histogram of elevations from these points is the response function. The simulated waveform is a convolution of the histogram and the ICESat transmitted pulse. The power distribution of the

ICESat transmitted pulse is approximately Gaussian. To incorporate the intensity distribution within the emitted pulse, weights are assigned according to the airborne points, depending on their distance from the footprint center. Comparison of the simulated waveform and the corresponding ICESat full waveform helps obtaining understanding the reflectivity and interception between laser pulse and the earth surface.

### 1.3.3 Validation studies

In order to evaluate the decomposition results from the previous step, derived waveform parameter values are validated. Validation is performed for two cases:

Comparison between ICESat full waveform data and accurate airborne laser scanning data over The Netherlands is studied. The vertical accuracy of ICESat products with respect to different land cover types like forest, urban areas, bare land and water will be assessed. On average, height differences between the ICESat and the airborne laser scanning data are below 25 cm over bare land and urban areas. Over forests, height differences are even smaller, but with slightly larger standard deviations of about 60 cm.

In addition, comparison of feature heights like forest, buildings and low vegetation from the ICESat waveforms and the airborne laser data is also discussed. A waveform based feature height comparison, resulted in feature height differences of 1.89 m over forest, 1.48 m over urban areas and 29 cm over low vegetation.

An internal comparison of waveform pairs obtained at approximately the same time and coinciding locations is an alternative tool to describe waveform quality and to quantify error sources. The differences in the values of waveform parameters like median energy, waveform extent, relative return energy and intensity distribution are determined and discussed. Three case studies of rare pairs with perfectly overlapping footprints are studied. The results of this study shows that occurring differences in waveform parameter values can be explained by errors in the reported footprint size, in geolocation and by small differences in acquisition time.

### 1.3.4 ICESat applications

The potential of applying ICESat data in research is demonstrated for different fields.

Full waveform data are applied for land cover classification and forest applications. These applications are currently mainly studied based on imagery. However, full waveform analysis provides an alternative way to classify land cover type using a suited set of waveform parameters. Four land cover classes studied in this applica-

tion are forest, urban areas, bare land/low vegetation and water. As a result, an overall accuracy of the classification of about 75% could be obtained. The region considered in this study includes a large part of The Netherlands, and some part of Belgium and France.

For feature height extraction over forests, a single waveform can be used to derive parameters describing forest structure like tree height and canopy depth. The extracted forest parameters are also directly linked to other notions like aboveground biomass. Moreover, waveform pairs that are acquired at nearly repeated footprint locations at different times contain information on forest change due to seasonal influences. A pair of waveforms is studied, consisting of one waveform recorded in winter and one corresponding waveform recorded in summer. As a result of a comparison of such waveform pairs it is shown that forest change occurs mostly for broadleaf (a 148% change, winter to summer) and least for conifers (a 36% change) as can be expected as conifers do not lose their foliage in winter.

## 1.4 Outline of the thesis

This thesis is organized in six chapters. Chapter 1 presents background and research questions. Available full waveform systems for capturing small and large footprints are described in chapter 2. In chapter 3, the complete chain of full waveform processing is described. Parameters derived from full waveform data as described in chapter 3, are utilized in chapter 4 for the purpose of validation of ICESat products, and in chapter 5 for new application development. Chapter 6 presents a thesis conclusion with achievements, remaining research issues and a discussion on the possibilities of future missions capturing a full waveform laser altimetry system.

# Sensors and instruments

## 2.1 Introduction

In this chapter, an overview of full waveform laser sensors, instruments and operating platforms is given. In section 2.2, the principal of a laser sensor system is described. Development of full waveform scanning systems is described in section 2.3. Moreover, a comparison between different full waveform scanning systems is given. Then, the ICESat mission is presented in more detail in section 2.4. Conclusions are stated at the end of this chapter.

## 2.2 Principal of operation of a laser scanning system

The use of laser as a remote sensing instrument started more than 40 years ago. Between 1960 and 1970, many experiments demonstrated the benefits of using lasers in remote sensing including Earth and planetary laser ranging, atmospheric monitoring, and oceanographic studies (Garvin et al., 1998; Bufton, 1989). The system used for this purpose is usually called LIDAR (Light Detection And Ranging). This technology involves transmitting pulses of laser light toward the ground and measuring the time of pulse return. The return time for a pulse back to the sensor is processed to calculate the distance between the sensor and the object. Using additional positioning information from the operating platform at the time of the shot leads to a 3D referenced position.

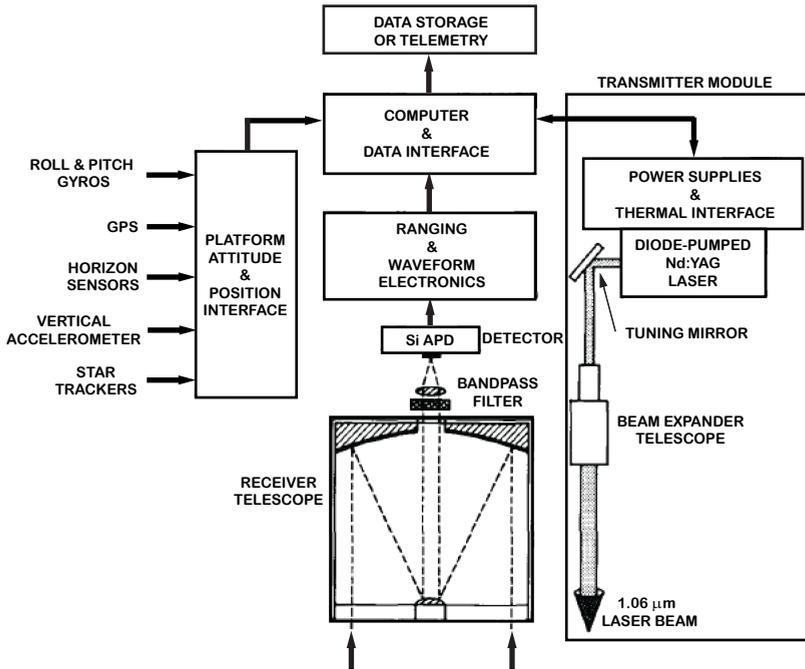


Figure 2.1: Laser altimeter subsystem components. Extracted from Bufton (1989).

The concept of a laser instrument is shown in Figure 2.1. It consists of a laser transmitter module, a receiver telescope, a detector package, ranging and waveform electronics, pointing attitude measurement components and a data storage module.

A short laser pulse of for example 5–10 nanosecond duration is generated by the pulse transmitter, typically by means of a high power, diode-pumped Nd:YAG laser (Neodymium Yttrium Aluminum Garnet). The wavelength of this pulse is mostly equal to 1064 nm. The laser pulse is reflected by the rotating mirror in order to direct the output laser pulses parallel to the receiver optical axis. The reflected pulse is further controlled by the beam expander telescope. The beam expander telescope controls the beam divergence in order to achieve the desired laser footprint size at the ground surface.

The backscattered radiation, i.e. the return signal from the target surface, is recorded by the receiver telescope. The return signal is enhanced by a bandpass filter in order to improve the Signal-Noise-Ratio that typically is spread by interaction with the target surface. The return signal is digitally sampled as a function of time at an interval of at least 1 nanosecond and saved in the data storage module. This return signal is called the full waveform signal. In addition, other

information of the system is obtained, e.g. platform altitude and position, time stamps of transmitted laser pulse and returning full waveform signal, and scanning angle. The full waveform is then processed and georeferenced to compute multiple elevations together with other attributes like pulse width and amplitude, Figure 2.2.

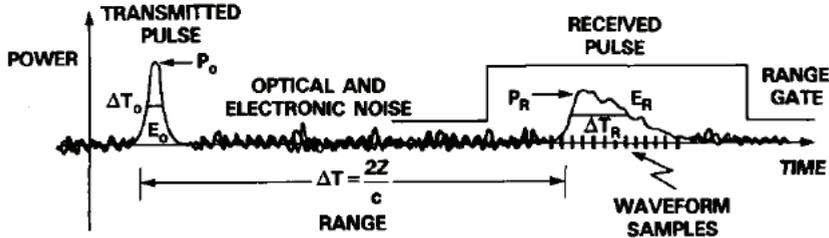


Figure 2.2: Laser pulse waveforms for range measurement and waveform digitization. Extracted from [Bufton \(1989\)](#).

Such system can acquire data by scanning either along or across the flight direction. The footprint resolution is defined accordingly. In an aircraft system, for example, the along-track footprint spacing is determined by the laser pulse repetition rate, the number of cross-track footprints, the aircraft ground speed, and pitch information. The pitch is the angle between the centerline of the aircraft and the horizontal plane. The cross-track footprint spacing is determined by the angular separation between successive transmitted pulses and the aircraft altitude above the ground. Typical settings of laser altimetry are described by the following parameters.

*Operating altitude:* the height of the spacecraft or aircraft above the ground when acquiring data.

*Pulse width at half maximum:* the length of the emitted laser pulse at half of its maximum energy level.

*Pulse energy:* the energy of the transmitted pulse .

*Pulse firing rate:* the frequency that describes how many pulses are transmitted from the laser source per second to the ground.

*Laser beam divergence:* the angle that describes the spread of the laser beam. Together with the operating altitude it determines the footprint size of an illuminated spot on the ground.

*Footprint size:* the area of the spot illuminated by the laser beam on the ground, typically parameterized by diameter (circular shape) or major- and minor axis

(elliptical shape).

*Scan angle range*: the range in degrees that expresses at what maximum scan angle the scanner can operate.

*Sampling rate*: the parameter that describes the vertical resolution of the receiver.

Significant differences between airborne and spaceborne systems are: altitude operation, footprint size, scanning angle and scanning pattern, and spatial coverage. They will be described in the following section.

## 2.3 Development of full waveform scanning systems

Laser altimetry is an emerging remote sensing technique with a wide variety of applications in the Earth and planetary sciences (Blair et al., 1999). Although laser altimetry has developed since 1960s, and was used on for example the Apollo Moon missions 15, 16 and 17 (Kaula et al., 1974), laser altimeter instruments have operated in the space environment to a very limited extent until 1990s. In the last decade the more advanced Mars Observer Laser Altimeter (MOLA-1) sensor was designed and launched in 1992 (Garvin et al., 1998). MOLA-1 was lost in 1993 and the follow-up mission MOLA-2 was launched in 1996. The MOLA-2 was designed to obtain global topographic mapping of Mars and aimed to acquire a data set consisting of about one billion measurements during the 687-day mission. Beside its primary goal of measuring the range, the MOLA-2 also measured the returned laser pulse width that contains roughness information within the 130 m footprint on the Martian surface (Zuber et al., 1992). The MOLA-2 has stopped its mission in 2001 and has been in space for 1696 days. Meanwhile, experiences obtained during the development of the MOLA-sensors served as a pathfinder to build up a new spaceborne system for capturing full waveform signals called the Shuttle Laser Altimeter (SLA). The SLA sensor was built up from the MOLA-1 design, and from experiences with a waveform analyzer. Moreover, a larger capacity flight data storage system was incorporated (Bufton, 1989; Blair et al., 1999).

The SLA was developed by the Laboratory for Terrestrial Physics at NASA's Goddard Space Flight Center (GSFC) and collected land surface elevation data during the Space Shuttle missions in January 1996 (SLA-01) and August 1997 (SLA-02), see also Figure 2.3. The SLA emits infrared laser pulses and measures the time of flight of the backscattered echoes in order to address global Earth System science issues like in particular land cover dynamics in arid and coastal areas (Garvin et al., 1998), and the shape of land surfaces and vegetation canopies (Gesch, 1998).

The SLA has flown in space at an altitude of 305 km and recorded a backscatter

laser echo illuminating a 100 m diameter spot on the ground with a footprint spacing of about 700 m. This system allowed to achieve a full waveform with a vertical resolution of 1.5 m. Figure 2.5 visualizes a footprint of the SLA system together with footprints of other systems.

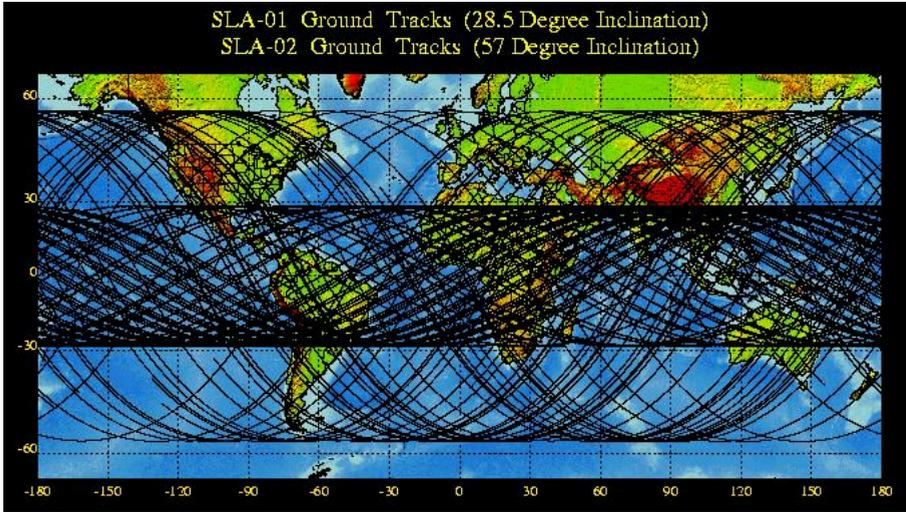


Figure 2.3: All ground tracks from the two SLA missions. SLA-01 operating with a 28.5 degree orbit inclination and SLA-02 with a 57.9 degree inclination. Extracted from [http://denali.gsfc.nasa.gov/lapf\\_web/](http://denali.gsfc.nasa.gov/lapf_web/).

The error budget associated with the geolocation process is summarized in Table 2.1.

Table 2.1: Preliminary Error Budget for SLA.

Error Source	Radial Error (m)	Horizontal Error at the Surface (m)
Orbit	<1.5	<5
Attitude	<3.0	<SLA footprint
Altimeter bias	~0.2	<1.5

With all experiences of design and implementation of above, new full waveform capturing systems were developed with two goals: (i) extending applications of full waveform systems to lower operating altitude airborne systems. (ii) Improvement of the vertical resolution and the coverage of spaceborne systems. The first goal resulted in the development of SLICER (Scanning Lidar Imager of Canopies by Echo Recovery) and LVIS (Laser Vegetation Imaging Sensor). The second goal emerged in VCL (Vegetation Canopy Lidar) and GLAS (Geoscience Laser

Altimeter System). Detailed information on SLICER and LVIS will be given in the following section.

### 2.3.1 Scanning Lidar Imager of Canopies by Echo Recovery (SLICER)

Scanning Lidar Imager of Canopies by Echo Recovery (SLICER) was one of the first airborne full waveform scanning systems. SLICER was launched onboard a National Aeronautics and Space Administration (NASA) C130-Q aircraft and operated on 5 days between July, 18 and July, 30 in 1996. It was designed to acquire canopy height, vertical structure and terrain elevation. SLICER acquired data in the northern boreal forests of Canada to support the project considering the BOREal Ecosystem-Atmosphere Study (BOREAS) (Harding, 1998). In nominal operation, the flight altitude of SLICER was 4500 m, yielding footprints of 9 m in diameter. The laser pulses were emitted in a direction perpendicular to the aircraft flight direction. Five laser pulses were typically fired for each cross-track scan with 9 m cross-track spacing and approximately 10 m along-track spacing, see Figure 2.4. Data coverage of this system is not so large, as it has acquired only between three to ten flight lines. A flight line is typically 40 km long.

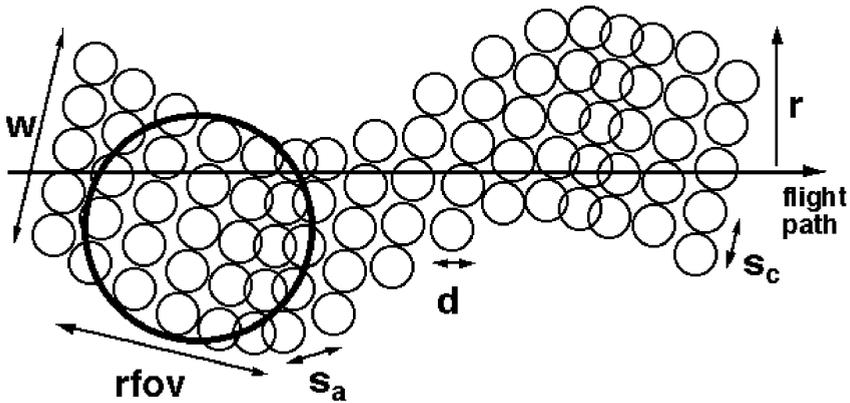
Canopy structure plays a fundamental role in controlling the exchanges of radiative energy, sensible heat, water, carbon dioxide, and trace gases between the surface and the lower atmosphere. Therefore the objectives of this system were to (1) provide a rapid and direct measurement of canopy structure and ground topography in the BOREAS study area which are not readily achievable by ground-based or other remote sensing techniques, and (2) utilize BOREAS ground-based canopy measurements to assess SLICER performance and capabilities.

The SLICER system has a bin resolution of about 0.11 m, and resampled at 0.44 m thereby taking a sum of four adjacent bins to improve the signal-to-noise ratio of the distribution. With 1200 bins in total, it yielded a waveform length of up to 120 m.

### 2.3.2 Laser Vegetation Imaging Sensor (LVIS)

Experience derived from SLICER was incorporated by Blair et al. (1999) in a completely redesigned, next generation scanning surface lidar system referred to as the Lidar Vegetation Imaging Sensor (LVIS). The LVIS was a second airborne full waveform laser altimetry system and incorporated a laser transmitter with a faster pulse repetition rate and a larger field of view. Therefore it could achieve much wider scan patterns. It stored also calibrated measurements of transmitted and received energy. The LVIS was primarily used for algorithm development and

## SLICER 5-beam Swath Geometry



$d$  = footprint diameter =  $\tan(\text{laser divergence}) \times \text{altitude above ground}$

$W$  = swath width = cross-track spacing  $\times$  cross-track footprints

$S_c$  = cross-track spacing =  $\tan(\text{scanner step angle} \pm \text{aircraft roll}) \times \text{altitude above ground}$

$S_a$  = along-track spacing =  $\text{ground speed} / (\text{laser rep rate} / \text{cross-track footprints})$

$r$  = roll-induced swath displacement =  $\tan(\text{aircraft roll}) \times \text{altitude above ground}$

$rfov$  = receiver ground field-of-view =  $\tan(\text{telescope angular fov}) \times \text{altitude above ground}$

Figure 2.4: SLICER 5-beam Swath Geometry. Extracted from [Harding et al. \(2000\)](#).

validation activities in support of the Vegetation Canopy Lidar (VCL) mission, the first mission selected by NASA's Earth System Science Pathfinder Program (ESSP). VCL was scheduled to be launched into Earth-orbit in 2000, however, this mission was canceled in the end due to problems with the science payload (see [space.skyrocket.de/doc.sdat/vcl.htm](http://space.skyrocket.de/doc.sdat/vcl.htm)).

LVIS was a medium-altitude imaging laser altimeter, designed and developed at NASA's Goddard Space Flight Center. In a flight configuration, it operated at an altitude of 8 km above the ground and obtained data in a swath of 1 km consisting of forty footprint shots, Figure 2.6. The footprints had 25 m diameter and were separated by 25 m along and 12.5 m across track. However, footprint diameters could vary from 1 to 70 m, and the footprint spacing could vary both along and across track. The transmitted pulse and the full waveform signal were digitally sampled at 30 cm vertical resolution along with the travel time of the pulse from the laser to the intercepted surface ([Blair et al., 1999](#); [GSFC, 2009](#)).

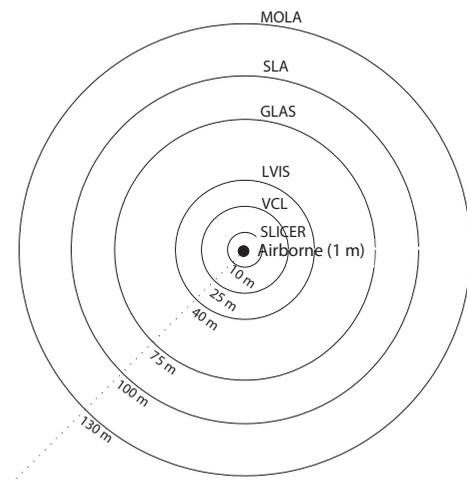


Figure 2.5: Relative footprint sizes of laser systems. A typical airborne footprint size of less than 1 m is displayed by the black dot.

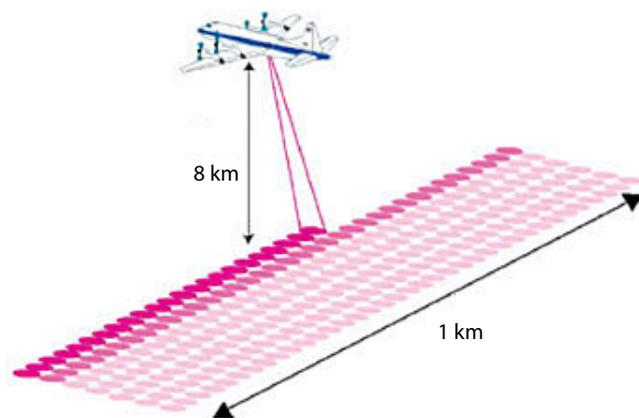


Figure 2.6: A flight configuration of the LVIS airborne laser altimeter. A 1-km wide swath was generated using forty 25-m wide footprints from 8 km above the ground. Smaller footprint sizes were possible, as were overlapping footprints or less dense sampling schemes. Extracted from Blair et al. (1999).

### 2.3.3 A typical commercial airborne full waveform scanning system (LMS-Q560)

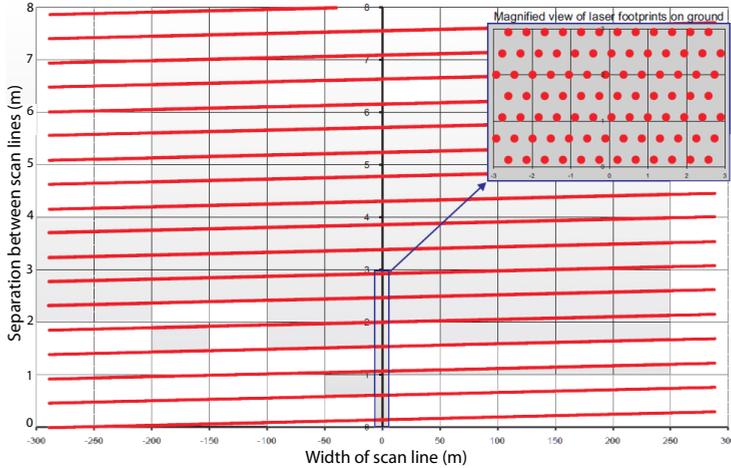


Figure 2.7: Riegl LMS-Q560 system.

A typical airborne full waveform scanning system is the Riegl LMS-Q560 system. In comparison to the previous full waveform sensors, this system has a lower operating altitude of up to 2 km above the ground, with a very high pulse repetition rate and especially a small footprint of about 0.5 m diameter. In addition, a digital camera is mounted on the aircraft to produce a very high resolution image of the scanned area. However, the usage of these systems is costly when data acquisition takes place over a large area, i.e. at regional or global scale. A scanning pattern from the Riegl airborne full waveform system is extracted from [http://www.riegl.com/uploads/tx\\_pxpriegldownloads/10\\_Datasheet\\_LMS-Q560\\_05-03-2009.pdf](http://www.riegl.com/uploads/tx_pxpriegldownloads/10_Datasheet_LMS-Q560_05-03-2009.pdf) and shown in Figure 2.7. The system results in a scan pattern of parallel lines. It is acquired at a flight altitude of 500 m with 120 scan lines/s and a field of view of 60 degree. This scan pattern results in a point spacing of 0.47 m, a width of scan line of 577 m, a distance between consecutive scan lines of 0.46 m and a point density of 4.2 points per squared meter.

Detailed characteristics of almost all full waveform systems are summarized in Table 2.2.

Table 2.2: Specifications of typical full waveform scanning systems.

Sensor	MOLA 1-2	SLA	SLICER	LVIS	GLAS	LMS-Q560	MarkII	ALTM3100
Operating altitude (km)	400	305	<5	<10	600	<1.5	<1	<2.5
Wavelength (nm)	1064	1064	1064	1064	1064	1064	1064	1064
Pulse width at half maximum (ns)	8	15-20	4	10	6	4	4	8
Pulse energy (mJ)	48	20-35	0.7	5	75	8	N/A	<0.2
Pulse firing rate (Hz)	10	10	<80	100-500	40	<10 <sup>5</sup>	<5×10 <sup>4</sup>	<5×10 <sup>4</sup>
Laser beam divergence (mrad)	0.4	0.25-0.40	2-10	8	0.11	0.5	1	0.3 or 0.8
Scan angle range(°)	Fixed at 0	N/A	N/A	±7	Fixed at 0 or 14	±22.5	Fixed at 20	±25
Scan rate (Hz)	-	-	-	-	-	5-160	-	-
Footprint size (m) at altitude (km)	130 305	100 5	≈10 5	40 8	50-100 600	0.5 1	1 1	0.3 or 0.8 1
Sampling rate(ns)	-	4-10	<1	2	1	1	1	1
Number of lasers	1	1	1	1	3	1	1	1
Operation Start date	1997	1995	July; 18-30	1995	2003	-	-	-
End date	2001	1998	1996	1998	present <sup>a</sup>	present	present	present

<sup>a</sup>This information is valid up to November 24, 2009.

## 2.4 ICESat/GLAS spaceborne laser scanning system

ICESat abbreviates Ice, Cloud and land Elevation Satellite. The ICESat spaceborne laser altimeter system was launched in January 2003 with the principal objective to measure polar ice-sheet elevation change, atmospheric profiles of cloud and aerosol properties, land topography profiles referenced to a global datum, and height of vegetation canopies (Zwally et al., 2002). These objectives, described in more detail below, are accomplished through the use of the Geoscience Laser Altimeter System (GLAS), in combination with precise orbit determination. GLAS uses a laser altimeter to measure the range between the satellite and the earth surface. The instrument time stamps each laser pulse emission, and measures the echo pulse waveform from the surface. Together with the spacecraft position and orientation, elevations are seen as a final product of ICESat mission. The ICESat system is visualized in Figure 2.8.

### 2.4.1 ICESat/GLAS Science objectives

ICESat/GLAS objectives are mainly focusing on three different aspects of the system Earth: polar regions, land and atmosphere. Objectives for every aspect are originally defined by Schutz et al. (1997) and presented verbatim in this thesis.

**Cryosphere (Polar regions):** The primary cryospheric science goals of GLAS are to measure long-term changes in the volumes (and mass) of the Greenland and Antarctic ice sheets to sufficient accuracy to assess their impact on global sea level, and to measure seasonal and inter-annual variability of the surface elevation in sufficient spatial and temporal detail to permit identification of long-term trends and to help explain those trends. A further goal is to provide precise elevation topography of these ice sheets and describe the nature of surface characteristics (e.g., roughness), including sea ice.

**Land Processes:** The primary land processes science goal of GLAS is to conduct topographic measurements of the Earth's land surface on a global basis in order to contribute to a global grid of ground control points for georeferencing of topographic maps and digital elevation models. The secondary land processes science goal is to detect topographic change at the meter per year level or better in selected regions of limited spatial extent.

**Atmospheric Science:** The primary atmospheric science goal of the GLAS cloud and aerosol measurement is to determine the radiative forcing and vertically resolved atmospheric heating rate due to cloud and aerosol by directly observing the vertical structure and magnitude of cloud and aerosol parameters

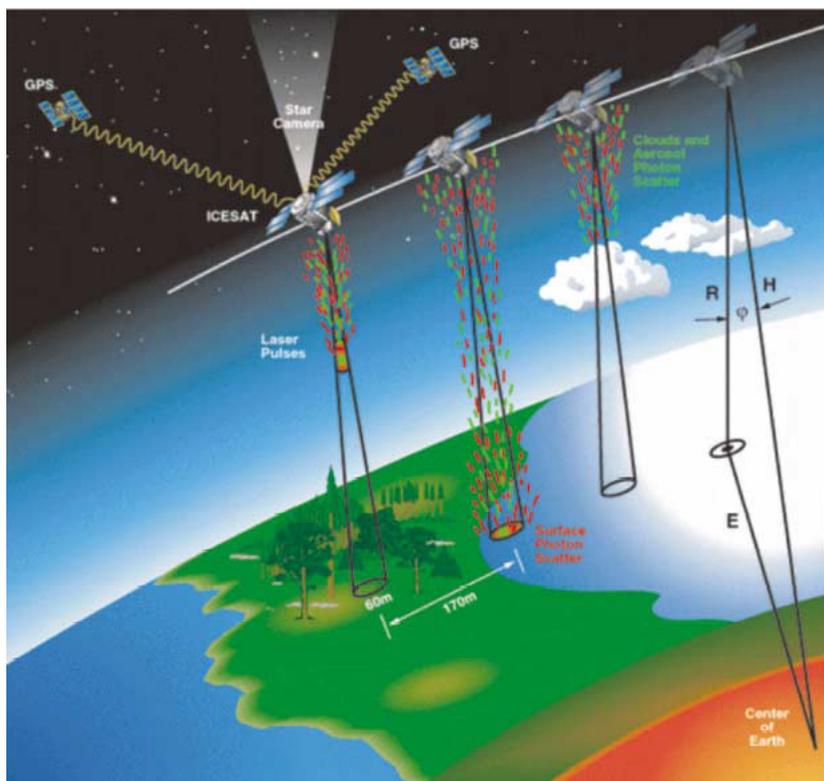


Figure 2.8: Principal of operation of ICESat. Extracted from Zwally et al. (2002).

that are important for the radiative balance of the earth-atmosphere system, but which are ambiguous or impossible to obtain from existing or planned passive remote sensors. A further goal is to directly measure the height of atmospheric transition layers (inversions) which are important for dynamics and mixing, the planetary boundary layer and lifting condensation level.

## 2.4.2 Overview on instruments/missions

The ICESat mission aims at fulfilling its science objectives using the GLAS instrument. GLAS was designed to operate at an altitude of 600 km above the ground with an inclination of 94 degrees. It should acquire elevation profiles of the entire earth along tracks that are revisited in a 183-day repeat cycle. GLAS has three identical lasers (designated Laser 1, 2 and 3) mounted on the ICESat satellite, with only one laser operating at a time (Zwally et al., 2002; Abshire et al., 2005).

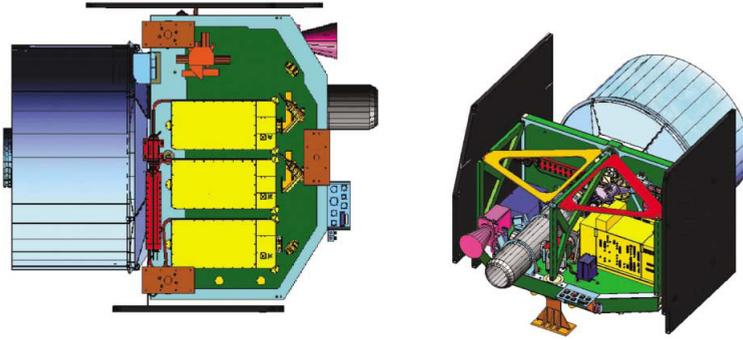


Figure 2.9: GLAS instrument showing on the left the 3 laser boxes (yellow) on the optical bench, the 1-m diameter telescope with heat pipe (red) and side radiators, and on the right the GLAS star tracker (pink), electronics boxes, and the small telescope (grey) of the stellar reference system. Figures extracted from [Zwally et al. \(2002\)](#).

A diode pumped Q-switched Nd:YAG laser produces a 1064 nm pulse for the measurement of surface topography, but a doubler crystal produces an additional 532 nm wavelength pulse for a more sensitive determination of the vertical distribution of clouds and aerosols ([Spinhirne et al., 2005](#)). The GLAS instrument is shown in Figure 2.9.

Laser 1 was commanded to start firing on February 20, 2003. However, this laser already failed on March 29, 2003. This failure resulted in a modified operating plan, which called for approximately 30 day operation periods, denoted laser “data campaigns” in the following, three times per year ([Abshire et al., 2005](#)). Laser 3 failed as well on October 19, 2008. Therefore, at the current stage of this thesis<sup>1</sup>, there is one working laser remaining, i.e. Laser 2.

The return pulse is first captured by a 1 m diameter telescope and directed to an analog detector. Then it is digitized by a 1 GHz sampler, along with a digitized record of the transmitted pulse. Each pair of digitized transmitted and corresponding echo pulse reflected from the ground is measured for analysis. These digitized pulses are referred to as laser waveforms. A waveform, recording laser back-scatter energy as a function of time, is digitized in 544 or 1000 consecutive bins at a temporal resolution of 1ns over land and 200 bins over sea for each footprint ([NSIDC, 2005](#)). The land waveform of 15 cm vertical resolution yields either an 81.6 m height range (544 waveform bins  $\times$  15 cm/bin) and or an 150 m height range (1000 bins  $\times$  15 cm/bin).

In general, GLAS transmits a laser pulse having a bell shape. The pulse has

<sup>1</sup>At November 24, 2009.

a maximum power at the center of the footprint and falls off at the footprint boundary by about  $1/e^2$ . Some typical ICESat footprints are shown in Figure 2.10.

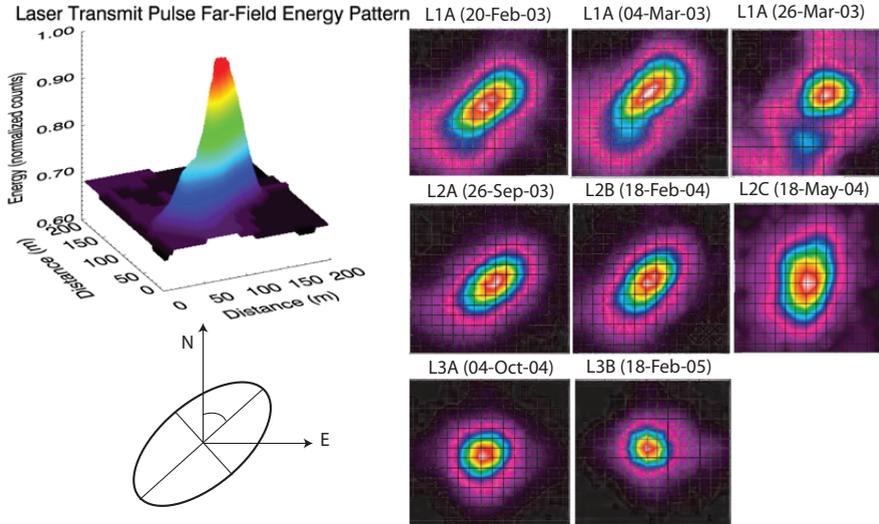


Figure 2.10: ICESat footprints are visualized. The 3D image (top left) is the power distribution within one footprint. The footprint shape and orientation (bottom left) are provided in some ICESat products like GLA06 and GLA14. Figures extracted from [Abshire et al. \(2005\)](#); [Lefsky \(2009\)](#).

### 2.4.3 ICESat track coverage

ICESat, operating at an orbit inclination of 94 degree, will acquire data from  $86^{\circ}\text{N}$  to  $86^{\circ}\text{S}$ . The GLAS acquires elevation profiles of the entire earth along tracks. The operating laser produces 40 transmitted pulses per second along a track. The resulting illuminated spots on the Earth's surface have an approximately 70 m diameter and the along-track spacing between two neighbouring spots is about 175 m. The separation between neighbouring tracks is a function of latitude as shown in Figure 2.11. The separation between two consecutive ascending or descending tracks is small when the ICESat mission moves to the poles and large when the ICESat passes above the equator. That means that data acquired at the poles are more dense than at the equator. For example, The Netherlands is located at a longitude of about  $53^{\circ}\text{N}$ . With a 91 day orbit, the smallest separation between tracks acquired by the ICESat mission is about 20 km. At the equator, the track separation is about 30 km.

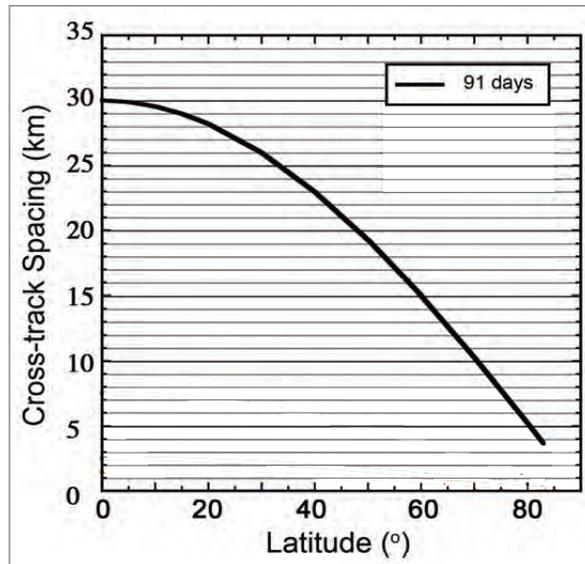


Figure 2.11: Track spacing (between ascending or descending track pairs) as a function of latitude. Figure extracted from (Abdalati et al., 2007).

ICESat elevation data acquired over the world shown in Figure 2.12. This data set has been acquired in 55 days between 25/09–19/11/2003 by Laser 2. It is also called data campaign L2a. A list of all data campaigns available for use is summarized in Table 2.4.

#### 2.4.4 ICESat data products and processing releases

##### ICESat data products

The ICESat mission has acquired a huge data set organized in 15 products. A description of each data product is summarized in Table 2.3. These products are created by the ICESat Science Investigator-led Processing System (I-SIPS) at the Goddard Space Flight Center. Within a few weeks after acquisition, the ICESat processed data are distributed at the national Snow and Ice Data Center (NSIDC). NSIDC will distribute data requested by the community of users in several ways. Users can order data by giving coordinates of regions of their interest. Other ways are described in Table A.1. Ordered data can then be downloaded via a given ftp link provided by the NSIDC or is delivered on DVD disk by post. The first case is the fastest way to obtain data. More information about data products can be accessed at <http://nsidc.org/data/icesat/data.html>.

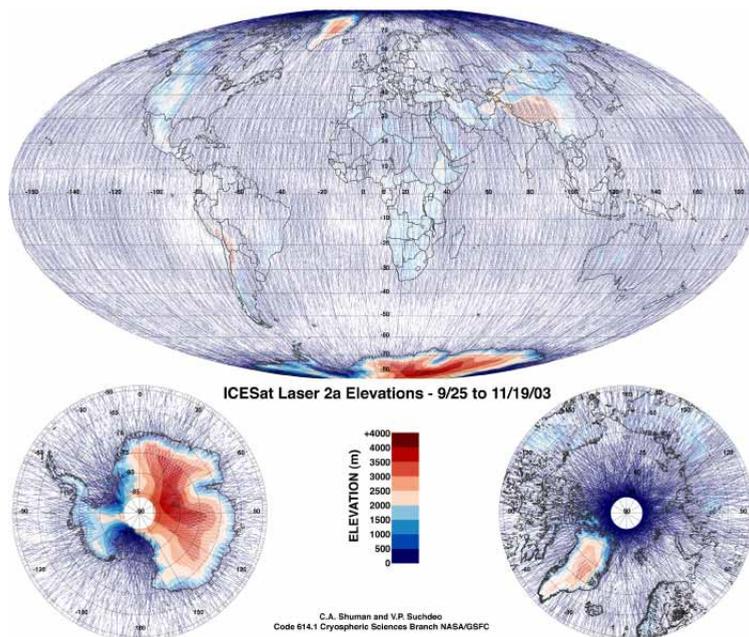


Figure 2.12: All ICESat tracks of data campaign L2a are visualized over the world, north and south polar regions. Figures are taken from [http://nsidc.org/data/icesat/laser2a\\_orbit.html](http://nsidc.org/data/icesat/laser2a_orbit.html).

### ICESat data processing releases

ICESat data are provided together with a version of the applied processing method. This is the so-called release number. A newer version has a higher release number. The newest release number 31 incorporates most comprehensive improvements and provides the most reliable data products for research activities. At the moment<sup>2</sup>, ICESat data products are provided in releases 12, 13, 14, 18, 19, 21, 22, 24, 26, 28, 29 and 31. The history of the releases is briefly described in the following, especially with respect to the full waveform (GLA01) and global elevation (GLA14) data products. Detailed information is given at [http://nsidc.org/data/icesat/past\\_releases.html](http://nsidc.org/data/icesat/past_releases.html). The changes in data processing are divided among the following major issues: (i) Saturation, (ii) Footprints, (iii) Return energy, and (iv) Waveform fitting.

**Saturation.** Saturation generally refers to the distortion of the received pulse shape due to saturation of the detector amplifier. This effect causes a range bias in the order of meters. In order to receive a good return signal, the gain

<sup>2</sup>At November 24, 2009.

Table 2.3: Standard GLAS data products.

Product name	Product long name	File size (MB)	Orbits per file	Files per day
GLA01	L1A Global Altimetry Data	9	1/4	56
GLA02	L1A Global Atmosphere Data	671	2	7
GLA03	L1A Global Engineering Data	19	2	7
GLA04	L1A Global Laser Pointing Data	2–386	2	4
GLA05	L1B Global Waveform-based Range Corrections Data	25	1/4	56
GLA06	L1B Global Elevation Data	7	1/4	56
GLA07	L1B Global Backscatter Data	827	2	7
GLA08	L2 Global Planetary Boundary Layer and Elevated Aerosol Layer Heights	7	14	1
GLA09	L2 Global Cloud Heights for Multi-layer Clouds	82	14	1
GLA10	L2 Global Aerosol Vertical Structure Data	289	14	1
GLA11	L2 Global Thin Cloud/Aerosol Optical Depths Data	13	14	1
GLA12	L2 Antarctic and Greenland Ice Sheet Altimetry Data	104	14	1
GLA13	L2 Sea Ice Altimetry Data	107	14	1
GLA14	L2 Global Land Surface Altimetry Data	209	14	1
GLA15	L2 Ocean Altimetry Data	279	14	1

mechanism is implemented to work as a signal adjustment/modulation within the GLAS instrument detector. When the energy of the return signal is low, the gain enhances the signal amplitude. If on the other hand the energy of the return signal is high, the gain will adjust the amplitude to a lower value. The gain provided in the ICESat products is in count units. The highest gain is 255 counts and the lowest gain is 0 counts. Saturation happens in recording a return waveform when the power of the return signal surpasses the maximum gain of the detector and stays there awhile before going down. Throughout the releases 12 to 31 methods for saturation identification and correction have been improved several times.

**Footprints.** The shape of footprints is defined by four parameters and provided in four flags: the transmit pulse intensity ([i\\_tpintensity\\_avg](#)), the footprint azimuth angle ([i\\_tpazimuth\\_avg](#)), the footprint eccentricity ([i\\_tpeccentricity\\_avg](#)), and the footprint major axis ([i\\_tpmajoraxis\\_avg](#)). This information is added to the waveform (GLA01) and elevation products (GLA12–GLA15) since release 19. However, incorrect footprint data are recognized in some campaigns and reported to the user community in release 28. For example, the reported footprint in data campaign L3a has a reported major axis that may be too small by a factor of 2 compared to other campaigns. Therefore it was advised to use the average size (55.8 m) and eccentricity (0.57) for this campaign. Such problems were announced to be

Table 2.4: GLAS data campaigns.

Data campaign	Year	Start date	End date	Durations (days)	Orbit (days)	Current Release*
L1A	2003	20/02	21/03	29	8	529
L1B		21/03	29/03	9	8	529
L2A		25/09	19/11	55	91	529
L2B	2004	17/02	21/03	34	91	529
L2C		18/05	21/06	35	91	428
L3A		03/10	08/11	37	91	428
L3B	2005	17/02	24/03	36	91	428
L3C		20/05	23/06	35	91	428
L3D		21/10	24/11	35	91	428
L3E	2006	22/02	27/03	34	91	531
L3F		24/05	26/06	33	91	531
L3G		25/10	27/11	34	91	531
L3H	2007	12/03	14/04	34	91	428
L3I		02/10	05/11	37	91	529
L3J	2008	17/02	21/03	34	91	529
L3K		04/10	19/10	15	91	N/A
L2D		25/11	17/12	19	91	N/A

\* This release is up to date of August 31st, 2009.

corrected in the next release. However, in release 31, the incorrectness of footprint data in campaign L1a was not solved yet.

**Return energy.** The return energy of an ICESat full waveform signal is basically the integral of the waveform. Two main problems with respect to the return energy occurred: (i) an error in the gain shift setting was found and reported in release 18. The gain used onboard for shot N+1 is contained in the telemetry for shot N. Therefore, energy parameters like [i\\_RecNrgAll\\_EU](#), [i\\_RecNrgLast\\_EU](#), [i\\_reflectUncorr](#), [i\\_reflectuncmxpk](#), are in error by an amount proportional to the shot-to-shot change in the received gain setting. (ii) effects of saturation on the computed received energy are not yet corrected. A correction method for future releases will consider the following concerns: (a) if the saturation energy correction is very large compared to the received energy, the data is questionable and often a large correction will result in reflectivity values greater than 1. (b) the saturation energy corrections that were determined using narrow waveforms may not be valid for wide over-land waveforms. (c) for about 20% of the land data there is saturation for which the current saturation energy correction does not provide a valid correction.

**Waveform fitting.** Two waveform fitting algorithms for ICESat products have been developed by the NASA processing team: a standard fitting method using one single Gaussian mode for the processing of data over polar regions, and an

alternate fitting method using multiple Gaussian modes for land regions. Some main issues concerning the alternative fitting method are: the threshold used to distinguish between noise and the actual waveform signal, the number of possible Gaussians, that is now controlled at a maximum of 6, and the reported standard deviation between waveform signal and fitted multi-Gaussian curve. Also the method of fitting a Gaussian to the transmitted pulse has been slightly adapted in newer releases. This change of method may result in a change of a few millimeter in the reported ranges and corresponding elevation results.

In general it is advised to sharply study the release information. Not only this guarantees that all available corrections will be applied, in also gives insight in ongoing, possibly yet unsolved issues concerning the processing of raw ICESat data.

## 2.5 Conclusion

In this chapter, an overview of full waveform laser systems has been presented. The general principal of laser full waveform systems and their sensor components are described. The ICESat mission and its GLAS laser system has been described in more detail as ICESat data products provide the major data source for the research performed in this thesis. Although ICESat data products have been processed carefully for about 5 years, some unexpected and unsolved issues occurred considering instruments (e.g. footprint determination and gain values), weather conditions (clouds) and surface characteristics (surface slope and roughness). To keep studying existing issues may help scientists to understand and obtain more accurate information from full waveform data.

At the moment, there are two types of full waveform systems in operation: those using airborne platforms (e.g. Riegl and TopEye Mark II) and one based on a spaceborne platform (ICESat). In near future, two follow up missions, ICESat-II and DesDyn-I will be designed and scheduled for launch. The ICESat-II will inherit knowledge and experiences of the current ICESat mission with the same measurement objectives, while the DesDyn-I mission will be a combination of two kinds of sensors, lidar and radar.



# Full waveform processing

## 3.1 Introduction

After the overview of full waveform instruments and their characteristics in chapter 2, in this chapter a step by step procedure is described, aiming at the processing of full waveform data. In this chapter, different scenarios of interaction between a laser pulse and the earth surface are discussed in section 3.2. In section 3.3, Gaussian decomposition is discussed. A mathematical model describing the full waveform signal and a detailed procedure to process a typical large footprint full waveform (ICESat) is presented. This procedure is based on decomposition of the waveform into suited Gaussian components. This procedure can also be applied to small footprint full waveform systems (Riegl). An alternative way to assess waveforms, so-called waveform deconvolution, is described in section 3.4. In section 3.5, a waveform simulation method using accurate airborne laser data is described. As a last step, a list of all waveform parameters that will be applied in following chapters is presented in section 3.6 and Table 3.1–3.3.

## 3.2 Interaction of the laser pulse with the surface

In general, a laser full waveform altimetry system transmits a signal having either a bell shape or a bit skewed shape with a sharp rise and a slow fall. The laser pulse of the ICESat system resembles a Gaussian function with a certain width, amplitude and peak location (mean value), see Figure 3.1. The signal intercepts with objects and the ground of the earth surface within the signal footprint. A

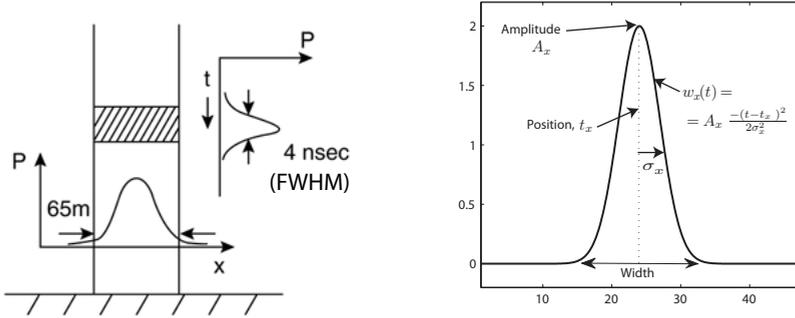


Figure 3.1: (a) An example of a transmitted pulse of the ICESat laser altimetry system. (b) A bell shape or Gaussian function used to describe the transmitted pulse with peak position  $t_x$ , width  $\sigma_x$  and amplitude  $A_x$ .

part of the energy reflects from non-terrain objects, like trees and buildings, and the remaining energy reflects from the ground. The return signal is the so-called full waveform.

A transmitted pulse is shown in Figure 3.1(a) and is modeled by a Gaussian function as shown in Figure 3.1(b). Let us have a look at some typical schematic examples, considering the shape of the full waveform signal as it is expected to return from different kind of surfaces. In Figure 3.2, it is assumed that the footprint covers all features in each scenario. It should be noted however that in reality the signal of the transmitted pulse is strongest in the middle and drops to the sides of the footprint. For the sake of simplicity it is assumed here that the emitted laser energy is distributed uniformly over the footprint location. When this emitted laser pulse intercepts the earth surface, the resulting return signal (full waveform) can be different in shape in the following cases:

- In Figure 3.2(a), a pulse having a Gaussian shape is intercepting a flat surface. The return waveform will have the same shape as the emitted pulse and can therefore be described by a Gaussian function as well. Because this surface is flat, the waveform width, denoted as  $\sigma_F$  is theoretically the same as the width  $\sigma_{Tx}$ , of the transmitted pulse. The return energy of the full waveform is always less than that of the transmitted pulse. The main reason is that only a very small part of the diffuse reflection is in the direction of the receiver.
- When a transmitted pulse intercepts a sloped/rough surface, the shape of the return waveform will theoretically be similar to the shape of the transmitted

pulse. However, the waveform width will be widened with respect to the width of the transmitted pulse. For example, if the transmitted pulse has a Gaussian shape as in Figure 3.1(b), the return waveform is assumed to have a Gaussian shape as well but with a larger width as shown in Figure 3.2(b)–(d). Let  $\sigma_S$  denote the width of the Gaussian resulting from the interception of the emitted pulse with a smooth sloped surface. Similarly  $\sigma_R$  denotes the width of the return waveform resulting from the interception with a rough surface, and, finally,  $\sigma_{SR}$  is the width of the waveform resulting from the interception with a surface that is both sloped and rough. Here it is assumed that the roughness and surface slope in these cases is the same in the sense that they result in the same widening effect. Then  $\sigma_{SR} > \sigma_S, \sigma_R > (\sigma_F = \sigma_{Tx})$ . In general, if it is assumed that the combined effect of slope and roughness results in a Gaussian of width  $\sigma_{surf}$ , it holds that  $\sigma_{SR}^2 = \sigma_{Tx}^2 + \sigma_{surf}^2$  which allows to quantify the combined effects of slope and roughness as  $\sigma_{Tx}$  and  $\sigma_{SR}$  are known.

From the previous examples, it follows that over simple surfaces the return waveform can be theoretically modeled as a single Gaussian function. The remaining question is what waveforms look like when a transmitted pulse intercepts more complex surfaces containing features like trees and buildings. When the feature size is smaller than the footprint size, the return waveform will contain additional peaks at the top part of the waveform resulting from the interception between the emitted pulse and the feature. The lower peak still describes the ground surface as in the previous examples in Figure 3.2(a)–(d). The additional peaks at the top contain information on the features like building and tree heights, Figure 3.2(e)–(l). In general, every additional peak can also be modeled as a single Gaussian function, and therefore the return waveform is assumed to be a sum of single peaks or Gaussian functions. In the following, some typical cases are described in more detail.

- For a tree over a flat surface as shown in Figure 3.2(e), the return waveform has two peaks, The ground peak is similar in shape to the transmitted pulse. The upper peak reflects the canopy structure (canopy depth) and the distance to the ground peak represents the tree height.
- For a tree over a sloped surface, the return waveform in fact depends on the location of the tree. If the tree is located at the top of the slope with respect to the footprint location, Figure 3.2(f), the ground peak has a larger width, compared to the ground peak in Figure 3.2(e). Moreover, the energy of the ground peak in Figure 3.2(f) and (e) is less than the ground energy in Figure 3.2(b) and (a). In this case the widening of the ground peak is mainly caused by the the effect of the surface slope on the peak width. If the tree, on the other hand, is located at the bottom of this slope, like in Figure 3.2(g), the separation between the surface slope and tree is somehow

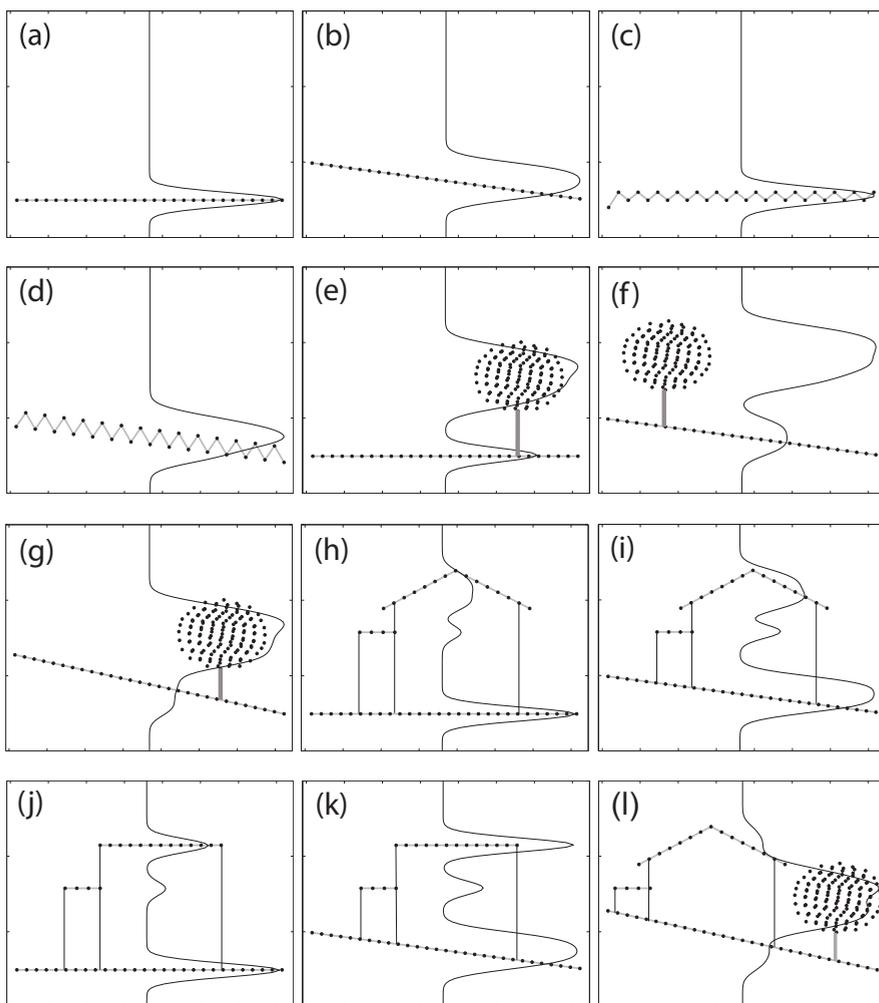


Figure 3.2: Different scenarios of waveform visualization. Waveforms are simulated from the elevation points (black dots).

not clear. Methods for extracting forest heights should therefore carefully consider the effect of the slope.

- For buildings, the slope of the roof structure has a strong influence on the shape of the top peaks. A tilted roof, Figure 3.2(h)–(i), results in a larger width compared to a flat roof, Figure 3.2(j)–(k). Therefore, one should be careful in interpreting a building height value extracted from a waveform.

- An example of a complicated case, Figure 3.2(1), is a waveform resulting from a footprint containing trees, buildings and a sloped surface. Ideally, a full waveform processing method can differentiate between the different surface components in the footprint location.

An alternative way of understanding full waveforms is by noting that the return waveform signal is in fact a convolution of the system response and the surface reflectivity function. The surface reflectivity function encodes both the reflectivity and the size of the reflectors. In this reflectivity function, the scattering properties according to a bidirectional reflectance distribution function are included. The system response is the transmitted pulse as recorded by the system. Therefore, in the following section, two waveform processing methods, Gaussian decomposition and waveform deconvolution are presented in order to extract a number of single peaks from a waveform.

### 3.3 Gaussian decomposition

Gaussian decomposition is studied by applying least square estimation to fit a curved line to the raw waveform signal that is assumed to be a sum of Gaussian components. The order of the processing steps is described in terms of ICESat full waveform data (GLA01 product). The decomposition method is applied to process both ICESat and Riegl full waveform data. Except for the decomposition result itself, also the quality of the decomposition is discussed.

#### 3.3.1 Mathematical model of transmitted pulse and return full waveform

##### Transmitted pulse

The transmitted pulse  $w_x(t)$  is assumed to have a bell shape (Brenner et al., 2003), therefore it is modeled as a Gaussian function as follows:

$$w_x(t) = A_x e^{-\frac{(t-T_x)^2}{2\sigma_x^2}} \quad (3.1)$$

where  $A_x$  is the amplitude of the transmitted pulse,  $T_x$  the mean value representing the peak location and  $\sigma_x$  the width of the transmitted pulse, compare Figure 3.1(b).

### Return full waveform

The return full waveform is assumed to be a sum of Gaussian components. Ideally, each Gaussian shape represents a signal component resulting from the interaction between the emitted pulse and the earth surface of a specific object within the footprint. Therefore the full waveform  $w(t)$  is modeled as sum of Gaussian functions and a noise component in the following way:

$$w(t) = \varepsilon + \sum_{m=1}^{N_p} w_m(t), \quad \text{with} \quad w_m = A_m e^{-\frac{(t-T_m)^2}{2\sigma_m^2}} \quad (3.2)$$

where  $w(t)$  is the waveform at time  $t$ ,  $\varepsilon$  a noise term,  $w_m(t)$  the contribution of the  $m^{\text{th}}$  Gaussian component,  $N_p$  the number of components,  $A_m$  the amplitude of the  $m^{\text{th}}$  component,  $t_m$  its position and  $\sigma_m$  its width. In order to extract every single Gaussian component from a full waveform, a fitting method will be employed.

However, some pre-processing steps need to be performed. The following steps are described for in particular ICESat full waveform data. How to order and obtain ICESat full waveform data is described in detail in appendix D.

### 3.3.2 Count to volt conversion

The binary data of the GLA01 full waveform product are first converted into ASCII format by an IDL program developed by the National Snow and Ice Data Center (NSIDC, 2006). The waveform data that is originally in counts (from 1 to 256) is converted into voltage units for further analysis. A lookup table between count and voltage units is available as a binary file, see appendix E for more details.

### 3.3.3 Waveform normalisation

The voltage waveform is first normalised to enable a comparison of waveforms captured in different epochs. For example, due to different atmospheric conditions or changes in the behaviour of the laser device, the amount of energy in the laser return pulse may vary with time, even if the terrain sampled by the laser pulse has not changed at all. These effects aggravated a comparison of absolute energy levels of particular constituents of different waveforms. The normalisation step requires a division of the received energy  $V_i$  by the total energy  $V_T$ , at moment  $i$  (see Equation 3.3). After normalisation, the area under any waveform equals 1 (see Figure 3.4).

$$V_N(i) = \frac{V_i}{V_T}, \quad \text{with} \quad V_T = \sum_{i=1}^N V_i \quad (3.3)$$

where  $N$  is the number of waveform bins. In this research  $N$  equals 544 or 1000.

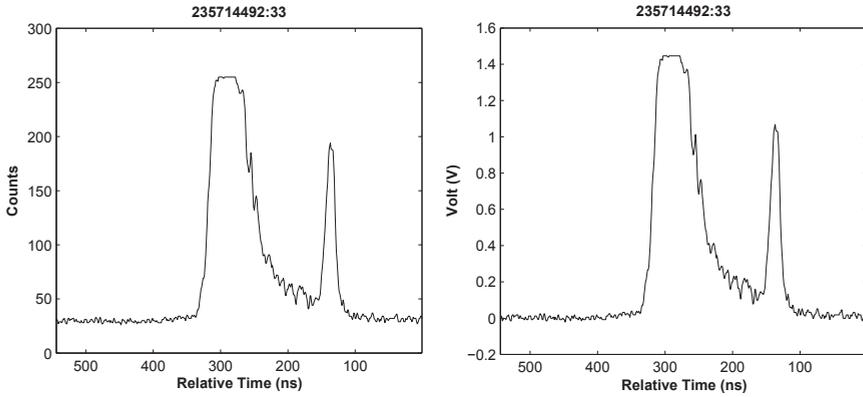


Figure 3.3: Waveform data in counts (left) and in voltage (right).

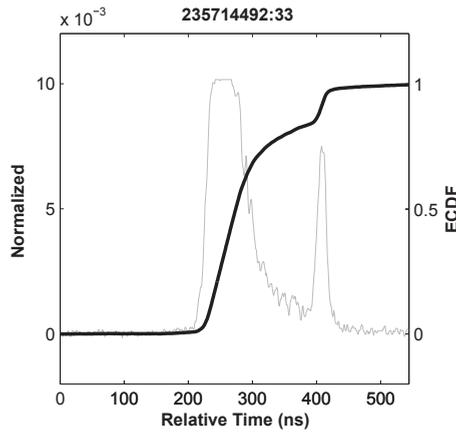


Figure 3.4: A normalised waveform (in grey) is displayed together with its cumulative distribution curve (in black).

### 3.3.4 Real waveform signal

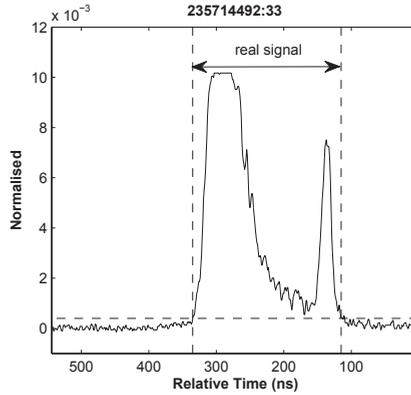


Figure 3.5: The real waveform signal is within the two vertical lines that are defined based on a threshold value (horizontal line).

The waveform detector system continuously measures the incoming signal level. Therefore a step is needed that extracts the actual waveform signal from this continuous time series of measurements. This step is implemented by considering when the signal level exceeds a noise level threshold. A waveform normally has 544 or 1000 bins over land that is

$$w(t) = V_t, \quad t = 1 \dots N \quad (N = 1000 \text{ or } 544) \quad (3.4)$$

where  $w(t)$  is the waveform,  $V_t$  is the amplitude of the  $t^{\text{th}}$  bin in the waveform, but by visualization of the data set, for example over The Netherlands, it can be seen that the actual waveform often starts after the 150th bin that is, after 150 ns. For this purpose, the mean  $m_N$  and standard deviation  $\sigma_N$  of the noise is determined from the first, say 150 ns range of the raw waveform signal, therefore,

$$m_N = \sum_{t=1}^{150} \frac{V_t}{150} \quad (3.5)$$

$$\sigma_N = \sqrt{\sum_{t=1}^{150} \frac{(V_t - m_N)^2}{150 - 1}} \quad (3.6)$$

Here it is assumed that the clipping of the received signal into individual waveforms is such that the first 150 ns of a waveform recorded only represents noise.

The threshold value  $n_N$  for distinguishing between noise and actual signal is determined as the mean plus four times the standard deviation (Lefsky et al., 2005) as described in Equation (3.7). The signal value at that part of the waveform that is below the threshold is set to zero (see Figure 3.5).

$$n_N = m_N + 4 * \sigma_N \quad (3.7)$$

Moreover, in this way the quality of waveforms can be evaluated and noisy/bad waveforms can be explicitly discarded. In Figure 3.6(a), the waveform is very noisy. When the threshold method described above is applied, there will not be any bins exceeding the threshold (dashed line). This waveform is then set to zeros and discarded. However, in Figure 3.6(b), the threshold value is smaller than the waveform values around the 500 ns bin. Although the waveform has a high noise level, still a dominant peak is visible that allows the extraction of a waveform signal using the threshold method.

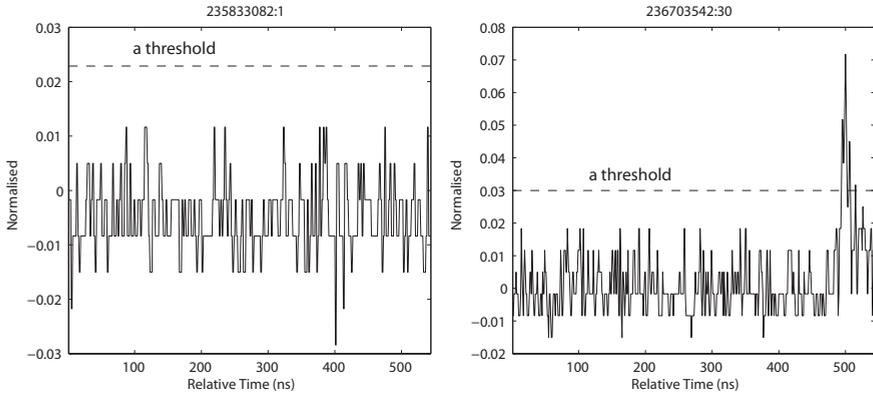


Figure 3.6: Example of a noisy waveform that will be discarded (a) and a rather noisy waveform that is still accepted for further processing (b).

### 3.3.5 Smoothing the waveform

Decomposition of waveforms using Gaussian components was proposed by Brenner et al. (2003). The methods for smoothing and the derivation of the number of modes as well as their approximate values were developed for this research. From the real waveform signal, initial parameters need to be identified first for the fitting step, more precisely the number of peaks or modes together with width, amplitude and location for each mode. Due to the noisy nature of many waveforms,

estimation of initial values from the raw signal results in a large number of modes with a low amplitude and a narrow width. Therefore, it is necessary to smooth the waveforms in order to get a smaller number of modes (Brenner et al., 2003).

The normalised waveforms are therefore smoothed, using a Gaussian filter. In this filter approach, weights for the available observations (waveform values) are obtained by the relative height of a Gaussian shape at an observation location (waveform bin). The Gaussian kernel is positioned such that its maximum coincides with the filtering location. Furthermore, the width of the Gaussian kernel is defined via the standard deviation ( $\sigma$ ) of the Gaussian function. However, when a Gaussian kernel is used for smoothing, it is common to describe the width of the kernel by the Full Width at Half Maximum ( $FWHM$ ). The  $FWHM$  is related to  $\sigma$  by the formula:  $FWHM = 2.35 \times \sigma$ .

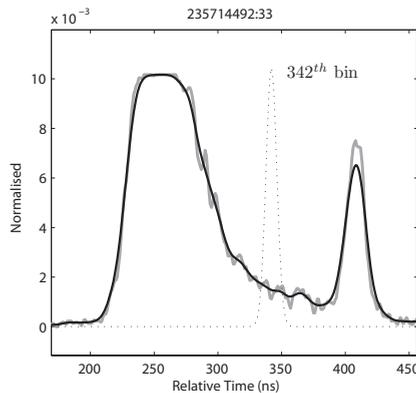


Figure 3.7: A waveform (grey line) smoothed by a Gaussian kernel (dotted line). The black line represents the smoothed waveform. The Gaussian kernel is located at the 342th bin to compute a smoothed value of the waveform at this position.

In Figure 3.7, a Gaussian function is used to smooth the waveform having a normal distribution with a  $FWHM$  value of 10 ns (dashed line). To smooth the raw waveform, this Gaussian kernel is moving from the beginning to the end of the raw waveform. For example, the Gaussian kernel (dashed line) at the 342th bin location is used to compute weight values for neighboring bins. The smoothed value of the 342th bin is obtained as a sum of the product of consecutive weight values and corresponding raw waveform values. The complete smoothed waveform is displayed by the solid black line. Following this step, the locations and amplitudes of the peaks in the smoothed waveform can be estimated.

### 3.3.6 Initial parameter estimation

To avoid small and noisy peaks, a search window of 75 cm wide (5 neighbouring height bins) is applied to estimate peak locations. The window moves from the beginning to the end of the waveform at steps of 1 vertical 15 cm height bin. If the waveform value at the middle of the window is higher than at the four other window positions, and if, in addition, the direct neighbours on the left and the right are higher than the two boundary points, then the centre position is considered as the location of a peak. The amplitude of the peak is extracted from the waveform at the peak location. Finally, an initial approximate mode width is calculated as half the distance between two neighbouring peaks. The minimal distance between neighbouring peaks is set at 1.5 m. Initial parameters of a waveform in the form of initial guesses for a Gaussian decomposition are shown in Figure 3.8.

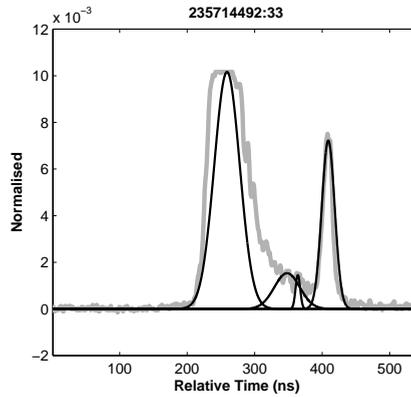


Figure 3.8: Waveform displayed in grey. Initial Gaussian decomposition into single modes is displayed in black.

### 3.3.7 Fitting algorithm

In the fitting step, Gaussian components as described in the previous sections are fitted to the normalised waveform  $w(t)$ . Every Gaussian component  $w_m$  corresponds to one Gaussian bell curve and the waveform  $w(t)$  is decomposed into a series of Gaussian components  $w_m$ , compare Equation (3.2).

A least-squares method is used to compute the model parameters. That is, the values for  $\varepsilon$ ,  $A_m$ ,  $T_m$ , and  $\sigma_m$ , for  $m = 1 \dots N_p$ , in Equation (3.2) are obtained by fitting the theoretical model to the observed waveform in such a way that the difference between model and observations is minimised in the least squares sense.

Alternative, (partly) similar fitting algorithms were described by Carabajal et al. (1999), Hofton et al. (2000), Brenner et al. (2003), Wagner et al. (2006), Persson et al. (2005), Jutzi et al. (2005) and Reitberger et al. (2006). Moreover, Mallet et al. (2009) describe methods using a larger choice of basis functions.

In order to fit Gaussian functions to the waveform, a nonlinear least squares estimation method is applied (Teunissen et al., 2005). Assume, that after determining the real part of the waveform, compare section 3.3.4, a raw waveform is given as

$$w_R(t) = V_t, \quad \text{for } t = t_1, \dots, t_n \quad (3.8)$$

with  $n \leq N$ , compare Equation (3.4), or equivalently, as

$$w_R(t) = \{(t_1, V_1), (t_1, V_2), \dots, (t_n, V_n)\} \quad (3.9)$$

If we write  $\underline{y} = (V_1, \dots, V_n)^T$  and

$$A(x) = \varepsilon + \sum_{i=1}^m w_i(t), \quad \text{with } w_i = A_i e^{\left(\frac{-(t-T_i)^2}{2\sigma_i^2}\right)} \quad (3.10)$$

then it is our wish, with

$$x = (\varepsilon, A_1, T_1, \sigma_1, \dots, A_m, T_m, \sigma_m) \quad (3.11)$$

to determine that solution  $\hat{x}$  such that

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|y - A(x)\|^2 \quad (3.12)$$

Solving system  $y \approx A(x)$  directly is not possible as  $A$  is non-linear, therefore, we consider the linearized system

$$\Delta y_0 \approx \partial_{x^T} A(x_0) \Delta x_0 \quad (3.13)$$

where  $\Delta y_0 = y - A(x_0)$ ,  $\Delta x_0 = x - x_0$  and  $x_0$  is an approximation of  $x$  that is the initial estimate derived in section 3.3.6. Here  $\partial_{x^T} A(x_0)$  is the  $n \times (3m+1)$  matrix of partial derivatives.

$$\begin{aligned} & \partial_{x^T} A(x_0) = \\ & \left[ \begin{array}{cccccccc} 1 & \frac{\partial w_1}{\partial A_1}(x_0) & \frac{\partial w_1}{\partial T_1}(x_0) & \frac{\partial w_1}{\partial \sigma_1}(x_0) & \dots & \frac{\partial w_1}{\partial A_m}(x_0) & \frac{\partial w_1}{\partial T_m}(x_0) & \frac{\partial w_1}{\partial \sigma_m}(x_0) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & \frac{\partial w_n}{\partial A_1}(x_0) & \frac{\partial w_n}{\partial T_1}(x_0) & \frac{\partial w_n}{\partial \sigma_1}(x_0) & \dots & \frac{\partial w_n}{\partial A_m}(x_0) & \frac{\partial w_n}{\partial T_m}(x_0) & \frac{\partial w_n}{\partial \sigma_m}(x_0) \end{array} \right] \quad (3.14) \end{aligned}$$

The partial derivatives of the  $i^{\text{th}}$  Gaussian function are described as follows.

$$\frac{\partial w_i}{\partial A_i} = e^{-\frac{(t-T_i)^2}{2\sigma_i^2}} \quad (3.15)$$

$$\frac{\partial w_i}{\partial T_i} = \frac{A_i(t-T_i)}{\sigma_i^2} e^{-\frac{(t-T_i)^2}{2\sigma_i^2}} \quad (3.16)$$

$$\frac{\partial w_i}{\partial \sigma_i} = \frac{A_i(t-T_i)^2}{\sigma_i^3} e^{-\frac{(t-T_i)^2}{2\sigma_i^2}} \quad (3.17)$$

From solving Equation (3.13),  $\Delta\hat{x}_0$  is obtained:

$$\Delta\hat{x}_0 = [(\partial_{x^T} A(x_0))^T (\partial_{x^T} A(x_0))]^{-1} (\partial_{x^T} A(x_0))^T \Delta y_0 \quad (3.18)$$

A least squares solution is  $x_0 + \Delta\hat{x}_0$ . Since  $A(x)$  is nonlinear, this solution is just expected to be closer to  $\hat{x}$ . This suggests that we take  $x_1 = x_0 + \Delta\hat{x}_0$  as our next approximation and that we solve the linear system  $\Delta y_1 \approx \partial_{x^T} A(x_1) \Delta x_1$  in the least squares sense again in order to obtain again an improved approximation of  $\hat{x}$ . By repeating this process a sufficient number of times, one hopefully will obtain a solution which is sufficiently close to  $\hat{x}$ . This iterative procedure will be terminated when the difference between two successful solutions is small enough. In general, iteration is described in Equation (3.19), and whole Gaussian decomposition procedure is summarized in Figure 3.9.

$$\Delta\hat{x}_k = [(\partial_{x^T} A(x_k))^T (\partial_{x^T} A(x_k))]^{-1} (\partial_{x^T} A(x_k))^T \Delta y_k, \quad k = 1, \dots \quad (3.19)$$

Our fitting method is implemented using the conditions described below.

1. The number of Gaussian components is limited to six.
2. The minimum distance between neighbouring peaks is 1.5 m.
3. The minimum amplitude of an individual peak equals the noise threshold value as described above.
4. The minimum sigma width of an individual peak is 30 cm.

For this research, the nonlinear curve-fitting (data-fitting) problem is solved in a least-squares sense and is implemented in Matlab (Coleman and Li, 1994, 1996;

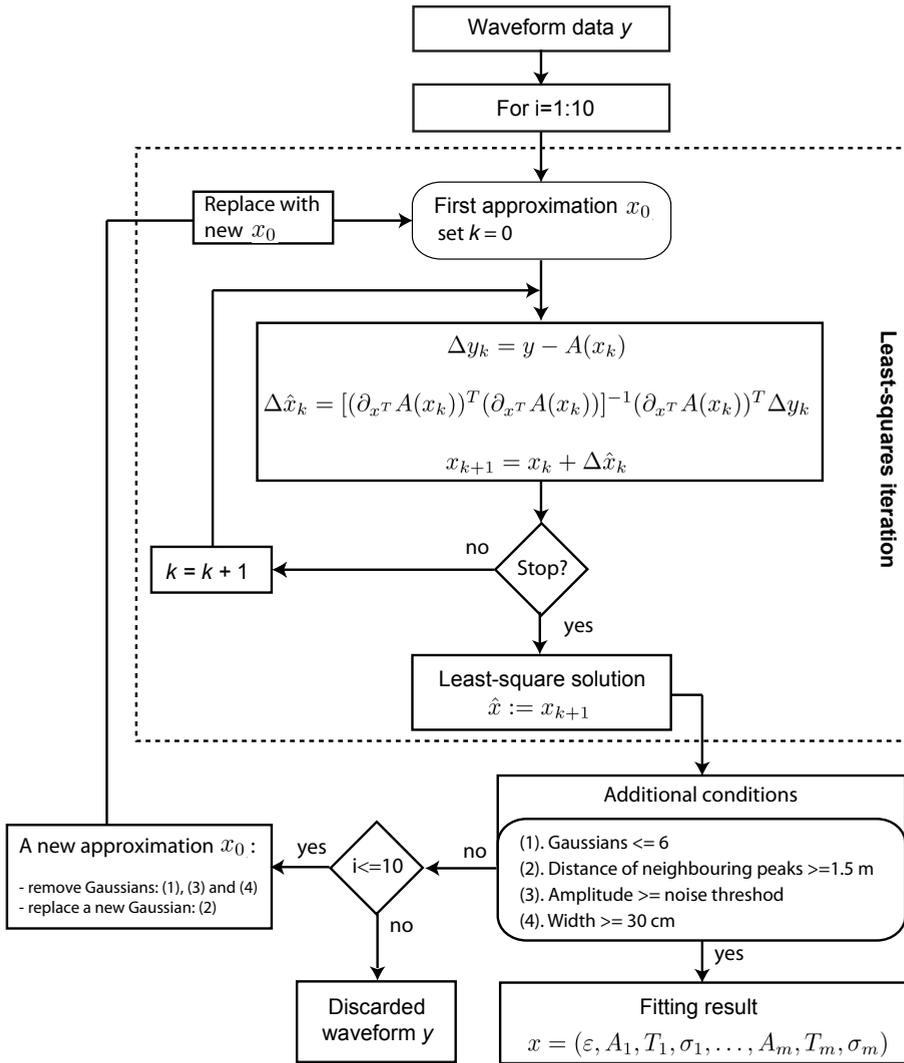


Figure 3.9: Gaussian decomposition is applied in a least square estimation sense.

Dennis, 1977; Levenberg, 1944; Marquardt, 1963; More, 1977). A maximum of 10 iterations is used in the fitting procedure, which is stopped when the result meets the criteria above. When the conditions are not met, the waveform is considered as not fittable and is discarded from further processing. Moreover, the residuals are used to quantify the quality of the fit. A result of the fitting algorithm is shown in Figure 3.10(a). Four modes were found in the waveform.

This fitting step can be also applied to the transmitted pulse described in Equation (3.1). Fitting a single Gaussian model to the transmitted pulse will result in the actual Gaussian parameters of this signal. A transmitted pulse and the fitted mode are shown in Figure 3.10(b). Moreover, the return waveform has sometimes a bit longer tail caused by cloud conditions. This effect is often referred to as the ringing effect (Fricker et al., 2005). In some cases this effect may result in modes that do not correspond to actual terrain features.

In order to demonstrate that the Gaussian decomposition method can be able to process small footprint airborne full waveform data, a test is made on airborne full

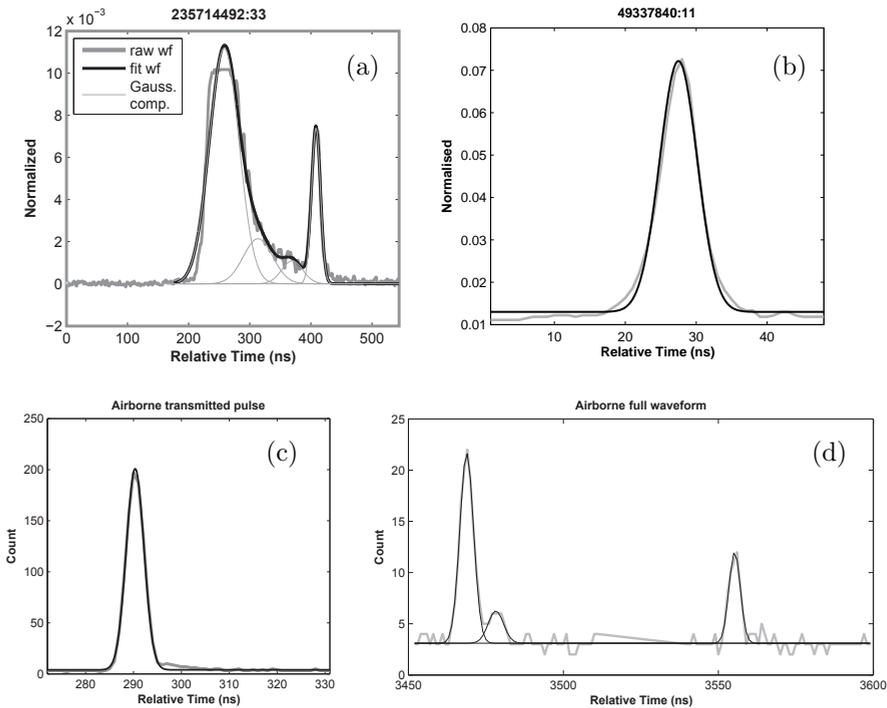


Figure 3.10: Fitting results: (a) The fitted ICESat waveform (black solid) is displayed together with the raw ICESat waveform (grey thick line) and the four Gaussian components (grey thin line). (b) The ICESat transmitted pulse is displayed in grey and the fitted Gaussian mode in black. (c) Airborne transmitted pulse and (d) Airborne return waveform (grey) with three Gaussian modes (black). The airborne data set is provided by Institute of Photogrammetry and Remote Sensing (IPF), Vienna University of Technology.

waveform data set acquired over the forest areas of Schoenbrunn, Austria in 2004 with the RIEGL LMS-Q560 airborne full waveform scanner. The flight height of this system is 500 m with 60% overlap between neighbouring trips. Using the decomposition conditions as described above, the Gaussian decomposition method still work well with this data set. Figure 3.10(c) and 3.10(d) show typical airborne full waveform data together with decomposed Gaussian modes.

### 3.3.8 Quality of fit

For validating the quality of fit, two noise levels are taken into account: local noise and global noise. For the local noise,  $\sigma_{N_i}$ , the noise in the first 150 bins of every raw waveform is used to calculate a standard deviation as described in Equation (3.6) (the index  $i$  stands for the  $i^{th}$  waveform). The global noise,  $\bar{\sigma}_N$ , is the mean local noise of the entire waveform data set,  $\bar{\sigma}_N = \frac{1}{M} \sum_1^M \sigma_{N_i}$ , where  $M$  is the number of waveforms in the data set.

Moreover, the remaining bins of the waveform are used to define two other notions that allow validation of the quality of fit:  $\sigma_{WF_i}$  and  $\bar{\sigma}_{WF}$ . The  $\sigma_{WF_i}$  is the local standard deviation of the  $i^{th}$  waveform and is defined as the mean of the bin-wise differences between the actual waveform  $w_i(t)$  and the sum of the fitted Gaussian components  $w_f(t)$ , as described in Equation (3.22).

$$w_d(t) = w_i(t) - w_f(t), \quad t = 1 \dots N \quad (N = 1000 \text{ or } 544) \quad (3.20)$$

$$m_{WF_i} = \sum_{t=1}^N \frac{w_d(t)}{N} \quad (3.21)$$

$$\sigma_{WF_i} = \sqrt{\sum_{t=1}^N \frac{(w_d(t) - m_{WF_i})^2}{N - 1}} \quad (3.22)$$

where  $w_d(t)$  is the bin wise difference between the raw waveform and the fitted waveform,  $\sigma_{WF_i}$  is the standard deviation of the  $i^{th}$  waveform.

The actual waveform is that part of the waveform between the begin point and end point in Figure 3.5, compare section 3.3.4. The parameters  $\sigma_{N_i}$  and  $\sigma_{WF_i}$  are depicted in Figure 3.11. Figure 3.11 shows the location of the actual waveform as bounded by the dashed lines. The  $\bar{\sigma}_{WF}$  is the mean waveform standard deviation of the entire data set,  $\bar{\sigma}_{WF} = \frac{1}{M} \sum_1^M \sigma_{WF_i}$ .

To access the quality of fit of the waveform, the four described parameters of above are used: the noise standard deviation  $\sigma_{N_i}$ , the waveform standard deviation  $\sigma_{WF_i}$ , the mean noise standard deviation, and the mean waveform standard deviation, where  $N$  is the number of waveforms in the entire data set.

For implementation of the quality of fit, an ICESat full waveform data set over The

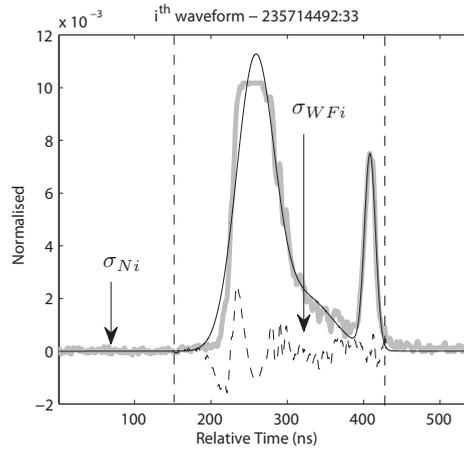


Figure 3.11: The noise level  $\sigma_{N_i}$  of a waveform is computed where the deviations occur in the range between 1 to 150 nanoseconds. The waveform level  $\sigma_{WF_i}$  is determined by the difference (dashed line) between the raw waveform and the fitted line.

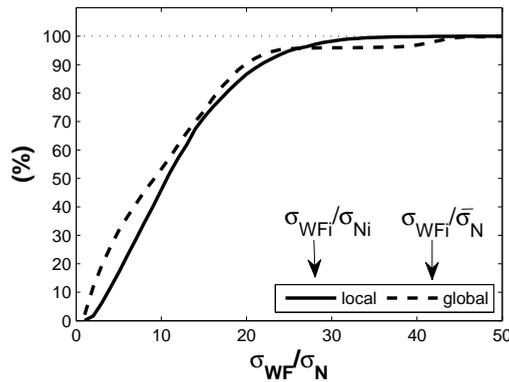


Figure 3.12: Quality of fit in terms of the noise standard deviation: the solid curve (local) and dashed curve (global).

Netherlands is studied. More than 3000 waveforms are fitted and their local and global standard deviations from the fitting results are computed. This validation result is displayed in Figure 3.12. The x-axis is the ratio between  $\sigma_{WF_i}$  and  $\sigma_{N_i}$ . The y-axis is the percentage of waveforms fitted. The continuous and dashed black curves represent the local and global level, respectively. Figure 3.12 shows that

the results for the quality of fit using either local or global noise are similar.

In the ideal case, when the local ratio,  $\frac{\sigma_{WF_i}}{\sigma_{N_i}}$ , is approaching 1, the residuals remaining after the waveform fitting are at the same level as the background noise, so one could say that the signal in the waveform could be well described by Gaussians. In practice, however, this ratio is much larger than 1. This may be caused by the following reasons: (i) In reality, the number of modes in the waveform may be larger than six. (ii) The modes may not be Gaussian. (iii) The noise level may not represent well the actual noise in the waveform. Figure 3.12 shows that in 95% of the cases,  $\sigma_{WF_i}$  is fitted well within 25 times  $\sigma_{N_i}$  or within 23 times  $\bar{\sigma}_N$ . Therefore, selection of the local and global noise will give the same result in validating the quality of fit. Moreover, in terms of mean standard deviations of waveform and noise, we get  $\sigma_{WF} = 10.8\bar{\sigma}_N$ . Strictly speaking, this result may not be that useful. But this analysis allows to compare different fitting methods using different base functions like generalized Gaussian, Wiebull, Nakagami or Burr (Mallet et al., 2009).

### 3.4 Waveform deconvolution

Waveform deconvolution is a technique to remove the system response for isolating the actual surface function. In this way it is possible to identify the time at which different object/scatterers are hit and consecutively their relative height (Jutzi and Stilla, 2006b; Neuenschwander, 2008; Molijn, 2009). In the time domain, the return waveform,  $w(t)$ , is a convolution of the surface response,  $h(t)$ , and the system response,  $w_x(t)$ :

$$w(t) = h(t) \bullet w_x(t) \quad (3.23)$$

The convolution operator is denoted by  $\bullet$ . Because the return waveform,  $w(t)$ , and the system response,  $w_x(t)$  are recorded (transmitted pulse), the surface response can be obtained. By transforming Equation (3.23) to the frequency domain, the function becomes a simple multiplication operation:

$$W(f) = H(f)W_x(f) \quad (3.24)$$

where  $W(f)$  is the return waveform,  $H(f)$  the surface response and  $W_x(f)$  the system response in the frequency domain.

In an optimal case without the presence of noise, the response surface can be obtained from above equation. In practice, however, noise is present in the return waveform and the system response, see Figure 3.13(a)–3.13(b). This noise

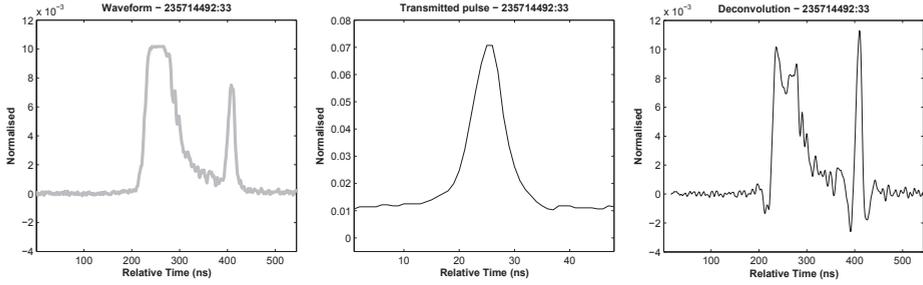


Figure 3.13: (a) ICESat waveform, (b) ICESat transmitted pulse and (c) Deconvolution result.

will cause a large numerical error or a infeasible solution in the surface response estimation, in case Equation (3.24) is directly used to estimate  $H(f)$ . To solve this problem, both noise terms and noise filtering techniques should be taken into account.

The noise term in the time domain,  $n(t)$ , can be obtained from the raw waveform,  $w(t)$ , from for example, the first 150 bins of the waveform as described in section 3.3.4. The noise term is then transformed to the frequency domain,  $N(f)$ . In order to filter the noise out of the waveform, a Wiener filter is applied. The Wiener filter function is described by Equation (3.25).

$$W_i(f) = \frac{|W(f)|^2}{|W(f)|^2 + |N(f)|^2} \quad (3.25)$$

Multiplying the Wiener filter and Equation (3.24), the surface response is obtained as follows:

$$H(f) = \frac{W(f)}{W_x(f)} W_i(f) \quad (3.26)$$

Here,  $H(f)$  again represents the estimated surface response function in the frequency domain. After applying an inverse fast-Fourier transformation, the estimated surface response function is transformed into the time domain,  $h(t)$ , see Figure 3.13(c).

One of the advantages of this deconvolution method is that no a priori information like initial parameters are required as in the situation described in section 3.3.6. In comparison to the Gaussian decomposition, the deconvolution method results in a more rough line, the so-called surface response function. Negative peaks may

occur in the result and a large number of separated peaks can be found. Negative peaks are caused by large values of  $N(f)$ . In contrast, if the value of  $N(f)$  is too high, the deconvolved waveform will be low and may miss some reflecting objects. The surface response resulting from the deconvolution method can be further used to extract information by defining other waveform parameters or by applying the Gaussian decomposition method to derive single Gaussian modes.

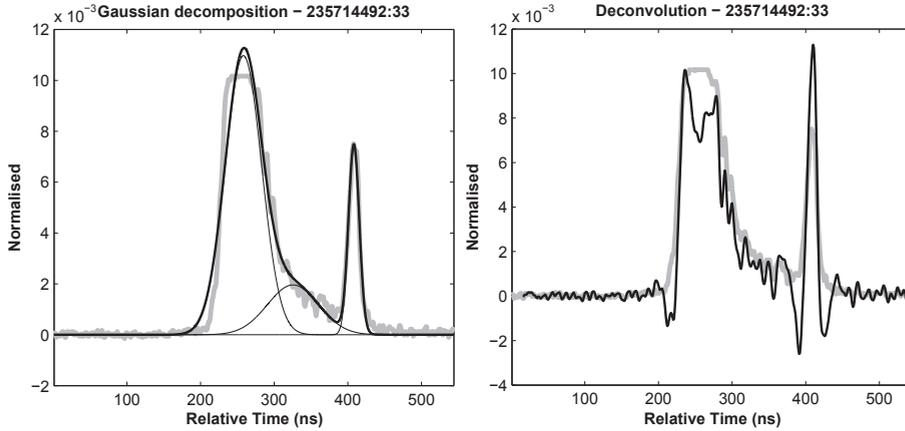


Figure 3.14: Results of a Gaussian decomposition (left) and deconvolution (right) are displayed in black line with the original ICESat waveform in grey.

In Figure 3.14, results of two decomposition algorithms are displayed. Figure 3.14(a) shows a Gaussian decomposition (thick black smooth line) with three Gaussian components (thin black lines). Meanwhile, Figure 3.14(b) shows a deconvolution result as a rough line with many separated peaks. According to (Neuenschwander et al., 2008), especially for applications in vegetation and urban areas, this deconvolution method is favorable for accurately measuring the difference between the top and the bottom of the object. For example, estimation of tree heights could be more accurate. Moreover, the separability between tree canopy and ground surface could be improved and the retrieval of vertical resolution is increased. A larger number of distinguished peaks provides a larger number of recognized objects along the line of sight within the illuminated footprint area. However, this method also depends on various complex relationships along the laser path's line-of-sight like surface topography, surface reflection, forward scattering, laser pointing, laser energy, footprint size, shape and orientation, and object height and position within the footprint.

## 3.5 Waveform simulation

As introduced in section 3.2, a full waveform signal is a convolution of a respond surface function and a system response. The system response is the transmitted pulse as recorded by the system. If a surface response is given or can be derived from other data sources like airborne laser data, a simulated waveform can be obtained.

For this thesis research, a waveform simulation method is implemented on Dutch national airborne laser data (AHN). The surface response can be approximated by the height histogram of the AHN points. A simulated waveform is obtained by a convolution of a given transmitted pulse and this histogram as in Equation (3.23).

With the purpose to develop a method to simulate ICESat full waveforms, a step by step waveform simulation procedure is described in the following. A bin size of 15 cm is used.

- ICESat footprint ellipses are used to select AHN points within the footprints, Figure 3.15(a)–(b). The footprint has an elliptical shape with major and minor size of approximately  $101 \text{ m} \times 46 \text{ m}$ , as extracted from the GLA01 data campaign L2a. Such footprint contains approximately 700 AHN points.
- The selected AHN points are used to generate a histogram of heights, Figure 3.15(c). In fact the histogram, representing the surface response, is convolved with the emitted pulse to result in a simulated waveform. However, the power distribution of the emitted pulse is a Gaussian function. The maximum energy is located at the footprint center and falls off at the footprint boundary by about  $1/e^2$ . Due to unavailability of intensity values for the airborne laser data, reflectivity of the laser points is assumed to be uniformly equal to one. Therefore, the energy reflecting back from the laser points is also following the Gaussian distribution. A power distribution corresponding to the Gaussian shape of the transmitted ICESat pulse within the footprint is applied to assign weight values to the AHN elevation points, Figure 3.15(d)–(e).
- The histogram of AHN weighted elevation frequencies of all AHN points within a footprint area is determined and denoted  $h(t)$ , Figure 3.15(f). The frequency of the ground peaks in the histogram in Figure 3.15(f) is decreased when the weight values are applied. This is because small weight values are assigned to the ground points at the footprint boundary. Therefore the total contribution of the ground points to the histogram became smaller than in case of Figure 3.15(c). If more attributes could be applied in this step, notably intensity data, the surface response approximation could be further improved.

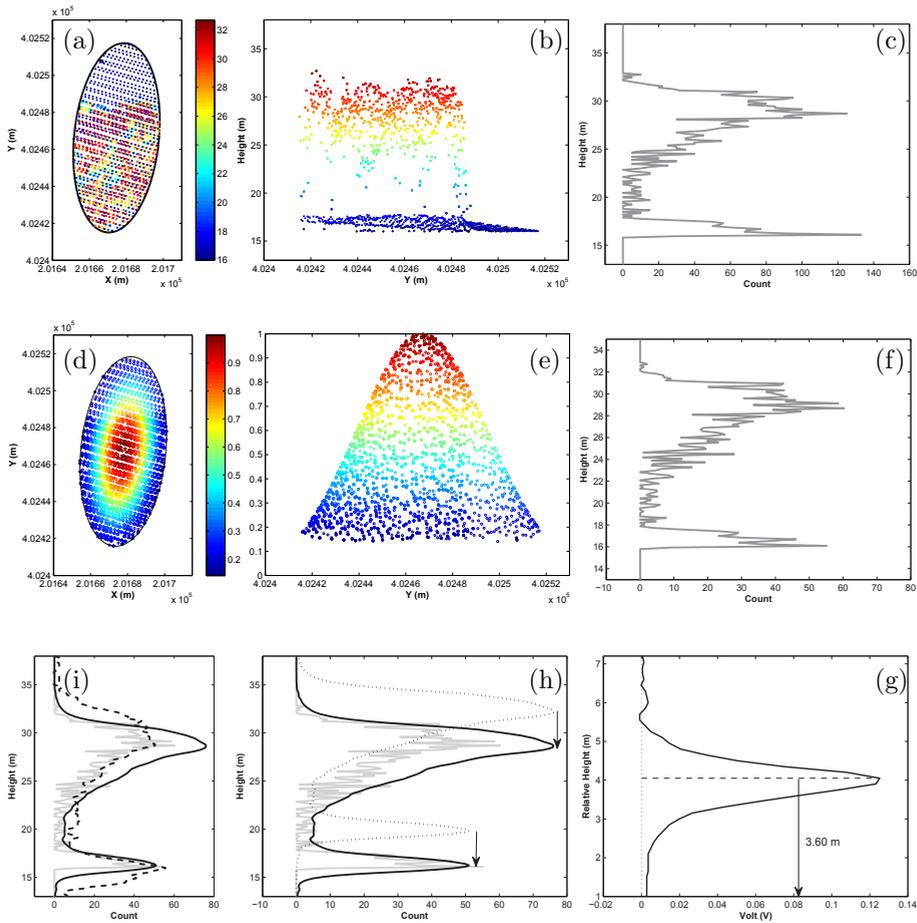


Figure 3.15: (a) Selected raw AHN data within ICESat footprint. (b) AHN data projected to YZ plane. (c) Histogram of raw AHN data. (d) Gaussian distribution of power of a transmitted pulse. (e) Power distribution projected to YX plane. The top color bar is the elevation scale for (a)–(b), and the middle one for (d)–(e). (f) The weighted histogram is derived by applying the Gaussian distribution of (d)–(e) to AHN points (a)–(b). (g) Emitted pulse as system response of the GLAS system. (h) Simulated waveform (thick black curve) displayed with the histogram (f) in grey and the convolved result (dotted curve). (i) ICESat waveform (dashed line) displayed with simulated waveform (solid line). Note that waveforms are scaled for the purpose of visualization.

- Next this histogram is convolved with the emitted pulse,  $w_x(t)$ , as shown

in Figure 3.15(g), resulting in a delayed signal displayed as a dotted curve in Figure 3.15(h). The georeferenced simulated waveform is the delayed waveform shifted downward by half the range of the transmitted pulse ( $24 \text{ ns} \times 0.15 \text{ m} = 3.6 \text{ m}$ ) as represented by the thick black curve in Figure 3.15(h). Compared to the weighted histogram, this simulated waveform is smoother.

In comparison to the ICESat waveform (dashed line) in Figure 3.15(i), the simulated waveform has two clear dominant peaks corresponding to canopy structure and ground. However, the width of the ICESat waveform is larger than the simulated waveform width. This can be explained by the conclusions from two case studies that will be presented below.

Two typical simulated waveforms are shown in Figure 3.16. The shape of the AHN simulated waveform (solid curve) over bare land is quite similar to the corresponding ICESat waveform (dashed curve), see Figure 3.16(a). However, when comparing ICESat and simulated AHN waveform examples over urban areas, Figure 3.16(b)–(c), differences in the modes can be observed respectively. The first mode represents the vegetation echo and the last mode the ground echo. Several factors may cause a difference between ICESat and AHN first mode: (i) Changes in the topography between two acquisition times; (ii) The AHN data contain only airborne laser points resulting from the first and last echo data. Reflections from intermediate scatterers were not available. This may explain why the energy of the ICESat waveform does not drop to the same low level as the simulated waveform; (iii) Contamination of the AHN data by multi-path effects or indirect reflections; (iv) The differences in the width of the last mode may also depend on the season. If most energy is reflected in the canopy, less is left for reflection from the ground. This also affects the amplitude of the last mode; (v) Albedo effects: in Figure 3.16(b), the result at a height of 5 m shows that the upper mode of the ICESat waveform is bigger than the corresponding mode of the simulated waveform. However, over inhomogeneous areas, differences in reflectivity of features or scatterers in the ICESat footprint may decrease or increase the amount of signal at a certain height in the ICESat waveform when compared to the AHN response, where laser points are only weighted with respect to the theoretical distribution of the energy in the corresponding emitted ICESat pulse. In Figure 3.16(b)–(c) two different kinds of objects are contributing to this first mode, the roof of a factory (horizontally aligned points at 5 m height) and trees. If the relative intensity of the AHN points would be available and could be taken into account, the roof points would contribute more to the first mode in Figure 3.16(b).

An alternative method of waveform simulation can be found in (Blair and Hofton, 1999). More advanced methods are found in (Abshire et al., 1994; Filin and Csathó, 2002). Moreover, an interesting simulation setup on a sphere is sketched by Jutzi and Stilla (2006a). Their simulation methods incorporate additional configuration parameters of the ICESat mission. For example, general input parameters like

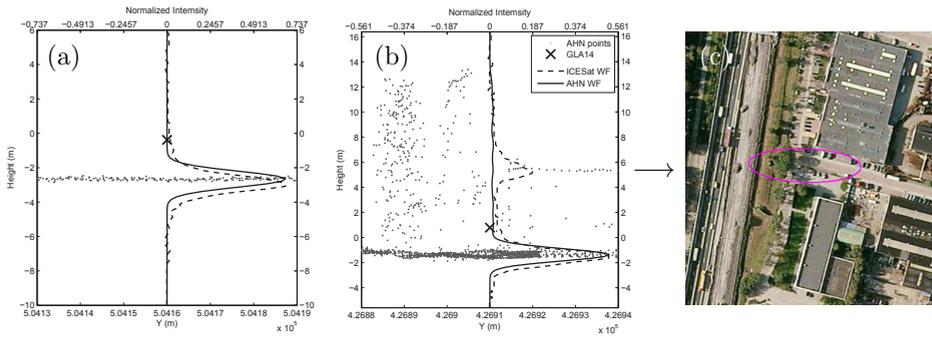


Figure 3.16: AHN simulated waveform (solid curve) and ICESat waveform (dashed curve) displayed with AHN points (grey dots) over: (a) bare land, (b) forest and (c) google image of (b) displayed with the ICESat footprint. The footprint partly covers the building (top) and high trees (left).

the flight altitude, the off-nadir pointing angle, the terrain elevation model and the surface albedo (optional) are included. At the receiver side, the area of the telescope, the bandwidth of the detector, gain, noise model, and sampling rate are also taken into account.

### 3.6 List of waveform parameters used

In section 3.3.7, it is described how waveform parameters like the number of modes, and attributes of each mode, as amplitude, width and location are derived during waveform processing. Several additional parameters derived from waveforms were identified in this research and in related work. Parameters of single waveform and parameters describing differences between corresponding waveforms are defined and listed in Table 3.1–3.2 and Table 3.3, respectively. Most used waveform parameters are the waveform extent ( $wEcho$ ), the height of median energy ( $HOME$ ) and the waveform distance ( $dEcho$ ). They are mainly used in forestry.

Table 3.1: Parameters of a single waveform and their physical explanation.

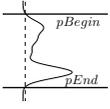
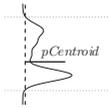
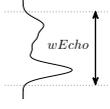
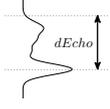
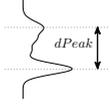
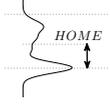
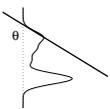
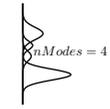
For a single waveform		
Parameters	Definition	Physical explanation & Visualization
Waveform begin ( $pBegin$ ), and waveform end ( $pEnd$ )	Position where the waveform first/last crosses above/below a threshold value (Lefsky et al., 2005; Sun et al., 2008).	The waveform begin presents the highest interception point between surface and transmitted pulse. The waveform end is the lowest elevation reflected from the earth surface. This parameter might be a good feature height estimator. 
Waveform centroid ( $pCentroid$ )	Position where the return energy is divided into two equal parts.	Mean elevation within the footprint (Carabajal and Harding, 2005b). 
Waveform extent ( $wEcho$ )	Distance between waveform begin and end.	Maximum tree height, maximum canopy height (Harding and Carabajal, 2005). 
Waveform distance ( $dEcho$ )	Distance from waveform begin to peak of the last Gaussian mode (e.g. ground return).	Top tree height, top canopy height (Sun et al., 2008; Drake et al., 2002). 
Peak distance ( $dPeak$ )	Distance between first and last peak.	Average tree height. 
Height of median energy ( $HOME$ )	Distance from peak of the ground return to the waveform centroid.	The HOME metric is sensitive to changes in the vertical arrangement of canopy elements and the degree of canopy openness (including tree density) (Drake et al., 2002), and in distinguishing between leaf-on/leaf-off condition (Duong et al., 2008). 
Height to Median ratio ( $rHTM$ )	$HOME$ divided by waveform distance ( $dEcho$ ).	$rHTM$ is also sensitive to changes in the canopy like $HOME$ (Drake et al., 2002).
Front Slope angle ( $aFSlope$ )	Angle from vertical to vector from waveform begin to peak of the canopy return energy (Ranson et al., 2004).	The variability of the upper canopy. 
Number of Gaussian fits ( $nModes$ )	The number of Gaussian modes derived from a decomposition method using non-linear least squares estimation	Describing number of height levels corresponding to objects and the earth surface. 

Table 3.2: Parameters of a single waveform and their physical explanation.

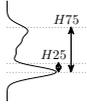
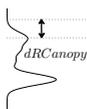
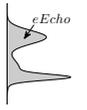
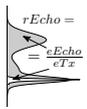
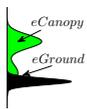
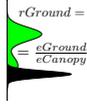
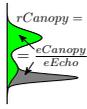
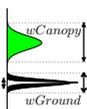
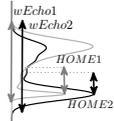
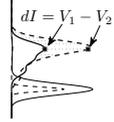
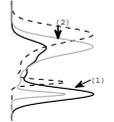
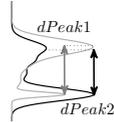
For a single waveform		
Parameters	Definition	Physical explanation & Visualization
$Hx$ , e.g. $H25$ , $H50$ , $H75$ and $H100$ .	$Hx$ is the x% quartile height and is calculated by subtracting the ground elevation from the elevation at which x% of the returned energy occurs. $H50$ equals the waveform centroid.	The heights (relative to the surface) below which 25%, 50%, 75% and 100% of the waveform energy is reflected (Sun et al., 2008). 
Roughness of Outer Canopy ( $d\_RouCanopy$ )Gaussian)	Distance from the waveform begin to the peak of the canopy (e.g. the first Gaussian)	Ruggedness of the uppermost canopy, spatial organization of plant surfaces within the canopy, and the decrease in laser energy with depth into the canopy as the pulse is intercepted by plant surfaces (Harding and Carabajal, 2005). 
Return waveform energy ( $eEcho$ )	The received energy i.e. the area below the waveform between begin and end.	Describing the surface characteristics in absolute values, e.g. reflectivity or land moisture conditions. 
Relative waveform energy ( $rEcho$ )	Return waveform energy ( $eEcho$ ) divided by the emitted waveform energy ( $eTx$ ).	Describing the surface characteristics in relative values, e.g. reflectivity or land moisture conditions. 
Ground energy ( $eGround$ ) and Canopy energy ( $eCanopy$ )	$eGround$ is the total intensity of the last mode (grey color). $eCanopy$ is the difference between return waveform energy ( $eEcho$ ) and ground return energy ( $eGround$ ).	Return energy of the ground and canopy. 
Ground return ratio ( $rGround$ )	Ground return energy ( $eGround$ ) divided by canopy return energy ( $eCanopy$ ).	An approximation of the degree of canopy closure (Drake et al., 2002). 
Canopy return ratio ( $rCanopy$ )	Canopy return energy ( $eCanopy$ ) divided by return waveform energy ( $eEcho$ ).	Measure of canopy cover (nadir-projected plant area versus total area) (Harding and Carabajal, 2005). 
Canopy width ( $wCanopy$ ) and Ground width ( $wGround$ )	Canopy and ground extent, relative to the transmitted pulse extent.	Measurement of crown depth (canopy top to lowest major branch) and topographic relief (Harding and Carabajal, 2005). 

Table 3.3: Parameters of repeated waveform pairs and their physical explanation.

Parameters	Definition	Physical explanation	Visualization
Difference	All waveform parameters above can be used in analyzing waveform pairs by taking a difference between parameter values.	<i>HOME</i> and <i>wEcho</i> were used by <a href="#">Duong et al. (2008, 2007)</a> .	
Canopy intensity difference ( $\Delta I$ )	The mean squared intensity difference of a pair of canopy returns.	Surface changes ( <a href="#">Duong et al., 2007</a> ).	
Correlation coefficients (CC)	Pearsons correlation coefficients ( <a href="#">Orfanidis, 1996</a> ).	Similarity of waveform shape pairs to detect forest fire or surface deformation.	
Ratio of last-mode to first-mode distances ( <i>rdPeak</i> )	The ratio between the distances from the last-mode to first-mode.	Detect changes of surface and forest heights ( <a href="#">Duong et al., 2006a</a> ).	

### 3.7 Conclusion

The main topic of this chapter is the processing and simulating of full waveforms. Two methods for waveform parameter identification are Gaussian decomposition and deconvolution. These methods can be applied to any kind of waveform data acquired from full waveform systems. Moreover, a large footprint waveform simulation method that is applied on airborne laser data, is also described. Therefore this chapter can be seen as a short tutorial for full waveform processing in general. In this thesis, ICESat data products are investigated, therefore some additional steps for processing in particular ICESat elevation products are described as well.

At the end of this chapter a complete list of waveform parameters is given, most of them directly derivable using Gaussian decomposition. Some parameters can in fact directly be derived from the raw waveform, like waveform extent. The list describes how waveform parameters are defined and physical meanings are also given. These parameters will be applied in the following chapters for two goals: validation and development of new applications.

The Gaussian decomposition method has been developed several years ago and has been applied in many research topics. On the other hand, the deconvolution method is just recently developed. Comparison of these two methods could help us to understand which method under which circumstance is the best to extract information from the waveform data, for example, forest structure.

# Validation of large footprint full waveform processing

## 4.1 Introduction

The topic of this chapter is the study of the quality of waveform derived parameters, as obtained from the waveform processing procedure described in chapter 3. This step gives insight into quality and quality bounds of waveform derived parameter values obtained under different circumstances. In section 4.2, an overview of existing literature on the calibration and validation of ICESat products is given. Next, in section 4.3, several ICESat waveform parameters and their physical meaning are compared to independent airborne laser scanning data with moderate point density over The Netherlands. For this purpose, elevation profiles of ICESat ground tracks (GLA14 and GLA01-derived data) are compared to corresponding elevation profiles of airborne laser data. The differences between the profiles are analyzed and displayed. It is shown that current state of the art ICESat full waveform processing is able to analyze waveforms at individual shot level. In section 4.4, analysis of ICESat full waveform pairs at overlapping footprints over the Europe continent is studied. How can changes in waveforms within one waveform pair be quantified and explained? For this purpose, a large database of tens of thousands of overlapping waveforms is presented. Differences in shape of waveforms obtained from the same location can be ordered in two categories: (i) if the acquisition time of the waveform pairs is approximately the same, shape differences can be caused by changes in settings in the GLAS system, changing weather condition and geolocation errors. (ii) if the time difference is large, the

shape difference can additionally be caused by seasonal influences and land cover changes.

## 4.2 Related work on ICESat data calibration and validation

### 4.2.1 Definition of calibration and validation

According to (Merriam-Webster, 1998), calibration is ‘*to standardize (as a measurement instrument) by determining the deviation from a standard so as to ascertain the proper correction factor*’ and according to (Random House Webster, 1997), validation is ‘*to make valid, substantiate, confirm; to give official sanction, confirmation or approval to*’. Therefore calibration primarily refers to the instrument configuration and its on-orbit operation by determination of calibration corrections of the on-orbit and prelaunch data. Validation refers to the science data products by making use of some additional data like GPS measurements, airborne laser data and imagery derived elevation data.

### 4.2.2 Error sources

A core product of the GLAS mission is a three-dimensional point, so-called geolocation. It is computed from three components: a spacecraft position  $G$ , a spacecraft attitude  $L$  and a round trip travel time of a laser pulse  $t_c$  (Figure 4.1). In principal, an elevation accuracy of about 13.8 cm can be obtained (Schutz, 2001), see details in Table 4.1. However, even after careful design and optimal on orbit operation, there are still underlying errors that are not fully described. Possible errors which affect the measurements of the three above components, are listed below (Schutz, 2001) and illustrated in Figure 4.1.

#### Instrumental and operational problems:

- Orbital variations as a function of latitude.
- Saturation of the return signal detector causes a negative bias in the elevation (Fricker et al., 2005), see Figure 4.9(c).
- Functional degradation of the oscillator/digitizer causes a drift, which contributes to the bias in the measured altitude.
- Range bias depending on altitude, time stamping, etc.

- Thermal influences on the measurements.
- Effects of yaw changes between sailboat and airplane mode. In the sailboat mode, the velocity vector is parallel to the solar panel axis; and in the airplane mode, the velocity vector is perpendicular to the solar panel axis.
- Effects of solar array induced jitter.
- Laser pointing angle.
- Changes in the instrumental configuration during operation. The operational laser has been switched from 1 to 2 to 3 and back to 2. Moreover, some changes occurred in the altimeter detector, the oscillator/digitizer and in the GPS receiver/antenna.

#### Atmospheric effects:

- Tropospheric delay by refraction. A correction is suggested by [Herring and Quinn \(1999\)](#).
- Forward scattering.

#### Surface conditions:

- Surface change due to solid Earth tides (proposed correction by [McCarthy \(1996\)](#)) and ocean-load tides ([Yi et al., 1999](#)).
- Slope and roughness.

Table 4.1: Single shot error budget for ICESat elevation measurements. Note that pointing error assumes 1.5 arcsec pointing error on  $1^\circ$  surface slope ([Schutz, 2001](#)).

Error Source	Error (cm, $1-\sigma$ )
GLAS range measurement precision	10
Radial orbit (POD)	5
Pointing determination (PAD)	7.5
Atmospheric delay	2
Atmospheric forward scattering	2
Other (tides, etc.)	1
RMS	13.8

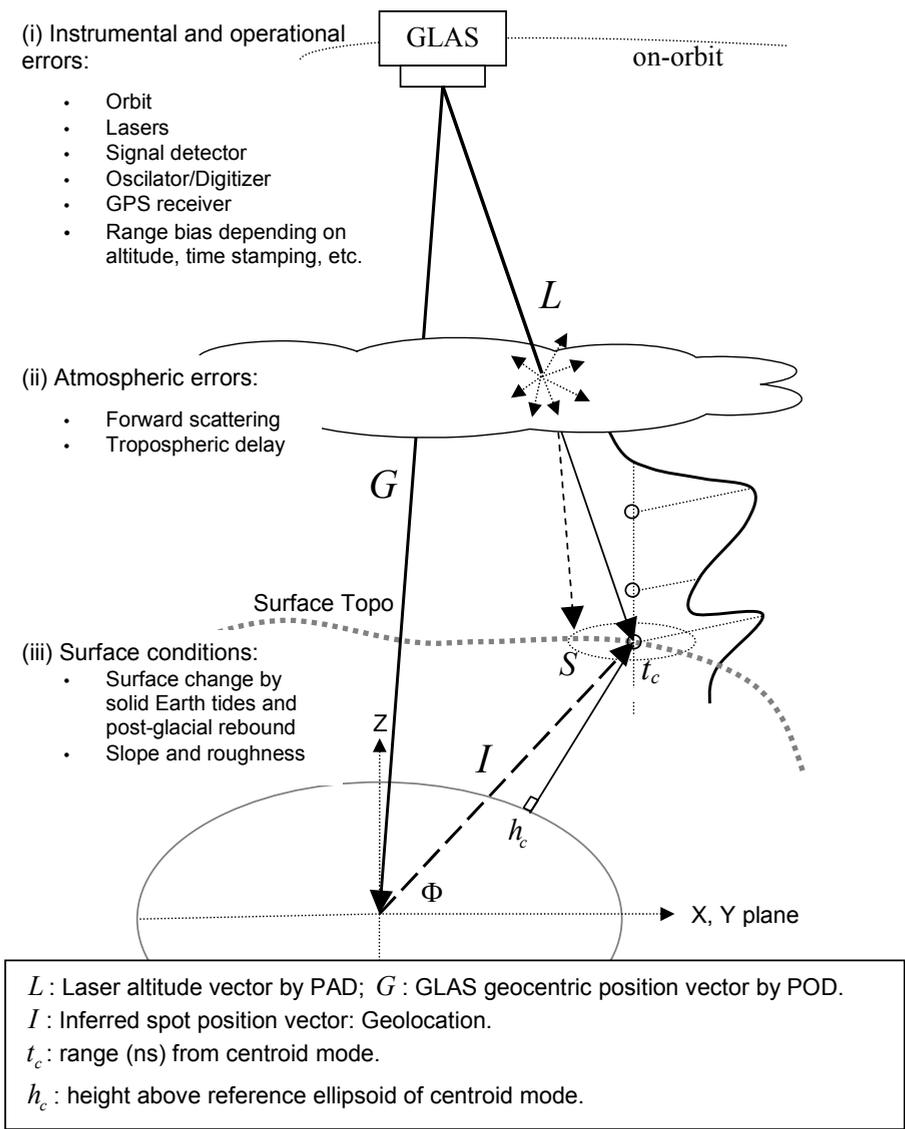


Figure 4.1: Three components contribute to the error budget of an individual laser shot.

Identifying and removing geophysical artifacts, and unexpected and anomalous effects from the data products are the fundamental objectives of the calibration/validation procedure. In fact, there are many relevant factors involved in

measuring the ICESat altitude as well as in determining its geolocation like (i) orbit, environmental and spacecraft conditions, and (ii) changes in instrument configuration. Some designed calibration/validation approaches are just able to minimize the influence of some contributions. There is no single method that can address all contributions. The error budget of ICESat measurements are summarized in Table 4.1 based on combination of errors from three above components.

### 4.2.3 Current calibration/validation approaches

To ensure appropriate geophysical interpretation of ICESat data products, independent calibration and validation techniques to analyze the above error sources were developed. However, other error sources still remain in the data set. Even after careful calibration of the participating instruments and after incorporating filtering constraints based on quality flags as available in the GLAS products, still possible and unexpected errors remain in the final data products. Validation procedures aim at identifying remaining systematic and random errors under different circumstances.

#### Systematic corrections

An approach for validation is performed to identify systematic errors. Luthcke et al. (2005) and Magruder et al. (2005, 2007) have developed methods to analyze orbital and instrument parameters for the reduction of ICESat systematic geolocation errors.

Luthcke et al. (2005) present a method to calibrate and correct for orbital variations together with long-term pointing errors of GLAS. An integrated residual analysis (Luthcke et al., 2002) which computes a range, using knowledge of the spacecraft position and pointing and timing parameters, is applied for the processing of ICESat range measurements. The difference between the computed range and the observed ICESat range from ocean scans and “round”-the-world scans is found as a function of the orbital variation and the bias trend of the systematic pointing angle. It is concluded that over ocean, for example, the systematic and random pointing error can be reduced from a mean value of 2.81 to 0 arcsec and a standard deviation of 3.99 to 0.94 arcsec. The resulting range bias is then stabilized at the centimeter level with a mean and standard deviation of  $23.37 \pm 0.61$  cm.

Another approach is to validate ICESat geolocation and time measurements (Magruder et al., 2005, 2007). At White Sands Space Harbor, an electro-optical detection system and a passive array of corner cube retro reflectors are arranged on the ground surface. This system validated the arrival time of the laser light of the ICESat footprint on the earth surface to  $3 \mu\text{sec} \pm 1 \mu\text{sec}$ . It satisfies the ICESat

mission requirements for timing precision, that were set at 0.1 msec. Moreover, a geolocation comparison between the GLA14 data and the computed data from the system results in a horizontal difference of  $10.5 \text{ m} \pm 4.5 \text{ m}$ . It corresponds to a pointing angle difference of  $3.5 \pm 1.5$  arcsec.

Such systematic validation is a good way to identify and if possible resolve errors in the calibration and processing of ICESat data products. However, in order to perform such validation, detailed technical information is required that is not always available in standard ICESat products. Therefore, it may not be possible to do such validation in practice. Instead, comparison between ICESat data and other reliable data sources is an alternative way of validation. Such validation of the quality of ICESat data can in general be performed in two directions: (1) by comparison to other reference data; (2) by internal comparison of ICESat elevation data. Below an overview of comparisons performed is given.

### **Comparison between ICESat elevation data (GLA06, GLA12 and GLA14) and high accuracy data**

The accuracy of ICESat elevation data is investigated by comparing/validating ICESat elevations to precise GPS measurements by Fricker et al. (2005), Atwood et al. (2007) and Braun and Fotopoulos (2007) as well as to airborne laser scanning data by Kurtz et al. (2008). A comparison between the resulting height differences is given in Table 4.2. The study areas mostly consist of relative flat topography. They consist of: (i) a polar coastal plain adjacent to the Arctic Ocean (Atwood et al., 2007), (ii) Arctic sea ice (Kurtz et al., 2008), and (iii) the largest salt flat in the world, Salar de Uyuni, Bolivia (Fricker et al., 2005). According to Fricker et al. (2005), under optimal conditions and after applying all corrections, ICESat elevations (GLA06) can be derived up to a vertical offset of less than 2 cm with a standard deviation of at least 3 cm. However, the accuracy of ICESat products depends also on other factors like the slopes and roughness of the terrain and its land cover type. In general it is expected that the accuracy of ICESat products will vary with respect to the land characteristics. Braun and Fotopoulos (2007) derived so-called ASCM elevations (Alberta Survey Control Monument) from different land based techniques including notably GPS leveling. They obtained a standard deviation of the difference between ICESat and ASCM of more than 2 m. There are two reasons for this large standard deviation. (i) Non-coinciding elevation pairs: within 100m radius of each ASCM point, the corresponding elevation of ICESat points (maximum of 2) is computed by taking the average, and (ii) This comparison takes place on arbitrary surfaces possibly including various land cover types.

Table 4.2: Differences in elevation between GLAS data and other data.

No.	Difference	Mean (m)	Std (m)	No. of footprints
Compared to high accuracy data				
1 <sup>a</sup>	GLA06–GPS	0.90	2.18	40
2 <sup>b</sup>	GLA06–GPS	−0.02	0.03	320
3 <sup>c</sup>	GLA06–GPS	−0.01	0.18	60
4 <sup>d</sup>	GLA12–ALS	−0.02	N/A	N/A
5 <sup>e</sup>	GLA06–ALS	0.07	0.01	288
6 <sup>f</sup>	GLA14–ALS centroid	0.12	1.01	2790
	Last mode–ALS lowest segment	−0.18	0.15	1249
Compared to moderate accuracy data				
1 <sup>g</sup>	GLA06–SRTM	1.60	5.00	40
2 <sup>h</sup>	SRTM–GLA06	−1.52	0.15	368
3 <sup>i</sup>	GLA14–SRTM	−0.14	3.65	7224
	Highest–SRTM	6.39	6.39	5371
	Lowest–SRTM	−6.06	5.73	5371
4 <sup>j</sup>	GLA12–InSAR	−0.49	2.86	1434162

<sup>a,g</sup> [Braun and Fotopoulos \(2007\)](#).

<sup>b</sup> [Fricker et al. \(2005\)](#).

<sup>c</sup> [Atwood et al. \(2007\)](#): The study area is Arctic Alaska. GPS measurements of five lakes are compared to ICESat GLA06 data. The study area is covered by snow and the data are acquired at the winter season. The large standard deviation is caused by snow conditions.

<sup>d</sup> [Kurtz et al. \(2008\)](#).

<sup>e</sup> [Magruder et al. \(2007\)](#).

<sup>f</sup> Duong et al. (200x): bare earth.

<sup>j</sup> [Bamber and Gomez-Dans \(2005\)](#): at a slope of 0.05 degree.

<sup>h</sup> [Bhang et al. \(2007\)](#): with respect to the bare earth surface.

<sup>i</sup> [Carabajal and Harding \(2005b, 2006\)](#): 0–20 percent tree cover, and  $\leq 5\text{m}$  roughness, Amazon and Western US data set.

### Comparison between ICESat elevation data (GLA06 and GLA12) and data of moderate accuracy with respect to land cover type

Notably, elevation differences between ICESat data and data from the Space Shuttle Radar Topography Mission (SRTM) were examined. Resulting height differences over open and flat terrain are summarized in Table 4.2. In addition, [Bhang et al. \(2007\)](#) show that the elevation difference vary with respect to land cover classes as derived from Landsat-7 images. As a first step, elevation differences with respect to land cover type (forest, agricultural land, the bare land, etc.) between ICESat and SRTM data were obtained. The elevation difference over bare land is about 1.5 m. In the second step, it was assumed that the difference over bare land equaled zero, thus this difference was subtracted from the other classes.

As a result, it was concluded that maximal differences of up to 2 m occur over forest and urban terrain. Differences decrease over wetland, agriculture and water in this order (0.74 m to  $-0.18$  m). Moreover, when comparing GLA12 data to two widely used digital elevation models of Antarctica, [Bamber and Gomez-Dans \(2005\)](#) show that the elevation difference increases with larger slope. Furthermore, [Carabajal and Harding \(2005b\)](#) state that height differences are notably large if the footprint location is located on high relief or densely vegetated surface. The state of relief is obtained from roughness estimates based on SRTM data, the amount of vegetation from the Moderate Imaging Spectrometer (MODIS) vegetation product ([Carabajal and Harding, 2006](#)). The SRTM elevation is usually located between the ICESat highest and lowest elevations, and on average is closely correlated with the ICESat centroid. Therefore, the standard deviation of the centroid differences is smaller than that of the highest and lowest elevation differences. Note that [Braun and Fotopoulos \(2007\)](#) reported a large standard deviation, compared to other authors. This is because this study did not compare elevations with respect to only the bare earth. In contrast, this comparison also included different land cover types, for example, high vegetation. Although ICESat accuracy can be described as a function of land cover type and relief (slope/roughness) after comparison to moderate accuracy elevations, the result is not fully satisfactory as ICESat elevations seem quite accurate by themselves. Therefore a comparison to high accuracy data is needed.

### **Comparison between ICESat full waveform data and simulated/synthesized full waveform data**

ICESat elevation is one processing result of the ICESat full waveform data (GLA01). For the purpose of the evaluation of the waveform quality, especially over complicated surfaces as vegetated areas and cities, a comparison between ICESat full waveform data and simulated waveforms derived from accurate airborne laser scanning data with high point density was employed by [Carabajal and Harding \(2005a\)](#), [Harding and Carabajal \(2005\)](#) and [Duong et al. \(2009\)](#). The airborne data within the ICESat footprint locations are selected and used to simulate a waveform by an approach of [Blair and Hofton \(1999\)](#), extended by [Harding and Carabajal \(2005\)](#). Similarity in shape of the corresponding waveforms is described by a cross-correlation method. Moreover, to assess the quality of ICESat geolocation, waveforms were simulated at 121 locations by [Harding and Carabajal \(2005\)](#). Waveforms were not only simulated from airborne data directly in the footprint: for each footprint 120 alternative locations were considered as well, as obtained by moving the original location to the North and the South by maximal 45 m at intervals of 9 m. At each location a waveform is simulated for a footprint centered at that location. The location where the correlation coefficient between ICESat waveform and simulated waveform is maximal, was defined as the best match. The distance between the ICESat location and the location of the best matching sim-

ulated waveform presents the geolocation difference. However, this method still revealed differences between corresponding waveforms. These can be caused by a combination of surface changes (e.g. vegetation coverage), weather conditions (e.g. clouded vs. cloud free), simulation method and differences in acquisition time (Harding and Carabajal, 2005).

In addition, waveform simulation is also performed on small footprint waveform data. Neuenschwander et al. (2006, 2008) synthesized small footprint waveform data to GLAS waveforms. Weight values are computed from the distance between all small footprint airborne full waveform locations within the ICESat footprint and the ICESat footprint center to account for the GLAS power distribution within the emitted pulse. The synthesized waveform is the sum of all the weighted small-footprint waveforms. Overall, the synthesized waveforms show a good qualitative fit to the ICESat waveforms, but the error of the ICESat waveform-derived elevation increases with the amount of wood cover. Moreover, Sun et al. (2008) show that the ICESat waveform is similar to the aggregation of LVIS 20m-footprint waveforms within the ICESat footprint. In fact, the simulation method only enables to show matches of good quality between GLAS waveforms and airborne synthesized waveforms over stable and simple surfaces like bare earth. Over for example urban areas, this comparison should solve for effects like varying reflectivity of the scatterers in the illuminated footprint, and for differences in configuration (e.g. looking angle) between the airborne and the ICESat system. Instead, analysis of ICESat full waveform pairs with completely/partly overlapping footprints possibly avoids those unsolved issues. A study of such ICESat waveform pairs may be a good way to identify both remaining errors in the ICESat products and unresolved issues in the processing of the ICESat products.

### **Analysis of ICESat elevation data at nearly repeated footprints:**

Slobbe et al. (2008) recently conducted a research on estimates of volume change rates of Greenland's ice sheet. Elevation differences of the repeated ICESat footprints are inferred to as the ice sheet change. Due to the not exact coincidence of footprint pairs, the footprint locations located on sloped surfaces result in an elevation bias. This slope effect is partly removed by reconstructing it from combining distances between the ICESat footprint locations with an available DEM of a resolution of 1 km. This research focussed at identifying differences at overlapping footprints corresponding to real elevation changes. Alternatively, differences found at locations that should be stable possibly point to unresolved processing issues.

## Goals of this chapter

To fully assess the application possibilities of especially the full waveform products, the next sections of this chapter will focus on two remaining research questions:

- Comparison between ICESat data (elevation and waveform) and accurate airborne laser scanning data over The Netherlands with respect to land cover type is firstly discussed. After the comparison, it is considered whether precise processing of waveform data at the individual laser shot used is possible.
- Can internal analysis of waveform pairs at overlapping footprints identify other errors from the ICESat system as well as temporal changes on the Earth surface?

## 4.3 Case study 1: ICESat full waveform altimetry compared to airborne laser scanning altimetry over The Netherlands

### 4.3.1 Background

For validating ICESat elevations a terrain with simple surface characteristics like sea ice, (Kurtz et al., 2008), or the Salar de Uyuni salt flats, Bolivia, (Fricker et al., 2005) is most suitable. Moreover, the ICESat elevations, i.e. the GLA06 and GLA12 products, are typically only obtained from the maximum peak of the full waveform. Consequently, the full waveform data itself and the techniques for processing such data have not been thoroughly investigated. Over land regions, the full waveform data were only processed and applied for estimation of forest structure and above ground biomass (Lefsky et al., 2002, 2005; Harding et al., 2001), canopy cover (Sun et al., 2008; Harding and Carabajal, 2005) and forest species inventory (Ranson et al., 2004). Therefore, in this case study, ICESat elevation (GLA14) data and ICESat full waveform (GLA01) data are processed and compared to the Dutch national high accuracy airborne laser scanning product AHN for a variety of land cover types.

The case study gives clear insight into the accuracies that can be expected over more complex terrain types. For understanding the nature of larger differences that occur in the comparison, an analysis of waveforms deviating at the individual shot level was necessary. Such analysis has enabled us to eliminate most outlying differences.

The objectives of this study are: (i) to propose an appropriate way to estimate the accuracy of ICESat data over complex land surfaces containing forests, buildings

and artificial objects; (ii) to estimate the accuracy with respect to different land cover classes: forest (broadleaf, mixed and needleleaf), urban and bare land/low vegetation; (iii) to investigate the choice of waveform parameters suited to compare feature heights of forests and buildings.

To do so, a new method of comparison between ICESat and laser scanning data (AHN over The Netherlands) is introduced. First, the study proposes an appropriate method to register the waveform data into a common coordinate system via information provided in the GLA14 product, leading to so-called georeferenced waveforms. From the georeferenced waveform later on the most suited position can be extracted to either represent the surface terrain height or the height of features like trees or buildings. On the other side, the many AHN laser points within an ICESat footprint are segmented to obtain one particular segment representing the surface terrain. Moreover, from the AHN laser points an AHN simulated waveform is created for two reasons: first, to enable a comparability check between ICESat and AHN waveforms and second, to enable the extraction of an AHN feature height to be compared to the corresponding ICESat waveform height.

Two comparisons between ICESat data and AHN data are studied in detail in this case, comprising of an accuracy assessment of complex terrain height and a feature height comparison. These comparisons are defined in detail in section 4.3.3. The results of the accuracy assessment are evaluated with respect to the different land cover classes according to the CORINE Land Cover 2000 database (CLC2000) (CORINE land cover 2000, 2006). In section 4.3.2, study areas and data sets are introduced. Then methodology is proposed. Results and future work are discussed at the end of the case study.

## 4.3.2 Study Area and Data

### (a) Study area

The area of study is The Netherlands, bounded approximately by  $3^{\circ}E$  to  $7^{\circ}E$  longitude and  $50^{\circ}N$  to  $54^{\circ}N$  latitude, which contains a large variety of land cover types. The total length of the six considered ICESat ground tracks is about 1290km. Figure 5.14 shows a map of the digital elevation model (AHN) of The Netherlands, colored by height together with the six tracks. In this case study, we differentiate results with respect to land cover type. For this purpose the CLC2000 data set is used.

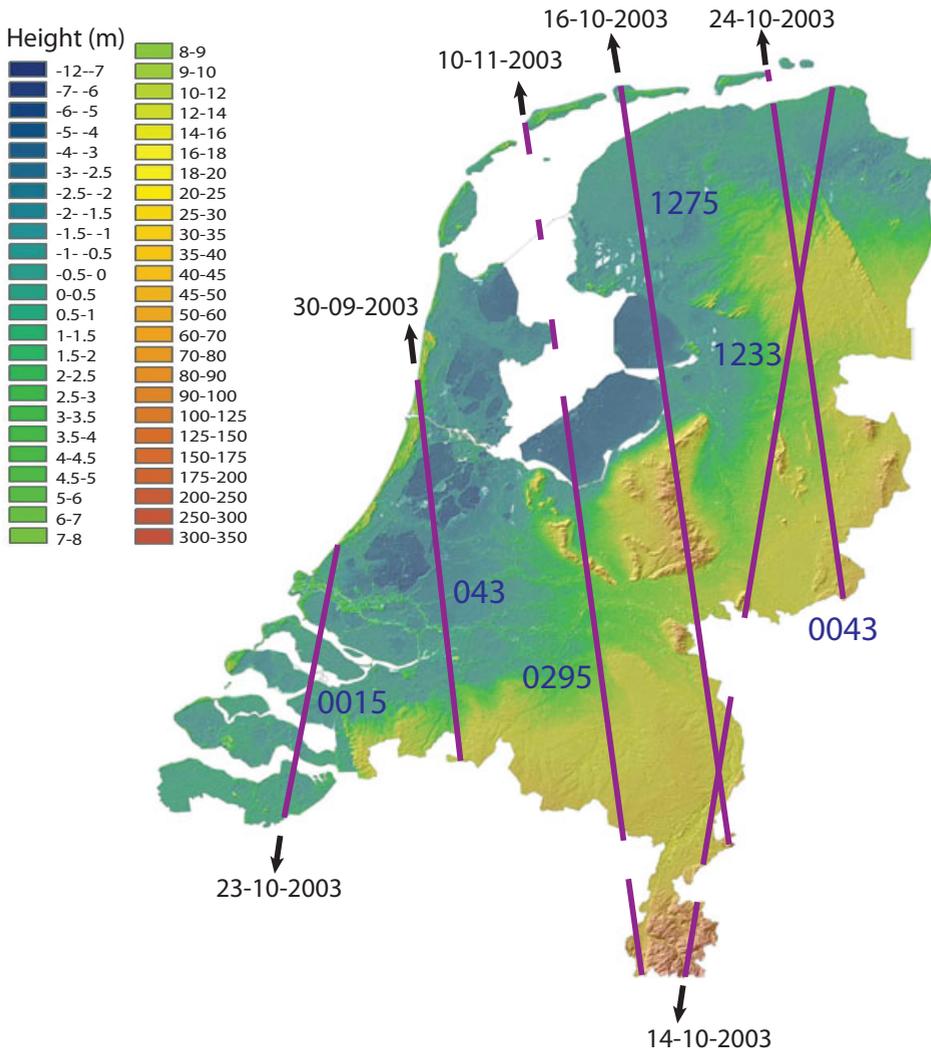


Figure 4.2: Study area: Six ICESat ground tracks (magenta) displayed with the actual height model of The Netherlands (AHN). The upward arrows indicate ascending tracks and the downward arrows descending tracks. ICESat reference track numbers are given in blue.

### (b) ICESat/GLAS

The GLA01 full waveform data are in general sampled as relative intensities in 200 bins for sea ice and ocean, and 544 bins over land and ice sheets. The bin

size is 1 nanosecond. The surface type is determined by the instrument from the on-board DEM.

The GLA14 global land surface elevations are obtained by combining precise orbit data (POD) (Rim and Schutz, 2002), precise altitude data (PAD) (Bae and Schutz, 2002) and range data. The range data are determined from a time stamp pair corresponding to the centroid of the transmitted pulse and the reference point, mostly the centroid, of the return waveform. After all instrumental, atmospherical and tidal corrections have been applied (Brenner et al., 2003), geolocated latitude, longitude and footprint elevation are computed (Schutz, 2002). The position of the reference point is stored as a range offset in GLA14. For coping with the potential complexities of land returns including possible combined influences of slope, roughness, vegetation and cultural features, offsets are provided in the GLA14 product, corresponding to alternative bin positions of the waveform, like the begin and end. Note that when the GLA14 elevation product is corresponding to the waveform centroid, it is representing the mean elevation within the illuminated footprint (Harding and Carabajal, 2005). In addition, ICESat waveforms that cause saturation of the ICESat detector, result in a lower elevation (Fricker et al., 2005). A saturation elevation correction, `i_satElevCorr`, is applied to all GLA14 data.

To avoid large changes in surface features and land cover because of acquisition time differences, the acquisition time of ICESat data needs to be close to the acquisition time of the AHN data (1996–2003) and the CLC2000 data (1999–2001). Therefore ICESat GLA14 and GLA01 products from campaign L2a, obtained in the period between 2003-09-25 and 2003-11-18, are chosen for this study. As a result, the difference in acquisition time between the data considered, varies from 0 to 7 years. All data are from release 428 and the waveform data were digitized in 544 bins. In Table 4.3 the orientation and the length of major and minor axes of the ellipses describing the footprint shape are given for each track.

These six ICESat ground tracks are chosen because of the following reasons: (i) To be well-spatially distributed over the study area, (ii) To cover all different land cover classes, (iii) ICESat measurements along these six tracks were relatively successful compared to other L2a tracks (cloud cover), (iv) For some of the tracks considered repeated tracks in subsequent campaigns are available (tracks 0015, 0043 and 0295) which allows to repeat this analysis for later campaigns. Moreover, waveforms from overlapping footprints from repeated tracks can be compared to assess terrain height changes, feature height changes and to identify possible remaining errors in the processing of ICESat data (Duong et al., 2007).

After applying filtering constraints as described later in section 4.3.3(i), a total of 3172 waveforms from six ICESat tracks were assigned to different land cover classes using the CLC2000 land cover database (Table 4.4). In Table 4.3 the transmitted energy dropped from 81 mJ to 66 mJ during campaign L2a. The average return

energy of the waveform from each track varies from 17 fJ to 316 fJ. Moreover, the nominal pointing angle is always about 0.3 degree. According to [Urban et al. \(2008\)](#), given the reported pointing error of  $0\pm 1.5$  arcsec in data campaign L2a, ICESat elevation data have a theoretically vertical accuracy of 2.25 cm per  $1^\circ$  incident angle and a horizontal accuracy of 4.5 m ([Urban et al., 2008](#)).

### (c) Actual Height Model of The Netherlands (AHN)

The actual height model of The Netherlands (AHN) was acquired between 1996 and 2003 in leaf-off conditions and is based on airborne laser altimetry, with a point density of at least  $1/16$  point/m<sup>2</sup>. There are four levels of detail available: raw point cloud data, and interpolated grid data at 5, 25 and 100 m resolution ([Heerd et al., 2000](#); [AHN, 2008](#)). In this study, the raw point cloud data are used. Over rural areas, the raw point cloud data are divided into non-ground points (so-called vegetation points) and ground points. Over urban areas no filtering is applied. Hence, both vegetation and buildings are present in the urban AHN data sets. All data is in ASCII format files with XYZ coordinates given in the Dutch coordinate system RDNAP (Rijksdriehoeksmeting and Normaal Amsterdams Peil, ([RDNAP, 2007](#))). The accuracy depends strongly on the amount of vegetation and topography. For solid surfaces (e.g., roads, parking lots), and soft but flat surfaces (e.g., beaches and grass-fields), the maximum systematic offset equals 5 cm with a standard deviation of 15 cm. Over wooded areas, the maximum systematic offset equals 10 cm with a standard deviation of 20 cm in case of at least one ground point per 36 square meter ([AHN, 2008](#)). Starting from 2007, AHN2 is being acquired

Table 4.3: Number of ICESat waveforms used with energy and footprint information.

Track	Date	Energy			Footprint	
No.	2003	Tx (mJ) <sup>a</sup>	Re (fJ) <sup>b</sup>	Mj (m) <sup>c</sup>	Mi (m) <sup>d</sup>	Ori (deg) <sup>e</sup>
043	30-Sep	81	158	87	42	184
1233	14-Oct	77	316	99	44	11
1275	16-Oct	76	88	102	43	184
0015	23-Oct	73	54	105	44	9
0043	24-Oct	72	17	106	44	185
0295	10-Nov	66	157	100	47	186

<sup>a</sup> Transmitted energy (mJ); <sup>b</sup> Return energy (fJ); <sup>c</sup> Major (m); <sup>d</sup> Minor (m); <sup>e</sup> Orientation (degree)

Table 4.4: Number of ICESat waveforms used: Forest (Broadleaf, mixed and needleleaf), Urban, Bare land, and Water.

Track	Date	Number of corresponding pairs between ICESat and AHN data						
No.	2003	Broad	Mixed	Needle	Urban	Bare land	Water	Total
043	30-Sep	29	5	15	100	298	6	453
1233	14-Oct	1	19	44	216	622	7	909
1275	16-Oct	29	73	154	61	379	6	702
0015	23-Oct	4	0	0	31	212	15	262
0043	24-Oct	0	1	0	12	253	0	266
0295	10-Nov	14	3	21	110	429	3	580
Total		77	101	234	530	2193	37	3172

over The Netherlands. AHN2 has even better specifications than the first version of AHN used in this study. The release of AHN2 will offer good possibilities to assess the quality of the most recent ICESat campaigns.

#### (d) CORINE Land Cover 2000 database (CLC2000)

The CORINE Land Cover 2000 database (CLC2000) was developed by the European Environment Agency and the European Joint Research Centre. The CLC2000 database originates from the year 2000 but is actually obtained during a 3-year period from 1999 to 2001, with a horizontal geolocation accuracy of 25 m based on satellite images of Landsat 7 ETM+ with 25 m pixel resolution. The CLC2000 data product is obtained from Landsat data via a computer-assisted visual interpretation of the satellite images, under the requirements of a scale of 1:100 000, a minimum mapping unit of 25 hectares and a pixel resolution of 100 m (Perdigão and Annovi, 1997). The CLC2000 classification is hierarchical and distinguishes 44 classes at the third level, 15 classes at the second level and 5 classes at the first level. Detailed information of land cover levels can be found on the metadata section on the European Environment Agency website (CORINE land cover 2000, 2006). The total thematic accuracy of the CLC2000 database was almost 95%. The database is geo-referenced in the European reference system (ERS) (Hazeu, 2003).

### 4.3.3 Methodology

ICESat full waveform data and AHN data are distributed in different coordinate systems and also have different spatial coverage. In order to compare terrain elevations and feature heights between these data, some processing steps are explicitly performed. To do this, both data sets are first converted to the same coordinate

system. Then an appropriate method is developed to extract terrain elevations from ICESat and AHN data. Also feature heights are derived from both data sets. Finally, to ensure a fair and reliable comparison, filtering constraints are defined and applied. Details of these processing steps are described in the following.

### (a) Datum transformation and coordinate systems

A critical step in the elevation comparison is the coordinate system conversion. For comparison between ICESat and AHN data with respect to land cover type as obtained from the CLC2000 product, data sets need to be available in the same georeferenced coordinate system. AHN and CLC2000 data are available in RDNAP and ERS coordinates, respectively. Because AHN data and CLC2000 are very big data sets, ICESat data, initially in the TOPEX/Poseidon reference frame, have been converted instead to both RDNAP and ERS coordinates. The conversion scheme is summarized in Figure 4.3. The ICESat data in the TOPEX/Poseidon ellipsoid are first converted to the WGS84 ellipsoid by IDL scripts provided by NSIDC (NSIDC, 2006). This conversion produced a very small error of less than 1 cm (Meeus, 1991). Then the ICESat data in WGS84 coordinates are converted to the ETRS89 reference system by the program PCTrans 4.0, (RNIN, 2008). The accuracy of this step is up to centimeter level. Next these data are transformed to the RDNAP system by the Coordinate Calculator developed by RDNAP (2007). This conversion is accurate within 1 cm (Hoog/NAP, 2009). The total accuracy of the previous steps is still restricted to the centimeter level, therefore this error component cannot be considered very significant in the comparison. Moreover, the ICESat GLA14 data in ETRS89 system are additionally converted to the ERS coordinates by ArcGIS 9.2 (EEA, 2008). The ICESat geolocation accuracy of about 4.5 m is well below the CLC2000 resolution of 100 m. Therefore this conversion has no significant effect either. Finally, the ICESat data is assigned to land cover type classes by comparison to the CLC2000.

### (b) Principal of determination of geolocation and surface elevation

A geolocated surface elevation,  $S$ , is determined as a sum of a laser altimeter vector,  $L$ , and an ICESat/GLAS geocentric vector,  $G$ , with respect to the center of mass of the earth (see Figure 4.4). The laser altimeter vector includes the GLAS laser pointing angle and a range,  $t_i$ , between the GLAS instrument and the surface as identified by measuring a travel time of a transmitted pulse until its return as a waveform. The range is then calculated as a half-travel time multiplied with the speed of light. The geocentric vector represents the orbit position of the ICESat satellite with respect to the center of mass of the earth. Therefore the laser spot or geolocation is inferred by the sum of these two vectors. The surface elevation is obtained by converting the geocentric laser spot position  $(r, \varphi, \lambda)$  to ellipsoidal

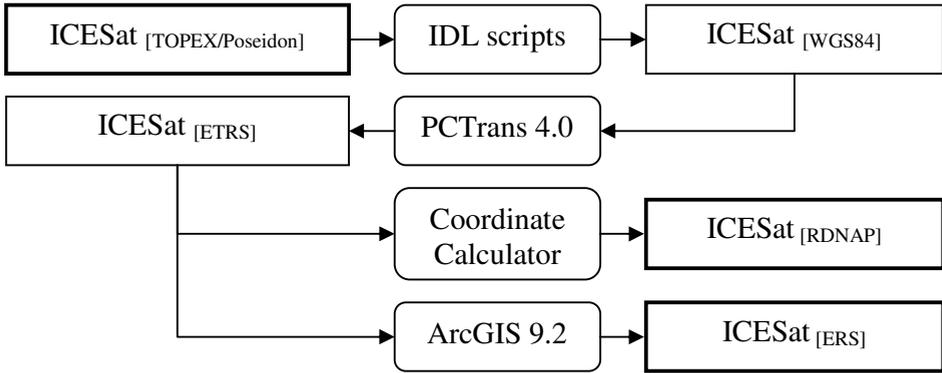


Figure 4.3: Procedure of coordinate system conversion.

height and geodetic latitude and longitude  $(h, \varphi_g, \lambda)$ .

In Figure 4.4, the land-specific range from GLAS to the ground surface can be calculated based on different waveform parameters like the waveform centroid or the height of the first or last mode of the waveform. Using the first mode gives a shorter range and results in a higher elevation point. The first mode results from elevation points of trees, forest or artificial features like buildings. Using the centroid of the waveform gives an average elevation while the last mode potentially represents the ground surface.

### (c) Georeferenced waveform

Typically, one position on the time axis of each waveform is used to compute a range between the GLAS sensor and the Earth surface, the so-called reference point. Together with the ICESat orbit position and orientation, elevation data, like available in GLA14, can be obtained (Schutz, 2002). In order to be able to use different positions in one waveform for different comparisons, this case study proposes a two step approach. First a waveform is registered into the RDNAP coordinate system using the reference point (the waveform centroid in most cases). In a second step a suitable position in the waveform is extracted. Now the height in RDNAP of this position is simply established by considering the distance to the reference point.

In Figure 4.5, the ICESat GLA14 elevation is represented by a cross; the square represents the mean AHN elevation of the ground points within the ICESat footprint. For georeferencing of the waveform (solid curve), the GLA14 reference point (here the waveform centroid displayed by the horizontal dotted line) is matched

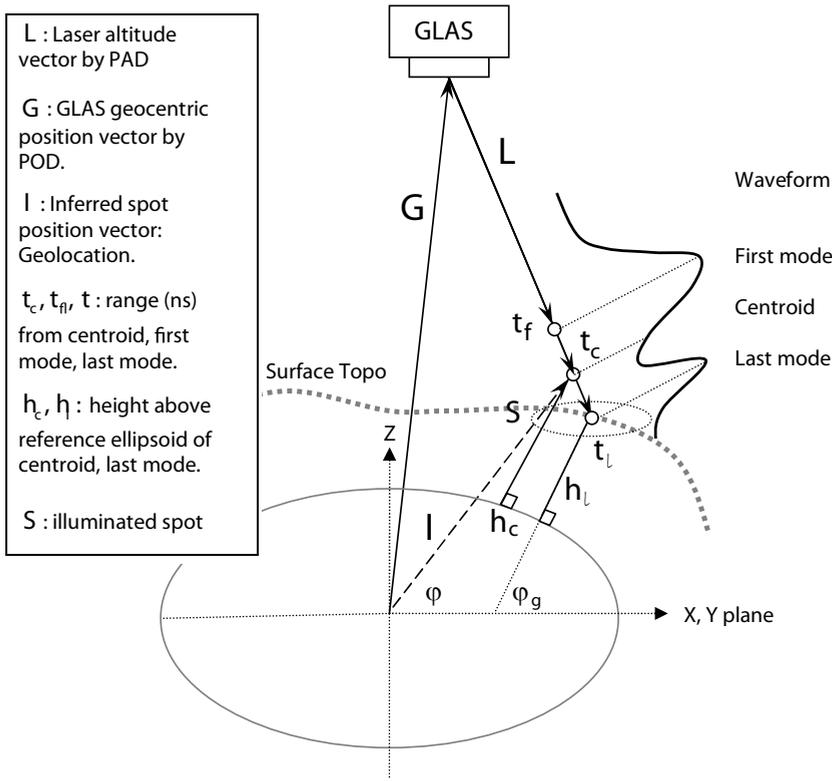


Figure 4.4: Principal of ICESat geolocation and surface elevation determination.

with the GLA14 elevation point. The reference point is obtained by adding the land range offset of `i_ldRngOff` to the reference range of `i_refRng` in the GLA14 product. Which position in the waveform is used as the reference point is indicated by the elevation definition flag of `i_ElvFlg` (NSIDC, 2008).

#### (d) Derivation of GLA01-based elevation data (ICESat last mode)

A georeferenced waveform is decomposed into a maximum of six Gaussian components which allows to derive waveform parameters as amplitude, width and location of each Gaussian mode. The waveform decomposition method described in detail in (Duong et al., 2006a), is applied in this case study. The first threshold crossing in the ICESat waveform usually corresponds to the highest intercepted surface within the footprint. The centroid of the complete waveform corresponds to the average height of the objects in the footprint, while the last Gaussian mode

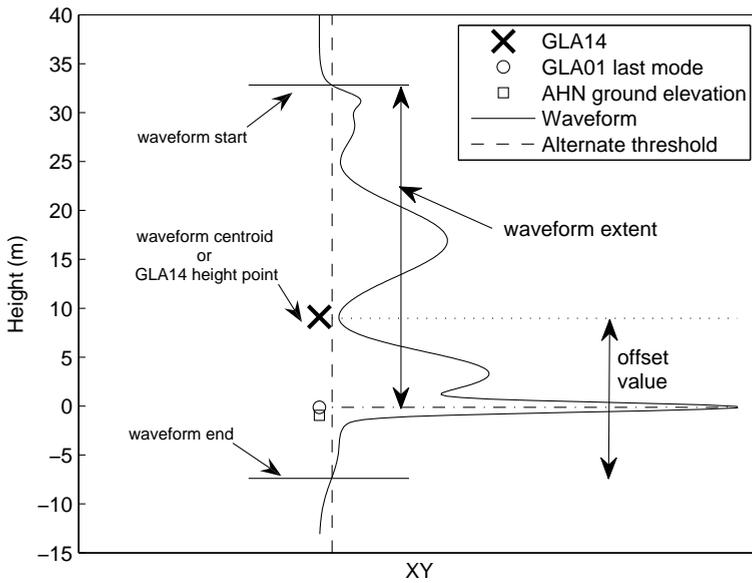


Figure 4.5: A waveform (solid curve) is georeferenced by matching the waveform centroid (horizontal dotted line) to a GLA14 elevation point (cross). The GLA01-derived elevation is the centroid of the last peak (circle). Waveform start and waveform end are defined by threshold crossing locations.

is resulting from the lowest elevation in the footprint (Brenner et al., 2003). Over flat terrain, the lowest elevation is the ground surface if the terrain is bare. As Dutch topography is in general flat, the last mode is the most suitable representation of the ground elevation.

#### (e) Derivation of lowest ground surface from AHN data (AHN lowest segment)

An average ICESat footprint from campaign L2a contains approximately 700 AHN data points with an average point density of 0.20 point/m<sup>2</sup>. As stated before, there exists no AHN ground point product over urban areas. Moreover, over complex topography the average height of the AHN ground points is often not representative for the height of the last mode of the ICESat waveform. Some typical examples are shown in Figure 4.6. Figure 4.6(a) shows an ‘urban’ waveform. In the corresponding AHN data, different height levels are visible, e.g. from trees (left), actual ground surface (black dots) and building roofs (middle). The mean elevation of this discontinuous data set ( $\sim 7.5$  m) is not representative for the

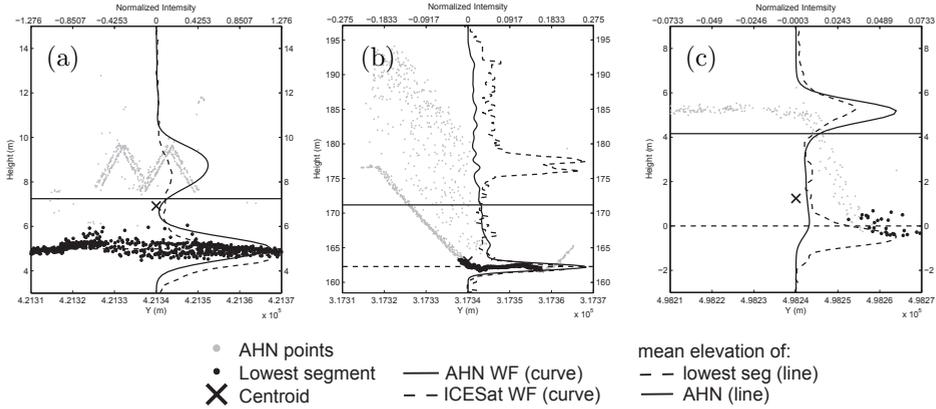


Figure 4.6: AHN points (grey dots) and its mean height (horizontal solid line), lowest segment points (black dots) and its mean height (horizontal dashed line) are displayed with AHN simulated waveform (solid curve) and ICESat waveform (dashed curve). (a) City with roofs and trees. (b) Canal with vegetation on the embankment. (c) Staired surface.

ICESat last mode height ( $\sim 5$  m). In Figure 4.6(b), the ground level data (grey dots) are continuous but varying in height due to the presence of a canal in the footprint. The elevation of the horizontal bottom (black dots) is about 163 m instead of 172 m for the mean ground surface. Similarly, Figure 4.6(c) has a series of steps in the surface elevation. The elevation of the lowest surface (black dots) is about 0 m instead of 4 m for the mean of the surface. Compared to the ICESat last mode, the segment containing the black points is most suitable for the comparison and therefore chosen as representative of the lowest surface.

This lowest surface is determined in two steps. First, the complete data set of AHN points (ground and non-ground points) within the ICESat footprint is subdivided into many small homogeneous segments by applying a segmentation method (Rabani et al., 2006). Then the segment containing at least 10 points, that has the smallest average height is selected as the lowest surface. The mean elevation of this segment is called ‘AHN lowest segment’. Figure 4.6 shows a visualization of the AHN points (grey dots), the lowest segment points (black dots), the mean height of the AHN points (horizontal solid line), the mean height of the lowest segment (horizontal dashed line), and both the ICESat waveform (dashed curve) and the waveform simulated (solid curve) from the AHN data as introduced in the next section.

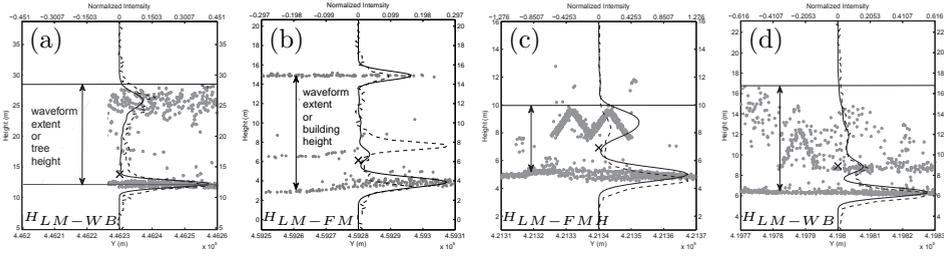


Figure 4.7: Definitions of feature heights are visualized: (a) forest height,  $H_{LM-WB}$ , (b) building height with a flat roof,  $H_{LM-FM}$ , (c) building height with a tilted roof,  $H_{LM-FMH}$ , and (d) building height in which the tree is higher than the building,  $H_{LM-WB}$ . Moreover, the ICESat waveform (dashed line) and AHN simulated waveform (solid line) are displayed together with the AHN points (black dots).

### (g) Definition of feature heights

Precise extraction of feature heights over forest and urban areas is a challenging task. Below it is described how to parameterize feature height using suitable waveform and point cloud parameters and how to consecutively derive these parameter values using the indicated waveform processing method. It is also shown however that it is very difficult to parameterize feature height in a uniform way such that feature heights can be extracted automatically giving satisfactory results for many individual waveforms. Feature heights are both extracted from ICESat full waveforms and AHN simulated waveforms.

When using airborne laser point cloud data, like the AHN data, over forested areas, the forest height can be defined with respect to the point cloud as the distance between the maximum elevation and the minimum elevation of the laser points. If the surface is flat or has a constant slope, the forest height can easily be extracted. In contrast, if the terrain topography contains, for example, a series of steps or is subject to a varying surface slope, the definition and determination of the ground surface are critical aspects for forest height extraction, (Lefsky et al., 2002), and more advanced methods of e.g. canopy height modeling, need to be applied for precise forest height extraction (Hollaus and Wagner, 2007; Harding et al., 2001). Moreover, over urban areas the feature height is more complicated to define. Even when the underlying terrain height can be sufficiently determined, (Rabbani et al., 2006; Kraus and Pfeifer, 1998), the building height extraction is still in question. The distance from the terrain to the highest elevation point in the AHN point cloud is expected to represent the building highest point. However, this highest AHN elevation point could also correspond to high trees or to other non-building objects.

When using the ICESat full waveform, it should be taken into account that the slope and roughness of the terrain surface, and variabilities in the upper parts of tree canopies are all having a widening effect on the full waveform. In order to enable a fair comparison to feature heights derived from the AHN point cloud, at least the slope/roughness effects need to be corrected in advance (Lefsky et al., 2005, 2007).

Because of these potential problems in comparing feature heights derived from the AHN point cloud to the ICESat full waveform, a simpler approach is proposed that directly compares waveform parameter values extracted from both the ICESat waveform and the AHN simulated waveform. Because the topography over The Netherlands in general is flat, the effect of slope is not expected to significantly contaminate the obtained feature heights.

In order to be able to compare feature heights in section 4.3.4, feature heights extracted from the ICESat waveform and AHN simulated waveform are introduced in this section. A feature height over forest is called and expected to represent forest height. Over urban areas, building heights are considered, and over bare land low vegetation heights (e.g. scrubs and bushes).

Over forest, the waveform extent, which is the vertical distance between waveform begin and waveform end as identified by threshold values, was used in Harding and Carabajal (2005) to estimate maximum forest height. Due to effects of surface slope and variability of upper canopy parts, an SRTM-derived terrain index and the leading edge of the waveform were applied to correct the waveform extent.

To facilitate the introduction of feature heights, we define the following abbreviations for the significant waveform parameters. The waveform begin,  $WB$ , is the position where the ICESat waveform firstly crosses the threshold value (Figure 4.5). The position  $FHM$  is the location in the ICESat waveform where the waveform reaches half of the first maximum ( $FHM$ , first half maximum) energy of the first mode (Figure 4.7(c)). Moreover, the peaks of the ICESat first and last mode are denoted as  $FM$  and  $LM$ , respectively, (Figure 4.7(b)).

As discussed in the previous section, the ICESat last mode ( $LM$ ) is most suited to represent terrain height. Therefore the forest height,  $H_{LM-WB}$ , is defined as the vertical distance between the ICESat last mode ( $LM$ ) and the waveform begin ( $WB$ ), Figure 4.7(a). For bare land, the low vegetation height  $H_{LM-WB}$  is defined in the same way.

Over urban areas, an ICESat footprint may represent parts of different buildings but also high trees. Three typical cases can be identified, consisting of footprints containing dominantly (i) a building with a flat roof (factory), (ii) buildings with tilted roofs (typically Dutch residence house) and (iii) a combination of high trees and buildings. In order to obtain a suitable descriptive parameter for a building height, three potential parameters of building heights are defined. For example, in

the first case, illustrated in Figure 4.7(b), a suited building height is the distance between the last mode ( $LM$ ) and the first mode ( $FM$ ),  $H_{LM-FM}$ . In the second case, Figure 4.7(c), the most suited building height can be the distance between the last mode ( $LM$ ) and the position  $FHM$ ,  $H_{LM-FMH}$ . In the third case, the feature height is  $H_{LM-WB}$  if the tree is higher than the building roof, Figure 4.7(d), or  $H_{LM-FMH}$  if the building is higher than the tree, Figure 4.7(c).

Feature heights of three land cover classes (forest, urban and bare land) are summarized below:  $H_{LM-WB}$  is applied for forest, urban and bare land areas. Moreover, over urban areas, results from two additional feature height parameters  $H_{LM-FM}$  and  $H_{LM-FMH}$  are compared. The parameter that is most suited to describe the feature height over urban areas will be discussed and chosen in section 4.3.4.

To enable a comparison to ICESat full waveform derived feature heights, feature heights are also extracted from the AHN data. For this purpose, the AHN points within an ICESat footprint are simulated to a waveform as described in section 3.5. Then feature heights from these simulated waveforms are extracted in the same way as for the ICESat waveforms. The resulting ICESat and AHN feature heights are compared in section 4.3.4.

#### (h) Derivation of ICESat incident angles

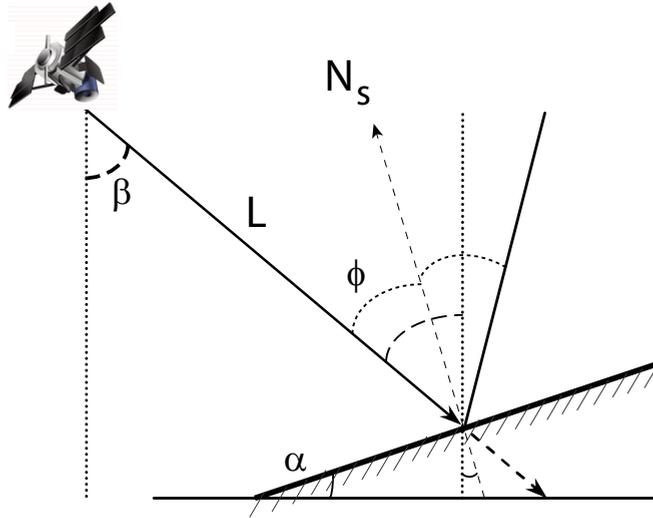


Figure 4.8: Procedure of incident angle determination of the ICESat laser pulse. Details of this procedure are found in [Molijn \(2009\)](#).

The incident angle  $\phi$  of a laser beam is the angle between the laser vector  $L$  and

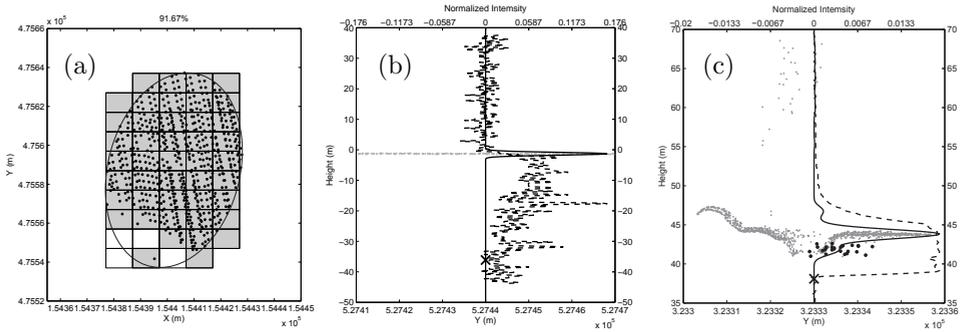


Figure 4.9: (a) Coverage of the ICESat footprint by AHN points. (b) Cloud presence resulting in a height difference  $\geq 10$  m. (c) Saturated waveform with a saturation flag value of 3. The range of saturation, between 39 m to 45 m, is clearly visible.

the normal  $N_s$  of the illuminated surface at the center of the footprint. Except on the surface slope and aspect it depends on the pointing angle; that is the angle  $\beta$  between nadir and the direction of the emitted laser beam. Figure 4.8 shows in an one dimensional view the laser pointing angle  $\beta$  and the surface slope  $\alpha$ . The laser pointing angle is obtained from the GLA14 flag [i\\_beam\\_coelev](#) and the surface slope and aspect are calculated by fitting a plane to the AHN lowest segment data within the ellipsoidal ICESat footprint.

### (i) Filtering Constraints

In order to use only reliable height values for a comparison between AHN and ICESat data, constraints on both data sets are taken into account. In this case study, both data sets must satisfy the following four conditions:

- C1. All ICESat data must be considered valid, as indicated by the elevation use flag [i\\_ElvuseFlg](#).
- C2. AHN points must be spatially well distributed within the ICESat footprint. For this purpose, the footprint extent area is regularly divided into  $10 \text{ m} \times 10 \text{ m}$  boxes. An ICESat footprint is removed from further comparison if (i) the number of AHN points in the ICESat footprint is less than 50 (about 7% of the point density in section 4.3.3(e)) or (ii) the ratio between the number of boxes containing AHN points, and the total number of boxes is less than 90%. Figure 4.9(a) shows a footprint where the coverage of the AHN points within the ICESat footprint is about 91%.

- C3. Cloudy sky conditions may cause errors in the elevation data by effects of absorption, forward scattering and signal delay (Herring and Quinn, 1999). Figure 4.9(b) shows a very noisy waveform (dashed curve) obtained under cloudy conditions, resulting in a low return energy. Therefore, only those ICESat waveforms are incorporated that are not effected by cloud cover. A gain value of less than 100 counts as indicated by the gain value of `i_gval_rcv` is used to identify clear sky conditions, (Nguyen and Herring, 2005).
- C4. ICESat waveforms that saturate the ICESat detector, result in a lower elevation (Fricker et al., 2005). Figure 4.9(c) shows an ICESat waveform (dashed curve) and an AHN simulated waveform (solid curve) where comparison to the AHN lowest surface (black dots) suggests that saturation results in this case in a height jump from 42 m to 40 m. Hence, ICESat waveform data should not suffer from saturation. A GLA14 flag, `i_satCorrFlg`, value of 0 indicates that no saturation has occurred while receiving the waveform signal. Only waveforms with a `i_satCorrFlg` value of 0 are kept for comparison.

In the next section, two comparisons are considered:

**Case 1:** ICESat last mode  $\leftrightarrow$  AHN lowest segment.

**Case 2:** ICESat feature height  $\leftrightarrow$  AHN feature height.

In all comparisons, the constraints C1–C4 are explicitly applied and an AHN elevation is always subtracted from the corresponding ICESat elevation to obtain an ICESat–AHN height difference.

### 4.3.4 Results and comparison

Table 4.5: Height difference and its standard deviation in parenthesis between ICESat last mode and AHN lowest segment.

Track	Date	Conditions C1–C4: ICESat last mode – AHN lowest segment, (m)					
No.	2003	Broad	Mixed	Needle	Urban	Bare land	Water
043	30-Sep	-0.16 (0.92)	-0.21 (0.11)	-0.21 (0.20)	-0.51 (0.51)	-0.25 (0.20)	-0.28 (0.58)
1233	14-Oct	1.27 (0.00)	-0.34 (1.21)	-0.09 (0.23)	-0.22 (0.20)	-0.22 (0.16)	-0.26 (0.03)
1275	16-Oct	-0.12 (0.15)	0.15 (0.67)	0.01 (0.34)	-0.18 (0.32)	-0.16 (0.22)	-0.17 (0.29)
0015	23-Oct	0.31 (0.63)	N/A	N/A	0.21 (0.53)	-0.09 (0.35)	0.21 (0.20)
0043	24-Oct	N/A	-0.24 (0.00)	N/A	0.04 (0.32)	-0.18 (0.20)	N/A
0295	10-Nov	-0.30 (0.82)	-0.24 (0.51)	-0.19 (0.30)	-0.28 (0.24)	-0.30 (0.21)	-0.59 (0.58)
Total		-0.14 (0.59)	-0.11 (0.69)	-0.07 (0.30)	-0.24 (0.28)	-0.21 (0.20)	-0.08 (0.34)
Number of pairs		77	101	234	530	2193	37

**(a) Case 1: Comparison of ICESat last mode and AHN lowest segment**

In this section we compare the ICESat last mode elevation to the AHN lowest segment elevation as described in section 4.3.3. Both elevations are considered to represent the terrain height. The ground surface is assumed to be stable between AHN and ICESat acquisition time although the difference in acquisition ranges between 0 and 7 years. It is moreover assumed that the AHN data are accurate when compared to ICESat data. Therefore this comparison describes the accuracy of ICESat data over terrain surface with respect to different land cover types. However, there are still outliers in the comparison like errors in waveform fitting (missing last mode, see Figure 4.10(a)), effects of slope or rough surfaces where the last peak is higher than the mean elevation of the lowest segment (see Figure 4.10(b)) and errors in segmentation (small segment in black dots is not representative for the sloped surface, see Figure 4.10(c)). Therefore the Median/MAD (Median Absolute Difference) robust technique (Muller, 2000) is applied to discard outlier effects in order to arrive at realistic statistics. After applying filtering constraints C1–C4, 3172 footprints remain. The final results are given in Table 4.5 and in the histograms in Figure 4.11.

The results show that it is possible to derive the terrain surface height with a bias of about 21 cm and a standard deviation of 20 cm over bare land from the ICESat altimetry data. Over forest, the height difference is on average 14 cm for broadleaf, 11 cm for mixed and 7 cm for needleleaf. The standard deviation value

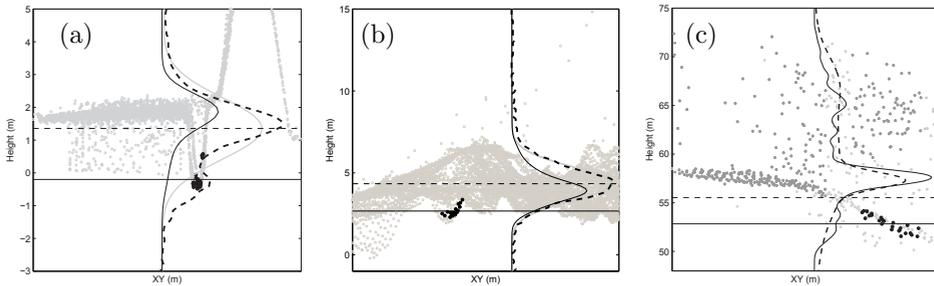


Figure 4.10: Examples of outliers in the ICESat–AHN comparison. The ICESat waveform is represented by the dashed black curve, the AHN simulated waveform by the solid black curve, the mean elevation of the ICESat last mode by the horizontal dashed line, the mean elevation of the AHN lowest segment by the solid black line. (a) the ICESat last mode at  $-0.2$  m is not extracted because of fitting errors. (b) the AHN lowest segment does not give a good corresponding elevation compared to the ICESat last mode. (c) the AHN lowest segment is not well defined for a slant surface.

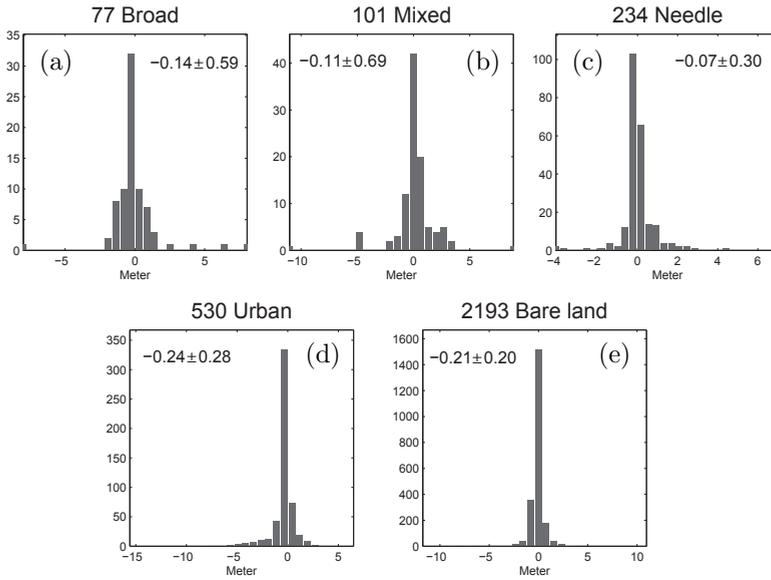


Figure 4.11: Histograms of height differences obtained after applying conditions C1–C4 over: (a) broadleaf, (b) mixed, (c) needleleaf, (d) urban and (e) bare land. Median and standard deviation values have been estimated by the Median/MAD robust statistics (Muller, 2000).

is, however, large over broadleaf (0.59 m) and over mixed (0.69 m), and smaller over needleleaf (0.30 m). Moreover in the urban case, the difference in terrain height is 0.24 m with a small standard deviation of 0.28 m.

In the case of water, a small case study shown in Figure 4.12 gives us several reasons to mistrust comparisons over footprints classified by CLC2000 as water: varying water levels, strongly varying numbers of AHN points over water due to specular reflectance, absorption, appearance of structures like boats in the water, and footprints on the border between water and land. Therefore we conclude that comparisons for footprints marked as water are not reliable. Some waveform locations over a water body are displayed superimposed on Google Earth imagery (bottom) together with a visualization of the corresponding ICESat waveforms and AHN data (top) in Figure 4.12. The green dots represent the AHN points within the ICESat footprint, and the grey dots are outside the footprint. The footprint size is indicated by the length of the horizontal black line. We make three observations.

(i). Occurrence of artificial objects: In Figure 4.12(a)–(b) and the AHN elevation

image (bottom panel), AHN data indicate the occurrence of an artificial object (e.g., a long boat or a construction work with a height of 5 m) of about 175 m length (ICESat footprint spacing) over the water surface while the corresponding ICESat waveform has a single peak at an elevation of  $-5.7$  m. This object causes a height difference of 4.7 m in case of Figure 4.12(a) and 0.86 m in Figure 4.12(b).

(ii). Figure 4.12(d) shows an ICESat footprint location that is classified as water. Still this footprint contains some AHN points that originate from the land surface. As there are no AHN points over the water part of the footprint, a difference of 1.31 m occurs between the ICESat and AHN centroid elevation.

(iii). On the other hand, Figure 4.12(c) shows a footprint that only covers water, but still AHN points are available, resulting in a height difference between the AHN and the ICESat centroid of 0.29 m. This difference is probably caused by a change in water level between AHN and ICESat acquisition time (0 to 7 years), given the consistent AHN and ICESat elevations in the neighboring footprints.

### ICESat incident angle effect:

Further insight into the observed differences in ICESat versus AHN derived terrain height is gained by additionally analyzing the influence of ICESat incidence angle and pointing error. According to (Urban et al., 2008), the maximum elevation error of the ICESat product equals about 7.5 cm per 1 degree incident angle with a laser pointing error of 1.5 arcsec. In order to limit the impact of incident angle on elevation errors to maximally a few centimeters, footprints were removed if the ICESat incident angle was larger than 1 degree. By using the AHN data, the ICESat incident angle is obtained from a combination of the ICESat laser pointing angle and the AHN surface slope. The AHN surface slope is obtained by fitting a plane to the AHN lowest segment data within the footprint, see section 4.3.3(h) in detail.

In comparison to the previous results (Table 4.5), the results show that the observed differences in terrain height are strongly changed in forested areas and slightly improved in bare land and urban areas. Over forested areas, the average difference of terrain heights slightly increases, i.e. by 13 cm over mixed forest, and by 5 cm over both needleleaf and broadleaf forest. The standard deviation is significantly reduced from a maximum of 69 cm (over mixed) to a maximum of 25 cm (over broadleaf). Moreover, over urban and bare land areas, only the standard deviation is reduced by 1–5 cm. The price to pay for the improvement in the statistics is that 873 pairs were removed from a total of 3172 pairs. Although this analysis was performed over relatively flat terrain, it demonstrates that there is an impact of the ICESat incidence angle on the accuracy of the terrain height.

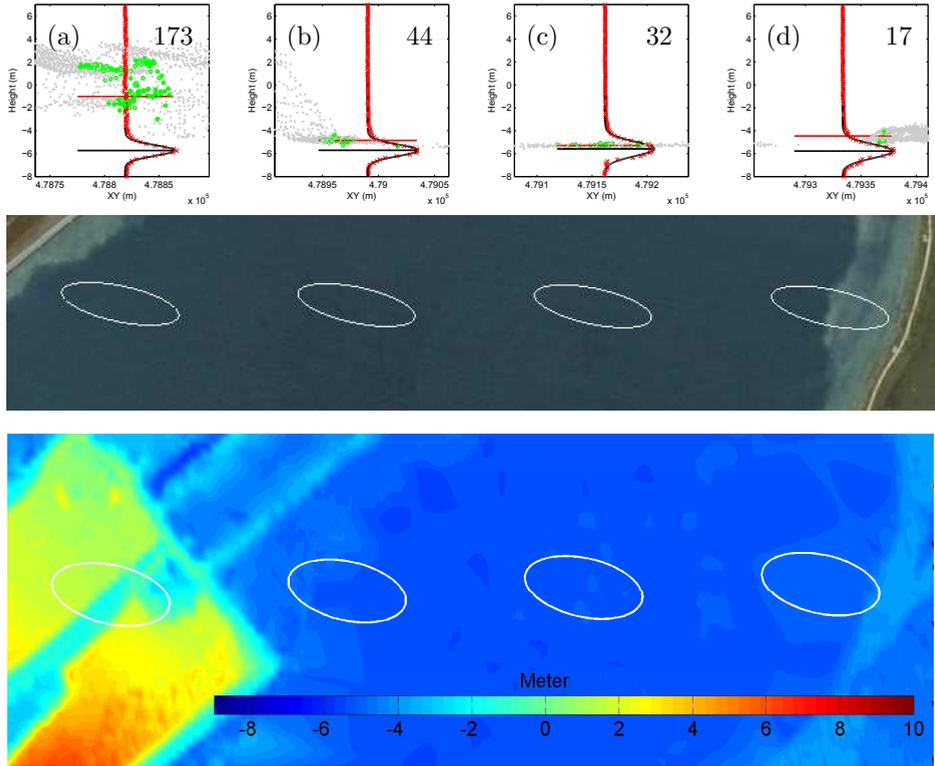


Figure 4.12: Water waveforms (top panel) with number of AHN points (top right) within ICESat footprints displayed together with the Google Earth image (middle panel) and the AHN elevation image (bottom panel). In Figure 4.12(a)–(d), the horizontal black line represents the last mode height and the horizontal red line is the mean height of AHN ground points. The green dots are the AHN points located inside the ICESat footprint and the grey dots are the AHN points outside the ICESat footprint.

### (b) Case 2: Comparison between ICESat and AHN derived feature heights

In this section, after applying filtering constraints C1–C4, 3172 feature heights derived from ICESat waveforms and AHN simulated waveforms as defined in section 4.3.3 are compared. Histograms of forest, building and low vegetation heights using the parameter  $H_{LM-WB}$  are shown in Figure 4.13(a)–(c) respectively. On average, the feature height is about 17–20 m over forest, 8–9 m over buildings, and

almost 3 m over low vegetation.

Table 4.6 shows height differences with respect to forest, buildings and low vegetation in terms of the feature height parameter  $H_{LM-WB}$ . Moreover, over urban areas, two other feature height parameters  $H_{LM-FM}$  and  $H_{LM-FMH}$  were also determined. According to the feature height  $H_{LM-WB}$ , the height difference between the ICESat data and AHN data equals 1.89 m over forest, 1.48 m over buildings and 29 cm over low vegetation. The standard deviation is larger over buildings (2.64 m) and forest (2.30 m), and small over low vegetation (67 cm). Compared to the mean feature heights as shown in Figure 4.13(a)–(c), respectively, the height difference over forest and bare land is 10%. However, over urban areas a height difference of 1.48 m is notably larger compared to a feature height in the order of 8-9 m.

Using the feature height parameters,  $H_{LM-FM}$  and  $H_{LM-FMH}$  over urban terrain, leads to a reduction of the mean height difference to 40 cm for  $H_{LM-FM}$  and 1.06 m for  $H_{LM-FMH}$ . However, the standard deviations are increasing to 3.75 m for  $H_{LM-FM}$  and 3.19 m for  $H_{LM-FMH}$ .

We conclude that none of the three feature height parameters defined for buildings give very satisfactory results. Several reasons explain this negative outcome. A first reason is that for different scenarios (like first intercept is flat roof, tilted roof or urban tree), the best feature height parameterizations are different. Therefore applying one parameterization for all possible scenarios will definitely result in suboptimal outcomes. Unfortunately it seems in practice impossible to classify an urban waveform as corresponding to a certain scenario. Still this reason does not explain differences in urban feature height as derived from either ICESat or AHN simulated waveforms.

The large standard deviation of the building height differences can be explained as follows: (i) Small geolocation errors. With a small shift in footprint location parts of other buildings may appear in the footprint, causing different height levels. (ii) Feature changes that occurred between ICESat and AHN acquisition time (0 to 7 years) are especially likely in urban areas, considering e.g. new built-up houses, removed/ renovated old houses, and artificially added or removed trees,

Table 4.6: Feature height difference between ICESat and AHN and its standard deviation in parenthesis.

By	Forest	Urban	Low vegetation
$H_{LM-WB}$	1.89 (2.30)	1.48 (2.64)	0.29 (0.67)
$H_{LM-FM}$	N/A	0.40 (3.75)	N/A
$H_{LM-FMH}$	N/A	1.06 (3.19)	N/A

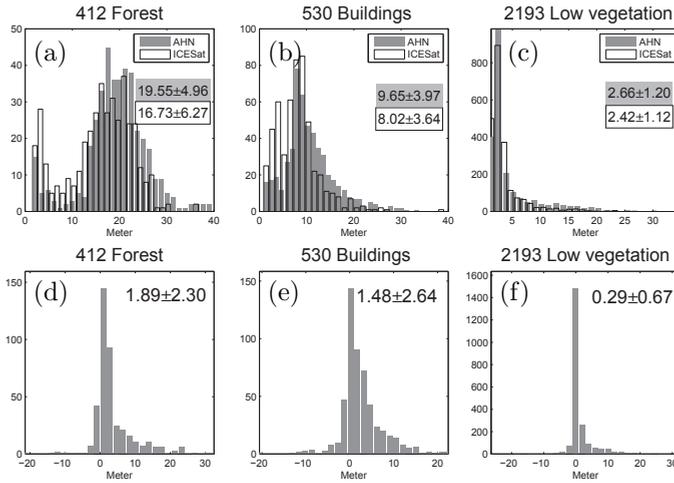


Figure 4.13: Histograms of feature heights extracted from AHN and ICESat data for (a) forest, (b) buildings, and (c) low vegetation. The histograms of the AHN feature heights are in grey, the ICESat histograms are bounded by a solid black line. For both AHN and ICESat, the mean value and standard deviation are given in the corresponding text boxes. Moreover, the histograms of  $H_{LM-WB}$  feature height differences between AHN and ICESat data are shown in figures (d)–(f) for different terrain classes. In each plot the median/MAD value is indicated.

see Figure 4.14. (iii) Due to the absence of intensity values in the AHN data set, the AHN simulated waveform is not always comparable to the ICESat waveform, see Figure 6(b), especially when the ICESat footprint is partly covering a building roof with a strong reflection. The energy return from such roof is recorded as the dominant peak in the ICESat waveform. However, the few reflecting points present in the AHN data are not enough to obtain such a dominant peak in the simulated waveform. The missing peak causes a large height difference in the feature height comparison.

In order to improve the feature height comparison, an additional condition could be implemented. Since the AHN data were acquired from 1996–2003, spanning a larger time window compared to the ICESat data acquired approximately in an 1-month period, it is likely that in some cases actual feature height changes occurred between AHN and ICESat acquisition. This could lead to anomalous feature height differences. The AHN and ICESat terrain heights as derived for the terrain height comparison can be applied to match ICESat and AHN simulated waveform corresponding to (actual) terrain height. Considerable differences between the two waveform shapes can now easily identified by applying a correlation coefficient

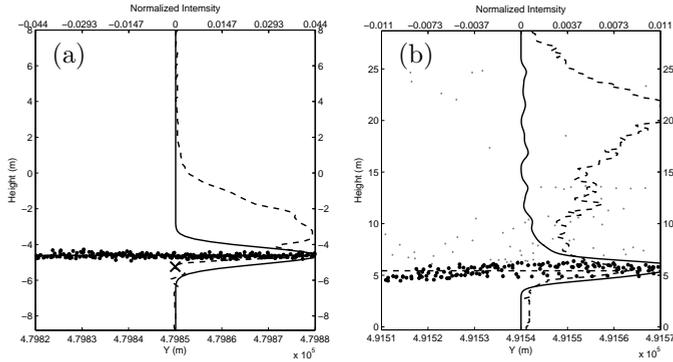


Figure 4.14: Big changes on the earth surface. The AHN waveform (solid line) is simulated from the AHN points (black dots) that does not contain any data points from above the terrain surface, Figure 4.14(a). However, the ICESat waveform (dashed line) properly includes other objects, e.g. new houses. Similarly, Figure 4.14(b), very few vegetation points in the AHN data correspond to a weak peak in the AHN simulated waveform, however there seems to occur, e.g. new trees in the ICESat waveform.

threshold.

In general, one can conclude that obtaining feature heights from ICESat waveforms makes sense for either well controlled footprint locations, where the objects represented by the ICESat waveform are known to the operator, or, on the other hand, for larger populations of waveforms, representing homogeneous features, that are likely to exhibit a similar change through time, as can be expected in agricultural or forest applications. Over urban regions, where the footprint contents is expected to change from footprint to footprint, analyzing feature height for groups of waveforms together seems doubtful.

### 4.3.5 Conclusions and further research

In this study we have compared ICESat derived elevations to high resolution airborne laser altimetry data over complex terrain. The airborne data is obtained from the Dutch national airborne laser altimetry product AHN. Two comparisons are performed: terrain height was assessed by comparing ICESat last mode against AHN mean lowest segment, and ICESat feature height was evaluated against AHN feature height.

For the terrain height, the difference between the elevation of ICESat's last mode

and the mean elevation of AHN lowest segment is on average  $-21$  cm over bare land,  $-24$  cm over urban areas, and from  $-14$  cm to  $-7$  cm over forested areas. The standard deviation of the differences ranges between 20 cm for bare land and 69 cm for forested areas. This comparison has been performed on a population of several thousands waveforms. These results show that ICESat and likely follow up missions have good potential for measuring terrain height, even over complex forested and urban terrain with an accuracy at the decimeter range.

For the feature height case, the difference between ICESat and airborne derived feature height is acceptable over forested and bare land areas. However, the result over buildings is not satisfactory. The main reason is that ICESat derived feature height parameters are sensitive changes in feature height that occur at spatial distances smaller than the size of the ICESat footprints. For homogeneous land cover type, ICESat waveform analysis is a suitable method for estimating feature heights. For the urban case, an additionally accurate DEM data may still enable the monitoring of feature height changes.

The two described comparisons demonstrate that not only good results can be obtained on average, for large populations of waveforms, but also that terrain and feature height extraction from ICESat data at the single shot level is feasible and accurate to the levels shown here. If the spatial coverage of future satellite laser altimetry missions could be improved by applying different acquisition patterns (e.g. different orbit configuration, multiple view lasers) such missions could be applied for obtaining large scale elevation and forestry/biomass products of unprecedented accuracies.

Analysis of single waveforms is important to obtain insights on differences that still occur between ICESat waveforms and in our case simulated waveforms as constructed from corresponding AHN points. The main reasons for the remaining differences in elevation in our comparison are actual changes within the footprint, explainable by temporal/seasonal acquisition differences, geolocation errors, inhomogeneous scattering distribution within the footprint and atmospheric disturbances.

Note that the signed terrain elevation difference between ICESat and AHN terrain height is negative. This means that ICESat slightly underestimates terrain height. The ICESat elevation accuracy depends on saturation, atmospheric forward scattering, surface roughness, pointing errors and field of view shadowing. The first two error sources were excluded in the comparison based on quality flags provided in the GLAS product. The third and fourth error sources are discussed in this case study. However, the last one, the field of view shadowing that is significant in campaign L2a data, is not analyzed yet. This factor causes distortion of the laser power distribution within the footprint, resulting in clipped/skewed waveform shapes. As a consequence, ICESat elevations can be too low by several cm, with a bias magnitude correlated with footprint size and laser energy level (**Urban**

et al., 2008). This error source could be the remaining problem for underestimated terrain heights, and, therefore, needs to be investigated and quantified in further studies. Finally, nearly synchronous data from airborne AHN2 and recent ICESat campaigns are becoming available. Repeating this analysis on these new data sets will enable better insight into error sources in ICESat height underestimation and terrain/feature height differences between ICESat and AHN data.

## 4.4 Case study 2: Error analysis by investigating overlapping pairs over Europe

### 4.4.1 Background

One way to identify errors is to compare waveforms that cover the same footprint location. Waveforms obtained at the same location and at approximately the same time should be very similar. Meanwhile, seasonal or annual influences can be tracked and quantified by comparing overlapping waveforms from repeated campaigns. As can be seen from the campaign schedule in Figure 4.15, bottom, these repetitions are all provided for by the ICESat mission. Waveform pairs within one epoch are obtained by considering crossovers between ascending and descending tracks, compare Figure 4.15, top, where the ICESat tracks over Europe of the first measurement campaign in winter 2003 are visualized.

A unique contribution of this study is that for the first time a large database of tens of thousands of repeated waveforms is presented that can be used to gauge a raw full waveform processing algorithm. This database will be used to answer the question: How can changes in waveforms from the same location be quantified and explained? A large benefit of choosing Europe as the Region of Interest is that many meta data are available to find explanations for inconsistencies as identified by the processing steps. This procedure is expected to result in a more robust waveform processing methodology to be used in e.g. future of-the-shelf processing of airborne full waveform laser altimetry and for the challenging task of processing large quantities of full waveforms over the polar regions.

The main research question that will be addressed in this study is as follows: ‘How can changes in waveforms within one waveform pair be quantified and explained?’ The focus of the research will be on waveform pairs from within one measurement campaign. In this case, waveforms should in principle be the same for all land cover classes. Differences may occur however because of partially overlapping footprints, incorrect slope estimation, changing weather conditions or changing surface moisture conditions. It will be investigated if it is possible to determine correlation between nearby waveform pairs in the same land cover class in order to separate, identify and quantify these type of error sources.

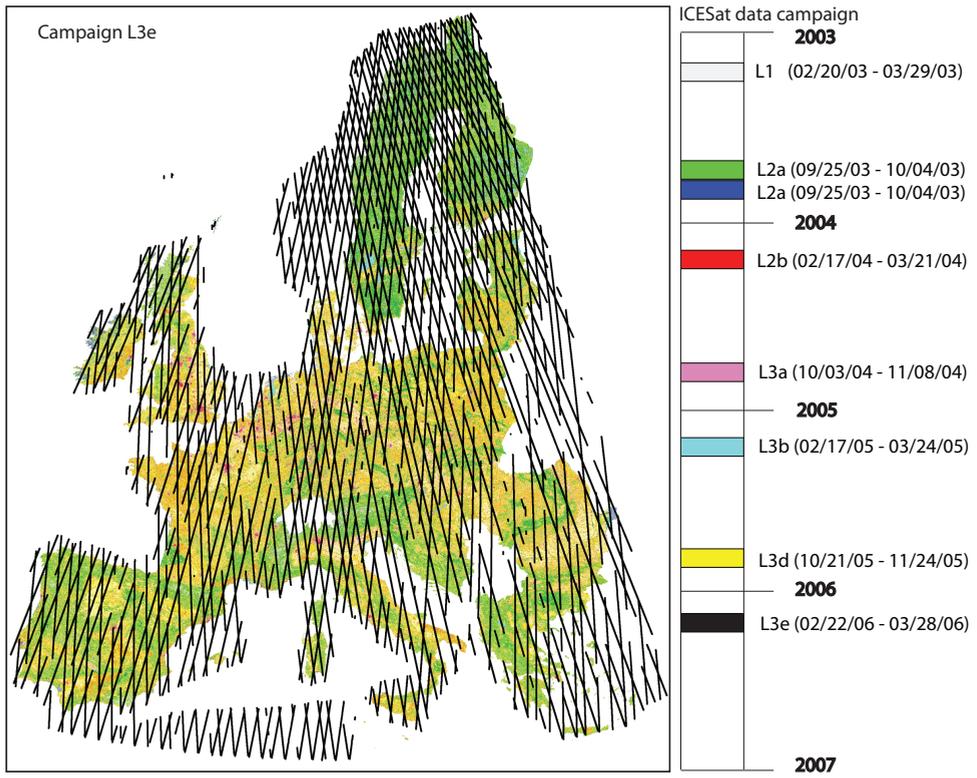


Figure 4.15: ICESat campaigns, right, and ICESat tracks over Europe during the Winter 2006 campaign (L3e), overlaid on CLC2000 land cover data, left.

In section 4.4.2, the study area and input data are described. In section 4.4.3 it is shown how to obtain a database of overlapping waveform pairs from ICESat tracks as available over Europe. Moreover, parameters are introduced describing differences between waveforms. In section 4.4.4 the resulting database of waveform pairs is described. As a first application, three case studies of almost perfectly overlapping waveform pairs are discussed, before reaching any conclusion.

## 4.4.2 Study area and input data

For this study data over the European continent from the ICESat products GLA14 and GLA01 are used. The data considered was acquired between 2003 and 2006. Data with a higher release number is processed according to newer insights. In this study, the release number of the Laser 3 campaigns was 28, of the Laser 2

Table 4.7: ICESat epochs, laser name, orbit repeat period, acquisition date, release, and number of points

Epoch	Data campaign	Repeat (days)	Dates	Release	Number of points
1	L1	08	02/20/03 - 03/29/03	118	496697
2	L2a	08	09/25/03 - 10/04/03	426	287146
3	L2a	91	10/04/03 - 11/19/03	426	1422863
4	L2b	91	02/17/04 - 03/21/04	428	1056698
5	L3a	91	10/03/04 - 11/08/04	428	1134105
6	L3b	91	02/17/05 - 03/24/05	428	1153022
7	L3d	91	10/21/05 - 11/24/05	428	1114930
8	L3e	91	02/22/06 - 03/28/06	428	1097114

campaigns it was 18, and for the Laser 1 campaign it was 7. At the time of writing already data of release number 31 is available. In the next section it is described how waveforms with overlapping footprints can be distilled from this data set.

### 4.4.3 Methodology

In order to investigate errors in the distribution and processing of ICESat products by analyzing corresponding waveform pairs, first overlapping waveform pairs over Europe are extracted as presented in section 4.4.3(a). Given the waveform pairs, the question remains how to compare them. For this purpose, in section 4.4.3(b), suitable waveform parameters are described. More details of these steps are presented in the following.

#### (a) Extraction of overlapping waveform pairs

Overlapping waveforms are extracted from the available ICESat full waveform data over Europe. Data of two campaigns (winter, i.e. February and March, and autumn, i.e. September, October, and November) each year (see Figure 4.15, bottom, and Table 4.7) are available from the National Snow and Ice Data Center. We identify two types of overlapping waveform pairs. A crossover pair consists of one ascending and one descending track waveform, Figure 4.16, left, a repeated track pair consists of two waveforms of the same ascending/descending track but from orbits performed at different moments, Figure 4.16, right. In both cases the footprint centers are within a threshold distance. This threshold is defined as the sum of the two half major axes of the two individual footprint ellipses, see Figure 4.16, left.

In this case study only waveform pairs from the same measurement campaign are considered. As a consequence, the time lag between the waveform pairs varies from a few days to a maximum of a few weeks. The procedure used to find waveform pairs is indicated in the flow chart in Figure 4.17.

### (b) Waveform pair comparison

In order to compare corresponding waveforms, initial processing steps for individual waveforms described in chapter 3 are explicitly applied like normalization and decomposition. Three parameters of individual waveforms were selected to compute a difference between corresponding waveform pairs. Three waveform parameters used in this study are height of median energy  $HOME$ , waveform extent  $wEcho$ , and relative waveform energy  $rEcho$ . The differences of these parameters are denoted as  $\Delta HOME$ ,  $\Delta wEcho$ ,  $\Delta rEcho$ .

The fourth parameter used in comparison of waveform pairs is defined as the mean squared difference,  $\Delta I$ , in normalized waveform energy per bin, (Duong et al., 2008).

$$\Delta I(X, Y) = \sum_{i=1}^N \frac{(X(i) - Y(i))^2}{N} \quad (4.1)$$

where  $N$  is the number of non-noise height bins in the waveform, i.e. non-ground portion of the normalised waveforms,  $X$  and  $Y$  are the repeated waveforms, and  $\Delta I$  is the average of squared differences in normalised amplitudes.

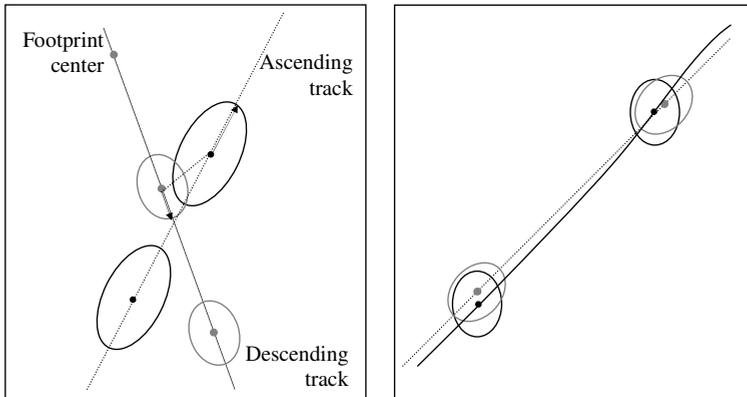


Figure 4.16: Crossover pair, left, and two repeated track pairs, right.

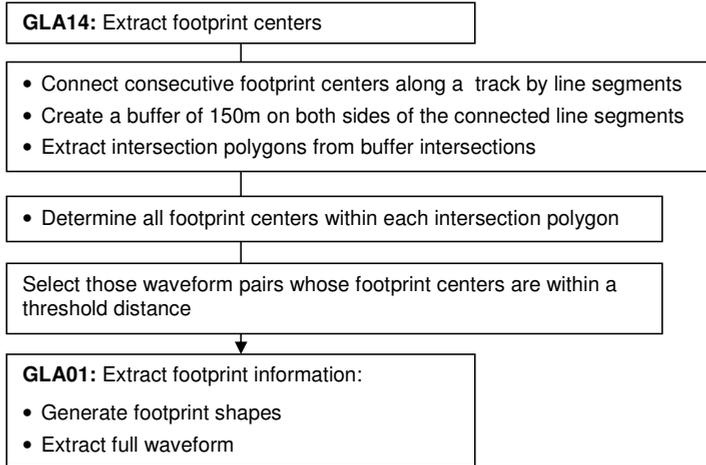


Figure 4.17: A flowchart of finding crossover and repeated pairs.

Table 4.8: ICESat epochs, total number of pairs, individual footprint diameters, distance between footprint centers and number of close by waveforms.

Epoch	Data campaign	Number of pairs	Footprint Diameter (m)	Distance (m)		Waveforms	
				Min	Max	<2m	<40m
1	L1	104809	52.0 - 95.0	0.20	95	60	15851
2	L2a	4592	79.3 - 111.9	1.98	108.56	1	949
4	L2b	1170	81.6 - 103.7	2.76	98.41	0	289
5	L3a	64	19.7 - 22.7	2.65	21.30	0	64
6	L3b	2119	50.8 - 99.0	0.93	92.37	7	472
7	L3d	815	48.7 - 55.9	1.73	52.88	1	205
8	L3e	386	47.7 - 58.7	1.83	54.30	2	208

## 4.4.4 Results and discussion

### (a) Waveform pairs

In Table 4.8 the numbers of overlapping waveforms for eight epochs of European ICESat data are listed. In total a number of 113 955 waveform pairs were found. When actually determining the area of intersection of two waveforms, the size and orientation of the footprint ellipse has to be taken into account. Approximate footprint diameters are given as well. The distance between footprint centers is in general in the order of tenths of meters, but some waveforms exists for which the footprint centers almost coincide.

The largest population of waveform pairs exists in data campaign L1. The distance between corresponding footprint centers varies from 0.2 m (L1) to 108.56 m (L2a).

Table 4.9: Differences between repeated waveforms

Epoch	$\Delta HOME$ (m)	$\Delta wEcho$ (m)	$\Delta rEcho$ ( $\times 10^{-11}$ )	$\Delta I$ ( $10^{-4} \times J^2$ )
1	0.2 $\pm$ 4.8	0.5 $\pm$ 9.0	1.7 $\pm$ 18	5.3 $\pm$ 529
2	0.1 $\pm$ 2.1	0.9 $\pm$ 4.5	1.9 $\pm$ 6	2.4 $\pm$ 38
3	0.1 $\pm$ 4.6	0.6 $\pm$ 9.2	0.3 $\pm$ 21	0.5 $\pm$ 1
4	0.6 $\pm$ 15.8	0.4 $\pm$ 23.5	0.1 $\pm$ 21	0.2 $\pm$ 0.3
5	5.3 $\pm$ 17.7	7.5 $\pm$ 23.9	0.1 $\pm$ 24	0.8 $\pm$ 3.8
6	0.3 $\pm$ 10.1	0.6 $\pm$ 15.5	0.9 $\pm$ 25	7.6 $\pm$ 151
7	0.5 $\pm$ 16.2	0.4 $\pm$ 27.6	0.3 $\pm$ 26	0.3 $\pm$ 1
8	0.7 $\pm$ 18.3	1.7 $\pm$ 27.4	2.1 $\pm$ 32	7.0 $\pm$ 84

If a maximal distance of 2 m is taken, 60 waveform pairs in campaign L1 and only a few pairs from other campaigns remain.

Note that in this case study an erroneous footprint size of 20 m for data campaign L3a was used, as was provided at the time in the GLA14 product release 428. Moreover, the footprint size for campaign L1a provided in the data is not reliable, the footprint size of 95 m  $\times$  52 m on average is used in this campaign (Abshire et al., 2005).

### (b) Differences between repeated waveforms

In Table 4.9 the average differences between the waveforms within the found waveform pairs are quantified, with respect to the waveform parameters as introduced in section 4.4.3. As expected the mean differences between waveforms within a pair are small, as both waveforms are obtained at approximately the same time and from the same location. The standard deviations of the differences are in most cases much higher. For one case this is illustrated in Figure 4.18, where the histogram of waveform extent differences for waveform pairs from epoch 1 are given. Differences are ordered in the sense that always the parameters of the more recent waveform are subtracted from the ones from the older waveform. No temporal trend can be observed but differences in waveform extent occur of up to 40m. Such differences can easily occur in areas with high buildings or steep rocks when the waveform footprints only partially overlap. More interesting are the many waveform pairs where only small differences occur. In a next step such pairs will be further analyzed to obtain possible relations with surface moisture changes, sea roughness changes or other more subtle changes that can be revealed by taking spatial correlation between changes into account.

#### 4.4.5 Case study: coinciding waveforms

Among the pairs of waveforms in the database some waveforms were found whose footprint centers were within 2m distance. As the acquisition time of these waveforms is almost coinciding as well, one would expect an almost perfect match between the waveforms. Here three cases are discussed where this reasoning does not hold true. In order to obtain insight into possible reasons for the unexpected differences in waveform characteristics, the three pairs under consideration are displayed in Figure 4.19 on the top, while the corresponding footprints, overlaid on images from Google Earth, are shown in the bottom.

##### (a) Case 1

The first waveform pair, Figure 4.19, left, is located nearby the city of Cottbus, Germany. The distance between footprints is 1.69 m. The acquisition date of the two waveforms differs by 16 days. The first raw waveform (cyan) shows a peak near the 340th nanosecond, that is absent in the second raw waveform (red). The location of this peak corresponds to a height above ground level of about 11m according to the GLA14 product. The Google image with the two similar and coinciding footprints overlaid, shows that a small misregistration of the footprint locations can explain the difference in waveform: apparently, the cyan waveform partly covers the building on the left of the photo. According to GLAS documentation, (NSIDC, 2007), the horizontal geolocation accuracy (i.e. distance between

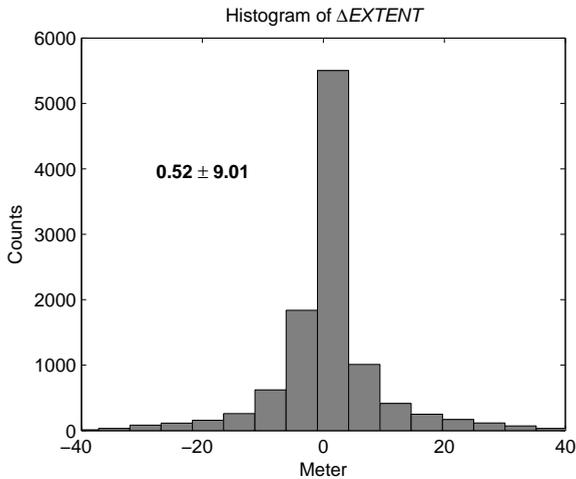


Figure 4.18: Histogram of waveform extent differences of epoch 1.

true and estimated footprint centers) has a mean of 4.6 m and a standard deviation of 9.3 m.

### (b) Case 2

The second pair is located in an open area in France. The distance between footprint centers is only 0.93 m (see Figure 4.19, middle). The extent of the second waveform (red) is wider than the extent of the first one (cyan). The wider waveform is displayed in red in the color image (bottom) and the other one in the cyan. These waveforms were acquired with a 8-day difference in 2005. The height difference between the waveforms is 30 cm, (GLA14). The wider waveform has a larger footprint size. As a consequence it contains more reflections from low vegetation. As a result the waveform has a widened extent. In this case, the accuracy of the footprint centering is reported to be  $2.9 \text{ m} \pm 3.7 \text{ m}$ , (NSIDC, 2007).

### (c) Case 3

The third pair is located over sea, south of Sardegna island, Italy. The footprint locations are identical, the shape is quite similar, but the amplitude is clearly different, Figure 4.19. The first waveform has a larger peak intensity and a bit smaller waveform extent. The difference in acquisition date is 12 days and the GLA14 height difference is 60 cm. The difference in waveform shape can be explained by changes in sea surface roughness or by changes in sea water parameters

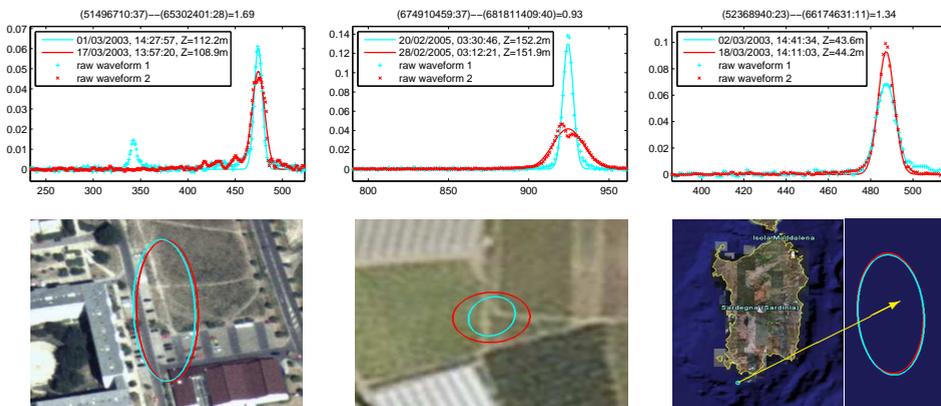


Figure 4.19: 3 case studies with 2m distance difference between footprint centers: waveform pairs (top) and footprint pairs over images of Google Earth (bottom).

like temperature or sediment concentration. A rougher sea can result in a wider waveform and in a lower energy return.

#### 4.4.6 Conclusion and further research

In this study a database of more than 100 000 waveform pairs over Europe is introduced. This database consists of waveform pairs, acquired within a period of a few weeks with footprints that at least partially overlap. For all pairs within the database, changes in waveform parameters are computed, showing small average changes, but with a large spread.

The heterogeneity of the intensively used space, which has to be measured with respect to GLAS footprint spacing and size, limits the possibilities of a change detection based on single waveform pairs. Because of geolocation accuracy this definitely holds for built up areas and assemblies of small agricultural fields, found in many areas throughout Europe.

A case study of three almost perfectly overlapping waveform pairs suggests that this database can be used to address issues like misregistration, full waveform water roughness and low vegetation parametrization.

### 4.5 Conclusion

In this chapter, different studies on the validation of ICESat data products have been described. These studies can be divided in two groups. First, studies on vertical accuracy of ICESat derived elevations, and second, studies on ICESat waveform shapes. In comparison to the previous validations described in section 4.2.3, the following is concluded:

In the first validation case, the comparison between ICESat and accurate airborne laser data over The Netherlands performed in this chapter confirms that the ICESat accuracy and precision strongly depends on land cover type. As expected, the accuracy decreases when the complexity of the surface increases (e.g., over forest and urban areas). With respect to different land cover type the accuracy is  $-21$  cm with a standard deviation of 20 cm over bare land/low vegetation,  $-24$  cm with a standard deviation of 28 cm over urban areas and  $-9$  cm with a largest standard deviation of about 45 cm on average over forest. Moreover, for the latest release version, i.e. 28<sup>1</sup>, the achieved horizontal accuracy of ICESat data is about 5 m. Moreover, this study indicates that highly accurate data and careful processing is needed to validate ICESat elevations: a previous comparison between ICESat and

---

<sup>1</sup>At September 30th, 2009

moderate quality SRTM data over bare land resulted in an average difference of 1.52 m. Here the difference is reduced to only 21 cm.

The second validation is performed in terms of waveform shapes. Internally repeated ICESat waveforms were studied in this chapter. However, at the current state of research, many problems are still unsolved. They consist mainly of first the lack of identification of completely overlaying footprints, and second of the lack of waveform pairs situated at stable areas. Unfortunately, the number of almost completely overlapping footprints from repeated ICESat measurements that could be obtained from campaign L1 to L3e in the region between 36N and 71N latitude and 11W and 33E longitude was very limited. Therefore, at this stage, it is still very challenging to identify the cause of changes in waveform parameters as occurring in (partly) overlaying waveforms. Performing a full world wide search over all campaigns is however expected to result in a quite large data set of suitable almost perfectly overlapping repeated footprints. Such data set could therefore be used to identify and resolve remaining ICESat processing issues.

Moreover, by applying the filtering constraints, a processing method for single ICESat full waveforms is developed. The last mode of the waveform is representing the terrain height well. Another parameter, the waveform extent, is suitable to describe forest height.



# Applications of large footprint full waveform data

## 5.1 Introduction

After applying all corrections derived from the calibration/validation procedure, the ICESat product can be used in different applications considering the surface of the Earth. How to develop applications by using large footprint full waveform data is discussed in this chapter. The ICESat products used can be the elevation data, the raw full waveform data or a combination of these two. In section 5.2, an overview of applications reported in literature is given first. By considering the full waveform parameters derived from the waveform processing procedures of chapter 3, two applications of ICESat full waveform data are studied in detail. In section 5.3, seasonal changes over broadleaf, mixed-wood, and needleleaf forests between winter and summer epochs of 2003, along near-coincident ground tracks are quantified. The results indicate that, although the maximum tree height barely changes over 6 months, i.e., less than 2.2% for three forest types, a suited waveform parameter can detect forest canopy change mostly for broadleaf (a 148% change, winter to summer) and less for conifers (a 36% change). Alternative waveform parameters to describe forest changes are also discussed. In section 5.4, it is demonstrated that ICESat full waveforms can be used for land cover classification. It is the first time that the possibility of using ICESat data is investigated for this purpose. Over The Netherlands ICESat footprint locations were classified into four categories: high vegetation (high trees or forest), urban, water, and bare land/low vegetation. The following waveform parameters are used as class attributes: return

energy, waveform extent, waveform start and number of Gaussian components. It is concluded that the accuracy of classification equals 73% in comparison to a confusion matrix based on the CORINE land cover database 2000 (CLC2000) covering the same study area.

## 5.2 Literature review

ICESat data products have been applied in many research topics. Due to the global coverage of ICESat data, applications can be found over ocean, polar and land surface regions but also in atmospheric research. However, in this section only applications over polar and land regions are discussed. We distinguish applications where single elevations are used (ICESat elevation product) from applications using the complete vertical distribution of elevation points (ICESat full waveform product).

### 5.2.1 Elevation applications

The ICESat elevation data consist of three kinds of elevation products. The GLA06 product contains global elevation data derived from the NASA standard fitting method covering both polar and land regions. In order to have a ready-to-use elevation product, the GLA06 product is further processed and categorized in two additional products: GLA12 for polar regions and GLA14 for land regions. The GLA12 product is obtained from the GLA06 product after applying all corrections for tidal, atmospheric and saturation effects. Similarly, the GLA14 product is also obtained from the GLA06 product with all corrections applied. However, in order to cope with the on average higher complexity of land surfaces, a maximum of six elevations per waveform as derived from the NASA fitting method is provided in the GLA14 product. In contrast, the GLA12 product provides only one elevation derived from fitting one single Gaussian mode.

Over the Antarctic ice sheet region, the quality of this GLA12 product was investigated by [Shuman et al. \(2006\)](#). For low slope (less than 2 degree) and clear sky conditions, the mean elevation difference of the GLA12 product at crossovers for campaign L2a and release 21 is about 2 cm with a standard deviation of about 20 cm. The ICESat elevation product was used to generate digital elevation models (DEM) ([DiMarzio et al., 2007](#)). [Brenner et al. \(2003\)](#) proposed to use the standard elevation product (GLA12) over polar regions to create a 500m DEM product over Antarctica and a 1km DEM product over Greenland, see [Figure 5.1](#). However, the quality of this DEM is not validated.

Elevation changes between co-located ICESat data sets were investigated in depth over East Antarctica by [Nguyen and Herring \(2005\)](#), using a combination of

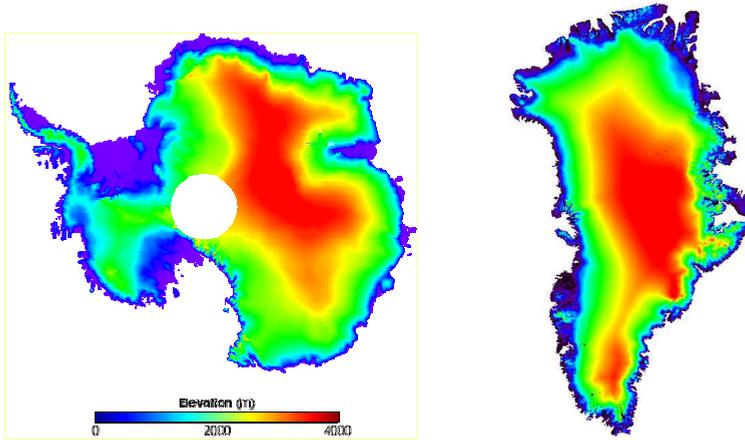


Figure 5.1: Digital elevation models over Antarctica (left, 500m resolution) and Greenland (right, 1km resolution). Figures downloaded from <http://nsidc.org/data/icesat/data.html>.

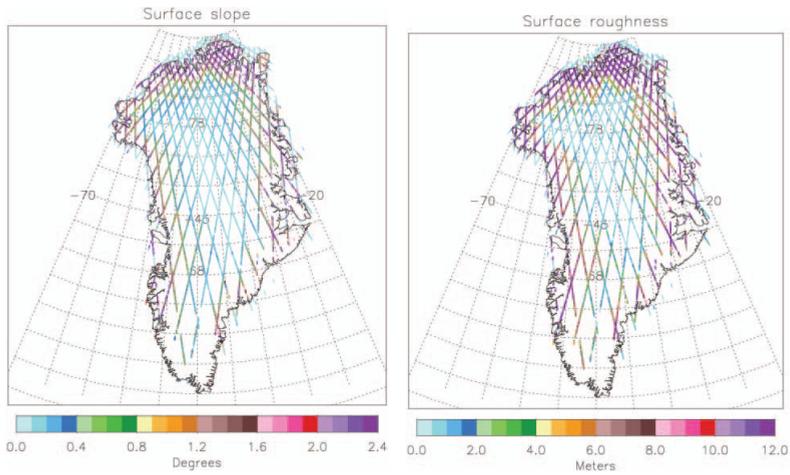


Figure 5.2: Maps of slope in degree (left) and roughness in meter (right) over the Greenland ice sheet. Figures are extracted from [Yi et al. \(2005\)](#).

Kalman filtering and Kriging. They concluded that at that time only two laser operational periods (L2a and L3a) of the GLA06 product had adequate pointing calibration to be used for height change detection. Detected elevation changes

are used in estimating the volume change rate of the Greenland ice sheet (Slobbe et al., 2008) and, as proposed by Wahr et al. (2000), in mass balance evaluation of the Antarctica and Greenland ice sheets in combination with GRACE gravity data (Slobbe et al., 2008b), and in determining snow accumulation on ice sheets (Bindschadler et al., 2005). Alternatively, ICESat elevations of repeated/adjunct and crossed ground tracks within a short duration (36 days) are used to infer the effect of surface slope and roughness (Yi et al., 2005). By applying a suited interpolation method, all ICESat observations within 1 km of the ICESat reference track are selected to make a slope and roughness map of the Greenland ice sheet, but restricted to the areas around the tracks, see Figure 5.2.

Moreover, ICESat elevation data can be used together with other sources in order to describe surface characteristics. Zwally et al. (2008) proposed to derive sea ice freeboard heights from ICESat elevation profiles. The freeboard height ( $F$ ) is defined as the total height of the combination of snow cover and above ocean sea ice with respect to a referenced ocean height, see Figure 5.3(a). The freeboards are determined relative to the ocean reference level detected over areas of open water and very thin ice within the sea ice pack. The detected freeboards are validated by available Envisat Synthetic Aperture Radar (SAR) images. Furthermore, the freeboards are applied to estimate sea ice thickness by using snow depth on the sea ice from AMSR-E passive microwave data together with nominal densities of snow, water, and sea ice.

According to Brenner et al. (2007), a comparison between time-coincident ERS-2, Envisat, and ICESat elevations at cross-mission and crossovers should enable the characterization of range errors in the Envisat and ERS-2 radar altimeters over the continental ice sheets. The radar elevation error budget was modeled as a function of the surface slope. Additionally, Fricker and Padman (2006) stated that ICESat elevation products contribute significantly to improving knowledge

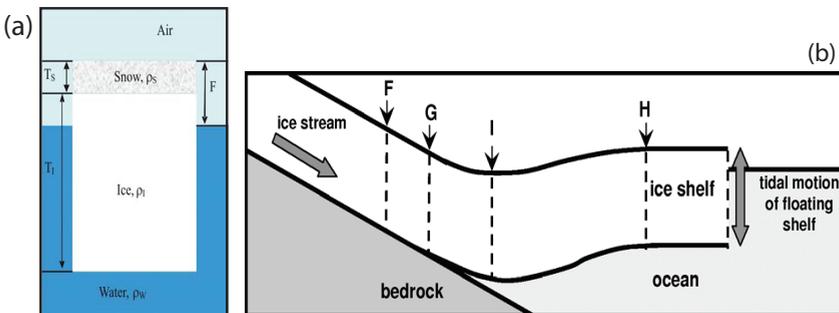


Figure 5.3: The definitions of freeboard (Zwally et al., 2008) and grounding zone (Fricker and Padman, 2006) are illustrated in (a) and (b) respectively.

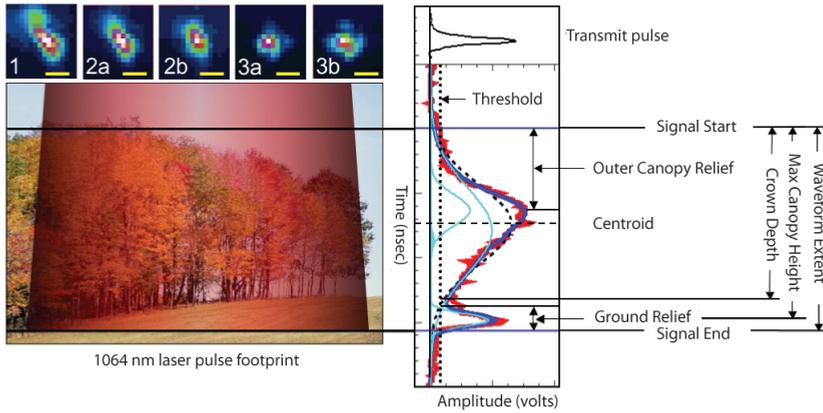


Figure 5.4: Biophysical parameters are defined and extracted from the waveform shape over low-relief surfaces. Figure extracted from [Harding and Carabajal \(2005\)](#)

of grounding zone structures and to studies requiring accurate grounding zone locations, compared to the traditional method using satellite imagery or digital elevation models. The grounding zone is defined as the transition region between the fully grounded ice sheet and the free-floating ice shelf, that is, the region between locations  $F$  and  $H$  in Figure 5.3(b). Moreover, ICESat elevation data were used to register a set of high resolution satellite images for measuring elevation changes on Byrd Glacier, Antarctica ([Schenk et al., 2005](#)) and to detect spatial patterns of surface elevation change on Siple Coast ice streams, Antarctica ([Csatho et al., 2005](#)).

In most cases, the ICESat elevation product (GLA12) is used for developing applications over polar regions. The elevation product is obtained by determining the peak location of a single Gaussian mode fitted to the return waveform. These applications show good achievements, however, so far no applications using GLA01 full waveform data over the polar regions have been discussed.

### 5.2.2 Waveform applications

The ICESat waveform data are able to reveal many previously unmapped features on ice shelves, such as crevasses, rifts, grounding zones and ice fronts. [Fricke et al. \(2005\)](#) show the possibility of detecting rifts and measuring ice and snow trapped inside the rifts. By applying the NASA waveform fitting product (GLA05), containing a maximum of six Gaussian components, a detailed vertical distribution of elevation points is obtained. Therefore the rifts are visibly enhanced and extracted from elevation profiles along ICESat ground tracks. The measured snow within

this rifts is then validated by in situ measurements.

The shape of ICESat full waveform data are well investigated for forestry applications. Over low-relief areas with tree cover, the waveforms and derived elevation products provide useful biophysical parameters, see Figure 5.4, including maximum canopy height, crown depth, outer-canopy ruggedness, and a measure of canopy cover (Harding and Carabajal, 2005; Sun et al., 2008). The maximum canopy height can consecutively be used to estimate the aboveground biomass. Results were compared to field measurement data over the Santarem study area, Para State, Brazil by Lefsky et al. (2005) and Boudreau et al. (2008). In high-relief areas, however, the waveform extent is greater than in low-relief areas due to slope effects. Therefore, the canopy height is not accurately estimated. In order to reduce this effect, existing DEM data are studied by Lefsky et al. (2007) and Rosette et al. (2008). They proposed to use terrain index (TI; maximum minus minimum elevation) from a 10 m DTM (Rosette et al., 2008) and 90 m SRTM data (Lefsky et al., 2005) to adjust the waveform extent. Moreover, Lefsky et al. (2007) stated that the leading and trailing edge extents of the waveform have a direct relationship to the influence of terrain slope and canopy height variability. Therefore, these parameters were applied to estimate more precisely the canopy height via an adjustment of the waveform extent, see Figure 5.5.

In addition, forest species like evergreen conifer, deciduous conifer and mixed forest, were classified using ICESat waveform parameters like front slope, and centroid locations (Ranson et al., 2004). Forest disturbance caused by fire could be identified by comparing corresponding waveforms before and after the fire (Ranson et al., 2006). The validation is performed using MODIS data. Simard et al. (2008) presented 3D mapping results of mangrove forests based on estimation of tree height and biomass distribution in the Cinaga Grande de Santa Marta, Colombia.

In the above described research projects, waveform data are mainly applied for analyzing forest structure. Applications of the full waveform product on other fields are somehow missing. An analysis of repeated waveform pairs, as potentially available from the repetition of ICESat ground tracks, is also not investigated. Therefore, in this chapter, waveforms at repeated ICESat ground tracks with partly co-located footprints will be applied to detect forest canopy changes in section 5.3. Such waveform pairs observed at different seasons are used to study the influence of leaf-on and leaf-off conditions over The Netherlands.

The other application based on waveform data that will be developed is land cover classification. Instead of using the spectral response at different bands of hyperspectral sensors/data, waveform classification applies the vertical distribution of all objects within the approximately 70 m footprint to distinguish between land cover classes. This new method is described and validated in section 5.4.

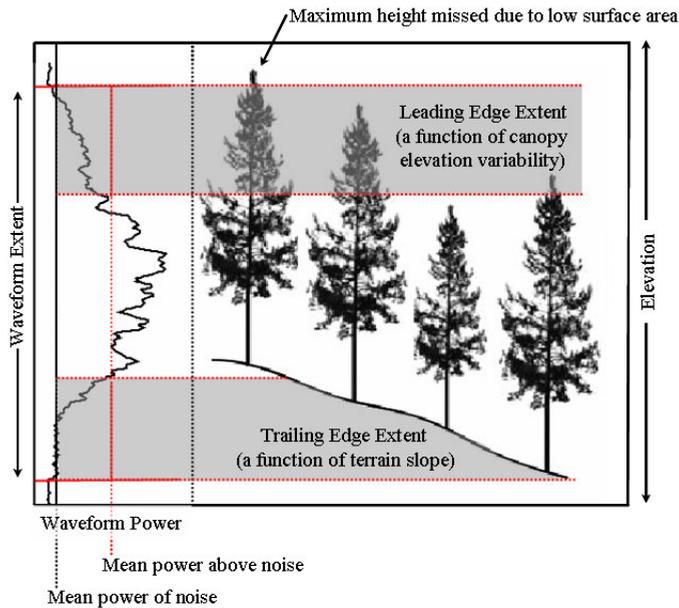


Figure 5.5: Biophysical parameters are defined and extracted from the waveform shape over high-relief surfaces. Figure extracted from [Lefsky et al. \(2007\)](#).

## 5.3 Case study 1: Single and two epoch analysis over forested areas

### 5.3.1 Background

Characterisation and quantification of forest canopy structure across extensive areas challenge remote sensing scientists ([Harding et al., 2001](#)). Forest canopy structure is defined as “the organisation in space and time, including the position, extent, quantity, and connectivity, of the aboveground components of vegetation” ([Parker, 1995](#)), and plays a key role in developing a better understanding of how forest ecosystems function ([Drake et al., 2002](#)). The forest canopy is also defined as “the collection of all leaves, twigs, and branches formed from the combination of all the crowns in the stand” ([Maser, 2001](#)), and is responsible for the majority of material and energy exchanges with the atmosphere ([Lefsky et al., 1999](#)). However, accurate estimation of aboveground forest biomass at regional/subcontinental scales remains a major obstacle when using conventional remote sensing techniques ([Dubayah et al., 1997](#)).

A new method for analysing individual waveforms and pairs of waveforms for the derivation of forest structure parameters is introduced in this case study. The purpose of this case study is twofold. First, it seeks to extract forest parameters from single shots of a satellite-borne full waveform laser altimeter. Second, it attempts to show the utility and limitations of analysing ICESat waveform pairs acquired over roughly the same footprint during leaf-off and leaf-on seasons. The first objective means a change of paradigm in extracting forest parameters from airborne laser scanning data. So far, high point densities (typically 1–10 points/m<sup>2</sup>) of discrete return laser altimetry data have been used (Næsset et al., 2005). Data are typically acquired in dedicated missions and information like mean tree height, basal area and timber volume is provided at the stand-level properties. The latter is necessary because a sufficient amount of information (discrete backscattering points) is required for deciding which points are above the ground and which are on the actual ground, and for linking forest parameters to point observations via statistical procedures. We aim to demonstrate the derivation of forest parameters on the basis of the footprint area of individual, full waveform space-borne laser-ranging measurements. Limitations imposed by GLAS data characteristics, e.g., spatial misregistration, changes in footprint size and laser power, can only be partially mitigated, making conclusions concerning forest structural change over time difficult, but not impossible.

The second objective, i.e. using repeatedly-measured ground spots for deriving forest change metrics, is an extension of the shot-based approach in a multi-temporal embedding. To achieve this objective, ICESat full waveform data from two different epochs are compared and combined. One data set was acquired by Laser 1, L1, in winter, February 2003, the other by Laser 2, L2, in summer, September 2003. The two data sets are obtained from repeated tracks and therefore many pairs of footprints from both data sets exist that overlap entirely in the ideal case and at least have some common topographic intersection in all other cases. Because the data are acquired from two lasers with different powers in the transmitted pulses, the received intensities of corresponding waveforms from winter and summer are not comparable and we consequently consider normalised waveforms. Also atmospheric attenuation is not necessarily the same. Using a full waveform analysis, forest parameters of both data epochs are calculated and compared: maximum canopy height (*dEcho*), height of median energy (*HOME*), ratio between total intensity of ground return and canopy return (*rGround*), and canopy intensity difference ( $\Delta I$ ). This study proposes a direct way of determining the seasonal change in forest structure from pairs of waveforms within at least partly overlapping footprints. As an application, but mainly as a method of validation, general forest waveform pairs are classified into forest types, resulting in a kappa value of  $\kappa = 0.57$ .

In section 5.3.2, the study area and data input are introduced. Then a method for preceding the actual inter-epoch waveform comparison is described in section 5.3.3.

In section 5.3.4, the results of the comparison between summer and winter waveforms are presented. Problems with the current data set, mainly caused by shot-to-shot misregistration and by the effects of topography on the waveform, are identified and discussed. Although the data used here are from ICESat, the methods that are described can also be generalised and applied to other waveform LiDAR data sets. At the end of this case study, recognised problems and future works are given in section 5.3.5 and 5.3.6.

## 5.3.2 Input forest data

### (a) ICESAT/GLAS data

In this study, we investigated the product GLA01, the global full waveform data, and GLA14, the global land surface altimetry data. The data we analysed belong to a track covering part of the Netherlands, Belgium and France. The GLA14 was used to visualise the geolocation of the waveform data, which appears as a straight line in Figure 5.6. The ICESat data used in this area stemmed from two epochs, one from 27-02-2003 (winter season, blue track) and the other from 30-09-2003 (end of summer season, red track). The ICESat footprint had elliptical shapes with an average size of  $95 \text{ m} \times 52 \text{ m}$  (Abshire et al., 2005) for the winter data. The azimuth of the major axis was  $162.5^\circ$ . The summer data had an average footprint size of  $87 \text{ m} \times 43 \text{ m}$ , with an azimuth of  $184.2^\circ$ . The attributes of the footprints, major and minor axes, and the azimuth angle of the major axis that described the summer data were provided as metadata from the NSIDC. However, the footprint attributes for the winter acquisition were taken from Abshire et al. (2005) because of operational errors associated with the winter data acquisition.

There were 2848 waveforms in the winter season data set and 4358 waveforms in the summer one (see Table 5.1). The waveform data of both epochs were decomposed into Gaussian components described in chapter 3. The number of fitted waveforms was 2775 (97% fitted successfully) in winter, and 4284 (98%) in summer. These numbers were comparable to the fitting success figure of 98% reported by Wagner et al. (2006) for small footprint airborne full waveform laser scanning from a flying height of 500m. The fitting and decomposition step did sometimes not give a solution for noisy waveforms that did not satisfy the condition outlined in section 3.3.7. Moreover, the number of intersecting waveform footprints between epochs was 2600 before and 2363 after the decomposition step. The 2363 waveforms were then categorised in land cover types.

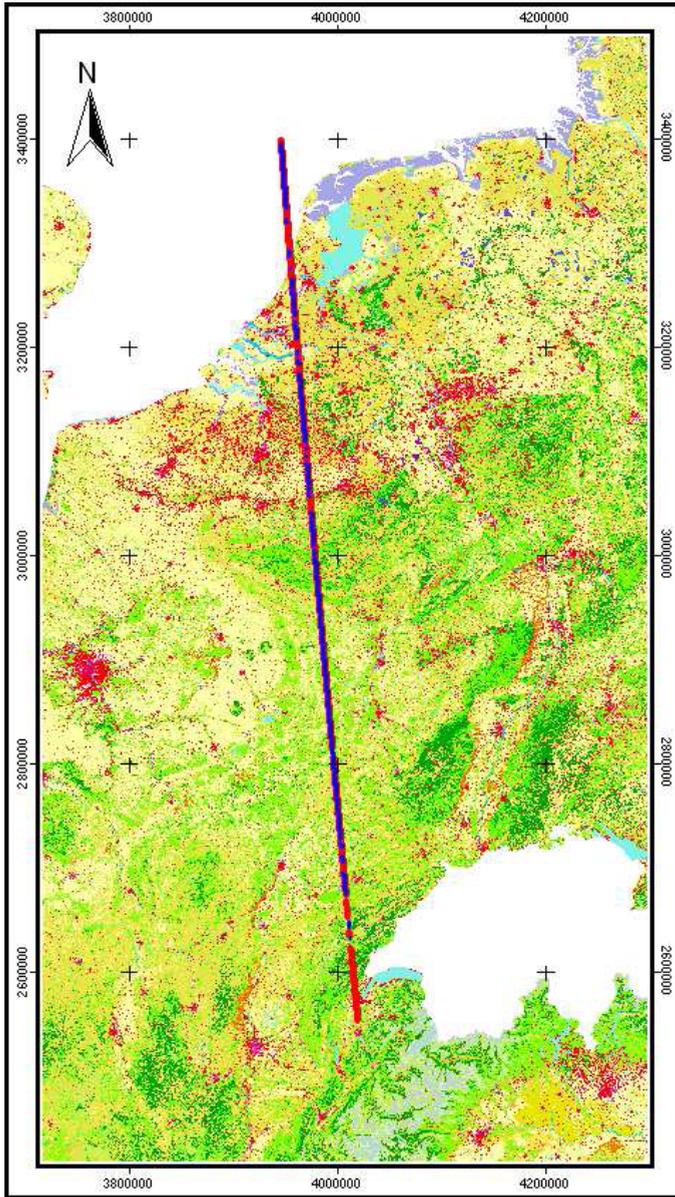


Figure 5.6: ICESat groundtracks from February (blue) and September 2003 (red) overlaid on the CORINE land cover image.

Table 5.1: Number of waveforms used

Data	Total	Fitted	Broad	Mixed	Needle	Urban	Bare	Water
Feb-03	2848	2775	559	102	111	204	1444	355
Sep-03	4358	4284	702	130	201	455	2091	705
Overlap	2600	2363	440	62	89	170	1248	354
$nModes$ (F/S) <sup>a</sup>	N/A	2.1/2.0	3.3/3.5	3.9/3.5	3.2/3.3	2.3/2.2	1.8/1.5	1/1
$eEcho$ (F/S) <sup>b</sup> $\times 10^{-15} J$	3.7/8.6	3.7/8.4	4.9/9.0	7.2/10.1	6.6/6.7	3.5/10.9	3.3/9.8	2.1/1.6
$wEcho$ <sup>c</sup> (F/S) $\times m$	N/A	4.3/3.2	23.0/28.1	30.3/28.4	26.7/24.8	13.4/12.8	11.9/9.3	6.7/2.6

<sup>a</sup> Number of modes (Feb/Sep).

<sup>b</sup> Total Energy (Feb/Sep).

<sup>c</sup> Waveform extent.

### (b) Footprint shifts

On average, the distance between the centre points of corresponding footprints was 67.83 m. The footprint pair distance varies from 40 m to 90 m. The majority of this footprint distance is either about 60 m or 75 m in this data set. This may cause an inaccuracy in the change detection of forest areas. However, for flat or low relief area covered by homogeneous forest, we assumed that similar waveforms were returned from all over the area. In case of significant topographic relief or nonhomogeneous forests, a comparison of a single waveform pair was not expected to give accurate results, but considering many waveform pairs simultaneously was likely to result in interpretable average waveform changes. We could therefore detect seasonal changes in the forest structure by looking at averaged changes across near-coincident, paired footprints. In Figure 5.7, the grey ellipse footprints represent data tracked in September 2003, while the black footprints were collected in February 2003.

### (c) Intensity difference

The intensity, or the full returned energy of the waveform data, in September, tended to be larger than in February. This is illustrated in Figure 5.8, where histograms of the intensities in February and in September are provided. The mean intensity differed by a factor of almost 2 (see Table 5.1). One of the reasons for this difference was the change from GLAS laser L1 to laser L2 aboard of the ICESat satellite. Another reason was maybe caused by the foliage and herbaceous vegetation that was more abundant in September resulting in larger reflected laser energy in the near infrared laser wavelength. This could explain, for example, the large differences in average intensity return for the broadleaf and the bare land classes. As a consequence, we could not directly compare waveforms from the two different seasons (confirmed by NSIDC). Therefore, relative intensities, which were obtained by the waveform normalisation step, will be considered in the following

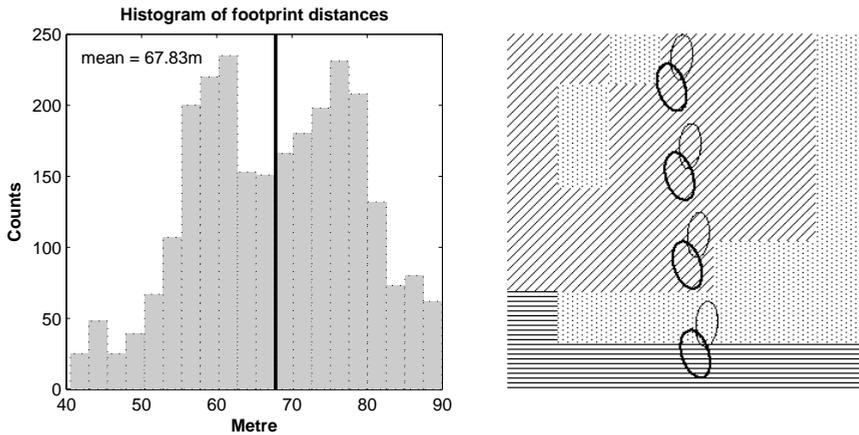


Figure 5.7: Left: histogram of inter pair distance and its mean (vertical black line). Right: visualisation of footprint locations in winter (thick black) and summer (thin black) and the 100m CLC2000 map with bare land (dot pattern), needleleaf (horizontal line pattern) and broadleaf (slant line pattern).

sections.

#### (d) CLC2000 land cover data

The CORINE Land Cover 2000 database (CLC2000) is introduced in section 4.3.2 and then reclassified into six land cover types: urban, bare land/low vegetation, water, broadleaf forest, mixed forest, and needleleaf forest. Pairs of decomposed waveforms were identified as belonging to one of these six land-cover types. The 2363 waveform pairs were divided into broadleaf (deciduous, 440), needleleaf (coniferous, 89), mixed (62), urban (170), bare land/low vegetation (1248) and water/sea (354). Before classifying waveforms, however, an additional step was needed to convert ICESat geo-location data (GLA14) to the coordinate system of CLC2000. The ICESat geo-location used the same ellipsoid as TOPEX/Poseidon and Jason-1 (radius of equator 6378136.30 m and reciprocal flattening ( $1/f$ ) 298.257 (NSIDC, 2005)). Therefore, the GLA14 geo-location data were converted to WGS84 by a tool developed by the NSIDC and were then transformed into the European coordinate system (CORINE land cover 2000, 2006) by ArcGIS 9.0 software (www.esri.com). Finally, the ICESat geolocations were overlaid with the CLC2000 to classify the waveforms according to the land cover types.

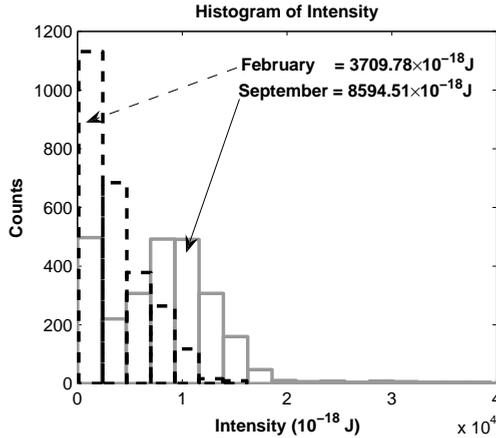


Figure 5.8: Two intensity histograms displayed together: February, 2003 (dashed) and September, 2003 (grey) with its mean intensity. The mean intensity of the summer data is 2 times the mean intensity of the winter data.

### 5.3.3 Methodology

In order to compare corresponding waveforms, the initial processing steps for individual waveforms as described in chapter 3 are applied. Notably the waveform intensity is normalised because of the observed differences in average intensity between epochs as discussed in section 5.3.2. Then two additional steps for processing waveform pairs are introduced: waveform shift computation in section 5.3.3(a) and determining waveform parameters for forest structure in section 5.3.3(b).

In the following, the right-most Gaussian component of the waveform decomposition will be referred to as the last mode, as this mode corresponds to the energy reflected by the surface hit last (see Figure 5.9). In forest applications, the last mode will generally correspond to the bare earth below the trees, as long as the earth surface is not completely occluded by vegetation. The left-most Gaussian component will be referred to as the first mode, as this component corresponds to the first feature in the laser footprint that reflects the laser light. Over forest areas, the first mode mostly originates from reflections by the tree canopy.

#### (a) Waveform shift computation

The raw waveform is distributed in the GLA01 product, but the GLA01 data are not georeferenced. GLA14 data are georeferenced, and the link between the two data sets is a common, unique index number associated with each shot. Each winter and summer GLA01 waveform was annotated with the corresponding

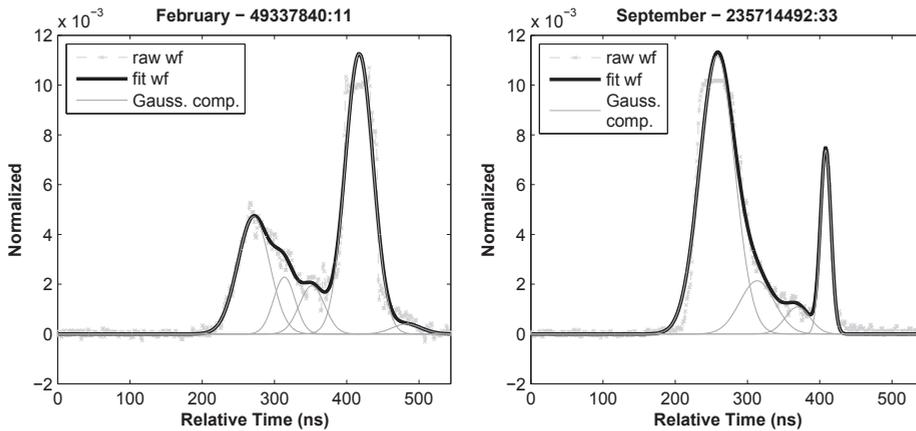


Figure 5.9: Two fitted waveforms (black solid) are displayed together with the raw waveform (grey starred line) and the Gaussian components (grey solid) for February, 2003 (left) and September, 2003 (right).

GLA14 shot location. Coincident or near-coincident winter and summer GLAS shots were identified based on their geographic coordinates. Once multitemporal (winter/summer) pairs were identified, the paired, normalised waveforms had to be aligned. Optimal shift parameters were computed in order to make the normalised waveforms comparable along the waveform amplitude axis (y-axis). To do this, the normalized waveforms are georeferenced in the same coordinate system first, see details in section 4.3.3(c).

The shift computation can be then performed on the complete waveform (Hofton and Blair, 2002) or on just its last mode. The complete waveform gave good matching results in case of similar waveform shapes in two epochs. The last mode method was useful in case of rather different waveform shapes in the two epochs. In the latter case, the ground surface was assumed to be stable, which implied that the last modes of the two epochs should be matching. However, the waveform sometimes had a ‘tail’ on the side of the last mode, which was a small Gaussian mode found by the fitting algorithm (see Figure 5.9, on the left). Such a tail mode was not considered to represent the ground surface. A modified last mode method was therefore developed, which was expected to result in a better match. If the amplitude ratio between the last mode and the second-to-last mode was less than 15 percent, the last mode was removed and the second-to-last mode was used as the new last mode. A new decomposition result was determined by removing the small last mode from the previous decomposition result without any need of repeating the decomposition step.

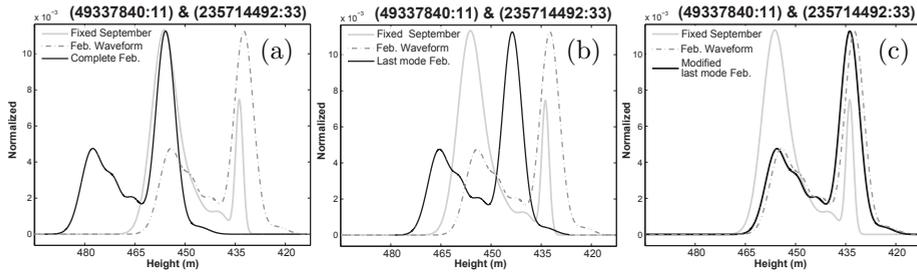


Figure 5.10: The original summer waveform is kept fixed (grey solid). The original winter waveform (dashed dot line) is shifted by: (a) complete waveform, (b) last mode, and (c) modified last mode. The shifted waveform is displayed in a black line.

For a given pair of coincident, winter/summer, normalised and georeferenced waveforms, the shift is found by determining that time shift  $m$ , for  $m = 1, \dots, 2N - 1$ , that maximizes the Pearson correlation between the georeferenced waveforms of February,  $\underline{x}$  and September,  $\underline{y}$ . Note that in our case,  $N$  can be the length of the entire waveform ( $N = 544$ ) or just the length of the last mode of the waveform ( $N \approx 100$ ). The Pearson correlation  $\rho$  for the time shift  $m$  is defined as

$$\rho(\underline{x}, \underline{y}) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5.1)$$

where  $\bar{x}$  and  $\bar{y}$  are the mean normalised amplitude of the winter and summer waveform, respectively.  $x_i$  and  $y_i$  are the amplitude of the  $i^{\text{th}}$  element of the winter and summer waveform.

The shift operation is illustrated in Figure 5.10. The two last modes (peaks) of the summer and winter waveforms, recorded at approximately the same location in the same year, did not match. Therefore, a shift was determined with the cross-correlation method. This method gave an incorrect match (dash) based on the complete waveform due to the strong change in waveform shape. The shifted, dotted grey waveform based on the last mode did not match well due to noise in the tail of the winter waveform. Employing the modified last mode approach, the noisy tail of the winter waveform was removed, producing a good match between the winter and summer waveforms.

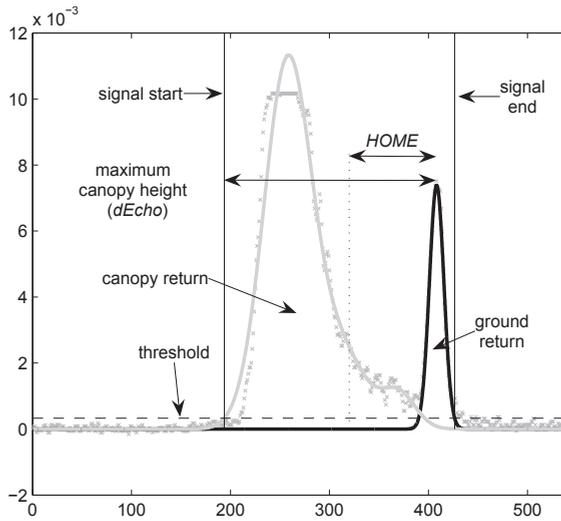


Figure 5.11: Waveform parameters: maximum canopy height ( $dEcho = 32.15$  m (214.3 ns)), canopy return, ground return, height of median energy ( $HOME = 20.60$  m (137.3 ns)). The  $rGround$  for this waveform is 0.16.

### (b) Waveform parameters for forest structure

After the normalisation and shift operation, we were able to quantify the difference in canopy between a corresponding summer and winter waveform by analysing the change in a number of parameters that describe the canopy structure.

Consider two waveforms, one waveform  $W_F$ , from February and its corresponding waveform  $W_S$ , from September, both normalised and, in addition, matched by the shift operation. A forest waveform normally has two dominant peaks that represent the canopy portion (the left side of the waveform) and the ground portion. The true (modified) last mode of the fitting result was used as ground return and the remaining waveform, the difference between the full waveform (fitted) and the ground return, was used as the so-called canopy return.

Three parameters that describe forest structure were used in this investigation (see Figure 5.11). First, a maximum canopy height ( $dEcho$ ), or tree height (Harding et al., 2001), was defined as the distance from the position of the signal start of the waveform to the peak position of the ground return. Second, a height of median energy ( $HOME$ ) (Drake et al., 2002) was defined as the distance from the peak of the ground return to the position of median energy of the fitted full waveform. Third, the ground return ratio ( $rGround$ ) was calculated as the total energy of the

ground return (the last mode), divided by the total energy of the canopy return.

The  $dEcho$  or the tree height was not expected to change much between the winter and summer seasons. Therefore, differences in  $dEcho$  between two epochs ( $\Delta dEcho$ ) should be small. In analysing the results, the  $\Delta dEcho$  was used as validation. However, the  $HOME$ , a measure for the distribution of the returned energy, was found to be quite sensitive to changes in the canopy. For example, Figure 5.9 shows a winter waveform (left) and a summer waveform (right) over a broadleaf forest. The winter waveform clearly has a much smaller canopy return. This is reflected in the position of the median energy that is much closer to the ground. Finally, the ground/canopy ratio value ( $rGround$ ) is sensitive to both the intensity return from the canopy and the ground. For the summer waveform, the intensity return is larger from the canopy and smaller from the ground. As a consequence, the  $rGround$  in Figure 5.9 small for summer and large for winter.

The canopy portion or canopy return of the waveform was defined as the difference between the fitted waveform and the last mode. Therefore the canopy intensity difference of a given pair of February/September waveforms was determined as the mean squared intensity difference of a pair of February/September canopy returns.

If  $VC_F$  and  $VC_S$  denote the intensities of the canopy return of the waveform in February and September, the canopy intensity difference was defined as follows:

$$\Delta I(WC_F, WC_S) = \sum_{i=1}^K \frac{(VC_F(i) - VC_S(i))^2}{K} \quad (5.2)$$

where  $K$  is the number of non-noise height bins in the canopy, i.e. non-ground portion of the normalised waveforms,  $WC_F$  and  $WC_S$  are the waveform canopy returns of February and September, and  $\Delta I$  is the average of squared differences in normalised amplitudes.

For the comparison of  $W_F$  and  $W_S$  we introduced four parameters: a difference in maximum canopy height ( $\Delta dEcho$ ); a difference in height of median energy ( $\Delta HOME$ ); a difference in the ratio of total intensity between ground and canopy return ( $\Delta rGround$ ); and the canopy intensity difference ( $\Delta I$ ). The differences between the four parameters were calculated under leaf-on and leaf-off conditions.

### (c) Confusion matrix: validation of classified forest species

As a first application, but especially as a validation, the changes of the forest parameter values were used to classify general forest waveform pairs into the three forest type classes (broadleaf, mixed, needleleaf). The forest parameters from section 5.3.4 were used to identify the forest type with waveform pairs.

The result of the forest classification was validated by a confusion matrix (Lillesand et al., 2004) against the CLC2000 data. A confusion matrix compares, on a category-by-category basis, the relationship between known reference data (the CLC2000 data) and the corresponding classification results (the ICESat data). The matrix shows how well a classification is categorized, but also how bad the classification errors of omission and commission are.

Several other descriptive measures can be obtained from the confusion matrix. The *overall accuracy* is the probability that a pixel randomly taken from the classified data has the same class as the corresponding pixel in the reference data and vice versa. The *overall accuracy* is computed by dividing the total number of correctly classified pixels (the sum of the elements along the major diagonal) by the total number of reference pixels. The *producer's accuracy* indicates how well the reference data of the class are classified. The *user's accuracy* indicates the probability that a pixel classified to a class actually represents that class in the reference data.

The most common parameter for the classification accuracy is kappa  $\kappa$  (Lillesand et al., 2004). The  $\kappa$  parameter is conceptually defined as in Equation (5.3):

$$\kappa = \frac{\text{observed accuracy} - \text{random agreement}}{1 - \text{random agreement}} \quad (5.3)$$

where the *observed accuracy* is the proportion of correctly classified entries and the *random agreement* is the proportion of classes from classification data that could be expected to be classified by chance. Actually,  $\kappa$  is computed according to Equation (5.4).

$$\kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \quad (5.4)$$

where  $r$  is number of rows in the confusion matrix;  $x_{ii}$  is an observation in row  $i$  and column  $i$  (on the major diagonal);  $x_{i+}$  is the total number of observations in row  $i$  (shown as marginal total to the right of the matrix);  $x_{+i}$  is the total number of observations in column  $i$  (shown as marginal total at the bottom of the matrix);  $N$  is the total number of observations included in the matrix.

### 5.3.4 Results and discussion

#### (a) Waveform shifts

As mentioned in Section 5.3.3, shifts occurred between pairs of georeferenced GLA01 waveforms. In order to avoid anomalous effects in the shift computation

and the further parametrization of the forest structure, two filtering constraints are defined as follows.

- Remove those paired GLAS shots where one or both shots are unimodal.
- Remove those paired GLAS shots where  $|\Delta dEcho| \geq 5$  as such change is considered unrealistically high.

First, 440 broadleaf, 62 mixed, and 89 needleleaf waveform pairs are identified from 2363 fitted, February-September waveform pairs. These pairs are then filtered according to the above constraints. After applying the constraints, 154 broadleaf, 40 mixed, and 44 needleleaf GLAS pairs remain. For those forest pairs, and for 170 urban, 1248 bare land and 354 water, the shifts are computed and summarized in Table 5.2.

The results in Table 5.2 indicate that the three shift approaches generate similar shifts for water, bare land/low vegetation, and urban areas, approximately  $-0.5 \pm 1.0$  m. However, the three methods give different outcomes for forest pairs. For all three methods, the average shift is largest for broadleaf, medium for mixed and smallest for needleleaf. This can be explained as follows: forest canopy structure is strongly affected by seasonal influences while needleleaf is hardly affected. The modified last mode results in the smallest shift. Moreover, the shift resulting from the modified last mode for all classes is quite comparable.

The shift calculated for forest waveform pairs is in general larger than for waveforms in open terrain. This might be caused by the following reasons:

- The returned waveforms or waveform pairs from the forest areas have more complicated shapes and a greater noise component relative to waveforms in open areas.
- Waveforms returning from high relief terrain or from terrain covered by dense forest have wide last modes.
- In some cases, the last mode is not identified properly by the fitting algorithm.

Table 5.2: Median and standard deviation values of the shift comparison are computed using Median/MAD robust statistics (Muller, 2000).

Based on	<i>Broad</i>	<i>Mixed</i>	<i>Needle</i>	<i>Urban</i>	<i>Bare</i>	<i>Water</i>
Complete	$-7.7 \pm 11.9$	$-1.2 \pm 4.1$	$-0.6 \pm 2.0$	$-0.5 \pm 0.9$	$-0.5 \pm 1.4$	$-0.5 \pm 0.0$
Last mode (LM)	$-0.5 \pm 5.4$	$-0.8 \pm 1.4$	$-0.0 \pm 1.8$	$-0.5 \pm 1.6$	$-0.6 \pm 2.0$	$-0.3 \pm 0.2$
Modified LM	$-0.5 \pm 3.2$	$-0.8 \pm 1.4$	$-0.2 \pm 0.9$	$-0.3 \pm 1.1$	$-0.5 \pm 1.4$	$-0.3 \pm 0.0$

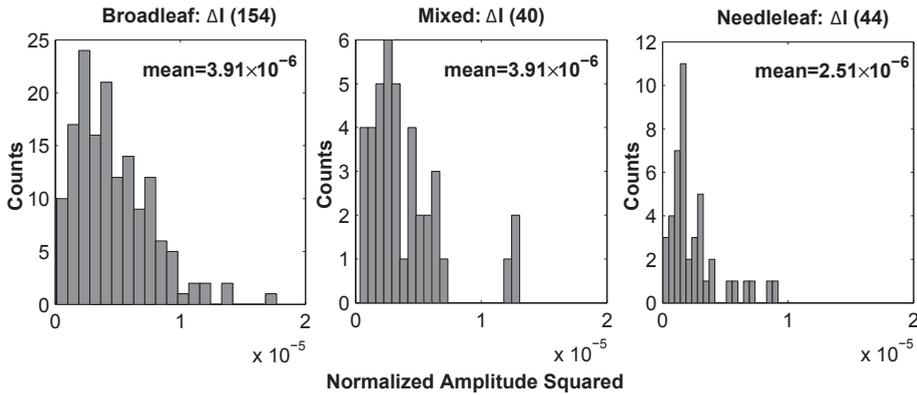


Figure 5.12: Histograms of canopy intensity differences ( $\Delta I$ ) with  $-5 \text{ m} < \Delta dEcho < +5 \text{ m}$  in broadleaf (left), mixed (middle) and needleleaf (right).

- The waveforms from areas of dense forest may only represent forest canopy, in case no laser light could penetrate to the ground.
- The larger normalised waveform shifts between paired broadleaf shots could be explained by the sensitivity of the centroid of a waveform to seasonal foliage changes.

In this case study, we assumed that the ground return is stable. As a consequence, the use of the last mode of the decomposition result is considered more reliable in the matching step. Because the complete waveform method occasionally results in a wrong match, and because a large second-to-last mode is sometimes trailed by a small last mode, the matching result from the modified last mode method is employed in subsequent analysis.

Shifts of the last mode between overlapping waveform pairs can be caused, for example, by a change in atmospheric or instrumental conditions between acquisition time, or by seasonal changes in the foliage cover. In that case the laser light cannot penetrate to the ground surface, and a mode representing the ground surface will be completely absent.

### (b) Waveform parameters for forest structure

The forest parameters derived from these waveform pairs are reported in Table 5.3. The intensity differences between the canopy waveform returns are given in Figure 5.12. On average, the largest change is detected for broadleaf forest, following

Table 5.3: Forest parameters derived from waveform with  $-5\text{m} < \Delta dEcho < +5\text{m}$ 

Forest Type	Season	$dEcho$ (m) (Tree height)	$HOME$ (m) (Median)	$rGround$ ( $\frac{Ground}{Canopy}$ )	$\Delta I$ (Intensity Distance)
<i>Broad-leaved</i>	Winter	24.65	4.85	1.80	N/A
	Summer	24.70	12.06	0.61	N/A
	<b>Difference</b>	0.05	7.22	-1.20	$4.63 \times 10^{-6}$
	<b>(%) compared to winter</b>	0.20	148.86	66.67	N/A
<i>Mixed</i>	Winter	26.46	6.13	1.25	N/A
	Summer	26.40	10.20	0.47	N/A
	<b>Difference</b>	-0.06	4.07	-0.78	$3.90 \times 10^{-6}$
	<b>(%) compared to winter</b>	0.22	66.39	62.40	N/A
<i>Needle-leaved</i>	Winter	24.44	8.54	0.77	N/A
	Summer	23.92	11.70	0.40	N/A
	<b>Difference</b>	-0.52	3.15	-0.36	$2.51 \times 10^{-6}$
	<b>(%) compared to winter</b>	2.13	36.88	46.75	N/A

by a moderate change for mixed forest, while the smallest change occurs for needleleaf forest, as expected.

Table 5.3 quantifies changes in forest canopy structure from winter to summer, measured through the use of GLAS. The maximum canopy height for the forest study area is approximately 25 m. The differences in  $dEcho$  from winter to summer are small on average over the three forest types (0.05 m for broadleaf,  $-0.06$  m for mixed,  $-0.52$  m for needleleaf). In contrast to  $dEcho$ , the  $HOME$  parameter is quite sensitive to canopy change. The change in  $HOME$  is largest (7.22 m) for broadleaf and decreased for mixed (4.07 m) and needleleaf (3.15 m) forests.

The ground/canopy ratio ( $rGround$ ) also changes most for broadleaf forest. For the summer waveform, the intensity of the ground return is noticeably less than the canopy return intensity. For the winter waveform, the ground return is much stronger than the canopy return. As a result, the difference of ground/canopy ratio between summer and winter is large (ratio differences of up to 66%). The difference of ground/canopy ratio becomes smaller for the other forest types, ( $rGround$  should be large in winter and small in summer), due to smaller changes in canopy morphology in mixed and needleleaf forests.

### (c) Classification of forest types based on waveform-derived forest parameters

The three forest parameters from section 5.3.4 used are  $\Delta I$ ,  $\Delta HOME$  and  $\Delta rGround$ .  $\Delta dEcho$  is not considered to be a classification variable because the differences between February and September are small.

Table 5.4: Confusion matrix and classification results

Classification Data	Reference Data (CLC2000)				Classification results			
	<i>Broad</i>	<i>Needle</i>	<i>Mixed</i>	<i>Total</i>	Percentage based accuracy		Waveform based accuracy	
					Producer	User	Producer	User
<i>Broad</i>	254	17	45	316	80.38	72.78	254/316	254/349
<i>Needle</i>	52	8	11	71	11.27	29.63	8/71	8/27
<i>Mixed</i>	43	2	10	55	18.18	15.15	10/55	10/66
Total	349	27	66	442				

Overall classification accuracy = 61.54% and  $\kappa = 0.57$

Based on the forest parameters from 442 pairs of summer-winter waveforms, we choose three critical threshold values:  $C_I$ ,  $C_{HOME}$  and  $C_{rGround}$ . These are all defined as the 0.20 quantile of the parameter values found. This procedure give a value of  $2.02 \times 10^{-6}$  for  $C_I$ , a value of  $-0.75$  m for  $C_{HOME}$  and a value of  $-1.53$  for  $C_{rGround}$ .

If an intensity difference  $\Delta I$  is smaller than  $C_I$ , it is classified as small. The same is applied to  $\Delta HOME$  and  $\Delta rGround$ . If the values of  $\Delta HOME$  and  $\Delta rGround$  or the values of  $\Delta HOME$  and  $\Delta I$  are both large, the forest type is classified as broadleaf. If the values of  $\Delta HOME$  and  $\Delta rGround$  or the values of  $\Delta HOME$  and  $\Delta I$  are both small, the forest type is classified as needleleaf. In all other cases, the forest type is classified as mixed forest.

The accuracy assessment and the classification result are shown in Table 5.4. The  $\kappa$  value of the classification result is 0.57. This kappa value is not high enough to ensure a reliable differentiation between forest types. Known limitations to this classification method will be explained in section 5.3.5. The use of footprint pairs with larger overlap between single footprints and improved waveform analysis methods, for example the inclusion of slope effects, are expected to improve current classification results. Moreover, additional waveform parameters, introduced by [Duong et al. \(2006b\)](#) and [Ranson et al. \(2004\)](#), can be added to the classification method to potentially improve classification results based on GLAS measurements alone.

### 5.3.5 Recognised problems

Known limitations of these two GLAS data sets may negatively influence the results. Distances between footprint centres within waveform pairs are between 40m

to 90m (see Figure 5.7). As such, the returned winter and summer waveforms are only directly comparable over homogeneous areas, but many forest waveforms are located in hilly areas where slopes of up to  $25^\circ$  occur (Figure 5.13). The locations of the forest waveforms are displayed along a topographic profile (grey) of the (repeated) ICESat track. Broad-leaf locations (according to CLC2000) are in green, needleleaf in blue and mixed in red (top panel). The slopes can, by first approximation, be derived from the GLA14 global land surface altimetry product, since both a footprint centre distance and a footprint centre height difference is given for each waveform pair. According to a procedure of [Lefsky et al. \(2005\)](#), waveform extent is a function of maximum canopy height and topographic relief. By overlaying the waveform footprint over an SRTM digital elevation model, the range of ground surface elevation within the footprint is determined. This range is the so-called terrain index. The maximum canopy height is then identified by estimating two coefficients based on both the terrain index and the waveform extent. It is clear that slopes directly influence the forest parameters obtained from single waveforms, while local changes in slopes within the area covered by the corresponding waveforms, will make the summer and winter returns less comparable.

Many waveform pairs are removed (see Figure 5.13) from the original 440 waveform pairs, because of unrealistic (changes in) forest parameters: 124 waveform pairs with only one mode and 162 waveform pairs with a reported tree height change exceeding the allowed  $-5\text{ m} < \Delta dEcho < +5\text{ m}$  range. The one-mode waveform pairs either occur in one season (113 pairs with a waveform extent/width and its standard deviation of  $11\text{ m} \pm 11\text{ m}$  for the winter, and  $23\text{ m} \pm 12\text{ m}$  for the summer) or in both (11 pairs,  $8\text{ m} \pm 6\text{ m}$  for the winter and  $4\text{ m} \pm 4\text{ m}$  for the summer). One reason for the occurrence of one-mode waveforms can be found in errors in the CLC2000 classification, or in changes in the landscape after the CLC2000 classification took place. Large changes in reported tree height may be caused by topographic height differences in the area covered by partially overlapping waveforms.

### 5.3.6 Conclusions and further research

In this case study we have introduced methods for the analysis of large footprint full waveform signals for describing forest structures. Notably, normalisation of waveform return energy and determining vertical shifts between repeated waveforms make waveforms of ICESat repeated tracks comparable. Based on CORINE land cover data, summer-winter waveform pairs were divided into three forest types. It proved that it is possible to derive several parameters describing (changes in) forest trees, starting with a Gaussian decomposition of the waveform into single modes. On average, the (change) values obtained for these parameters matched expectation.

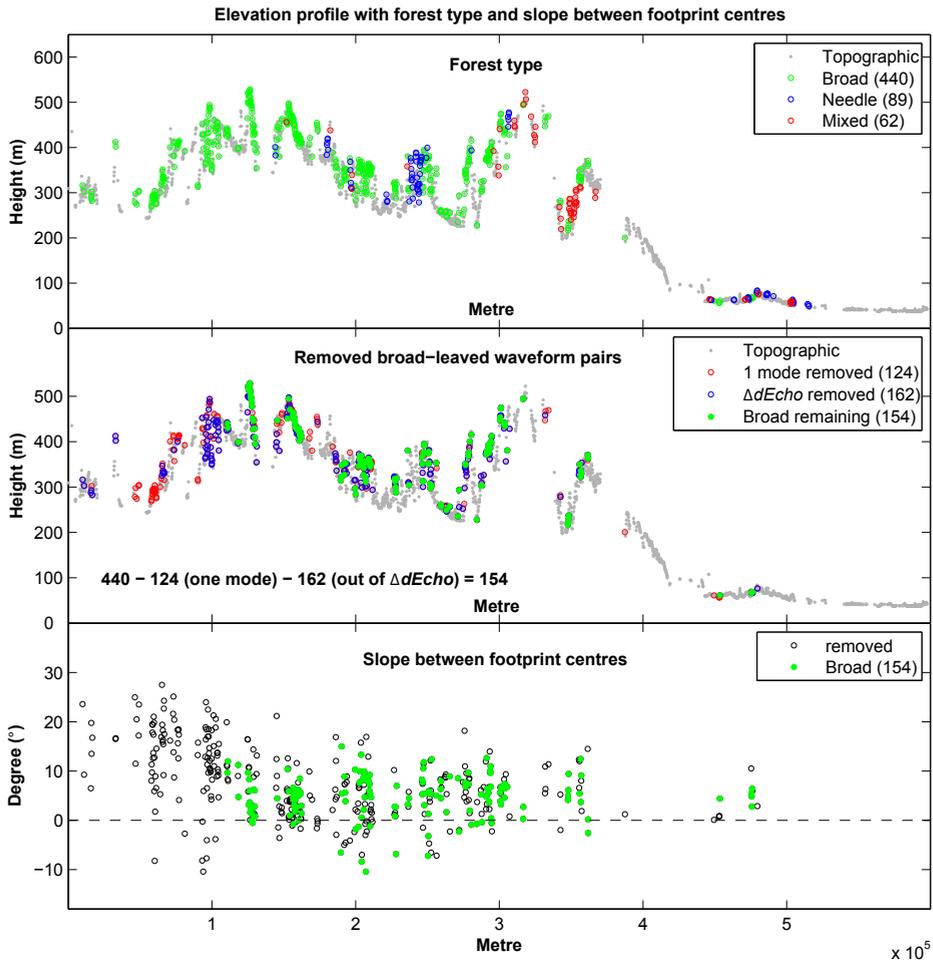


Figure 5.13: Along track topographic profile (grey) with the three forest types given by the CLC2000 (top). Remaining and removed broadleaf waveform pairs (2<sup>nd</sup> panel). The third panel shows the slopes between the paired footprint centres.

We first identified some parameters that can be extracted from waveform data like maximum canopy height, ( $dEcho$ ), height of median energy, ( $HOME$ ), ground/canopy return ratio, ( $rGround$ ), and canopy intensity difference, ( $\Delta I$ ). As part of the next step we analysed changes in these parameter values. The method showed that we were able to detect and extract the change in forest, or more specifically canopy structure, between different seasons, here winter (leaf-off) and summer (leaf-on). The results indicate that large-footprint, full waveform laser altimetry data may be able to infer changes in, for instance, leaf area index, between different epochs.

Another approach that was employed in this study included the use of tree parameters derived from waveform pairs to classify footprint areas directly into forest type classes. Preliminary results, with a kappa value of 0.57, provided a baseline against which improvements in the data and methodology can be gauged.

Future works should include a method for correcting slope-induced changes in the waveform results, both for individual waveforms and for waveform pairs. In addition, we expect further improvement by including a neighbourhood analysis, e.g. considering correlation between waveforms and their changes. An individual quality descriptor of the tree parameter values could be obtained from quantifying and propagating errors encountered during the waveform processing. Moreover, tree parameter values should be validated against either data from field measurements or data from other sensors. Finally, the parameter definitions should be refined and validated in order to improve agreement with biophysical characteristics of the forest.

The data set we considered was far from ideal. The ground surface was not flat for most waveform pairs, while the footprint locations from different epochs only partly coincided. In addition, the error sources used for data acquisition differed between the two epochs. Still, we were nevertheless able to find considerable and significant differences between the summer and winter data, with difference values matching the expectations.

## 5.4 Case study 2: ICESat laser data for land cover classification

### 5.4.1 Background

Creating land cover databases is one of the most important targets in remote sensing. Land cover assessment and monitoring of its dynamics are essential requirements for sustainable management of natural resources and for environmental protection. They provide the foundation for environmental, food and humanitarian programmes of international and national. Accurate, detailed, and reliable land cover information remains in demand (GLCN-LCTC, 2006).

Remotely sensed images are the main source of data used for land cover classification. The images are acquired from different platforms as satellite, airplane or ground, from different systems both passive and active, and from various sensors with different resolutions. In this study, we consider the possibilities that the ICESat mission has as a new source for land cover data. For this purpose we develop a method of classification. The classified results of ICESat data are validated by comparison to the CORINE land cover database.

ICESat data can therefore only be used for classifying profiles, as compared to regions mapped by imaging systems. The full waveform data gives new possibilities to extract more information about land cover of the earth surface.

In this study, the full waveform analysis is investigated in the context of land cover classification. The study suggests a new potential way to extract land cover classes from the full waveform data. Firstly, the waveform is assumed to be a sum of Gaussian components. The waveform is then decomposed into different Gaussian components. Next, the waveform parameters derived from the decomposition are applied in the classification step to identify land cover type. Moreover, the quality of the fitting step is also analyzed. The footprints are classified independently of each other.

In the next section, it will introduce our study area, the ICESat/GLAS data set and a reference land cover map. Then it will introduce new methodology for analyzing full waveform data with the purpose of land cover classification. In the results section it will give the total accuracy of the classification based on computing a confusion matrix between the ICESat data and the reference data. Finally, conclusions will be stated at the end.

## 5.4.2 Study area and data input

The area of study is mostly the Netherlands, bounded approximately by  $3^{\circ}E$  to  $7^{\circ}E$  longitude and  $50^{\circ}N$  to  $54^{\circ}N$  latitude which contains a large variety of land cover. There are 15 products of ICESat data, however, only the products GLA14 and GLA01 in release 24 are used for this research. The GLA14 contains precise geolocation of footprint centers with height information. The GLA01 contains the full captured waveform. Release 24 means that these products are processed using currently the most recent version of the NASA processing procedure. A GLA01 waveform is linked to a GLA14 location by index and shot number. The index and shot number are computed by relating the shooting time of an individual pulse to the starting time of the ICESat operation and the shooting frequency. The waveform locations and the footprint sizes are displayed together with a CORINE land cover map in Figure 5.14.

The CORINE Land Cover 2000 database (CLC2000) was described in section 4.3.2 and then reclassified into four land cover types: *high vegetation*, *urban areas*, *bare land/low vegetation* and *water*.

The full waveform data set was acquired in the period from 2003-09-25 to 2003-11-18. There are 3277 waveforms in the study area. The footprints of these waveforms are elliptical, its power distribution has a central maximum, while energy decreases to the boundary. The size of the ellipse is  $95\text{ m} \times 52\text{ m}$  on average (Harding and Carabajal, 2005). The footprint separation is 175 m along track. The horizontal

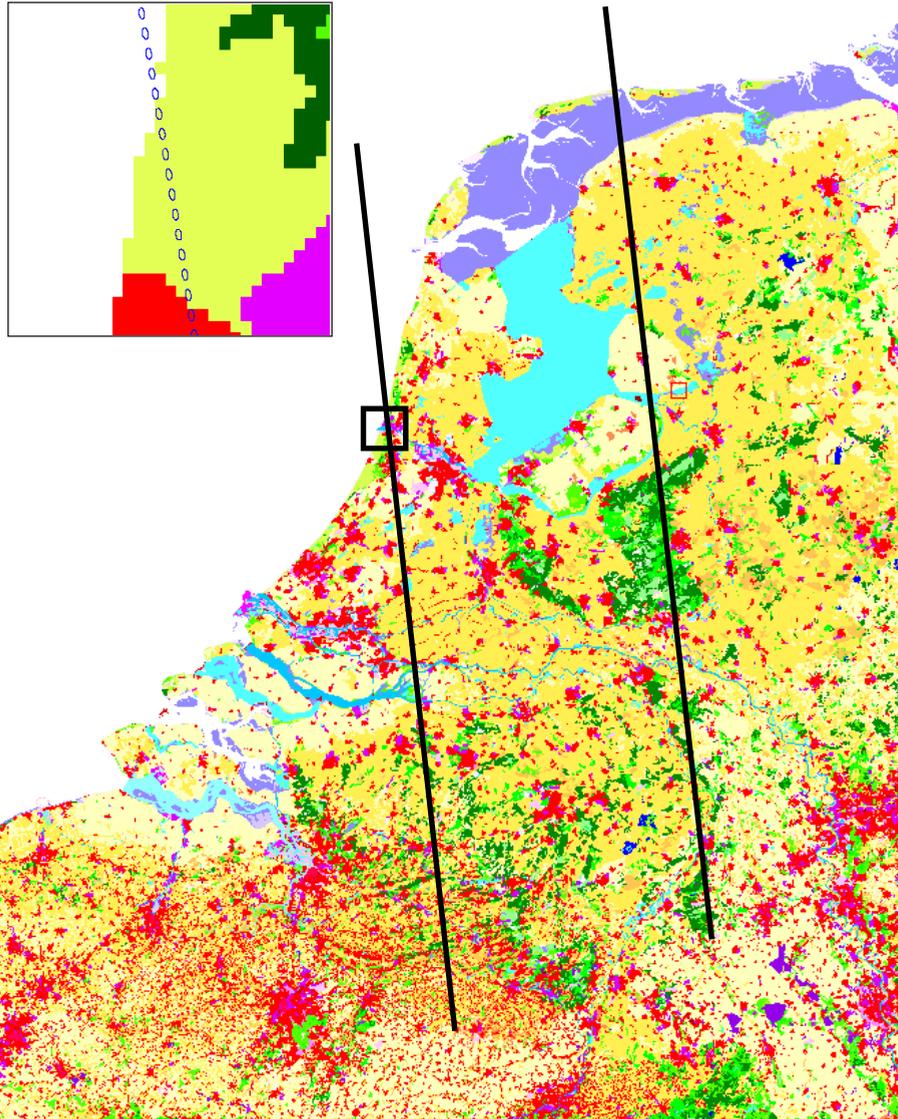


Figure 5.14: Study area: ICESat ground tracks (blue) and CLC2000 land cover map (100m resolution). A close up of the black rectangle is shown in top left corner box, with elliptical footprints of 95 m x 52 m. The track passes from the North Sea (in white) onto land (in color).

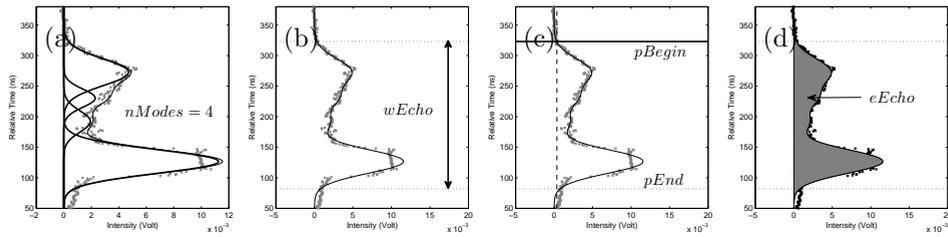


Figure 5.15: Waveform parameters derived from a typical waveform: (a) number of modes, (b) Total return energy, (c) waveform begin and (d) waveform extent. The crossed-line represents the normalized waveform and the solid thin line represents the fitted waveform.

geolocation accuracy of the ICESat footprints is 3.7 m. The waveform is digitized in 544 nanoseconds over land area and 200 nanoseconds over sea or ocean. That means that the waveform can acquire a height of up to 81.6 m over land and 30 m over sea with the vertical resolution of 15 cm. Both data from ascending and descending orbits are used in this study.

### 5.4.3 Methodology

First the initial processing steps for individual waveforms as described in chapter 3 are applied. The fitting step results in a number of Gaussian components with Gaussian parameters. However, for the classification purpose, other parameters are also helpful: waveform extent ( $wEcho$ ) and total energy ( $eEcho$ ).

#### Waveform parameters for classification

Due to the existing noise in the waveform, the actual width of the waveform or the actual waveform needs to be identified by finding the locations of the actual begin and end of the waveform. The threshold value for each waveform is defined by taking the maximum value of the (noise) intensity in the first 150 bins. This threshold value is applied to truncate the waveform since it determines the locations of the actual begin and end of the waveform. Therefore, the begin of the waveform is the intersection of the horizontal threshold value line and the first rising edge of the waveform (top). The end point of the waveform is the last intersection point (bottom). The width of the waveform is the distance between the begin and the end location of the waveform. A visualization of the width and the actual waveform is shown in Figure 5.15.

Four waveform parameters are used in the classification step:

- Number of modes,  $nModes$ : total number of Gaussians as found by the fitting algorithm. Over simple surfaces (bare land and water), the number of modes theoretically equals 1 or maximal 2. However, over complex surfaces (forest and urban areas), the number of modes can be equal or larger than 2. In Figure 5.15(a), the number of modes equals 4.
- Total energy,  $eEcho$ : Total energy received back by the GLAS instrument. This is calculated by taking the area under the actual waveform. The result is converted in energy units (attojoules =  $10^{-18}$  joules), see Figure 5.15(b).
- Waveform begin,  $pBegin$ : Location where a waveform firstly passes a signal threshold, see Figure 5.15(c). The waveform begin can be the highest point within the ICESat footprint. Over forest and urban areas, the waveform begin is higher compared to the waveform begin over water and bare land.
- Waveform extent,  $wEcho$ : The distance between the begin and the end of the actual waveform, see Figure 5.15(d). Over forest, urban and mountainous areas, the waveform extent is larger than compared to areas with little vertical structure.

### Waveform classification scheme

The four waveform parameters described in the above section are used in the classification step. Waveforms are classified into 4 different classes: *high vegetation*, *urban*, *bare land/low vegetation* and *water*. The flowchart for this classification is shown in Figure 5.16.

In general, water absorbs more energy of the emitted pulse than other land cover types and caused by specular reflection directs the radiation away from the observer. Urban areas may reflect more energy to the GLAS sensor due to, for example, white roofs of buildings. Therefore the energy of the recorded waveform over water is mostly lower than over bare land, urban or high vegetation. Therefore the total energy is used to extract water waveforms.

Bare land has a flat open surface, so it mostly gives uni-mode waveforms. Moreover, bare land can also cause the recorded waveform to have a wider width due to roughness or slope of the surface. Low vegetation can result in waveforms having multi-modes. However the extra modes have low amplitude and will be few. In this particular step, the bare land waveforms are classified by having only one mode. Moreover, by regarding the effect of low vegetation, bare land waveforms are selected in the next step by having a narrower waveform width.

Urban and high vegetation are known as complex areas with different height levels. Therefore, the land cover type usually results in multi-mode and wide-width waveforms. This is caused by artificial objects and high trees. However, waveforms

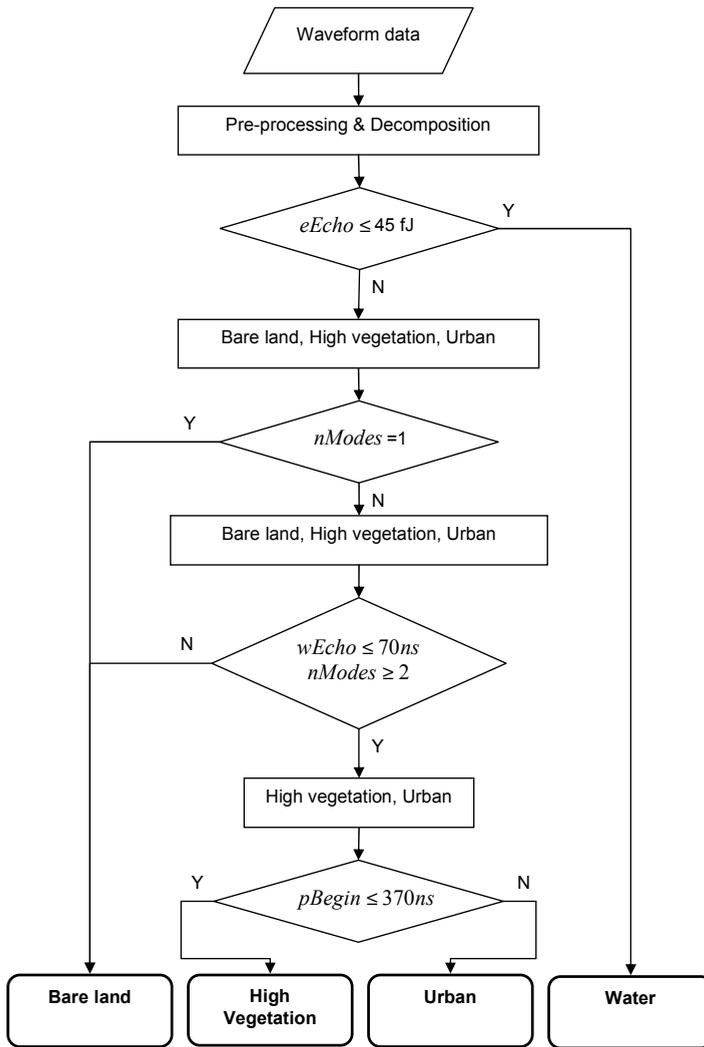


Figure 5.16: Classification scheme.

over high vegetation often have a wider first mode width due to the scattering from the tree crowns. The urban waveform has a narrower width for all modes due to the reflectance from the sharp shape of artificial objects like buildings, etc. In other words, the wider first mode results in an earlier rising of the first edge of the actual waveform. In this case, the location of the waveform start is used to separate urban from high vegetation waveforms. Four representative waveforms

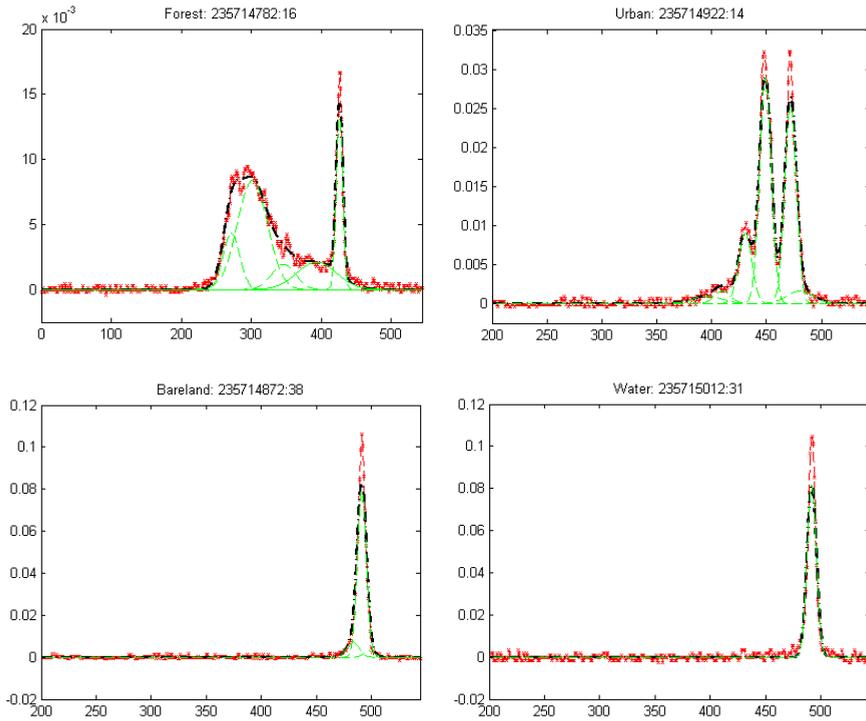


Figure 5.17: Top Left: High vegetation. Top Right: Urban. Bottom Left: Bare land or low vegetation. Bottom Right: Water.

for four different land cover classes are shown in Figure 5.17.

#### 5.4.4 Classification results and discussion

From the point of view of image processing, a waveform data with ellipsoidal footprint size of  $95\text{ m} \times 52\text{ m}$  can be approximated by a polygon. The polygon of the classified waveform represents the polygon of pixels in the classified image. The CLC2000 reference data with 100m resolution, is used for validation and represents a known land cover type. Therefore, the classified polygon can contain only a few pixels of the CLC2000. In order to improve the reliability of the classification results, the CLC2000 is splitted into 25 m pixels with the same pixel values as the full 100 m pixel. In this way, we obtain a larger number of pixels within the footprint polygon, which helps in the accuracy assessment. For example, suppose that the footprint polygon overlaps two reference pixels of 100

m resolution. Assume that the first reference pixel represents land cover type A and the second land cover type B. Moreover, assume that the first reference pixel has a large common area with the ICESat footprint area and the second a small common area. When using reference data of 100 m resolution, only two reference pixels are selected and it is therefore difficult to decide which land cover type, A or B, should be assigned to the ICESat footprint. However, if we use 25 m resolution reference data, it will have several 25 m pixels in the first reference pixel and less in the second one. In this case, land cover type A is assigned to the ICESat footprint.

One of the common means of expressing classification accuracy is the preparation of a classification *confusion matrix* (Lillesand et al., 2004) described in section 5.3.3. In this research, the commission error for water consist of 1394 water pixels that are falsely classified as not water. 644 are classified as bare land, 706 as high vegetation and 44 as urban. An omission error occurs when a pixel is falsely omitted from a class by the classifier method. Thus, the classification based on full waveform laser altimetry analysis fails to recognize and correctly identify 589 from 3162 as not water.

The CLC2000 data described in section 4.3.2 consists of 5 classes at the first level such as *artificial surfaces*, *agricultural areas*, *forest and semi natural areas*, *wetlands*, and *water bodies*. Therefore, the five classes of the CLC2000 data are then reclassified for the purpose of the ICESat validation. Here, the *artificial surfaces* is classified as urban class, the *agricultural areas* and *wetlands* as bare land/low vegetation class, the *forest and semi natural areas* as high vegetation class, and the *water bodies* as water class.

Table 5.5: Confusion matrix

Classification Data	Reference Data				
	Water	Bare land	High Vegetation	Urban	Total
Water	2573	398	185	6	3162
Bare land	644	1175	457	21	2297
High Vegetation	706	1873	7410	216	10205
Urban	44	81	507	4438	5070
Total	3967	3527	8559	4681	15596

Table 5.5 shows the confusion matrix between the ICESat and the CLC2000 classification. The columns contain reference data with known land cover type, and the rows are the classified waveform data. Consider for example the water column containing a total number of 3967 water pixels. 2573 are classified correctly as water, 644 are classified incorrectly as bare land, 706 are classified incorrectly as high vegetation, and 44 are classified incorrectly as urban.

Table 5.6 shows the classification result. The producer's accuracy of the water is

81.37% and the user' accuracy for the water is 64.86%. Moreover, from Table 5.5, we obtain the overall accuracy of the classification of about 75.22%.

Table 5.6: Classification results

Class	Producer		User	
	Accuracy (%)	Accuracy (%)	Accuracy (pixels)	Accuracy (pixels)
Water	81.37	64.86	2573/3967	2573/3162
Bare land	51.15	33.31	1175/3527	1175/2297
High Vegetation	72.61	86.58	7410/8559	7410/10205
Urban	87.53	94.81	4438/4681	4438/5070

In this research, the classification result  $\kappa$  equals 0.73. It can be thought of as an indication that an observed classification is 73 per cent better than one resulting from chance.

The  $\kappa$  of 0.73 is a quite promising result for a classification method based on waveform analysis. Moreover, we can see more interesting points in Table 5.6. The urban and high vegetation results are quite high in both producer's accuracy and user's accuracy (87.53% and 94.81% for urban, and 72.61% and 86.58% for high vegetation). These results could lead to the conclusion that our method is adequate for the purpose of mapping urban areas and high vegetation.

The classified ICESat waveform data are displayed together with the Landsat-7 image shown in Figure 5.18 (right top corner). Two ICESat tracks are displayed with corresponding tracks of CLC2000 reference data. These tracks are colored with respect to the land cover type. The high vegetation class is colored in green, the urban class colored in red, bare land/low vegetation in white and water in blue. In each pair of tracks, ICESat data are on the right and CLC2000 on the left. In comparison between the classified ICESat tracks and the CLC2000 reference tracks, it is clear that the classified tracks are quite matched to the reference tracks. However, there are three typical areas caused misclassification results. These misclassification areas are displayed in yellow box with respect to three different land cover types according to the CLC2000 data: bare land/low vegetation (case 1), water (case 2) and urban areas (case 3).

Case 1 considers an area located at the North of The Netherlands containing water, sediments and prove to tidal change. According to the Landsat-7 image, it should be classified as water, however in CLC2000 it is reported as bare land. In contrast, the waveform at this location is classified as urban because the number of modes is larger than 3. A possible explanation is as follows: during acquisition of the ICESat data, the water tide is low such that most water has disappeared and reflections are mostly caused by the topography of the underlying mud flats. The classification result could be improved if the tide data were incorporated. The waveform and its parameters are displayed with the CLC2000.

For case 2, water was classified as bare land/low vegetation at a water way between

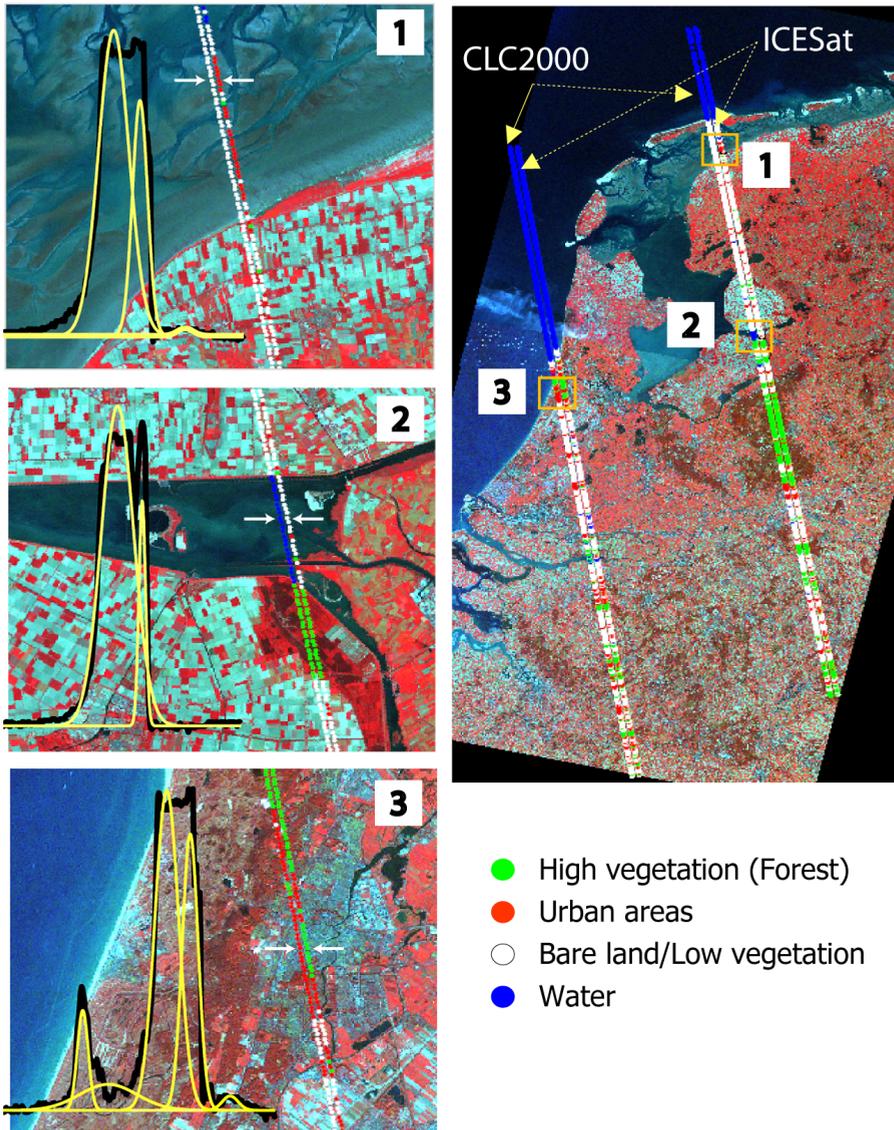


Figure 5.18: Right: land cover classification results displayed with two ICESat ground tracks and Landsat-7 image. Left: typical ICESat waveforms (black) and its Gaussian components (yellow) are displayed over the Landsat-7 image considering three case studies. The white arrows indicate the waveform location.

the Noordoost polder and Flevoland. This area could also be affected by a change in water level. However, the number of waveform modes (maximal 2) is less than in case 1. Moreover, this area is located inside a land region in contrast to the area of case 1. Occurrence of some floating vegetation could be an explanation for the larger observed return energy in combination with a small number of modes.

For case 3, an urban class (at Haarlem city) was classified as high vegetation. This area has a combination of high trees and buildings. An explanation for the misclassification may be the trees in this area are sometimes taller than the buildings, therefore this misclassification could be improved by more advance methods that are able to distinguish between buildings and high trees.

### 5.4.5 Conclusion and further research

In this study, we have developed a new method for land cover classification based on full waveform laser altimetry analysis. By applying the waveform analysis for land cover classification an accuracy of classification was found of 73% compared to the CORINE classification CLC2000. The group of high vegetation and urban waveforms and the group of bareland/low vegetation and water are well separated. The classification step is done by automatic processing of the waveforms. This research suggests a new and promising way to determine land cover information. Moreover, the result derived from the waveform analysis is also useful for comparison, validation or updating of classification data obtained by other methods.

However, the method can still be improved. There is some confusion for classifying individual classes in each group. In the group of high vegetation and urban areas, the classification method can be improved to get better discrimination of high trees and buildings. The solution is to consider the width of every Gaussian models as compared to the width of all Gaussian models. This is a potential way because the waveform reflected from buildings and other artificial objects returns a narrower signal. For high vegetation, the width is expected to be quite wider.

Furthermore, the estimated waveform parameters are not yet good enough for discrimination between the bare land/low vegetation and water. In theory, the water waveform has one mode due to the flat water surface while the bare land waveform should have more than one mode with a wider width waveform caused by slope, roughness or low vegetation. However, in practice this does not seem to hold. A further step for this discrimination is to consider other parameters such as the ratio between first mode and last mode. Moreover, water waveforms over lakes or rivers somehow show multi-modes where in theory only one mode should be present. This can be explained by the presence of waves and artificial objects like boats or ships.

High Altitude Long Endurance (HALE) UAVs will maybe in future also carry

lasers. With scanning mode or densely arranged profiles also laser ranging from space can be expected to provide area coverage type data in that case (Biesemans et al., 2005). The results of this case study show the feasibility of classifying land cover based on laser satellite altimetry data alone and suggests thus improved results for the combined classification of active range and passive optical imagery.

Disadvantages of ICESat data are that only ground tracks are mapped and thus ICESat data are not adequate for mapping land cover types over entire regions like The Netherlands. However, in general, full waveform analysis contributes a new promising way for land cover classification applications in near future.

## 5.5 Conclusion

ICESat elevation and full waveform data have been applied in various applications. Most used data are elevation products. Most studies using ICESat data are focusing on polar regions like Greenland and Antarctica. Repeatedly acquired near coincident ICESat elevations are used to derive elevation changes that can be directly applied in determining volume changes and snow accumulation over ice sheets and therefore indirectly contribute to the ongoing discussion on mass balance change of Greenland and Antarctica.

Waveform data are also used in recent research applications. They were used to reveal previous features on ice shelves like crevasses and rifts. They were however mainly used in forest applications for extracting biophysical information like forest structure and aboveground biomass.

A new contribution described in this chapter is studying repeated ICESat waveform pairs to assess forest change. By analyzing corresponding waveforms acquired in different campaigns at nearly coinciding footprints, canopy changes caused by seasonal influences were detected. In general, the new aspect of this chapter is to analyze waveform parameters in terms of a pair of full waveforms that are acquired at approximately the same location. Comparison of such pairs allows for example the detecting of seasonal influences on forest type. Moreover, it may become possible to further detect and estimate forest changes, e.g. forest growth and forest deforestation.

In addition, it is shown that ICESat waveforms can be used to build a feature vector consisting of suited waveform parameters that can be consecutively applied to classify the land cover class of the footprint location. The full waveform parameters derived from the decomposition step together with total return energy were able to discriminate between high vegetation, urban areas, bare land/low vegetation and water. This result showed the feasibility of land cover classification with spaceborne lasers. As the ICESat satellite has a near polar orbit, coverage is global, but still a main disadvantage of ICESat data is that only tracks are mapped. With the

development/launch of new systems, an area-wise coverage may become possible in future. In that case, automatic classification of large footprint full coverage full waveform laser data may lead to land cover classification results of the same high quality as can be obtained from optical data.



## Conclusion and future work

In this thesis the processing and applications of large footprint full waveform LIDAR data have been studied. Mainly ICESat elevation (GLA14) and full waveform data (GLA01) products were investigated. For obtaining parameters from the full waveform data the Gaussian decomposition method was used. Moreover, techniques for waveform deconvolution and waveform simulation were implemented. The accuracy of full waveform parameters was validated using the CORINE 2000 land cover database over Europe and accurate airborne laser scanning data over The Netherlands were applied for additional validation. From the results of two validation cases and two new applications based on ICESat data, the following conclusions and recommendations are drawn.

### 6.1 Data handling

Processing of ICESat products was well managed. ICESat data (GLA14 and GLA01) were primarily converted from their original binary format to ASCII format and successfully visualized. Full waveform data were georeferenced to several ground-based coordinate systems. Moreover, the datum transformation and coordinate conversion between ICESat and other coordinate systems, e.g. the Dutch coordinate system RDNAP and the European reference system, were performed. In order to avoid mistakes in using ICESat data it is recommended that users always stay up-to-date with the changes in the ICESat products.

## 6.2 Full waveform processing techniques

The principal processing technique is based on Gaussian decomposition . This method was developed to process both large footprint data (ICESat) and small footprint full waveform data (Riegl). The result of this method is a number of Gaussian components together with their amplitude, position and width. In addition, the methods of waveform deconvolution and waveform simulation were investigated.

However, several issues still remain. Sometimes the Gaussian fitting method cannot fit a Gaussian to a part of the waveform, for example, a canopy peak or a small ground peak. This could be explained by the fact that those peaks are not well described by a Gaussian, and therefore, the residuals resulting from the fitting are locally much larger than the background noise level. Besides a Gaussian function, new base functions, e.g. generalized Gaussian (Mallet et al., 2009) might possibly lead to a better fitting result. The benefits of those functions, however, should be carefully studied and proved in future.

Another issue in this processing method is the need for a thorough comparison between Gaussian decomposition and waveform deconvolution. The comparison of these methods, as two alternative ways to extract waveform parameters, is expected to give insight into the strong points of both methods. In other words, a combination of these two methods may result in a more precise description of feature height and object structure.

Slope and roughness effects on ICESat data should also be further studied. The widening of the waveforms caused by slope or roughness of the ground surface is still not well understood. Moreover, the widening effect also causes a bias in the estimation of the height of features such as buildings and forest. This thesis mainly compared ICESat data to high resolution laser data over The Netherlands, where changes in topography are very small. A similar analysis over a more hilly/mountaineous area, where high quality airborne data are available, may shed light on such effects and lead to improved estimates of feature heights. Modern laser ranging satellite systems, like the LOLA system onboard of the recently launched NASA Lunar Reconnaissance Orbiter, employ a beam splitter that allows to scan the surface in a two-dimensional spot pattern. As a consequence, local slope and roughness information can be directly obtained from the laser observations.

Alternatively, simulating a large footprint full waveform at different (simulated) footprint sizes from high resolution airborne laser data may help to understand the influence of slope and roughness on the waveform. With a smaller footprint, slope and roughness effects on the waveform may be largely discarded and isolated, which would enable the extraction of more accurate feature information.

## 6.3 Cloud detection and saturation

Two known issues in the processing of waveforms from ICESat and similar satellite missions should be considered in future: cloud detection and elevations derived from saturated waveforms.

In the thesis, as a support for the data handling and the full waveform processing method, filtering constraints were determined such that suspicious full waveform data, affected by for example cloud cover or saturation, can be eliminated.

The identification and elimination of ICESat shots affected by clouds is still problematic however. In the current processing method, the gain value ( $<100$ ) is mostly used to remove waveforms affected by clouds. Cloud filtering flags, such as the cloud multiple scattering warning flag (`i_cld1_mswf`) and the medium resolution cloud availability flag (`i_MRC_af`), should not be used at the moment because they tend to remove a lot of actual good waveform data when compared to reliable external meteorological cloud data. These cloud flags should be studied and improved in future releases. More reliable cloud flags could help to discard error sources, avoid the elimination of valid ICESat data and improve the ICESat elevation products.

ICESat elevation products derived from saturated waveforms were corrected for polar areas (GLA12 product), but not for land regions (GLA14 product). Over land regions, the saturation correction method with respect to the diversity of land cover type for the ICESat elevation product should be studied in future.

## 6.4 Validation

The validation of ICESat data allowed to draw conclusions on the quality of ICESat derived terrain elevations and feature height estimates over forest, urban areas and low vegetation regions. Moreover, a new method for identifying and eliminating errors in the ICESat products and the associated processing flow has been introduced, considering waveform pairs acquired at the same time over the same location.

From the first validation it is concluded that the vertical accuracy of ICESat products varies with the complexity of the earth surface. On average, the differences in the terrain height over The Netherlands derived from ICESat and accurate airborne laser scanning data are smallest over bare land and urban areas with a mean difference below 25 cm and a standard deviation of about 23 cm. Over forests, the variation in terrain height differences is in general larger. The terrain height difference is on average only 10 cm but with a relatively large standard deviation of about 60 cm.

Moreover, as part of this validation, a waveform based comparison of feature height (of forest, buildings and low vegetation) between ICESat and airborne laser data resulted in average feature height differences of 1.89 m over forest, 1.48 m over urban areas and 29 cm over low vegetation.

The feature height difference is quite large over forested and urban areas. There are two main reasons for this large difference: (i) A lack of intensity values in airborne laser data that prohibits the construction of a good simulated waveform, and (ii) A large temporal difference (about 3 years) between airborne laser data and ICESat data which makes actual surface changes quite likely. In order to discard anomalous effects in this comparison, the temporal difference should be small. Moreover, the waveform simulation method should incorporate intensities when available. New airborne laser data (AHN2) are systematically acquired over The Netherlands since 2007. Using ANH2 data in combination with recent ICESat campaigns could largely eliminate temporal differences and is therefore expected to result in more reliable feature height comparisons.

The second validation compared repeated waveforms over the whole Europe. It considered differences in the values of waveform parameters (e.g. median energy, waveform extent, relative return energy and intensity distribution) derived from ICESat waveforms obtained at approximately the same time and location. For three particular case studies, it was shown that the differences in the values of waveform parameters could be explained by the errors in the reported footprint size and the location, and by small differences in the acquisition time.

Over the European continent, it turned out that the number of repeated waveforms with almost perfectly overlapping footprints was rather limited. This is because (i) at lower latitudes not so many ICESat tracks are available, and (ii) only a maximum of three data campaigns is available per year. As a consequence, the number of pairs found was not sufficient to perform an elaborated study on the ICESat error budget. The derivation of a larger data set of overlapping waveforms over the entire Earth may be possible and helpful to investigate the ICESat error budget.

## 6.5 New applications

Two new applications of ICESat data products were identified.

The first application is land cover classification, which is traditionally carried out based on image data. In this thesis for the first time waveform parameters (e.g. waveform extent, total energy, begin waveform and number of modes) were applied to distinguish between different land cover types like forest, urban areas, bare land/low vegetation and water. It resulted in a very promising overall accuracy of classification of about 75%.

In addition to the land cover classification over The Netherlands, such a classification is also applied to polar regions. Molijn (2009) applied waveform parameters (waveform extent, return energy, etc.) to classify different land cover types over Greenland and parts of Antarctica into four classes: snow, rock, ice and water. Although the overall accuracy of such a classification method is promising, more advanced methods that could fully exploit the potential of waveform parameters are demanded.

Full waveform analysis can even be applied to analyze high resolution airborne laser scanning data. In places where full waveform data are not available full waveform analysis can still be applied. Airborne laser scanning data can be used to simulate large footprint full waveforms and can then be studied in a full waveform analysis sense. Antonarakis et al. (2008) used the histogram of elevations derived from airborne laser data to define parameters like skewness and kurtosis for the classification of forest species in France.

Moreover, the simulated waveforms resulting from different footprint sizes may help to understand which footprint size is optimal with respect to a certain land cover type, for example, forest and urban areas. Over homogeneous forests, the simulated waveforms from large footprint (70–100 m) and small footprint (2–5 m) data are to some extent (small and large) the same. Using a larger footprint (e.g. 70 m) will help to reduce the processing time. Moreover, such large footprint full waveform simulation can be used to determine tree species. Over more urbanized areas, a large variety of objects with varying characteristics is present. Most man made objects can be successfully identified and parameterized by a segmentation of high resolution airborne laser scanner data. To localize and classify remaining, mostly natural objects as grass, shrubs or trees, parameters derived from simulated waveforms with footprints in the order of meters could be exploited using methods described in this thesis.

The second application demonstrated the use of repeated large footprint full waveform as a new technique to detect forest canopy changes between winter and summer. For this purpose, waveform pairs acquired over the same footprint location were compared. In each pair, one waveform was recorded in winter and one in summer. As a result of the comparison, it was shown that forest change occurs the most for broadleaf (a 148% change, winter to summer) and the least for conifers (a 36% change). This is to be expected as conifers do not lose their foliage in winter.

The analysis of full waveform pairs may be applied to monitor forest biomass changes. Canopy changes and forest height growth are extracted by analyzing waveform parameters. These detected changes can be a good estimator for forest biomass data and could therefore be applied to assess the impact of climate change.

## 6.6 Upcoming missions

Planned missions, notably DesDyn-I and ICESat-II, will open interesting research applications. ICESat-II is the follow-up mission of ICESat and is expected to continue its measurements with the same scientific objectives. This system will provide a higher chance of acquiring a large database of repeated/overlapping footprint data that may be applied to understand error sources as well as to improve the quality of both ICESat-I and -II data products. Moreover, the DESDyn-I mission, which combines a radar sensor with a multiple beam lidar system may offer (i) a very good data set for validation of ICESat data products over areas where no reliable field data are available, (ii) a chance to produce extensive and accurate elevation change products by fusing ICESat and DESDyn-I data of the same region.

## Ordering ICESat data

ICESat data are distributed for free over the web site of the NSIDC at a link of <http://nsidc.org/data/icesat/index.html>. An easiest way to order data is to go to the link [http://nsidc.org/forms/glas\\_subset\\_form.html](http://nsidc.org/forms/glas_subset_form.html). Please specify the GLA01 product for full waveform data. By this product, the transmitted pulse and return full waveform are both provided.

Table A.1: Four methods to order ICESat data.

Data Source Link	Data Source Information
Data Pool	Quickly access data granules and corresponding browse files
Search 'N Order Web Interface	Access all ICESat/GLAS data and corresponding browse files
Warehouse Inventory Search Tool	Access all ICESat/GLAS data
ICESat/GLAS Data Subsetter	Subsetting is available for selected ICESat/GLAS data products



Appendix **B**

**Waveform parameters are displayed  
along ICESat tracks**

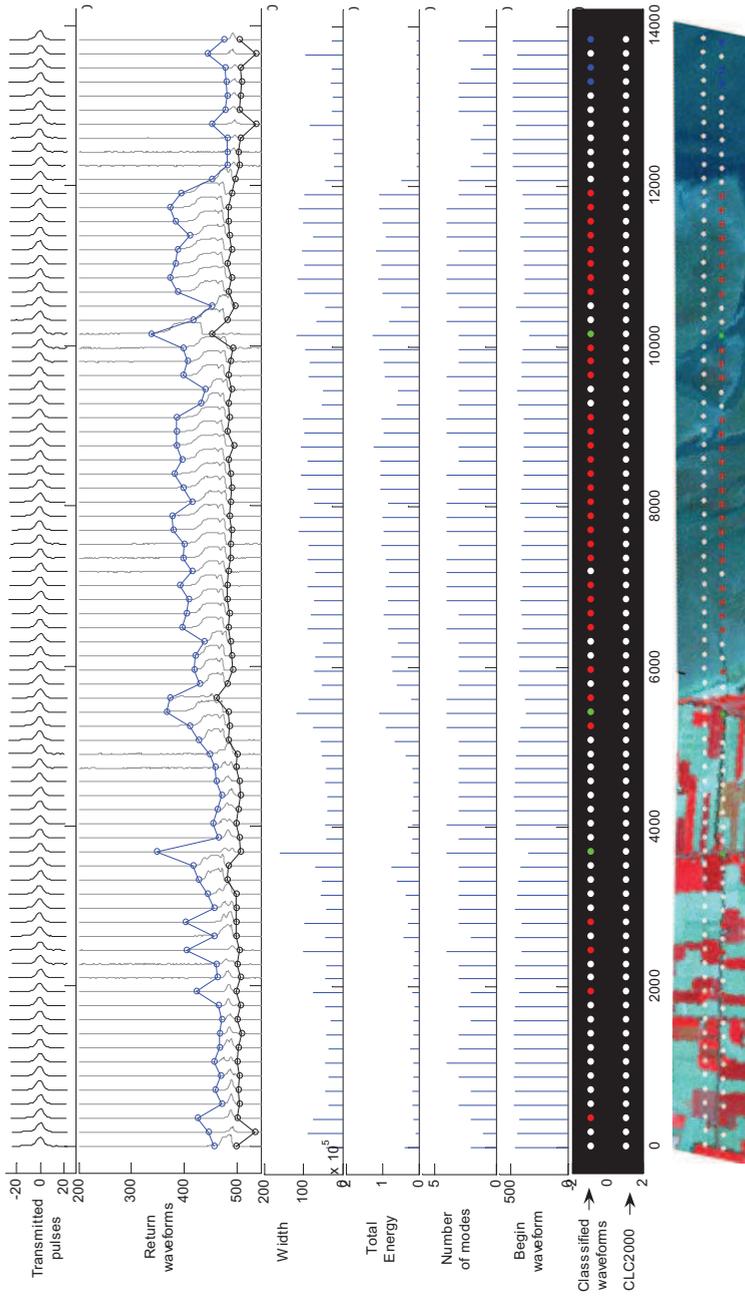


Figure B.1: Over the Waddenzee mud flat region.

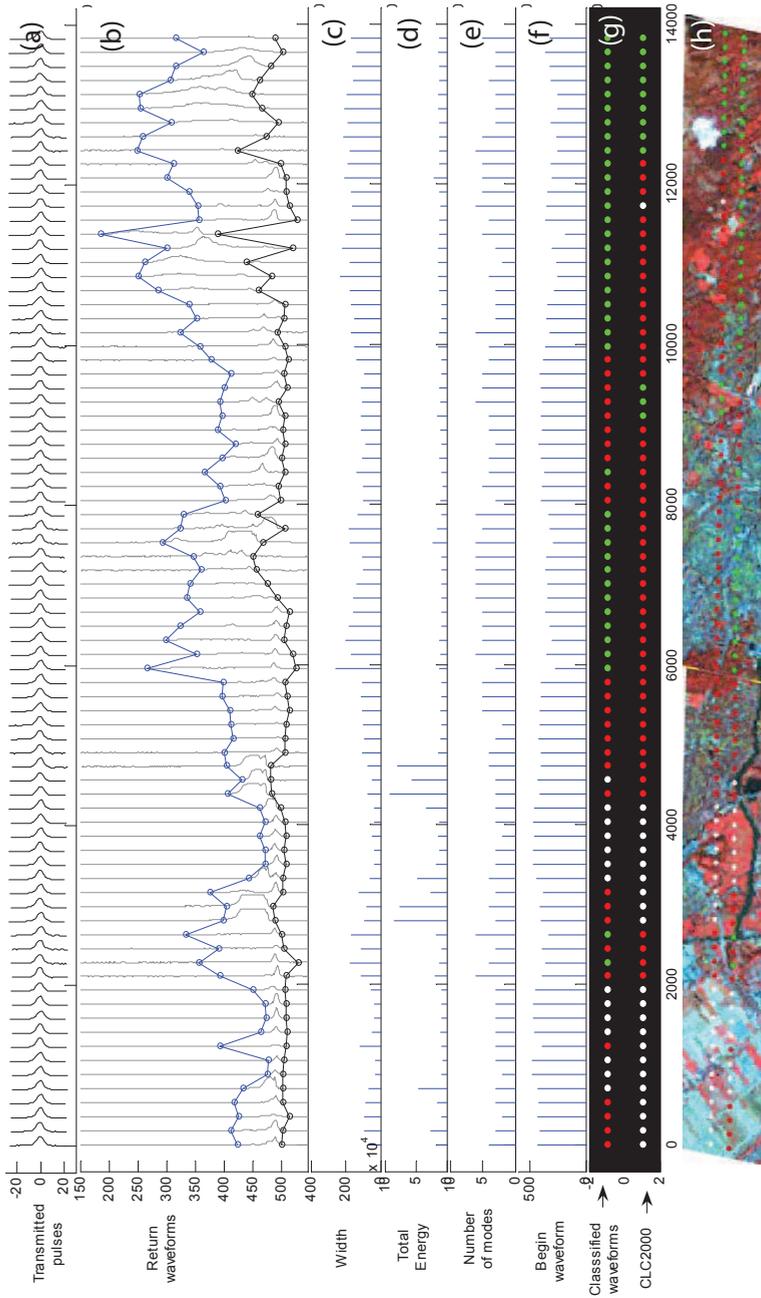


Figure B.2: Over an area near the Haarlem city.



Appendix **C**

## **Classification results**

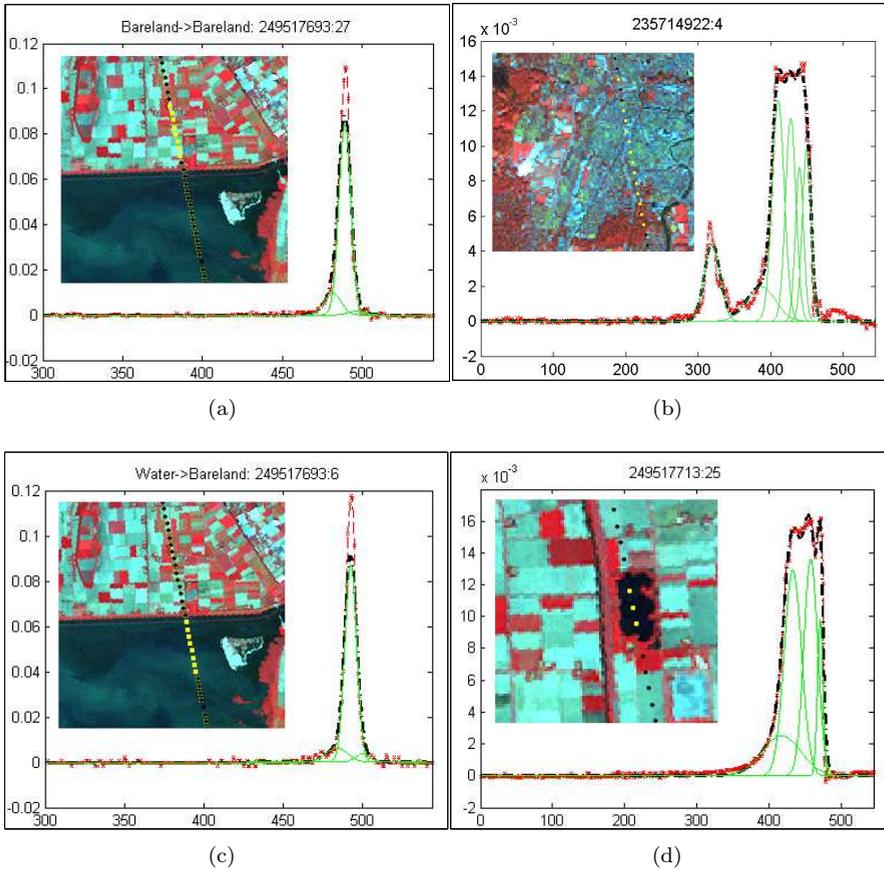
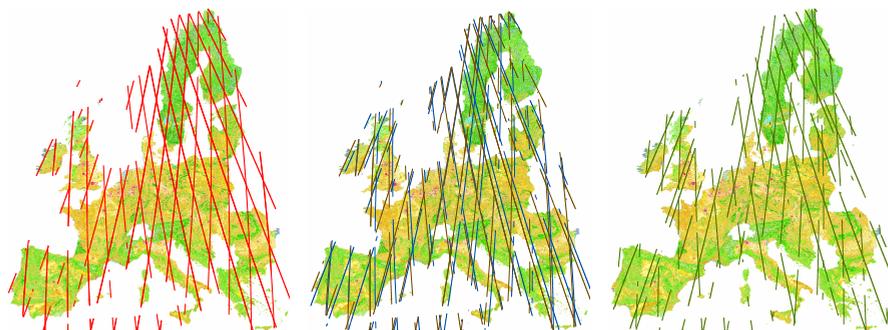


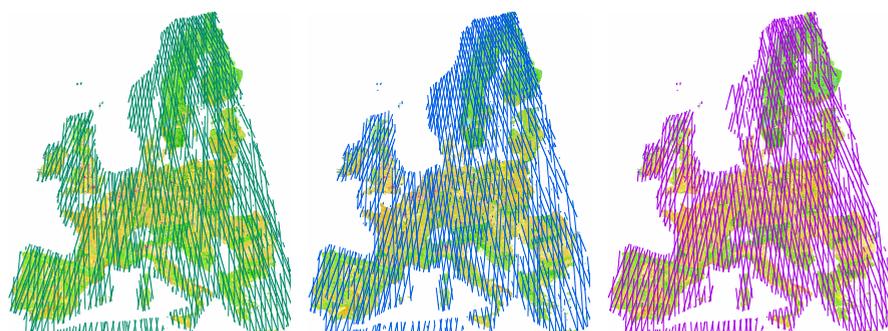
Figure C.1: (a) A correct classification result of bare land. However, there are still some misclassification results: (b) Urban class (Haarlem city) is classified as high vegetation; (c) Water classified to bare land at between Noordoost polder and Flevoland regions; (d) Water (with planted trees) classified to high vegetation at Noordoost polder.

Appendix **D**

## **ICESat data over Europe**



(a) L1 (02/20/03–03/29/03) (b) L2a (09/25/03–10/04/03) (c) L2b (02/17/04–03/21/04)



(d) L3a (10/03/04–11/08/04) (e) L3b (02/17/05–03/24/05) (f) L3d (10/21/05–11/24/05)



(g) L3e (02/22/06–03/28/06) (h) All campaigns

Figure D.1: ICESat data tracks over Europe.

Appendix **E**

## **Look-up table: Count to Voltage**

Table E.1: Look-up table: Count to Voltage for ICESat full waveform data.

Count	Voltage	Count	Voltage	Count	Voltage	Count	Voltage
1	-0.195279	65	0.231921	129	0.658924	193	1.055596
2	-0.188604	66	0.238596	130	0.665122	194	1.061794
3	-0.181929	67	0.245271	131	0.671320	195	1.067992
4	-0.175254	68	0.251946	132	0.677518	196	1.074190
5	-0.168579	69	0.258621	133	0.683716	197	1.080388
6	-0.161904	70	0.265296	134	0.689914	198	1.086586
7	-0.155229	71	0.271971	135	0.696112	199	1.092784
8	-0.148554	72	0.278646	136	0.702310	200	1.098982
9	-0.141879	73	0.285321	137	0.708508	201	1.105180
10	-0.135204	74	0.291996	138	0.714706	202	1.111378
11	-0.128529	75	0.298671	139	0.720904	203	1.117576
12	-0.121854	76	0.305346	140	0.727102	204	1.123774
13	-0.115179	77	0.312021	141	0.733300	205	1.129972
14	-0.108504	78	0.318696	142	0.739498	206	1.136170
15	-0.101829	79	0.325371	143	0.745696	207	1.142368
16	-9.52E-02	80	0.332046	144	0.751894	208	1.148566
17	-8.85E-02	81	0.338721	145	0.758092	209	1.154764
18	-8.18E-02	82	0.345396	146	0.764290	210	1.160962
19	-7.51E-02	83	0.352071	147	0.770488	211	1.167160
20	-6.85E-02	84	0.358746	148	0.776686	212	1.173358
21	-6.18E-02	85	0.365421	149	0.782884	213	1.179556
22	-5.51E-02	86	0.372096	150	0.789082	214	1.185754
23	-4.84E-02	87	0.378771	151	0.795280	215	1.191952
24	-4.18E-02	88	0.385446	152	0.801478	216	1.198150
25	-3.51E-02	89	0.392121	153	0.807676	217	1.204348
26	-2.84E-02	90	0.398796	154	0.813874	218	1.210546
27	-2.17E-02	91	0.405471	155	0.820072	219	1.216744
28	-1.51E-02	92	0.412146	156	0.826270	220	1.222942
29	-8.38E-03	93	0.418821	157	0.832468	221	1.229140
30	-1.70E-03	94	0.425496	158	0.838666	222	1.235338
31	4.97E-03	95	0.432171	159	0.844864	223	1.241536
32	1.16E-02	96	0.438846	160	0.851062	224	1.247734
33	1.83E-02	97	0.445521	161	0.857260	225	1.253932
34	2.50E-02	98	0.452196	162	0.863458	226	1.260130
35	3.17E-02	99	0.458871	163	0.869656	227	1.266328
36	3.83E-02	100	0.465546	164	0.875854	228	1.272526
37	4.50E-02	101	0.472221	165	0.882052	229	1.278724
38	5.17E-02	102	0.478896	166	0.888250	230	1.284922
39	5.84E-02	103	0.485571	167	0.894448	231	1.291120
40	6.50E-02	104	0.492246	168	0.900646	232	1.297318
41	7.17E-02	105	0.498921	169	0.906844	233	1.303516
42	7.84E-02	106	0.505596	170	0.913042	234	1.309714
43	8.51E-02	107	0.512271	171	0.919240	235	1.315912
44	9.17E-02	108	0.518946	172	0.925438	236	1.322110
45	9.84E-02	109	0.525621	173	0.931636	237	1.328308
46	0.105096	110	0.532296	174	0.937834	238	1.334506
47	0.111771	111	0.538971	175	0.944032	239	1.340704
48	0.118446	112	0.545646	176	0.950230	240	1.346902
49	0.125121	113	0.552321	177	0.956428	241	1.353100
50	0.131796	114	0.558996	178	0.962626	242	1.359298
51	0.138471	115	0.565671	179	0.968824	243	1.365496
52	0.145146	116	0.572346	180	0.975022	244	1.371694
53	0.151821	117	0.579021	181	0.981220	245	1.377892
54	0.158496	118	0.585696	182	0.987418	246	1.384090
55	0.165171	119	0.592371	183	0.993616	247	1.390288
56	0.171846	120	0.599046	184	0.999814	248	1.396486
57	0.178521	121	0.605721	185	1.006012	249	1.402684
58	0.185196	122	0.612396	186	1.012210	250	1.408882
59	0.191871	123	0.619071	187	1.018408	251	1.415080
60	0.198546	124	0.625746	188	1.024606	252	1.421278
61	0.205221	125	0.632421	189	1.030804	253	1.427476
62	0.211896	126	0.639096	190	1.037002	254	1.433674
63	0.218571	127	0.645771	191	1.043200	255	1.439872
64	0.225246	128	0.652446	192	1.049398	256	1.446070

# Bibliography

- Abdalati, W., C. Carabajal, B. Csatho, R. Dubayah, D. Fowler, H. A. Fricker, F. Hall, D. Harding, I. Joughin, D. Kimes, R. Kwok, S. Laxon, M. Lefsky, S. Luthcke, S. Martin, P. Minnett, J. Ranson, J. Sauber, T. Scambos, J. Shrinirne, R. Thomas, T. Urban, W. Wiscombe, and J. Zwally (2007). Report from the ICESat-II workshop. Technical report, National Aeronautics and Space Administration, June 27-29, Maryland.
- Abshire, J. B., X. Sun, H. Riris, J. M. Sirota, J. F. McGarry, S. Palm, D. Yi, and P. Liiva (2005). Geoscience laser altimeter system (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters* 32(21), 1–4. L21S02, doi:10.1029/2005GL024028.
- Abshire, J. M., J. F. McGarray, L. K. Pacini, J. B. Blair, and G. C. Elman (1994). Laser altimetry simulator, version 3.0, user’s guide. *NASA Technical Memorandum 104588*, 66pp.
- AHN (2008). Actual Height model of The Netherlands.
- Antonarakis, A., K. Richards, and J. Brasington (2008). Object-based land cover classification using airborne lidar. *Remote Sensing of Environment* 112, 2988–2998.
- Atwood, D., R. Guritz, R. Muskett, C. Lingle, J. Sauber, and J. Freymueller (2007, November). DEM control in Arctic alaska with ICESat laser altimetry. *IEEE Transactions on Geoscience and Remote Sensing* 45(11), 3710–3720.
- Bae, S. and B. E. Schutz (2002). Precision attitude determination (PAD). *Algorithm Theoretical Basis Documents (ATBD)*, 94pp.

- Bamber, J. and J. L. Gomez-Dans (2005). The accuracy of digital elevation models of the Antarctic continent. *Earth and Planetary Science Letters* 237(3-4), 516–523.
- Bhang, K. J., F. W. Schwartz, and A. Braun (2007). Verification of the vertical error in C-band SRTM DEM using ICESat and Landsat-7, Otter Tail County, MN. *IEEE Transactions on Geoscience and Remote Sensing* 46(1), 36–44.
- Biesemans, J., J. Everaerts, and N. Lewyckyj (2005). PEGASUS: Remote sensing from a HALE-UAV. *ASPRS annual convention*.
- Bindschadler, R., H. Choi, C. Shuman, and T. Markus (2005). Detecting and measuring new snow accumulation on ice sheets by satellite remote sensing. *Remote Sensing of Environment* 98(4), 388–402.
- Blair, J. B. and M. A. Hofton (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters* 26, 2509–2512.
- Blair, J. B., D. L. Rabine, and M. A. Hofton (1999). The laser vegetation imaging sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing* 54(2-3), 115–122.
- Boudreau, J., R. F. Nelson, H. A. Margolis, A. Beaudoin, L. Guindon, and D. S. Kimes (2008). Regional aboveground forest biomass using airborne and spaceborne lidar in qubec. *Remote Sensing of Environment* 112(10), 3876–3890.
- Braun, A. and G. Fotopoulos (2007). Assessment of SRTM, ICESat, and survey control monument elevations in Canada. *Photogrammetric Engineering & Remote Sensing* 73, 1333–1342.
- Brenner, A. C., J. R. DiMarzio, and H. J. Zwally (2007). Precision and accuracy of satellite radar and laser altimeter data over the continental ice sheets. *IEEE Transactions on Geoscience and Remote Sensing* 45(2), 321–331.
- Brenner, A. C., H. J. Zwally, C. R. Bentley, B. M. Csatho, D. J. Harding, M. A. Hofton, J. B. Minster, L. A. Roberts, J. L. Saba, R. H. Thomas, and D. Yi (2003). Geoscience laser altimeter system algorithm theoretical basis document: Derivation of range and range distributions from laser pulse waveform analysis. *Algorithm Theoretical Basis Documents (ATBD)*, 92pp.
- Bufton, J. L. (March 1989). Laser altimetry measurements from aircraft and spacecraft. In *In Proceedings of The IEEE*, Volume 77.
- Carabajal, C., D. Harding, S. Litheke, W. Fong, S. Rowton, and J. Frawley (1999). Processing of shuttle laser altimeter range and return pulse data in support of SLA02. In *International Archives of Photogrammetry and Remote sensing*, 32(3W14), pp. 67–52.

- Carabajal, C. and D. J. Harding (2006, March). SRTM C-band and ICESat laser altimetry elevation comparisons as a function of tree cover and relief. *Photogrammetric Engineering & Remote Sensing* 72(3), 287–298.
- Carabajal, C. C. and D. J. Harding (2005a). Evaluation of geoscience laser altimeter system (GLAS) waveforms for vegetated landscapes using airborne laser altimeter scanning data. *Geophysical Research Letters* 32(L21S02), doi:10.1029/2005GL024028.
- Carabajal, C. C. and D. J. Harding (2005b). ICESat validation of SRTM C-band digital elevation models. *Geophysical Research Letters* 32, L22S01.
- Coleman, T. and Y. Li (1994). On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds. *Mathematical Programming* 67 (2), 189–224.
- Coleman, T. and Y. Li (1996). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization* 6, 418–445.
- CORINE land cover 2000 (2006). European environment agency.
- Csatho, B., Y. Ahn, T. Yoon, C. J. van der Veen, S. Vogel, G. Hamilton, D. Morse, B. Smith, and V. B. Spikes (2005). ICESat measurements reveal complex pattern of elevation changes on siple coast ice streams, Antarctica. *Geophysical Research Letters* 32. L23S04, doi:10.1029/2005GL024289.
- Dennis, J. E. J. (1977). Nonlinear least squares. *State of the Art in Numerical Analysis ed. D. Jacobs, Academic Press*, 269–312.
- DiMarzio, J., A. Brenner, R. Schutz, C. A. Shuman, , and H. J. Zwally (2007). GLAS/ICESat 1 km laser altimetry digital elevation model of greenland. *Boulder, Colorado USA: National Snow and Ice Data Center. Digital media..*
- Drake, J. B., R. O. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, R. L. Chazdon, J. F. Weishampel, and S. Prince (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment* 79(2-3), 305–319.
- Dubayah, R., J. Blair, J. Bufton, D. Clark, J. Jaja, R. Knox, S. Luthcke, S. Prince, and J. Weishampel (1997). The vegetation canopy lidar mission. land satellite information in the next decade II: Sources and applications. In *In Proceedings of the American Society for Photogrammetry and Remote Sensing*, Bethesda, MD., pp. 100–112.
- Duong, H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2007). Error analysis of icesat waveform processing by investigating overlapping pairs over Europe. In *Proceedings: Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International: Sensing and Understanding Our Planet*, 4753–4756. 23–27 July, Barcelona, Spain.

- Duong, H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2009). ICESat full waveform altimetry compared to airborne laser scanning altimetry over The Netherlands. *IEEE Transaction on Geoscience and Remote Sensing* 47(10), 3365–3378.
- Duong, H., N. Pfeifer, and R. Lindenbergh (2006a). Analysis of repeated ICESat full waveform data: methodology and leaf-on / leaf-off comparison. In *Proceedings: Workshop on 3D Remote Sensing in Forestry*, 239–248. 14–15 February, Vienna, Austria.
- Duong, H., N. Pfeifer, and R. Lindenbergh (2006b). Full waveform analysis: ICESat laser data for land cover classification. In *proceedings: ISPRS Mid-term Symposium, Remote Sensing: From Pixels to Processes XXXVI*, 7, 31–35. 8–11 May, Enschede, The Netherlands.
- Duong, V. H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2008). Single and two epoch analysis of ICESat full waveform data over forested areas. *International Journal of Remote Sensing* 29(5), 1453–1473.
- EEA (2008). Eea reference grids.
- Filin, S. and B. Csathó (2002). An efficient algorithm for the synthesis of laser altimeter waveforms. *BPRC Technical Report*, 28pp..
- Fricker, H. A., J. N. Bassis, B. Minster, and D. R. MacAyeal (2005). ICESats new perspective on ice shelf rifts: The vertical dimension. *Geophysical Research Letters* 32. L23S08, doi:10.1029/2005GL025070.
- Fricker, H. A., A. Borsa, B. Minster, C. Carabajal, K. Quinn, and B. Bills (2005). Assessment of ICESat performance at the Salar de Uyuni, Bolivia. *Geophysical Research Letters* 32. L21S06, doi:10.1029/2005GL023423.
- Fricker, H. A. and L. Padman (2006). Ice shelf grounding zone structure from ICESat laser altimetry. *Geophysical Research Letters* 33. L15502, doi:10.1029/2006GL026907.
- Garvin, J., J. Bufton, J. Blair, D. Harding, S. Luthcke, J. Frawley, and D. Rowlands (1998). Observations of the earth’s topography from the shuttle laser altimeter (sla): Laser-pulse echo-recovery measurements of terrestrial surfaces. *Physics and Chemistry of The Earth* 23(9-10), 1053–1068.
- Gesch, D. B. (1998). Accuracy assessment of a global elevation model using shuttle laser altimeter data. In *Proceedings: Geoscience and Remote Sensing Symposium - IGARSS '98* 2, 840–842. Seattle, WA, USA.
- GLCN-LCTC (2006). Historical overview of glcn-lccs. *Land cover topic centre*.

- GSFC (2009). Laser vegetation imaging sensor (LVIS). *GODDARD SPACE FLIGHT CENTER*.
- Harding, D. (1998). Airborne lidar observations of canopy structure at the BOREAS tower flux sites. *In Proceedings: Geoscience and Remote Sensing Symposium - IGARSS '98, Seattle, WA*, 1550–1552.
- Harding, D., J. Blair, D. Rabine, and K. Still (2000). Slicer airborne laser altimeter characterization of canopy structure and sub-canopy topography for the boreas northern and southern study regions: Instrument and data product description. Technical report, In F. Hall & J. Nickeson (Eds.), Technical report series on the BOREAS ecosystem-atmosphere study (BOREAS): Remote Sensing Group. NASA/TM-2000-209891.
- Harding, D. J. and C. C. Carabajal (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters* 32. L21S10, doi:10.1029/2005GL023471.
- Harding, D. J., M. A. Lefsky, G. G. Parker, and J. B. Blair (2001). Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment* 76(3), 283–297.
- Hazeu, G. (2003). CLC2000 land cover database of The Netherlands: monitoring land cover changes between 1986 and 2000. Technical report, Green World Research, Wageningen, Alterra, Alterra-rapport 775/CGI-rapport 03-006.
- Heerd, R., E. Kuijlaars, M. Teeuw, and R. 't Zand (2000). Produktspecificatie AHN. *Rijkswaterstaat, Adviesdienst Geo-informatie en ICT*.
- Herring, T. and K. Quinn (1999). Atmospheric delay correction to GLASlaser altimeter ranges. Technical report, GLAS algorithm theoretical basis document. Version 1.0. Lanham, MD: Science Systems and Applications, Inc.
- Hofton, M. A. and J. B. Blair (2002). Laser altimeter return pulse correlation: a method for detecting surface topographic change. *Journal of Geodynamics* 34(3-4), 477–489.
- Hofton, M. A., J. B. Minster, and J. B. Blair (2000, jul). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing* 38, 1989–1996.
- Hollaus, M. and W. Wagner (2007). Operational use of airborne laser scanning for forestry applications in complex mountainous terrain. *9th International Symposium on High Mountain Remote Sensing Cartography*, 19–26.
- Hoog/NAP (2009). Nlgeo2004: Geoid model over The Netherlands.

- Jutzi, B., J. Neulist, and U. Stilla (2005). Sub-pixel edge localisation based on laser waveform analysis. In *Proceedings of the ISPRS WG III/3, III/4, V/3 Workshop Laser scanning 2005*, 39 (3/W19), pp. 109–114.
- Jutzi, B. and U. Stilla (2006a). Precise range estimation on known surfaces by analysis of full-waveform laser. In *Proceedings of the Photogrammetric Computer Vision PCV*.
- Jutzi, B. and U. Stilla (2006b). Range determination with waveform recording laser systems using a wiener filter. *ISPRS Journal of Photogrammetry & Remote Sensing* 61, 95–107.
- Kaula, W., G. Schubert, and R. Lingenfelter (1974). Apollo laser altimetry and inferences as to lunar structure. In *In Proceedings of the 5th Lunar Science Conference. Geochim. Cosmochim. Acta Suppl. 5*, 3049–3058.
- Kraus, K. and N. Pfeifer (1998). Determination of terrain models in wooded areas with airborne laser scanning data. *ISPRS Journal of Photogrammetry & Remote Sensing* 53, 193–203.
- Kurtz, N. T., T. Markus, D. J. Cavalieri, W. Krabill, J. G. Sonntag, and J. Miller (2008). Comparison of ICESat data with airborne laser altimeter measurements over Arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing* 46, 1913–1924.
- Lefsky, M. (2009). Forest height and biomass from spaceborne lidar. *Technical report*.
- Lefsky, M., D. Harding, M. Keller, W. Cohen, C. Carabajal, F. Espirito-Santo, M. Hunter, and R. Oliveira (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters* 32. L22S02, doi:10.1029/2005GL023971.
- Lefsky, M. A., W. B. Cohen, D. J. Harding, G. G. Parker, S. A. Acker, and S. T. Gower (2002). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography* 11, 393–399.
- Lefsky, M. A., W. B. Cohen, G. G. Parker, and D. J. Harding (2002). Lidar remote sensing for ecosystem studies. *BioScience* 52(1), 19–30.
- Lefsky, M. A., D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment* 67(1), 83–98.
- Lefsky, M. A., M. Keller, Y. Panga, P. B. de Camargod, and M. O. Hunter (2007). Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms. *Journal of Applied Remote Sensing* 1, 013537.

- Levenberg, K. (1944.). A method for the solution of certain problems in least squares. *Quarterly Applied Math.* 2, 164–168.
- Lillesand, T. M., R. W. Kiefer, and J. W. Chipman (2004). *Remote sensing and image interpretation* (5th ed.). New York: Wiley. 2004270046 GBA3-80608 Thomas M. Lillesand, Ralph W. Kiefer, Jonathan W. Chipman. ill. (some col.), maps (some col.) ; 25 cm. Includes bibliographical references and index.
- Luthcke, S. B., C. C. Carabajal, and D. D. Rowlands (2002). Enhanced geolocation of spaceborne laser altimeter surface returns: parameter calibration from the simultaneous reduction of altimeter range and navigation tracking data. *Journal of Geodynamics* 34(3-4), 447–475.
- Luthcke, S. B., D. D. Rowlands, T. A. Williams, and M. Sirota (2005). Reduction of ICESat systematic geolocation errors and the impact on ice sheet elevation change detection. *Geophysical Research Letters* 32. L21S05, doi:10.1029/2005GL023689.
- Magruder, L., E. Silverberg, C. Webb, and B. Schutz (2005). In situ timing and pointing verification of the ICESat altimeter using a ground-based system. *Geophysical Research Letters* 32. L21S04, doi:10.1029/2005GL023504.
- Magruder, L. A., C. E. Webb, T. J. Urban, E. C. Silverberg, and B. E. Schutz (2007). ICESat altimetry data product verification at white sands space harbor. *IEEE Transaction on Geoscience and Remote Sensing* 45(1).
- Mallet, C., F. Lafarge, F. Bretar, M. Roux, U. Soergel, and C. Heipke (September 1–2, 2009). A stochastic approach for modelling airborne lidar waveforms. In *In Proceedings: ISPRS Workshop, Laser scanning'09, Paris, France*.
- Marquardt, D. (1963). An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal Applied Math* 11, 431–441.
- Maser, C. (2001). *Forest primeval : the natural history of an ancient forest* (1st OSU Press ed.). Corvallis: Oregon State University Press. 2001016364 Chris Maser. ill. ; 23 cm. Originally published: San Francisco : Sierra Club Books, c1989. Includes bibliographical references (p. 247-267).
- McCarthy, D. e. (July 1996.). Iers conventions (1996), international earth rotation service technical note 21. Technical report, Observatoire de Paris.
- Meeus, J. (1991). *Astronomical Algorithms*. Willmann-Bell, Inc.
- Merriam-Webster (1998). *Merriam-Webster's collegiate dictionary* (10th ed.). Springfield, Mass.: Merriam-Webster. 97041846 Collegiate dictionary ill. ; 26 cm. Includes index.

- Molijn, R. A. (2009). ICESat full waveform signal analysis for classification of land cover types for cryospheric studies. Master's thesis, Delft University of Technology.
- More, J. J. (1977). The levenberg-marquardt algorithm: Implementation and theory. *Numerical Analysis*, ed. G. A. Watson, *Lecture Notes in Mathematics 630*, Springer Verlag, 105–116.
- Muller, J. W. (2000). Possible advantages of a robust evaluation of comparisons. *Journal of Research of the National Institute of Standards and Technology 105*, 551–555.
- Næsset, E., O. Bollandsås, and T. Gobakken (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment 94*(4), 541–553.
- Neuenschwander, A., R. Gutierrez, B. Schutz, and T. Urban (2006, 31 2006-Aug. 4). Comparison of small-footprint and large-footprint waveform lidar for terrestrial surface characterization. *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, 3758–3761.
- Neuenschwander, A. L. (2008). Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data. *Canadian Journal of Remote Sensing 34* (2), S240–S246.
- Neuenschwander, A. L., T. J. Urban, R. Gutierrez, , and B. E. Schutz (2008). Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: Implications for vegetation mapping. *Journal of Geophysical Research 113*. G02S03, doi:10.1029/2007JG00055.
- Nguyen, A. T. and T. A. Herring (2005). Analysis of ICESat data using Kalman filter and Kriging to study height changes in East Antarctica. *Geophysical Research Letters 32*(23).
- NSIDC (2005). Frequently asked question. *ICESat/GLAS Data at NSIDC*.
- NSIDC (2006). Tools for working with ICESat/GLAS data.
- NSIDC (2007). Laser operational periods - attributes.
- NSIDC (2008). GLAS altimetry data dictionary.
- Orfanidis, S. (1996). *Optimum Signal Processing. An Introduction. 2nd Edition*. Prentice-Hall, Englewood Cliffs, NJ.
- Parker, G. (1995). Structure and microclimate of forest canopies. in m. lowman and n. nadkarmi (eds). *Forest Canopies A Review of Research on Biological Frontier (San Diego: Academic Press)*, 73–106.

- Perdigão, V. and A. Annovi (1997). Technical and methodological guide for updating corine land cover database. *Joint Research Centre/European Environment Agency*.
- Persson, Å., U. Söderman, J. Töpel, and S. Ahlberg (2005). Visualisation and analysis of full-waveform airborne laser scanner data. In *In Proceedings of the ISPRS WG III/3, III/4, V/3 Workshop Laser scanning 2005, 39 (3/W19)*, pp. 103–108.
- Rabbani, T., F. A. van den Heuvel, and G. Vosselman (Dresden, Germany, September 25–27, 2006). Segmentation of point clouds using smoothness constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI, PART 5*, 248–253.
- Random House Webster (1997). *Random House Webster's unabridged dictionary* (2nd ed.). New York: Random House. 97017702 Random House compact unabridged dictionary. ill., maps ; 29 cm. + 1 computer laser optical disc (4 3/4 in.) Rev., updated ed. of: Random House dictionary of the English language. 2nd ed., unabridged.
- Ranson, K., G. Sun, K. Kovacs, and V. Kharuk (2004). Landcover attributes from ICESat GLAS data in Central Siberia. *International Geoscience and Remote Sensing Symposium, Anchorage, AK, SEP 20-24, 2004, IGARSS 2004: IEEE International Geoscience and Remote Sensing Symposium Meetings 1-7*, 753–756.
- Ranson, K. L., G. Sun, K. Kovacs, and V. I. Kharuk (2006). Use of ICESat GLASdata for forest disturbance studies in Central Siberia. *Geophysical Research Letters* 33. L07501, doi:10.1029/2005GL025227.
- RDNAP (2007). Rijksdriehoeksmeting and normaal amsterdams peil. *Dutch Geometric Infrastructure*.
- Reitberger, J., P. Krzystek, and M. Heurich (2006). Full-waveform analysis of small footprint airborne laser scanning data in the bavarian forest national park for tree species classification. In *In Proceedings of the International Workshop 3D Remote Sensing in Forestry, 14-15 February 2006, Vienna, Austria*.
- Rim, H. J. and B. E. Schutz (2002). Precision orbit determination (POD). *Algorithm Theoretical Basis Documents (ATBD)*, 111pp.
- RNIN (2008). Pctrans 4.2.3. Technical report, Hydrographic Service of the Royal Netherlands Navy.
- Rosette, J. A. B., P. R. J. North, and J. C. Suárez (2008). Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing* 29(5), 1475–1493.

- Schenk, T., B. Csatho, C. J. van der Veen, H. Brecher, Y. Ahn, and T. Yoon (2005). Registering imagery to ICESat data for measuring elevation changes on byrd glacier, Antarctica. *Geophysical Research Letters* 32(23).
- Schutz, B. E. (2001). Glas altimeter post-launch calibration/validation plan (version 1.0). Technical report, Center for Space Research, University of Texas at Austin.
- Schutz, B. E. (2002). Laser footprint location (geolocation) and surface profiles. Technical report, Center for Space Research, The University of Texas at Austin.
- Schutz, B. E., H. J. Zwally, J. Bufton, C. Bentley, T. Herring, J. Minster, J. Spinhirne, and R. Thomas (1997). Geoscience laser altimeter system GLAS science requirements. Technical report, NASA Earth Observing System.
- Shuman, C. A., H. J. Zwally, B. E. Schutz, A. C. Brenner, J. P. DiMarzio, V. P. Suchdeo, , and H. A. Fricker (2006). ICESat Antarctic elevation data: Preliminary precision and accuracy assessment. *Geophysical Research Letters* 33, L07501. L07501, doi:10.1029/2005GL025227.
- Simard, M., V. H. Rivera-Monroy, J. E. Mancera-Pineda, E. Castaeda-Moya, and R. R. Twilley (2008). A systematic method for 3d mapping of mangrove forests based on shuttle radar topography mission elevation data, icesat/glas waveforms and field data: Application to cinaga grande de santa marta, colombia. *Remote Sensing of Environment* 112(5), 2131–2144.
- Slobbe, D. C., P. Ditmar, and R. C. Lindenbergh (2008b). Estimating the rates of mass change, ice volume change and snow volume change in greenland from ICESat and GRACE data. *Geophysical Journal International* 176(1), 95–106.
- Slobbe, D. C., R. C. Lindenbergh, and P. Ditmar (2008). Estimation of volume change rates of Greenland’s ice sheet from ICESat data using overlapping footprints. *Remote Sensing of Environment* 112(12), 4204–4213.
- Spinhirne, J. D., S. P. Palm, W. D. Hart, D. L. Hlavka, and E. J. Welton (2005). Cloud and aerosol measurements from GLAS: Overview and initial results. *Geophysical Research Letters* 32. L22S03, doi:10.1029/2005GL023507.
- Sun, G., K. Ranson, D. Kimes, J. Blair, and K. Kovacs (2008). Forest vertical structure from GLAS: an evaluation using LVIS and SRTM data. *Remote Sensing of Environment* 112(1), 107–117. doi: 10.1016/j.rse.2006.09.036.
- Teunissen, P., D. Simons, and C. Tiberius (2005). *Probability and Observation theory*. Delft University of Technology.
- Urban, T., R. Gutierrez, and B. Schutz (2008). Analysis of ICESat laser altimetry elevations over ocean surfaces: sea state and cloud effects. In *In Proceedings of the IEEE International Geoscience & Remote Sensing Symposium*. July 6-11, Boston, Massachusetts, U.S.A.

- Urban, T. J., B. E. Schutz, and A. L. Neuenschwander (2008). A survey of ICESat coastal altimetry applications: Continental coast, open ocean island, and inland river. *Terrestrial, Atmospheric and Oceanic Sciences* 19, 1–19.
- Wagner, W., A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitizing airborne laser scanner. *ISPRS Journal of Photogrammetry & Remote Sensing* 60, 100–112.
- Wahr, J., D. Wingham, and C. Bentley (2000). A method of combining ICESat and GRACE satellite data to constrain Antarctic mass balance. *Journal of Geophysical Research* 105(B7), 16279–16294.
- Yi, D., J. Minster, and C. Bentley (1999). Ocean tidal loading corrections. Technical report, GLAS algorithm theoretical basis document. Version 1.0. Greenbelt, MD: Goddard Space Flight Center.
- Yi, D., H. J. Zwally, and X. Sun (2005). ICESat measurement of Greenland ice sheet surface slope and roughness. *Annals of Glaciology* 42(1), 83–89.
- Zuber, M. T., D. E. Smith, S. Solomon, D. O. Muhleman, J. W. Head, J. B. Garvin, J. B. ABSHIRE, and J. L. Bufton (1992). The mars observer laser altimeter investigation. *Journal of Geophysical Research* 97 (E5), 7781–7797.
- Zwally, H. J., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhirne, and R. Thomas (2002). ICESat’s laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics* 34(3-4), 405–445.
- Zwally, H. J., D. Yi, R. Kwok, and Y. Zhao (2008). ICESat measurements of sea ice freeboard and estimates of sea ice thickness in the weddell sea. *Journal of Geophysical Research* 113. C02S15, doi:10.1029/2007JC004284.





# List of Abbreviations

Abbreviations are divided into two groups. First a list of relevant quality flags of ICESat data products is given. Then additional abbreviations used in this thesis listed.

## A. ICESat products quality flags

<b>i_satCorrFlg</b>	Saturation Correction Flag
<b>i_satElevCorr</b>	Saturation Elevation Correction
<b>i_pctSAT</b>	Percent Saturation
<b>i_tpintensity_avg</b>	Transmit Pulse intensity - frame average
<b>i_tpazimuth_avg</b>	Transmit pulse azimuth - frame average
<b>i_tpeccentricity_avg</b>	Transmit Pulse eccentricity - frame average
<b>i_tpmajoraxis_avg</b>	Transmit Pulse major axis - frame average
<b>i_satNdx</b>	Saturation Index
<b>i_wfFitSDev</b>	The received echo fit standard deviation
<b>i_reflctUncorr</b>	Reflectivity not corrected for atmospheric effects
<b>i_reflctuncmxpk</b>	Reflectivity Not Corrected For Atmospheric Effects from max peak

<b>i.RecNrgAll_EU</b>	1064 Laser received Energy from all signal above threshold
<b>i.RecNrgLast_EU</b>	1064 nm Laser Received Energy (max peak)
<b>i.GainShiftFlg</b>	Gain Shift Flag
<b>i.satNrgCorr</b>	Saturation Energy Correction
<b>i.gval_rcv</b>	Gain Value used for Received Pulse
<b>i.ElvuseFlg</b>	Elevation use flag
<b>i.beam_coelev</b>	Co-elevation
<b>i.cld1_mswf</b>	Cloud multiple scattering warning flag
<b>i.refRng</b>	Reference range
<b>i.lDRngOff</b>	Land range offset
<b>i.MRC_af</b>	Medium resolution cloud availability flag
<b>i.ElvFlg</b>	elevation definition flag

## B. Additional abbreviations

<b>AHN</b>	Actueel Hoogtebestand Nederland (in Dutch): Actual Height model of the Netherlands
<b>RDNAP</b>	Rijksdriehoeksmeting and Normaal Amsterdams Peil: Dutch coordinate system
<b>CLC2000</b>	CORINE Land Cover 2000 database
<b>ATBD</b>	ICESat/GLAS Algorithm Theoretical Basis Documents
<b>SRTM</b>	Shuttle Radar Topography Mission
<b>LIDAR</b>	Light Detection And Ranging
<b>MOLA</b>	Mars Observer Laser Altimeter
<b>Nd:YAG</b>	Neodymium Yttrium Aluminum Garnet
<b>SLA</b>	Shuttle Laser Altimeter
<b>GSFC</b>	Goddard Space Flight Center

<b>SLICER</b>	Scanning Lidar Imager of Canopies by Echo Recovery
<b>LVIS</b>	Laser Vegetation Imaging Sensor
<b>VCL</b>	Vegetation Canopy Lidar
<b>GLAS</b>	Geoscience Laser Altimetry System
<b>POD</b>	Precision Orbit Determination
<b>PAD</b>	Precision Attitude Determination
<b>BOREAS</b>	BOReal Ecosystem-Atmosphere Study
<b>ICESat</b>	Ice, Cloud and land Elevation Satellite
<b>FOV</b>	Field of view.
<b>NASA</b>	National Aeronautics and Space Administration
<b>NSIDC</b>	National Snow and Ice Data Center





# Summary

## Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data

In the last two decades, laser scanning systems made the transition from scientific research to the commercial market. Laser scanning has a large variety of applications such as digital elevation models, forest inventory and man-made object reconstruction, and became the most required input data for flood plain and hydraulic models. This system is generally called as a discrete laser scanning system. A discrete laser scanning system sends a pulse to the ground surface and records the return signal resulting from the illumination of the surface. The area of the illuminated surface is defined as the footprint size of the laser shot. The two-way travel time of the laser pulse allows to determine the distance of the laser system to the surface. Traditional systems are unable to record the complete return signal, but typically store only one to four distances to objects in the laser footprint.

A new system developed to overcome the above limitations is the so-called full waveform scanning system. The system sends out a pulse of a certain width and amplitude. After reflection of the pulse on the objects surface, the system records the complete returning pulse signal. This complete signal is the so-called full waveform. Compared to traditional scanning systems, a full waveform system retrieves more information that should still be extracted though from the waveform shape. The shape of the full waveform contains information on the characteristics of the illuminated footprint, like object information (tree and building height), forest structure and ground surface characteristics (e.g. forest species surface roughness and slope) as well as a land cover type (water, bare earth, or urban areas).

Harding (1998), and Blair et al. (1999) showed that using observations from a full waveform laser system it is possible to achieve accurate forest structure and biomass estimates. However, the system considered in this case was just operated from an airplane flying at a low altitude of a few kilometers above the ground surface with a medium footprint size of about 10-20 m. Moreover, data acquisition could only be performed in a small area. Some typical systems used in these days were the Scanning Lidar Imager of Canopies by Echo Recovery system (SLICER), the Laser Vegetation Imaging Sensor (LVIS), and the commercial airborne full waveform scanning system from RIEGL the LMS-Q560 (2003).

In 2003, moreover, NASA launched the first satellite full waveform system, the so-called Ice Cloud and land Elevation Satellite system (ICESat), carrying the Geoscience Laser Altimetry System (GLAS) instrument. The purpose of ICESat was to collect, among others, measurements concerning the Antarctica and Greenland ice sheets and their mass balance, concerning land vegetation and concerning the atmosphere. This space borne system acquired data between 2003 and 2009 over the entire earth from 600 km altitude, with a footprint size of about 70 m and a distance between consecutive footprints of approximately 175 m. However, due to for example the high operational altitude, this system was also affected by many error sources, like instrumental and operational problems, atmospheric effects and surface conditions. Processing such data is a challenging task; and development of a validation method together with the development of new applications of large footprint full waveform data are the main targets of my thesis work.

For this purpose, the contents of this thesis is organized in six chapters In chapter 1, the background and scope of the research is introduced. In chapter 2, full waveform sensor and instrument development is presented. The main content starts in chapter 3. This chapter focuses on processing and parameterization of large footprint ICESat full waveform data.

Data handling is described in detail. It is presented how a pulse interacts to objects on the earth surface and what information a return pulse contains. Moreover, it is shown how waveform parameters can be obtained by fitting a number of Gaussian numbers to a waveform using least squares. waveform in a sense of least square estimation. Two alternative methods to obtain information on waveforms, i.e. waveform deconvolution and waveform simulation were also implemented and are discussed. At the end of this chapter, an overview is given of waveform parameters and their possible physical interpretation and application. . This chapter aims to provide a tutorial to wider audiences/readers who would like to use satellite laser full waveform data for their own purpose.

In order to evaluate the accuracy and the precision of elevation and height estimates that can be obtained using the ICESat full waveform system, the topic of validation is studied in chapter 4. In this chapter, first background and related work are discussed. Two validating cases discussed in detail : (i) Comparison

between the bare land elevations derived from ICESat full waveform data and airborne laser scanning data over The Netherlands; and (ii) Identification and investigation of error sources of the ICESat full waveform system by comparing waveform pairs that have overlapping footprints.

For the first validation case, filtering constraints have been investigated and further developed to avoid influence of data anomalies. Examples of factors having a negative influence on a comparison include on one hand waveforms affected by saturation or cloudy conditions and on the other hand footprints that are not covered well by the available airborne data points. The comparison between ICESat and accurate airborne laser data over The Netherlands confirms that the ICESat full waveform accuracy and precision strongly depends on land cover type. With respect to different land cover type, the accuracy was about -21 cm with a standard deviation of about 20 cm over bare earth. The accuracy was -24 cm with a standard deviation of 28 cm over urban areas and -9 cm with a largest standard deviation of about 45 cm over forest. As expected, the accuracy decreases when the complexity of the surface increases (e.g. from bare land to urban). The accuracy of ICESat derived elevations over water could not be assessed because of few effects: changes in water level, lack of airborne data points over water surfaces, etc., It is concluded that if a) proper filtering is applied, and b) the terrain relief is small, the last mode of an ICESat waveform represents the terrain/bare land height with both an accuracy and precision at the decimeter level.

Also feature height estimation of features like trees or buildings were studied. The difference between ICESat and airborne derived feature heights are acceptable over forested and bare land areas. However, the result over buildings are not satisfactory. The main reason is that ICESat derived feature height estimations are sensitive to feature height variations occurring at spatial distances smaller than the size of the ICESat footprints. For features that are homogeneous at the scale of ICESat footprints, ICESat waveform analysis is a suitable method for estimating feature heights. For urban environments, incorporation of an additionally accurate Digital Elevation Model might still enable the monitoring of feature height changes.

For the second validation case a database was constructed, consisting of more than one hundred thousand (>100000) repeated, partly overlapping ICESat waveforms over Europe. The aim is to identify the cause of changes in waveform parameters obtained from (partly) overlaying waveforms. Even in such a size of database it turned out almost impossible to identify suited waveform pairs. The first problem was the lack of almost completely overlaying footprints, and second, the lack of waveform pairs situated at stable areas. Unfortunately, the number of almost completely overlapping footprints from repeated ICESat measurements that could be obtained from campaign L1 to L3e in the region between 36N and 71N latitude and 11W and 33E longitude was very limited. Performing a full worldwide search over all campaigns is however expected to result in a quite large data set of suitable almost perfectly overlapping repeated footprints. Such data set could therefore be

used to identify and resolve remaining ICESat processing issues.

After applying all corrections identified in the calibration/validation procedure, the ICESat product is ready to be used in different applications considering the surface of the Earth. How to develop applications by using large footprint full waveform data is discussed in chapter 5. By considering the full waveform parameters derived from the waveform processing procedures of chapter 3, two applications of ICESat full waveform data were studied in detail.

A new contribution described in this chapter compares repeated ICESat waveform pairs to assess forest change. By analyzing corresponding waveforms acquired in different campaigns at nearly coinciding footprints, canopy changes caused by seasonal influences were detected. A general new aspect introduced in this chapter is to analyze waveform parameters in terms of a pair of full waveforms that were acquired at approximately the same location. Comparison of the waveform parameters of such pairs not only allows the detection of seasonal influences on forest type, it may also become possible to further detect and estimate important forest change parameters, like forest growth and deforestation.

As a first step seasonal changes over broadleaf, mixed-wood, and needleleaf forests between winter and summer epochs of 2003, along near-coincident ground tracks were quantified. It was found that although the maximum tree height barely changes over 6 months, i.e., less than 2.2% for the three forest types, a suited waveform parameter can detect forest canopy change mostly for broadleaf (a 148% change, winter to summer) and less for conifers (a 36% change). Alternative waveform parameters to describe forest changes are also discussed.

An application of the seasonal change in waveform parameters is to use these parameters to classify footprint areas directly into forest type classes. Preliminary results, with a kappa value of 0.57, provide a baseline against which improvements in both data and methodology can be gauged in future. Future work should include a method for correcting slope-induced changes in the waveform results, both for individual waveforms and for waveform pairs. In addition, further improvement is expected by including a neighbourhood analysis, that is by incorporating spatial correlation between close by waveforms and their changes. An individual quality descriptor of the tree parameter values could be obtained by quantifying and propagating errors encountered during the waveform processing. Moreover, tree parameter values should be validated against either data from field measurements or data from other sensors. Finally, the parameter definitions should be refined and validated in order to improve agreement with biophysical characteristics of the forest.

In the second application, it is demonstrated how ICESat full waveforms can be used for land cover classification. It was the first time that the possibility of using ICESat data for this purpose was investigated. Over The Netherlands, ICESat footprint locations were classified into four classes: high vegetation (high trees

or forest), urban, water, and bare land/low vegetation. The following waveform parameters were used as class attributes: return energy, waveform extent, waveform start and number of Gaussian components. It is concluded that the accuracy of classification equals 73% in comparison to a confusion matrix based on the CORINE land cover database 2000 (CLC2000) covering the same study area.

In addition, it is shown that ICESat waveforms could be used to build a feature vector consisting of suited waveform parameters that can be consecutively applied to classify the land cover class of the footprint location. The full waveform parameters derived from chapter 3 together with total return energy able to discriminate between high vegetation, urban areas, bare land/low vegetation and water. This result showed the feasibility of land cover classification with spaceborne lasers. As the ICESat satellite has a near polar orbit, coverage is global, but still a main disadvantage of ICESat data is that only tracks were mapped. With the development of new systems, an area-wise coverage may become possible in future. In that case, automatic classification of large footprint full coverage full waveform laser data may lead to land cover classification results of the same high quality as can be obtained from optical data.

At the end of the thesis, significant information is summarized and remaining steps but also potential new applications are described.

*Hieu Van Duong*





# Samenvatting

## Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data

In de laatste twee decennia heeft laser scanning de overgang gemaakt van wetenschappelijk onderzoek naar de commerciële markt. Laser scanning wordt gebruikt voor de constructie van digitale hoogte modellen en 3D modellen van gebouwen, maar ook voor bijvoorbeeld bosinventarisatie en het bepalen van ruwheid van de ondergrond in gebieden met een groot overstromingsrisico. Een typisch laser scanning systeem stuurt een puls naar het aardoppervlak en vangt vervolgens het teruggekaatste signaal op. Uit de dubbele reistijd van de laser puls wordt de afstand naar het oppervlak bepaald. De voetafdruk van een laser puls is het gebied op de grond dat door de puls geraakt wordt. Traditionele systemen slaan niet het complete teruggekaatste signaal op, maar registreren alleen de afstand tot één tot vier objecten in de voetafdruk van de laser.

Zogenaamde ‘full waveform’ systemen kennen deze beperking niet. Het systeem verstuurt een puls van een bepaalde breedte en amplitude. Na kaatsing op het oppervlak, registreert het systeem het volledige terugkerende signaal als functie van de tijd. Dit complete signaal wordt de ‘full waveform’ oftewel complete golfvorm genoemd. Vergeleken met traditionele laser scan systemen, registreert een ‘full waveform’ systeem in principe meer informatie. De golfvorm bevat allerlei informatie over kenmerken van objecten binnen de verlichte voetafdruk, zoals hoogte van bomen en gebouwen, de dichtheid van het bladerdak en eigenschappen van de ondergrond, zoals helling en ruwheid. Deze informatie moet echter nog wel uit de golfvormen worden geëxtraheerd.

Harding (1998) en Blair et al. (1999) toonden aan dat met behulp van waarnemingen van een ‘full waveform’ systeem het mogelijk is om nauwkeurige schattingen te verkrijgen van bosstructuur en biomassa. Echter, de systemen die voor dit soort onderzoek gebruikt werden, waren gemonteerd op een vliegtuig. Daarmee was de vlieghoogte beperkt tot een paar kilometer en de data inwinning alleen mogelijk binnen een relatief klein gebied. De grootte van de voetafdrukken was in de orde van 10 tot 20 m. Typische systemen die gebruikt werden in deze begin dagen waren het ‘Scanning Lidar Imager of Canopies by Echo Recovery’ systeem (SLICER), en de Laser Vegetation Imaging Sensor (LVIS).

In 2003 lanceerde NASA de eerste laser ‘full waveform’ satelliet, de zogenaamde ‘Ice Cloud and land Elevation Satellite’ (ICESat), met aan boord het ‘Geoscience Laser Altimetry System’ GLAS. Het doel van ICESat was om waarnemingen te verzamelen met betrekking tot de staat van de ijskappen op Antarctica en Groenland en om bovendien metingen te doen aan vegetatie en atmosfeer. ICESat verzamelde wereldwijd gegevens vanuit de ruimte op 600 km hoogte tussen 2003 en 2009. ICESat verzamelde alleen data direct onder zich met een voetafdruk van ongeveer 70 m en een afstand tussen twee opeenvolgende voetafdrukken van 175 m. Vanwege onder meer de grote hoogte van de satelliet, spelen vele aspecten een rol in het verwerken van de data, zoals bijvoorbeeld problemen met de lasers in de ruimte, bewolking, maar ook snelle overgangen tussen verschillende soorten landoppervlak. Daarom is het optimaal verwerken van zulke data een grote uitdaging. Dit, tezamen met het ontwikkelen van nieuwe methodes ter controle en toepassing van gegevens die uit golfvorm data met een grote voetafdruk gehaald kunnen worden, vormen de belangrijkste doelstellingen van mijn promotieonderzoek.

De resultaten van mijn onderzoek worden in dit proefschrift beschreven in zes hoofdstukken. In hoofdstuk 1 wordt de achtergrond en de scope van het onderzoek beschreven. In hoofdstuk 2 wordt in detail ingegaan op de ontwikkeling en eigenschappen van volledige golfvorm systemen. De eigenlijke inhoud begint in hoofdstuk 3. Dit hoofdstuk beschrijft om te beginnen de verwerking en parametrisering van ICESat full waveform data.

Er wordt beschreven hoe een puls verandert, afhankelijk van de objecten in de voetafdruk van de puls, en, daarmee samenhangend, welke informatie de teruggekaatste puls, zoals die door de ICESat satelliet wordt opgevangen, bevat. Bovendien wordt beschreven hoe een golfvorm gemodelleerd kan worden als een geschikte som van Gaussische klokvormen die worden verkregen met behulp van een kleinste kwadraten vereffening. Twee alternatieve methodes om informatie uit golfvormen te verkrijgen worden ook besproken: golfvorm simulatie, waarin een golfvorm als het ware wordt gereconstrueerd uit een gedetailleerde terrein beschrijving, en golfvorm deconvolutie, waarin de invloed van het terrein op de inkomende golf direct wordt bepaald, zonder tussenkomst van klokvormen. Aan het eind van dit hoofdstuk wordt een overzicht gegeven van golfvorm

parameters en hun fysische interpretatie en toepassing. Dit hoofdstuk kan als handleiding dienen voor lezers die zelf met laser golfvorm satelliet data willen werken.

Het onderwerp van hoofdstuk 4 is de validatie van de nauwkeurigheid en precisie van hoogte schattingen gebaseerd op ICESat golfvormen. Twee verschillende methodes van validatie worden besproken: (i) Vergelijking tussen terrein en object hoogtes verkregen uit ICESat golfvormen enerzijds en uit het zeer gedetailleerde AHN, Actueel Hoogtebestand Nederland, archief (lucht laser scanning) anderzijds, en, (ii) Vergelijking van steeds twee ICESat golfvormen waarvoor geldt dat de voetafdrukken ten minste gedeeltelijk overlappen.

Eerst is uitgebreid geanalyseerd welke golfvormen mogelijk onbetrouwbaar zijn. Op grond van deze analyse zijn regels ontworpen die worden toegepast om verdachte golfvormen in een vroeg stadium automatisch te kunnen verwijderen. Voorbeelden van factoren die een negatieve invloed kunnen hebben op ICESat golfvormen zijn het vollopen van de GLAS sensor, of bewolking. Een vergelijking met AHN data is ook niet fair als de beschikbare AHN data niet de volledige ICESat voetafdruk bedekt. De resultaten van de vergelijking bevestigen dat de nauwkeurigheid en de precisie van terreinhoogte afleidbaar uit ICESat golfvormen sterk afhankelijk is van het type bodembedekking. Op open Nederlands terrein is zowel de nauwkeurigheid als de standaard deviatie in de orde van 2 decimeter. Boven stedelijk gebied zijn deze getallen nauwelijks hoger, maar hoewel boven bosgebied het verschil tussen ICESat en AHN terreinhoogte gemiddeld kleiner dan 10 cm is, loopt de standaard afwijking op naar 45 cm. Zoals verwacht, neemt de nauwkeurigheid af naarmate de complexiteit van het oppervlak toeneemt. De kwaliteit van waterstanden verkregen uit ICESat golfvormen is om diverse redenen buiten beschouwing gelaten. De conclusie uit de vergelijking met de AHN data is dat het mogelijk is de terreinhoogte van eenvoudig terrein te bepalen met zowel een nauwkeurigheid als een precisie van 1 decimeter, door a) het toepassen van de juiste filter regels, en b) gebruik te maken van de onderste Gaussische klokform zoals die verkregen kan worden uit de ICESat golfvorm.

Er is ook na gegaan of het mogelijk is om hoogtes van bijvoorbeeld bomen en gebouwen uit ICESat golfvormen te schatten. Weer zijn de resultaten gevalideerd met AHN data. Het verschil tussen boomhoogtes verkregen uit ICESat en AHN data is aanvaardbaar, de resultaten voor gebouwen zijn echter onder de maat. De voornaamse reden is dat object hoogtes afgeleid uit ICESat data gevoelig zijn voor hoogte verschillen die optreden op een spatiële schaal die kleiner is dan de grootte van de ICESat voetafdrukken. Voor objecten die niet sterk veranderen binnen een ICESat voetafdruk, is golfvorm analyse een geschikte methode voor het schatten van de object hoogte.

De tweede methode van validatie is intern, omdat steeds ICESat golfvormen met elkaar worden vergeleken, in plaats van met andere data. Ten behoeve van de

deze methode is een database gebouwd, bestaande uit meer dan 100 000 paren van golfvormen met ten minste gedeeltelijk overlappende voetafdruk. Het betreft allemaal ICESat golfvormen uit de meetcampagnes L1 tot L3e, die boven Europa zijn verkregen, in het gebied tussen de 36 en 71 graden noorderbreedte en tussen 11 graden westerlengte en 33 graden oosterlengte. Het doel is het identificeren van de oorzaak van veranderingen in golfvorm parameters verkregen op (bijna) dezelfde locatie. Zelfs in een database van dergelijke grootte bleek het echter bijna onmogelijk geschikte paren te identificeren. Paren zijn geschikt als ten eerste de voetafdrukken bijna volledig het zelfde zijn, en ten tweede verkregen zijn in een stabiel gebied. Een database samengesteld uit alle wereldwijd beschikbare ICESat golfvormen verkregen tussen 2003 en 2009 zou waarschijnlijk wel voldoende geschikte paren opleveren. Zo'n dataset zou dan wel succesvol toegepast kunnen worden voor het identificeren en oplossen van problemen met betrekking tot het verwerken van ICESat data.

Na het doorvoeren van alle correcties die in de kalibratie/validatie procedure geïdentificeerd zijn, zijn de ICESat data klaar om toegepast te worden in het analyseren van het oppervlak van de Aarde. In hoofdstuk 5 worden de verschillende toepassingen besproken van golfvorm data met een grote voetafdruk, zoals verkregen met de ICESat satelliet. Door gebruik te maken van de parameters die uit laser golfvormen bepaald kunnen worden, zoals besproken in hoofdstuk 3, konden ook twee nieuwe toepassingen ontwikkeld worden, die in detail besproken worden. Een nieuwe bijdrage die beschreven wordt, laat zien dat herhaald ingemeten ICESat golfvormen gebruikt kunnen worden om veranderingen in bossen te monitoren. Parameter waarden verkregen van golfvormen, gemeten op dezelfde locatie, maar in verschillende seizoenen, kunnen gelinkt worden naar veranderingen in de kruin van bomen door met name bladverlies. Dit principe van het monitoren van parameters van golfvormen verkregen op ongeveer dezelfde locatie kan breder toegepast worden: zo kan informatie over het type bos verkregen worden, maar ook kan de groei van bomen of ontbossing op deze manier vanuit de ruimte gedetecteerd worden.

Als een eerste stap worden in hoofdstuk 5 seizoensgebonden veranderingen gequantificeerd en gedifferentieerd naar bostype. Voor deze analyse worden golfvormparameters uit de winter van 2003/2004 vergeleken met dezelfde parameters maar dan verkregen uit data van de zomer van 2003. Het kon worden vastgesteld dat, hoewel de boomhoogte nauwelijks verandert in 6 maanden, (volgens de analyse minder dan 2,2%), er geschikte golfvormparameters bestaan die veranderingen in de kruin beschrijven: voor loofbomen kon zo'n 148% verandering vastgesteld worden, terwijl voor naaldbomen, die minder gevoelig zijn voor seizoensinvloeden, ook een veel kleinere verandering van 36% gevonden werd.

Een directe toepassing van de mogelijkheid tot het detecteren van seizoensgebonden veranderingen in bossen is bostype kwalifikatie: een grote verandering duidt op loofbos, enz. Dit is geprobeerd en gevalideerd met externe data. Voorlopige

resultaten, met een kappa waarde van 0.57, kunnen dienen als referentie voor verbeteringen in de toekomst, zowel qua data als qua methode. Toekomstig werk dat veranderingen in golfvormen analyseert, moet wel rekening houden met de invloed van bijvoorbeeld een helling binnen de voetafdruk. Verdere verbeteringen in de methodiek kunnen worden aangebracht door ook te analyseren welke veranderingen in naburige golfvormen optreden, die wellicht nog hetzelfde bos beslaan. Een beschrijving van de kwaliteit van de resultaten kan worden verkregen door in elke stap de mogelijke (onvermijdelijke) fouten te identificeren en deze rigoureus naar het eindresultaat voort te planten. Ten slotte zal een uitgebreide validatie met gegevens van bijvoorbeeld in situ metingen duidelijk kunnen maken welke biofysische parameters automatisch uit ICESat golfvorm data kunnen worden geëxtraheerd.

In de tweede toepassing wordt aangetoond hoe ICESat golfvormen kunnen worden gebruikt voor classificatie van landgebruik. Het is de eerste keer dat de mogelijkheid van het gebruik van ICESat gegevens voor dit doel werd onderzocht. ICESat voetafdrukken uit heel Nederland werden ingedeeld in vier klassen: hoge vegetatie (hoge bomen of bos), stedelijk gebied, water en open terrein. Op basis van golfvorm parameters als lengte van de golfvorm en het aantal componenten van een Gaussische decompositie is een classificatie schema opgezet. Geconcludeerd wordt dat de nauwkeurigheid van de indeling gelijk is aan 73%, in vergelijking met een bestaand landgebruik product, de CORINE Land Cover database 2000 (CLC2000). Dit resultaat toont aan dat het mogelijk is landgebruik automatisch te detecteren op grond van satelliet laser golfvorm data. Omdat de ICESat satelliet een bijna perfecte polaire baan volgde, is de dekking van de data wereldwijd. Een belangrijk nadeel is echter dat ICESat alleen direct onder zich meet, wat betekent dat de beschikbare gegevens niet gebiedsdekkend maar alleen in smalle stroken beschikbaar zijn. Met de ontwikkeling van nieuwe satelliet systemen, zoals bijvoorbeeld ICESat II, wordt gebiedsdekkende inwinning in de toekomst wellicht wel haalbaar. In dat geval kan met deze nieuwe methode van automatische classificatie van laser golfvormen een classificatieresultaat verkregen worden van dezelfde kwaliteit als uit optische data mogelijk is.

Aan het eind van dit proefschrift worden de belangrijkste resultaten samengevat en wordt een overzicht gegeven van resterende stappen. Met name wordt ingegaan op potentiële nieuwe toepassingen, mogelijk gemaakt door enerzijds de ontwikkeling van nieuwe satelliet systemen, anderzijds door nieuwe methodologie, waarvan een aantal voorbeelden in dit proefschrift beschreven worden.

*Hieu Van Duong*



# List of Publications

Duong, H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2009). ICESat full waveform altimetry compared to airborne laser scanning altimetry over The Netherlands. *IEEE Transaction on Geoscience and Remote Sensing* 47(10), 3365–3378.

Duong, V. H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2008). Single and two epoch analysis of ICESat full waveform data over forested areas. *International Journal of Remote Sensing* 29(5), 1453–1473.

Duong, H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2007). ICESat full waveform altimetry compared to airborne laser altimetry over The Netherlands. *In Proceedings of the ISPRS Workshop on Laser Scanning*. 12–14 September, Espoo, Finland

Duong, H., R. Lindenbergh, N. Pfeifer, and G. Vosselman (2007). Error analysis of icesat waveform processing by investigating overlapping pairs over Europe. *In Proceedings: International Geoscience and Remote Remote Sensing Symposium: Sensing and Understanding Our Planet*. 23–27 July, Barcelona, Spain.

Duong, H., N. Pfeifer, and R. Lindenbergh (2006b). Full waveform analysis: ICESat laser data for land cover classification. *In proceedings: ISPRS Mid-term Symposium, Remote Sensing: From Pixels to Processes*, 31–35. 8–11, Enschede, The Netherlands

Duong, H., N. Pfeifer, and R. Lindenbergh (2006a). Analysis of repeated ICESat full waveform data: methodology and leaf-on / leaf-off comparison. *In Proceedings: Workshop on 3D Remote Sensing in Forestry*, 239–248. 14–15 February, Vienna





# Curriculum Vitae

Hieu Van Duong was born in 1976 in Ho Chi Minh City, Vietnam. In 1994, he enrolled at the Ho Chi Minh City University of Technology to obtain a BE degree in Telecommunication in the Department of Electrical and Electronic Engineering. After four and a half years, he graduated in 1999 as one of the best students in his year with the thesis ‘Development of a Geographic Information System for Telecommunication System Management and Surveillance at Bien Hoa City, Dong Nai Province, Vietnam’. He was then immediately hired as a project researcher at the Center for IT and Geographic Information System (DITAGIS). Over the next 3 years, he worked as a GIS professional in GIS training, student supervision, project management and as head of the spatial database group.

In 2002, he was awarded a full scholarship of the National Space Development Agency of Japan (NASDA) for pursuing a master degree at the Asian Institute of Technology (AIT), Thailand. The title of his master thesis was ‘A real time Building Identification System’. In 2003, he graduated as the best student and immediately joined the Geoinformatics Center, AIT. He worked as a researcher and daily supervisor for a mini-project assigned to the Vietnam group, Institute of Physics. He organized a GIS training in Laos, and went for field campaigns in Bangladesh and Vietnam.

In 2005, he joined the Section of Optical and Laser Remote Sensing, and the Section of Mathematical Geodesy and Positioning as a PhD student at the Delft University of Technology, The Netherlands. The research contents and result of his PhD work are all presented in this thesis. Currently, he is working as a postdoctoral fellowship at the Colorado State University, United States.