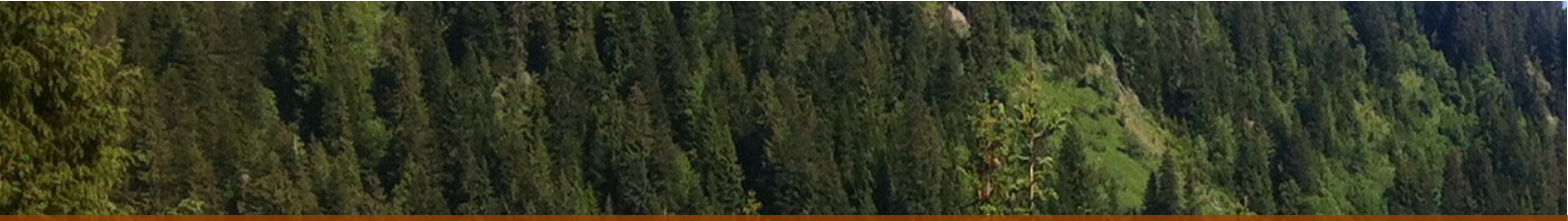


Master of Science Thesis

Boreal forest change detection with the spaceborne Geoscience Laser Altimeter System

Wildfires and canopy height changes derived from waveform pairs

S.N. Vollenhoven

December 2016

Delft University of Technology

Master thesis

Boreal forest change detection with the spaceborne Geoscience Laser Altimeter System

Wildfires and canopy height changes derived from waveform pairs

Author:

S.N.Vollenhoven
(student number: 4189647)

Thesis committee:

prof. dr. M. Menenti
dr. R.C. Lindenbergh
R.A. Molijn, MSc

Delft University of Technology
Faculty of Civil Engineering and Geosciences
Department of Geoscience and Remote Sensing
Section of Optical and Laser Remote Sensing

December 27, 2016

TU Delft

Acknowledgements

First of all I would like to thank dr. R.C. Lindenbergh. He made it possible for me to graduate in time before relocating to New Zealand and for that I am most grateful. Furthermore, many thanks to prof. dr. M. Menenti for his valuable feedback, as well as his guidance in my search for a graduation topic. Thanks to R.A. Molijn for being so generous with his time.

Additionally I would like to thank the employees of the National Snow and Ice Data Center for supplying the data and quick and clear responses to my questions. Correspondence with ms. Russo and mr. Chen of Natural Resources Canada has been very helpful.

Finally I would like to express my gratitude towards my life-partner. Thanks to her, one of my two big dreams will be fulfilled in about half a month from now when we move to New Zealand.

Abstract

The Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat) has provided measurements of vegetation structure and together with upcoming follow-up missions such as ICESat-2, vegetation changes may be studied globally in the near future. In this research project, ICESat's individual ability to estimate changes in maximum canopy height was evaluated over the boreal forests of northern Alberta, Canada. The changes were estimated from waveform pairs with one and two-year time intervals (2004 to 2006), and although footprint misalignments caused random errors on the order of meters, average decreases in maximum canopy height were observed over wildfires in northern Alberta. Only measurements from the three late-May to late-June campaigns were used since snow cover might have complicated interpretation of the results otherwise. Footprint center distances of waveform pairs were up to 150 m and the upper limit of the standard error of an individual change measurement was estimated at 8 m. Maximum canopy height was estimated to decrease with about one meter on average within wildfire perimeters and the standard error of the mean was estimated at ± 0.3 m.

Changes in waveform extent were used to estimate the changes in maximum canopy height, and a revised method of delineating the waveform extent is presented. The revision is shown to improve the reproducibility of the waveform extent measurements and is expected to improve the accuracy of change estimation and any biophysical variable retrieval methods that depend on the waveform extent.

The attempt made to study canopy changes with the GLAS was successful in detecting changes due to wildfires and has shed light on the many issues involved. With further research on changes in the transmitted laser pulse characteristics and implementation of biophysical variable retrievals, the GLAS may prove to be a valuable tool for studies on global vegetation dynamics.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter serves as an introduction to the research project. The main point of the chapter is to define the project in terms of research questions, scope and objectives. However, in order to make the reasoning behind the project definition clear to the reader, we will first provide some background information in the following section.

1.1 Research context

1.1.1 Tree growth and wildfires in the Boreal Zone

Tree growth can be defined as the increase in size and numbers of leaves, stems and roots of a tree [11]. It is usually quantified in terms of oven-dry biomass [36]. The energy that a growing tree stores in the form of biomass is taken from solar radiation via photosynthesis. In this biochemical process, which lies at the basis of tree growth, carbon-dioxide and water are converted to glucose and oxygen under the influence of light absorbed by chlorophyll molecules. Due to the use of atmospheric carbon-dioxide in the production of biomass, it is a form of carbon sequestration and of great interest for climate scientists [9]. Within the global carbon cycle, forests function as carbon sinks when growing, though they can become sources of atmospheric carbon-dioxide when burning and/or rotting processes get the upper hand [30]. Tree growth on a global scale is thus a process of great importance in understanding climate change.

The world's forests can be divided into at least three forest types: tropical, temperate and boreal. Of these three, boreal forests (sometimes called taiga [1]) are the most northernly and represent one of the largest biogeoclimatic areas [2]. These *"snow forests"* are roughly located between 50°N and 70°N, south of the circumpolar tundra and north of temperate regions. They are therefore mostly located in Russia and Canada (figure 1.1).

The biome is characterized and dominated by evergreen conifers. The growing season is relatively short due to the limited amounts of solar radiation and low temperatures during winter. Yearly height-growth is typically of the order of decimeters [17] and decreases towards the edge of the tundra, where the climatic conditions do not facilitate tree growth [14].

The development of boreal forest is heavily influenced by large-scale forest disturbances. In much of the Boreal Zone, fire is the major stand-renewing agent [13]. Wildfire cycles are roughly 100 years long for the boreal forests of Canada [19]. Other disturbances include diseases, insect outbreaks, wind throw and logging. After a disturbance, the affected forest usually undergoes a natural process of regeneration. Due to the high frequency of wildfires, most boreal forest stands are young and in the process of regeneration [17].

[1] The term taiga is usually used for the high-latitude forests of Russia, whereas the term boreal forest is used more often for North America. However, both terms are used in the definition of Canada's ecozones; taiga then refers to the open forest stands on the northern edge of the Boreal Zone and boreal forest to the closed forest stands at lower latitudes. [12]

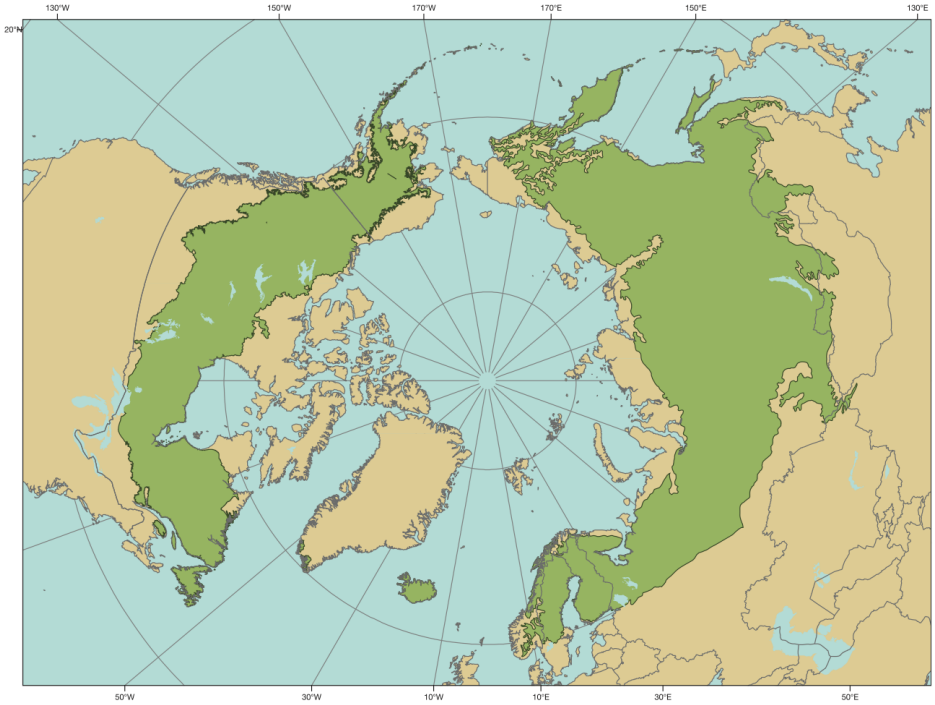


Figure 1.1: The boreal zone. *Extracted from Brandt et al. [2]*

1.1.2 ICESat and canopy height changes

ICESat (Ice, Cloud,and land Elevation Satellite) was an earth-observation satellite launched by NASA (National Aeronautics and Space Administration) in January 2003. Its main objective was to measure ice sheet elevation changes. Measuring vegetation characteristics globally, thereby supplying a reference for remeasurement during future follow-up missions, was one of its secondary objectives. The satellite operated in a near-polar orbit and collected global surface elevation measurements by laser ranging for about seven years. The LiDAR (Light Detection And Ranging) instrument onboard was named GLAS: Geoscience Laser Altimeter System. LiDAR is an active remote sensing technology that uses a laser to illuminate a target and retrieves information from the reflected light, most notably the distance to the target (the *range*). ICESat carried three lasers within the altimeter system. All three lasers emitted pulses of near-infrared light with a wavelength of 1064 nm. Pulses with a wavelength of 532 nm were also emitted, via a frequency doubler, which is an useful addition for measuring atmospheric properties. The pulse signal as a function of time is called the waveform[2] and transmitted waveforms had a gaussian shape. The received waveform holds information on the vertical structure of the reflecting surfaces within portion of the target area that is illuminated by the laser pulse. This illuminated portion is usually called the laser *footprint* in altimetric applications. In the case of satellite LiDAR measurements over forests, the received waveform often shows two major peaks: one broad peak corresponding to the forest canopy[3] and one narrow peak which indicates the signal reflected from the ground (the ground return). This scenario is illustrated in figure 1.2. When the ground beneath the forest is sloped, the ground return produces a broader peak. For GLAS, the nominal footprint size was 60 m in diameter and it therefore falls in the category of large footprint full waveform LiDARs. The term *full waveform* indicates that the system records the waveform with high temporal resolution, as opposed to single return and multiple return LiDAR systems, which only record the range and intensities of the largest peak(s).

The ground track was mostly kept within ±1 km of the reference ground track, which repeated itself initially every 8 days during the calibration period and subsequently every 91 days for the rest of the mission. Coverage is nearly global: between 86°S to 86°N. Starting near the end of the calibration period, the track repeat precision at latitudes above 59°N and south of 59°S was improved to the point that the ground tracks were generally within ±150 m of the reference ground tracks. The precision was unchanged in between these latitudes.

[2] "Waveform" refers to a received waveform unless indicated specifically otherwise

[3] Canopy will in this document refer to the collection of above-ground vegetation components of any plant community [24] unless explicitly stated otherwise as is the case here with "forest canopy".

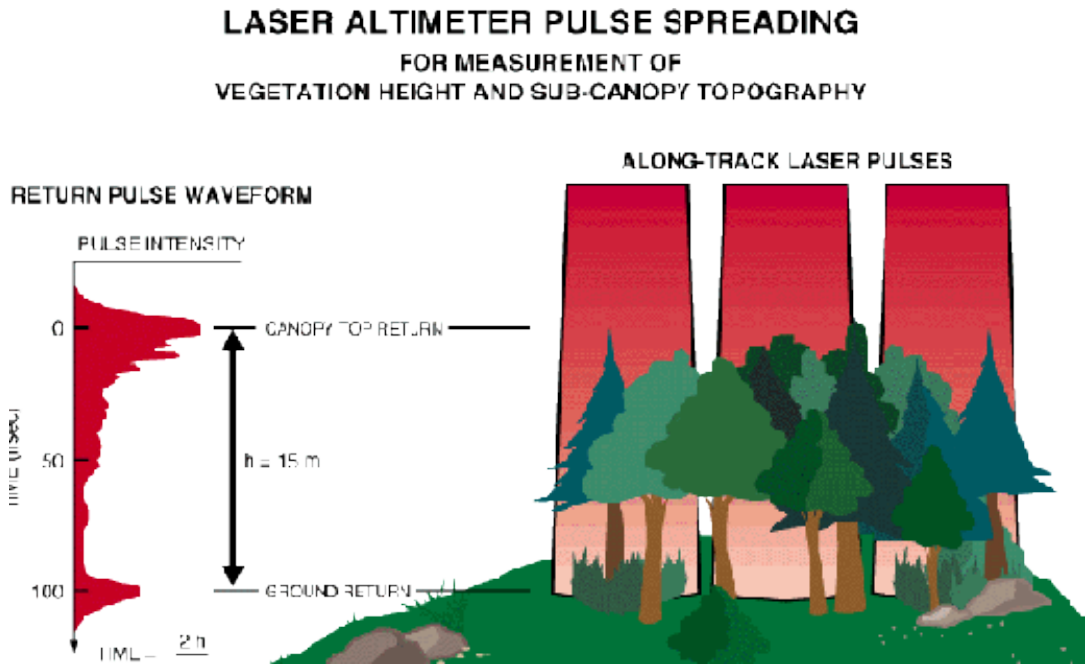


Figure 1.2: Illustration of ICESat laser pulses over forest. *Source: icesat.gsfc.nasa.gov*

Laser energy distributions over the near-infrared footprints approximate elliptical Gaussians with a major axis (to the $1/e^2$ relative energy points) between roughly 50 m and 150 m, and an eccentricity between roughly 0.5 and 0.9 [1] [26]. An example footprint energy distribution as detected by the Laser Profile Array onboard ICESat is shown in figure 1.3. Footprint centers were 172 m apart along the ground tracks. Figure 1.4 illustrates repeating and crossing tracks. For the 91-day repeat orbit, ground track separation was about 20 km at a latitude of 50°and about 10 km at 70° [10]. Track separation reached a maximum of 30 km at the equator and a minimum of roughly 5 km at the tracks maximum latitude of 86°.

GLAS has been operational from February 2003 to October 2009, though the last year of data is of significantly lower quality due to laser degradation. Data was not collected continuously but during 18 operational periods. Most of these "campaigns" lasted about 35 days and therefore cover about a third of the reference ground track repeat cycle.

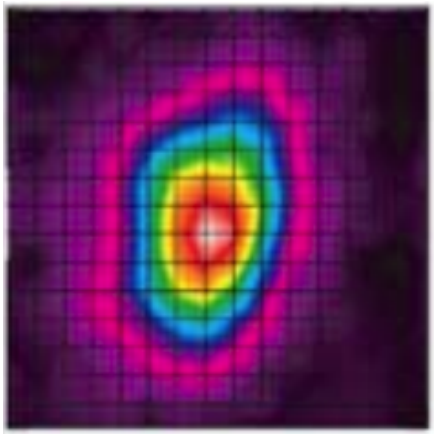


Figure 1.3: Example of a footprint laser energy distribution as detected by the Laser Profile Array onboard ICESat. *Extracted from Abshire et al. [1]*

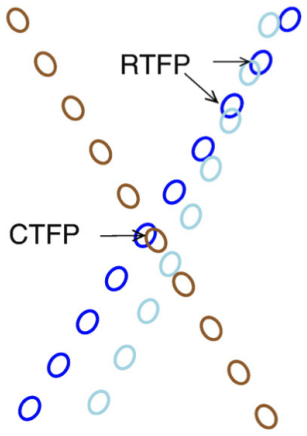


Figure 1.4: Illustration of three ICESat tracks with Crossing Track Footprint Pairs (CTFPs) and Repeat Track Footprint Pairs (RTFPs). *Extracted from Slobbe et al. [34]*

The ICESat mission has been successful in its main objective of quantifying elevation changes of ice sheet surfaces. As for forests, one of the main achievements is a global canopy height map [22]. The main purpose of canopy height maps and the biomass maps they feed is to serve as a reference point for carbon stock monitoring. ICESat's followup mission, ICESat-2, is scheduled to launch in 2018 and together these missions should provide estimates of long term changes in vegetation canopy height. However, we are currently not aware of any publications on ICESat's individual ability to estimate canopy height changes (or any other non-seasonal changes in canopy characteristics for that matter, but here we focus on canopy height). The GLAS datasets span almost seven years and canopy height changes during these seven years may be at a detectable level at specific locations (e.g. forest stands affected by wildfires) if not in the majority of forest stands (undisturbed forests may grow in height). This research project will investigate the possibilities of estimating canopy height changes in Canada's boreal forests, where forest structure is highly dynamic due to the stand-renewing disturbances such as wildfires [19].

1.2 Project definition

1.2.1 Research questions

This research project will focus on answering the following question:

> **To what extent is it possible to detect canopy height changes with the Geoscience Laser Altimeter System?**

The research may thereby result in supplying a (preliminary) method of quantifying canopy height changes using GLAS data, and/or the technical limitations that hinder their detection might be identified.

The following aspects of the main research question above are identified as subquestions:

i How can we efficiently identify and extract useful GLAS waveforms from the data sources?

ii How can we estimate changes in maximum canopy height from repeated waveform measurements?

iii Given a study site and a tolerance for measurement alignment, how are the repeated measurements spatially and temporally distributed?

iv To what extent are the measurements reproducible?

v Is diminishment of maximum canopy height as a result of wildfires detectable with the GLAS?

vi Do change measurement statistics suggest tree growth?

1.2.2 Scope and objectives

The scope of this project is defined as follows:

- The focus will be on canopy height changes. While biomass changes may be more relevant ecologically, it can only be derived from GLAS via assumptions concerning tree allometry and wood density [23]. Analyses of changes in other aspects of forest structure, such as canopy closure, are left for further research. (Some work was done with changes in canopy closure, but it is not presented in this report for the sake of brevity.)
- The main objective will be to explore methods of change detection and analysis, and not to improve upon the accuracy of GLAS-derived static biophysical parameters.
- The focus will be on the boreal forest plains of northern Alberta, Canada. This gives us the following advantages for forest change detection:
  - converging ICESat tracks at high latitudes
  - short wildfire cycles (about 50-100 years for western Canada) [19]
  - a limited amount of footprints on steep slopes
  - high availability of auxiliary data for Canada

Chapter 2

Theory

In this chapter we give a brief theoretical description of the waveform. We discuss two methods of delineating the waveform extent, which will be used to estimate changes in the maximum vegetation canopy height within a footprint.

2.1 Waveform interpretation

A waveform represents the backscatter intensity as a function of time, and since the travel time is directly and linearly related to the distance from the instrument (the fraction of multiply scattered photons is commonly held to be negligible), the waveform is proportional to the illumination and reflectance weighted vertical distribution of reflective surfaces within the laser pulse footprint [3]. Assuming nadir-pointing, the elevation z corresponding to the two-way travel time t is equal to the elevation S of the instrument above a reference surface (e.g. a geoid or ellipsoid) minus the range:

$$z = -\frac{1}{2}c \cdot t + S \tag{2.1}$$

where c is the speed of light. When interested only in elevations differences between reflective surfaces within the footprint, the constant S may be dropped[1]:

$$\Delta z = -\frac{1}{2}c \cdot \Delta t \tag{2.2}$$

Now, this concept is complicated by the non-zero duration of the emitted laser pulse: the travel time of a specific photon that is reflected back to the instrument's detector cannot be known exactly. The waveforms *transmitted* by the GLAS have an approximately Gaussian shape with a nominal full width at half maximum (FWHM) of 4 ns [3]. Thus, if a particular footprint would hypothetically yield a waveform $h(t)$ in response to a transmitted waveform of infinitesimal duration and unit integral, described by a Dirac-delta function $\delta(t)$, the response to the transmitted Gaussian pulse $g(t)$ would be the convolution of $g(t)$ and $h(t)$:

$$r(t) = g(t) * h(t) \tag{2.3}$$

Convolution with a Gaussian has a smoothing effect on the scale of the width of the Gaussian. That is not to say that the received waveforms are smooth: these are affected by detector noise. Yet the actual physical signal of backscatter energy as a function of time is expected to be smooth on such time scales.

In the context of forested land and assuming that part of the transmitted pulse reaches the ground, the received waveform is the sum of a ground component (or *ground return*), which has a Gaussian shape for flat land, and a vegetation component, which' shape depends on the spatial distribution of plant material. The two components may overlap significantly when the ground is not level or the understory is densely vegetated.

[1] Furthermore, whereas equation 3.5 requires several corrections to the range $\frac{1}{2}c \cdot t$ (e.g. for atmospheric delays), these are not needed for in-footprint elevation differences.

With the ground component presumably removed, the shape of the remaining vegetation-part of the waveform can provide information on the vertical vegetation structure. The ratio of the vegetation-part of the waveform energy over the total waveform energy can be interpreted as the reflectance-weighted canopy cover, where the canopy cover is the fraction of ground obscured by significantly higher layers of vegetation. Lefsky et al. found the reflectance of the ground at a wavelength of 1064 nm to be approximately equal to that of the vegetative surfaces for a boreal forest in Saskatchewan [22], implying that the fraction of vegetation-signal energy is equal to the canopy cover.

By making the assumption that all vegetative surfaces have the same reflectance at 1064 nm, we can extend the idea of a total canopy cover measurement to measurements of canopy cover (or *canopy closure*) per horizontal layer. The fraction of backscattered photons that have been intercepted at height H or higher then gives the cover of this horizontal layer. Under the assumption that the horizontal distribution of plant surfaces and their angular distribution is constant with height, the illumination weighting, which varies with height due to occlusion, may be corrected for [22]. However, the research presented in this report does not involve such methods, so they will be not be discussed any further.

2.2 Waveform-derived changes in maximum canopy height

In this report, we will only concern ourselves with the extent of the waveform: the time-difference between the beginning and the end of the signal, and the elevation difference that corresponds to it according to equation 2.2. Delineation of signal beginning and end is a prerequisite for the derivation of other waveform parameters and its accuracy will affect the accuracy of subsequent waveform parameter estimations and any biophysical variables retrieved from these (Duong gives an overview of parameters that may be estimated from a waveform [10]). The waveform extent itself reflects the elevation difference between lowest and highest detected surfaces within the footprint, although it may be slightly larger due to the convolution with the transmit pulse. So there are three causes of temporal broadening of the return pulse:

- Non-zero duration of the transmit pulse. The nominal FWHM of the transmit pulse (4 ns) corresponds to an elevation difference of 60 cm (equation 2.2) and this contributes to the waveform extent.
- In-footprint relief. Any ground elevation differences within the footprint contribute to the waveform extent (assuming that a part of the laser energy penetrates the vegetation layers all the way to the ground level).
- Vegetation. The maximum canopy height above ground level contributes to the waveform extent as well, on the order of tens of meters for treed land.

Accordingly, Yang et al. [38] and Selkowitz et al. [32] suggest to estimate canopy height h_{canopy} over terrain with an average slope θ from the waveform extent E as

$$h_{canopy} = E - r \cdot \tan\theta - c \cdot T \tag{2.4}$$

where r is the footprint diameter (assuming a circular footprint) and T is the FWHM of the transmitted pulse. So for forested plains (in-footprint ground elevation differences negligible with respect to the maximum canopy height, $\theta = 0°$), the waveform extent can be used as an estimator of maximum canopy height above ground level. The estimator is slightly biased by the non-zero duration of the transmit pulse. Yet a *difference* in waveform extent for a repeated measurement can be interpreted as an estimator for a *change* in maximum canopy height without the estimator being biased by a non-zero transmit pulse duration (but differences between the transmitted pulses may still introduce a bias to the estimator). Ground elevation differences within the footprint that are non-negligible with respect to the maximum canopy height do not pose a problem for the estimation of a change in canopy height, though it will not be applicable to the *maximum canopy height above ground* but to the height of vegetation surfaces located at the highest elevation. For the sake of simplicity, the difference in waveform extent for a repeated measurement will be taken as an estimator of change in maximum canopy height.

2.3 Delineating the beginning and end of the signal

The waveform extent is defined as the elevation difference that corresponds to the time interval between the beginning and the end of the received signal. Locating the signal beginning and end is complicated by detector noise. The common approach is to take the first and last samples above a threshold value as the signal beginning and end. Duong takes the four background noise standard deviations over the mean noise level as the threshold value [10]. Alternatively, one may additionally include any signal between the first threshold-crossing and the last zero-crossing before it, as well as any signal between the last threshold-crossing and the first zero-crossing after it. Here we will illustrate the differences. Only the delineation of the signal beginning is treated, but the same holds for the signal end except for a reversal of the time-axis.

Assume a hypothetical linearly increasing impulse response (at its beginning) for a certain footprint (figure 2.1a). The signal received in response to a Gaussian pulse would be its convolution with a Gaussian. (In this example it has a width of 50 ns and integrates to one). The convolution smooths the beginning of the impulse response, which is located at exactly 500 ns in the figure. Furthermore, detector noise is added to the original noise-free signal. The noise is taken from an actual waveform far before any sign of a signal and its mean is set to zero. In this example, we use a threshold of five background noise standard deviations. The first threshold-crossing and the last zero-crossing before it are both indicated in the example in figure 2.1a, as well as the true signal beginning without the smoothing effect of the convolution.

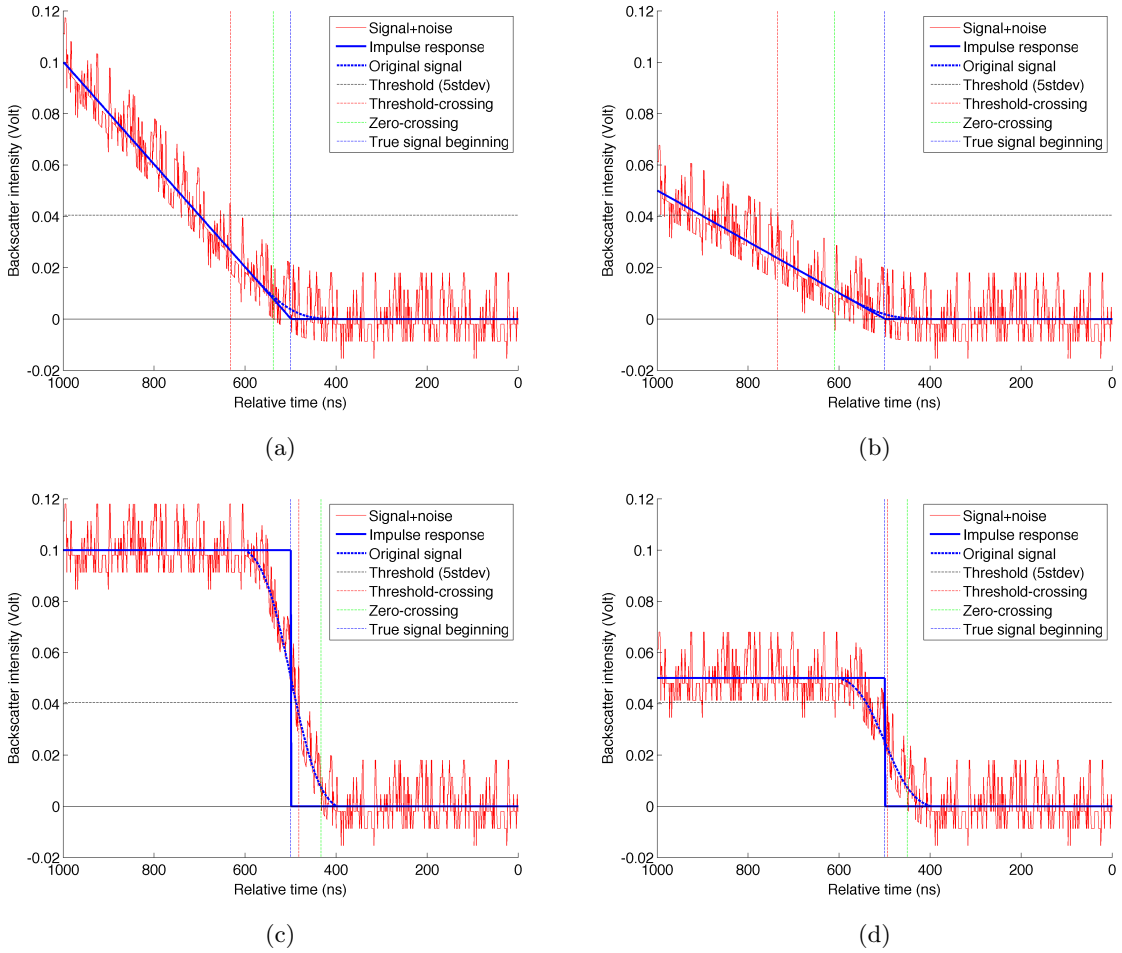


Figure 2.1: To four different hypothetical signals, which are convolutions of an impulse response with a Gaussian transmit pulse, detector noise is added. Two estimators of the signal beginning are tested.

We assume the noise is stationary (the mean of the noise does not vary in time) and independent of the signal amplitude. The zero-crossing approach more accurately estimates the true beginning of the signal than the threshold-crossing approach would. Concerning sensitivity to a change in the signal-to-noise ratio: both estimators depend directly on the noise level compared to the slope of the true signal. If the true signal is halved due to laser degradation, but the noise level remains unchanged (figure 2.1b), both estimators are located at a later point in time and their error in estimating the true signal beginning is approximately doubled. So, since the zero-crossing estimator of the true signal beginning has a smaller error, it is also less sensitive to the change in the signal-to-noise ratio.

Should the signal not increase linearly, the reproducibility of both estimators depend on the slopes of the true signal near the respective estimators. Had the slope been higher near the threshold-crossing estimator than near the zero-crossing estimator, it may have been less sensitive to a change in signal-to-noise ratio than the other. Therefore, if the impulse response increases abruptly and the true signal is strongly influenced by the convolution with a Gaussian, as in figures 2.1c and 2.1d, the threshold-crossing estimator might yield a higher reproducibility. The slope of a Gaussian curve first increases and then decreases, so the estimator providing better reproducibility for a Gaussian-convoluted signal with significant width depends on the local signal amplitude compared to the noise standard deviation (signal-to-noise ratio). If the local signal amplitude is large compared to the threshold level, the slope will monotonically increase up to the threshold-crossing and the threshold-crossing estimator may be less sensitive to changes in the signal-to-noise ratio than the zero-crossing estimator. Such may be the case for a Gaussian return from flat terrain without vegetation. Yet the signal returned from a forest will not be Gaussian, and as we might expect that the canopy closure steadily decreases in the uppermost part of the canopy, the return signal would be steadily increasing at its beginning.

To conclude, the zero-crossing approach more accurately estimates the beginning of the signal and, depending on the shape of the signal, may also be less sensitive to change in the signal-to-noise ratio. A complete mathematical treatment of the problem is outside of the scope of this project.

Apart from changes in the total energy of the emitted laser pulse due to laser degradation, misalignment of two footprints within a pair may also result in significant differences in the horizontal distribution of laser intensity. The tallest tree within the first footprint may contribute to a smaller degree to the second received waveform when it is further from the footprint center, where laser intensity reaches its maximum. Thus, the signal-to-noise ratio for specifically the beginning part of the waveform (corresponding to the largest elevations) will be lower.

Chapter 3

Methods

This chapter will describe the experimental method developed to use GLAS data for the analysis of changes in forest structure. The chapter is divided into two parts: preparations and experiment. The first part covers practical aspects of working with GLAS data: choosing a data product and format, ordering subsets, reading binary data and waveform pre-processing. The second part presents the experiments performed with the data. A reader that is only interested in the geo-scientific aspects may skip the first part of this chapter. Any scientifically relevant details mentioned in the first part will be repeated and treated more extensively in the second part of the chapter.

3.1 Preparations

3.1.1 Data acquisition

GLAS data products

The ICESat mission has resulted in 15 GLAS data products, GLA01-GLA15. Two of these are particularly relevant for measurements of forests. The raw transmitted and received waveforms are stored only in the data product GLA01 [39]. GLAH14 [40] contains parameters that are relevant for the analysis of land measurements. It contains a number of parameters that are important for forest structure assessment, of which the coordinates of the footprint centers are the most fundamental. GLA05 might be considered as an alternative for GLAH14 - it contains many of the same parameters. However, we will be using Shuttle Radar Topography Mission (SRTM) elevation values that have only been included in GLAH14. The parameters of the product GLAH14 are gathered in table 3.1 and the parameters of product GLA01 in table 3.2.

parameter name	**unit**	**description**	**application in this project**
d_maxSmAmp	Volts	peak amplitude of received echo after smoothing to remove high frequency noise	signal-to-noise ratio used during selection of high-quality measurements, lower limit set to 40
d_sDevNsOb1	Volts	standard deviation of the background noise for shots over land	
i_satNdx	gates	number of gates above the detector saturation threshold	removal of saturated waveforms; maximum allowed value of 2 gates
d_DEMhiresArElv	meters	DEM values (SRTM) at footprint center and to its N, NW, W, SW, S, SE, E and NE	total range defined as the terrain index; maximum allowed terrain index is 30 m
d_elev	meters	land surface elevation	used for removal of shots intercepted by clouds; maximum allowed absolute difference is 100 m
d_DEM_elv	meters	SRTM elevation at footprint center	
d_lat	°N	latitude coordinate of the footprint center	geolocation of footprints; footprint center distance calculations for measurement pair forming
d_lon	°E	longitude coordinate of the footprint center	
DS_UTCTime_40	seconds	laser shot transmit time in seconds after 01-Jan-2000 12:00:00 UTC	date and time of the measurement
i_rec_ndx	n/a	the record number; a record contains 40 shots	together a unique measurement identification, used to search for the corresponding waveforms
i_shot_count	n/a	the shot number, ranging from 1 to 40	

Table 3.1: GLAH14 parameters used in the project [40].

parameter name	unit	description	application in this project
i_rec_ndx	n/a	the record number; a record contains 40 shots	used to identify the records containing the waveforms that are searched for
i_gla01_rectype	n/a	type of the (sub)record; can be MAIN, LONG, or SHORT	determines how the record should be interpreted
i_InstState	n/a	describes several aspects of the instrument state	indicates which digitizer was used and thus which count-to-volt table should be used
i_compRatio	unitless	compression ratios p and q	waveform decompression
i_N_val	gates	after this sample, the compression ratio switches from p to q	
i_r_val	unitless	compression ratio for all samples	
i_4nsBgMean	0.01 counts	mean value of the background noise	subtracted from the waveform after conversion to volts
i_4nsBgSDEV	0.01 counts	standard deviation of the background noise	the threshold that determines the beginning and end of the signal is set at 5 times this value
i_comp_type	n/a	compression type, either Npq or R	indicates which compression ratios should be used for waveform decompression
i_rng_wf	counts	the received 1064 nm waveform, digitized in 1000 gates and then compressed to 544 samples (for land)	waveform extent calculations; waveform inspection for case studies

Table 3.2: GLA01 parameters used in the project [39].

Formats and ordering subsets

GLAS data is distributed by the NSIDC. Spatial/temporal subsets of these data products can be ordered via the NSIDC website [27]. The subsetting of GLAS data at the NSIDC has some restrictions that complicate the process. The data are distributed in binary format and in an easy-to-read HDF5 format (Matlab contains HDF reading functions; the HDF variants of the data products are denoted GLA01-GLA15 and are directly based on their respective binary source products GLA01-GLA15). While the GLAH14 data products can be subsetted and ordered via NASA REVERB (an online earth science data discovery tool [28]), GLA01 can not be spatially subsetted with this data tool. The option has been disabled since the footprint center coordinates in this data product are rough estimates based on ICESat's orbit.

Unlike the NASA REVERB tool, the NSIDC binary data subsetter is capable of handling spatial subset requests of the GLA01 product, though it does not support HDF5-data. Therefore, we have chosen to work with a combination of HDF-format GLAH14 and binary format GLA01. For spatial subsetting we have chosen a boundary box with corner coordinates (53.5 °N, 120 °W) and (60 °N, 110 °W). Neither of the ways to order data support geometries more complicated than boundary boxes. Temporal subsetting was only possible for GLAH14 and the annually repeating dates option has been used to select only late-May to late-June campaigns (L2C, L3C and L3F, see table 3.3; for brevity referred to a summer campaigns from this point on).

3.1.2 Waveform extraction from binary GLA01 files

The NSIDC provides IDL and Fortran tools to read the binary data and export the parameters to an ASCII file which might be read into any computing environment. But, when working with large subsets, the ASCII files are too inefficient in storage. The file sizes become unmanageable for a typical desktop computer (process aborted at a file size of 4GB for one campaign). Therefore, we have opted for the most direct route and read the data directly from the binary subset files (one per campaign) into Matlab. The binary-GLA01 reading program has been written based on the record format descriptions in the GLAS standard data products specification, level 1 version 9, which states how the bytes should be interpreted. The program also directly applies some pre-processing manipulations. For these parts of the program, the NSIDC Fortran reader has been an example (in particular one of its source files: [29]). It was written by the GLAS Science Algorithm Software (GSAS) Development Team.

Once the GLAH14 dataset has been used to select useful measurements, the identifiers of these measurements (record index ± shot ID, which are the input variables) are used to search the binary GLA01 subset files for these measurements. The waveform-reading program works through the following steps to read the corresponding waveforms (the output):

- Make a list of relevant records by finding all unique entries in the list of record indices.
- After skipping header records, read first bytes of the record which give the record index.

- If the record index read from the current record is not within the list of relevant records, skip to the next record and try again.
- Once a relevant record has been encountered, we read the record type.
- If the record type is MAIN, we read and temporarily store compression parameters and digitizer number.
- If the record type is LONG, we read the 8 compressed waveforms of 544 samples each, the respective noise parameters and the compression type.
- If the record type is SHORT, we read the 20 compressed waveforms of 200 samples each, the respective noise parameters and the compression type.
- Waveform pre-processing is performed: decompression depending on compression parameters and type, conversion to volts depending on digitizer number, adjustment for noise depending on noise parameters (mean and standard deviation).
- The waveforms are stored in such a way that each has an index matching the indices of its respective identifiers.
- When all useful data from a record has been read, the program skips to the next record and resumes the search for the next relevant record.

Skipping to the next record is possible because each record has a fixed length of 4660 bytes. Thus, the number of bytes to skip in order to arrive at the start of the next record can be calculated. Each MAIN record is followed by either 5 LONG records or 2 SHORT records (subrecords) with the same record index. The LONG records are used for measurements over land and we have encountered no SHORT records in the data used for this study (SHORT records are used over oceans). Neither have we encountered any use of the second digitizer.

3.1.3 Waveform pre-processing

The waveforms in GLA01 are stored in digitizer counts. Furthermore, they are time-compressed. Before proceeding, it is essential to decompress the signal and convert the values from digitizer counts to volts. These steps are performed directly during the waveform extraction process as to avoid storing large arrays of compression parameters. With the same logic, a third pre-processing step is performed during waveform extraction: adjustments for noise based on noise parameters in the GLA01 records.

Decompression

Two types of compression are applied on board the satellite for GLAS waveforms: *Npq*-compression and *R*-compression. For *Npq*-compression, the first N compressed samples have been generated by averaging over p subsequent gates. The remaining compressed samples (N+1 to last sample) are formed by averaging over q subsequent gates. Following the example of the Fortran reader, we decompress the signal by repeating each of the first N compressed samples p times and repeating each of the remaining samples q times. In Matlab, the most efficient way we have found to perform this manipulation is by using the Kronecker tensor product. If A is an mxn matrix, then the Kronecker tensor product of matrices A and B is given by:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix} \tag{3.1}$$

Taking advantage of this efficient built-in Matlab operation, we define A as an $N \times k$ matrix of the first N samples of k compressed waveforms and B as a $p \times 1$ matrix of ones to find the first $N \cdot p$ decompressed waveform samples $A \otimes B$. In a similar manner we find the remaining $1000 - N$ samples of the decompressed waveform by redefining A as an $(M - N) \times k$ matrix of the remaining samples of the length M compressed waveforms and B as a $q \times 1$ matrix of ones.

Whereas *Npq*-compression involved two compression ratios p and q, *R*-compression uses only one. The compression ratio R is applied over the entire waveform. Thus, we decompress an *R*-compressed waveform by calculating the Kronecker tensor product of an $M \times k$ matrix A of the compressed waveform samples, and B, an $R \times 1$ matrix of ones.

The original signal is of course not fully reproducible from the compressed signal. The decompression method above approximates the original signal and most importantly makes sure that the waveform samples are again separated by 1 ns intervals. Alternatively, one could determine the center-time for each averaged set of gates and not expand the number of samples. This approach is explained in the GLAS waveform Algorithm Theoretical Basis Document [3]. We have followed the example set by the writers of the Fortran reader program, since having a 1000 samples per waveform and a 1 ns sample interval consistently for all waveforms adds to the simplicity of further processing.

Conversion from counts to volts

The counts-to-volts conversion is a matter of reading or generating a look-up table for the voltages corresponding to each digitizer count. These look-up tables, as well as the linear digitizer calibration functions that have been used to generate them, can be found in The GLAS Science Algorithm Software (GSAS) User's Guide Version 7 [21]. There were two waveform digitizers on-board of GLAS, although all the waveforms used in this study have been digitized with digitizer 1. The look-up tables can be generated with:

$$f(y) = \begin{cases} a_1 \cdot y + b_1 & \text{if} \quad 0 \leq y \leq 127 \\ a_2 \cdot y + b_2 & \text{if} \quad 127 < y \leq 255 \end{cases} \tag{3.2}$$

where y represents a waveform value in digitizer counts and $f(y)$ is the waveform value in volts. Depending on which digitizer was used, the calibration constants are:

$$\text{digitizer 1} \begin{cases} a_1 = 0.006675 & \text{V/count} \\ a_2 = 0.006198 & \text{V/count} \\ b_1 = -0.19528 & \text{V} \\ b_2 = -0.13442 & \text{V} \end{cases} \tag{3.3}$$

$$\text{digitizer 2} \begin{cases} a_1 = 0.006625 & \text{V/count} \\ a_2 = 0.006128 & \text{V/count} \\ b_1 = -0.19383 & \text{V} \\ b_2 = -0.13044 & \text{V} \end{cases} \tag{3.4}$$

Alternatively, the tables can be copied from the GSAS users guide [21]. (Since Matlab does not support zero-indices, the counts-to-volt tables are stored at indices 1-256. Upon applying the conversion, the waveform counts are increased with 1 count.) These tables are used for the conversion of the waveforms. The use of a pre-calculated look-up table stored in temporary memory is computationally more efficient than performing a voltage calculation using the equation directly.

Adjustments for noise

The noise parameters are given in the LONG and SHORT records per waveform. These give the mean and standard deviation of the background noise (measured in between received pulses) in units of 0.01 digitizer counts. They are converted to volts using the digitizer calibration functions above (due to decimal counts, the look-up tables do not suffice). The mean of the background noise is subtracted from the decompressed waveform in volts.

For the adjusted signal, we define the signal threshold as 5 times the standard deviation of the background noise. The signal beginning is usually placed at the first gate that has a voltage larger than the signal threshold (for example, by Duong [10]). We additionally include the signal between this gate and the last zero-crossing before it. Similarly, the signal end is usually placed at the first gate that has a voltage smaller than the signal threshold. We also include the signal between this threshold-crossing and the first zero-crossing after it. All gates before the signal beginning and after the signal end are set to zero. Since the adjustments for noise directly influence any waveform-derived parameters, the reader is referred to sections 2.3 and 3.2.3 for a more extensive treatment of the concept.

3.2 Experiment

3.2.1 Data selection

Study region

For this project we have conducted experiments with GLAS measurements over northern Alberta. The study region is defined as a boundary box with corner coordinates 53.5 °N, 120 °W) and (60 °N, 110 °W (see figure 3.1). It falls largely within the Boreal Plains ecozone, described as a "nearly level to gently rolling plain" with both broadleaf and coniferous tree species [12].

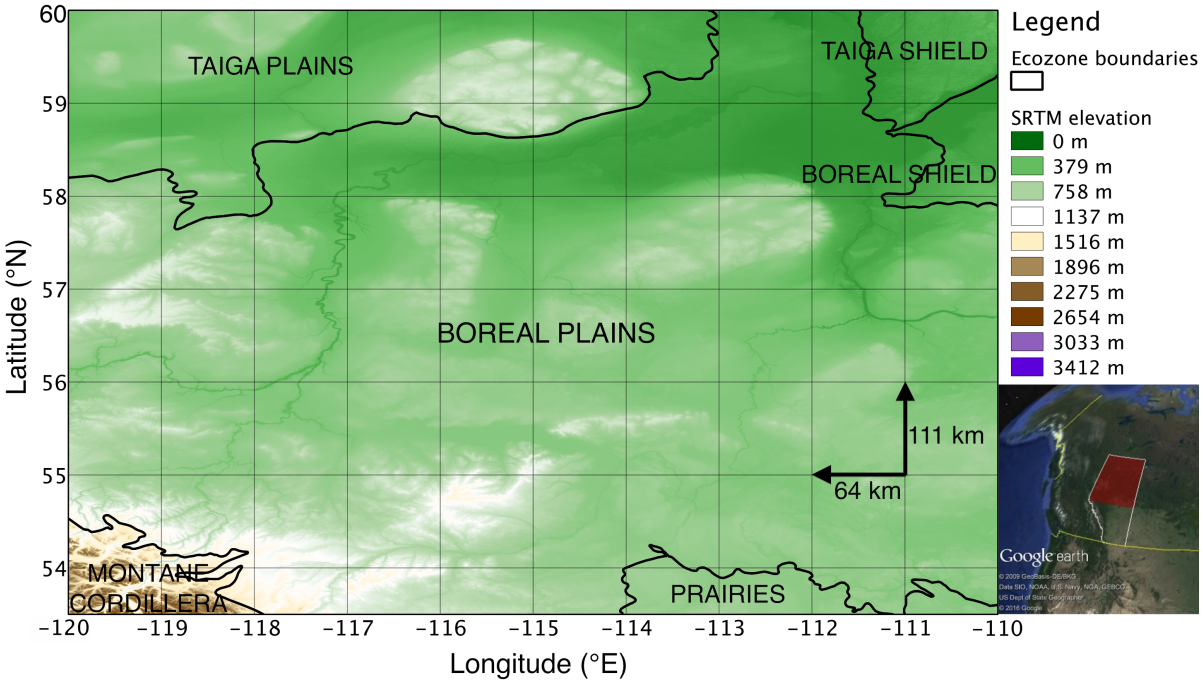


Figure 3.1: Map of the study region (northern Alberta) with SRTM elevations with respect to the Earth Gravitational Model 1996 (EGM96) vertical datum [20]. Ecozone boundaries are also indicated [6].

ICESat campaigns

Only GLAS measurements from summer campaigns are included in the analysis. By selecting summer campaigns, we avoid complications of waveform-interpretation due to possible snow cover, as well as the expected dominance of seasonal foliage changes on change measurements in (partly) deciduous forests. However, this also reduces the maximum time difference between measurements to approximately two years and only short-term changes may be investigated with this selection of campaigns.

Of the eighteen campaigns, three were conducted during summer: L2C, L3C and L3F (see table 3.3) in the years 2004, 2004 and 2006, respectively. The start and end day of the year approximately coincide for these three campaigns, as do the repeat orbit cycle and track number. The latter indicates that the roughly 1/3rd of the 91-day ground repeat track that was completed for each campaign has covered approximately the same parts of it. Reference-track pointing, assuring close proximity between measurements of different campaigns, was only enabled at the latitudes of the study region for campaign L3C and L3F. The laser power was higher during the L3C campaign than for the other two, by roughly a factor 1.5. The L2C campaign contains measurements with a low transmit energy (5 mJ), the other two do not. The laser footprints were significantly larger during the first of the summer campaigns and had a higher eccentricity, compared to the other two summer campaigns. While the latter two summer campaigns do not differ much in footprint major axis length, the eccentricity was smaller for the third summer campaign (L3F) than for the second (L3C). For example, the areas corresponding to ellipses with the mean major axis lengths and mean eccentricities as given in table 3.3 would be 11.1×10^3 m^2, 7.5×10^3 m^2 and 7.2×10^3 m^2 for campaign L2C, L3C and L3F, respectively. Horizontal geolocation accuracy is in the order of meters for all three summer campaigns.

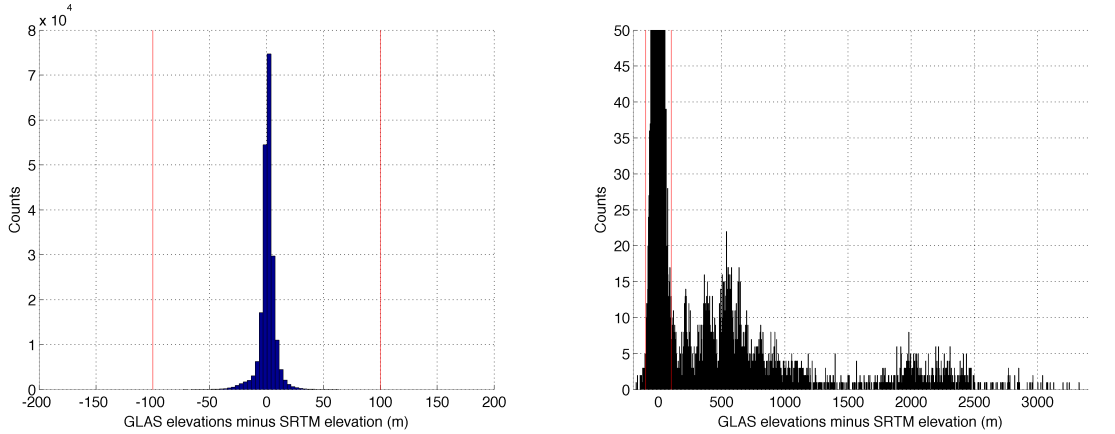
	laser 2C	laser 3C	laser 3F
start date	18-05-04	20-05-05	24-05-06
end date	21-06-04	23-06-05	26-06-06
start day of year	139	140	144
end day of year	173	174	177
number of orbit revolutions (tracks) per repeat orbit cycle	1354	1354	1354
start repeat orbit cycle and track number	002 - 1283	002-1275	002 - 1283
end repeat orbit cycle and track number	003 - 434	003 - 421	003 - 421
number of repeat orbit cycles completed	0,38	0,37	0,36
latitude range [°] for reference-track pointing (within ∼± 100 meters) (ascending)	>61.3S, >59.1N	>61.3S, >46.1N	>61.3S, >46,1N
latitude range ° for reference-track pointing (within ∼± 100 meters) (descending)	>59.1S, >61.0N	>59.1S, >51.8N	>59.1S, >51.8N
laser pule maximum to minimum transmit energy [mJ]	33 to 5	49 to 44	33 to 30
elliptical footprint major axis mean and standard deviation [m]	88.37±19.12	55.41±1.84	51.20±1.63
elliptical footprint eccentricity mean and standard deviation [m]	0.892±0.044	0.633±0.034	0.480±0.023
horizontal geolocation accuracy mean and standard deviation [m]	0.37±10.30	0.29±2.92	1.35±4.42

Table 3.3: Attributes of the campaigns involved in this study.

Removing cloud-waveforms

The 1064 nm laser pulses from GLAS do not penetrate far through clouds. As a result, some waveforms only show the intensity of photons backscattered by clouds and contain no information on the land surface beneath. These measurements are therefore not relevant for this study and need to be identified and removed.

There is a cloud flag included in the GLAS data products (GLAH14 parameter *elv_cloud_flg*). However, Molijn et al. noted that using the cloud flag at the time rejected too many useful measurements [25]. Therefore, we will use a threshold value for the absolute difference between GLAS and SRTM elevations from GLAH14 to filter out cloud-waveforms. The minimum value for a waveform to be included in the analysis is set at 100 m. Small elevation differences are expected for forested regions. A threshold that is too low would result in excluding important forest data. A threshold set too high would include waveforms that did not penetrate the clouds to reach the land surface or vegetation layers. See figure 3.2a and 3.2b.



(a) The main peak.

(b) Axes readjusted to highlight extreme values.

Figure 3.2: Histogram of GLAS elevation minus SRTM elevation.

Removing noisy waveforms

To accurately determine the waveform extent a high signal-to-noise ratio (SNR) is required. The SNR is here quantified as the amplitude of the maximum peak of the waveform after removal of high frequency noise (GLAH14 parameter *d_maxSmAmp*, [40]) divided by the standard deviation of the background noise (GLAH14 parameter *d_sDevNsOb1*, [40]). The minimum value for a waveform to be included in the analysis is set at 40.

Removing saturated waveforms

Saturation of the waveform, causing a peak to be flattened at the maximum detectable backscatter intensity (which depends on the gain setting) and oscillations after the peak [1], will affect any parameters derived from the waveform. With a relatively complicated approach, one might still be able to estimate some of the parameters used in this study from saturated waveforms, but this does not fall within the scope of this research project. Saturated waveforms, identified by saturation indices higher than 2, are not included in the analysis. Examples of saturated waveforms are given in Duong [10] for example.

In-footprint relief

Any reflective surface within the laser footprint contributes to the resulting waveform. Vegetation can provide reflective surfaces with varied heights, thus increasing the extent of the waveform, but in-footprint relief may also contribute to the waveform extent. In the case of flat or gently and steadily sloping terrain, the ground-return may still be recognized as a separate unimodal peak approximating a Gaussian distribution. We will apply a strict requirement on terrain flatness to minimize misinterpretation of the return signal.

In-footprint relief is most readily estimated from the arrays of high-resolution DEM values that are included in the GLAH14 dataset (parameter *d_DEMhiresArElv*). The 3×3 arrays give the DEM values at the footprint center and at positions to the NW, N, NE, W, E, SW, S and SE of the footprint center. The DEM source is the SRTM C-band 90 m DEM by the National Geospatial-Intelligence Agency (NGA). From the elevation array we will take the highest elevation minus the lowest elevation as the terrain index (TI). The maximum allowed terrain index is set at 30 m.

3.2.2 Forming waveform pairs

After the data selection process, the remaining measurements are searched for two-measurement combinations that yield a footprint center distance smaller than 150 m. The search is performed over tiles of approximately 1° x 1° separately to lower the amount of great-circle distance evaluations. For each tile, the footprint center distances are evaluated for all possible combinations of two measurements out of all measurements within the tile. The tiles overlap by at least 150 m so that valid pairs at the tile borders are not overlooked. After merging the pairs found in different tiles, identical pairs, present due to the overlap, are identified using the record index and shot number of both measurements in the pair. The copies are removed.

The resulting set of measurement pairs includes:

- repeat-track pairs of measurements from different campaigns
- crossover pairs of measurements from different campaigns
- crossover pairs of measurements from the same campaign

Repeat-track pairs of measurements from the same campaign are not found within the dataset because none of the campaigns have covered a full cycle of the 91-day ground repeat track. Pairs of measurements from different campaigns are separated by a time interval of either approximately one year (2004 to 2005 or 2005 to 2006) or approximately two years (2004 to 2006). A single measurement may be used in more than one pair. The location of the pair is set between the two measurements, at the averaged latitude and longitude, which closely approximates the midpoint of the great-circle arc between the two points for small distances such small distances (<150 m).

3.2.3 Waveform extent and delineating the beginning and end of the signal

The received waveforms corresponding to the measurement pairs are read from the GLA01 data product and pre-processed as discussed in section 3.1.2. The resulting waveforms all consist of 1000 samples corresponding to the 1000 gates of 1 ns width each. The extent of the waveform, the elevation difference corresponding to the duration of the whole backscatter signal, depends entirely on the delineation of the beginning and end of the signal. After subtracting the mean of the background noise (in volts) from the waveform, parts of the waveform consisting of only noise and no signal average to and fluctuate about zero. A common approach to locating the beginning and end of the signal uses a threshold value and finds the first and last crossings of the threshold. All samples before/after the first/last crossing are interpreted as noise. For the adjusted waveform, we define the signal threshold as 5 standard deviations of the background noise. This was previously set to 4 standard deviation as suggested by Duong [10], but with this threshold we still encountered crossings which were unlikely to reflect any presence of vegetation: one single gate crossing the threshold far above any other signal-over-threshold. So the signal threshold was increased.

Now the common method is to set all gates before/after the first/last threshold-crossing to zero [10]. Then the waveform extent is determined directly by the location of these threshold-crossings. We propose a slight modification to this method by moving the signal beginning from the first threshold-crossing to the last zero-crossing before it. Similarly, the signal end is moved from the last threshold-crossing to the first zero-crossing after it. All gates before/after the adjusted signal beginning/end are then set to zero. Both the original method and the modified method are implemented for comparison, using 5 standard deviations of the background noise as the threshold value in both cases.

The relative elevations corresponding to the waveform samples are calculated using equation 2.2 (the Δ to indicate differences is dropped here):

$$z = -\frac{c \cdot t}{2} \tag{3.5}$$

where t is the relative time of the sample and ranges from -999 ns for the first gate to 0 ns for the last gate. The waveform extent is then equal to

$$E = z(t_{begin}) - z(t_{end}) \tag{3.6}$$

where t_{begin} is the relative time of the first non-zero gate and t_{end} is the relative time of the last non-zero gate.

3.2.4 Wildfires and the GLAS measurement pairs

Wildfire perimeters are acquired in ESRI shapefile vector format, using the Canada Lambert Conformal Conic projection referenced to the North American Datum 1983 (NAD83) [5]. The NAD83 geodetic coordinates are acquired by inverting the projection using the GDAL tool *ogr2ogr* [15]. The transformation between NAD83 and the Topex/Poseidon (T/P) ellipsoid is skipped and the potential effect on the coordinates is held to be negligible compared to the GLAS footprint sizes and geolocation accuracy.

Using the wildfire perimeters, each GLAS measurement pair is provided with a date of most recent wildfire event at its location before the second laser shot, assuming no unreported wildfires. Pairs that are not located within the perimeters of any of the wildfires before the first laser shot are provided with the date of the first wildfire occurrence after the second laser shot. The wildfire dates usually represent starting dates. In some cases, the reported date represents the date of discovery.

We will analyze the relation between changes in waveform parameters and wildfire events and perform case studies over a selection of measurements. LandSat optical satellite imagery will be used for closer inspection a number of wildfires sampled by GLAS measurement pairs.

3.2.5 A moving average filter for GLAS measurement pairs

To map the spatial variation of local mean changes in waveform extent, a moving average filter is applied to the pairs. A test is performed for waveform extent changes from summer 2004 to summer 2005. The goal is to replace each change measurement i with the average change of 101 measurements (101 waveform pairs):

$$\Delta extent(i) = \frac{1}{101} \sum_{i-50}^{i+50} (extent_{05}(i) - extent_{04}(i)) \tag{3.7}$$

where i indicates the number of the pair in a list of pairs sorted on the transmit time of the first shot (summer 2004). This is implemented with a moving average filter with a window size of 101 pairs. The filter is applied over pieces of the ground track separately. When two measurement pairs i and *i+1* differ in transmit time of the 2004-shot by more than 2.5 seconds, pair *i+1* is said to start a new piece of ground track. Since GLAS transmitted 40 pulses per second and the footprints are separated by about 172 m, 100 pulses are emitted during 2.5 seconds, covering approximately 17.2 km. Of each piece of ground track, the first and the last 50 pairs are removed from the filtered dataset, since they do not yield averages of 101 pairs but less.

Chapter 4

Results

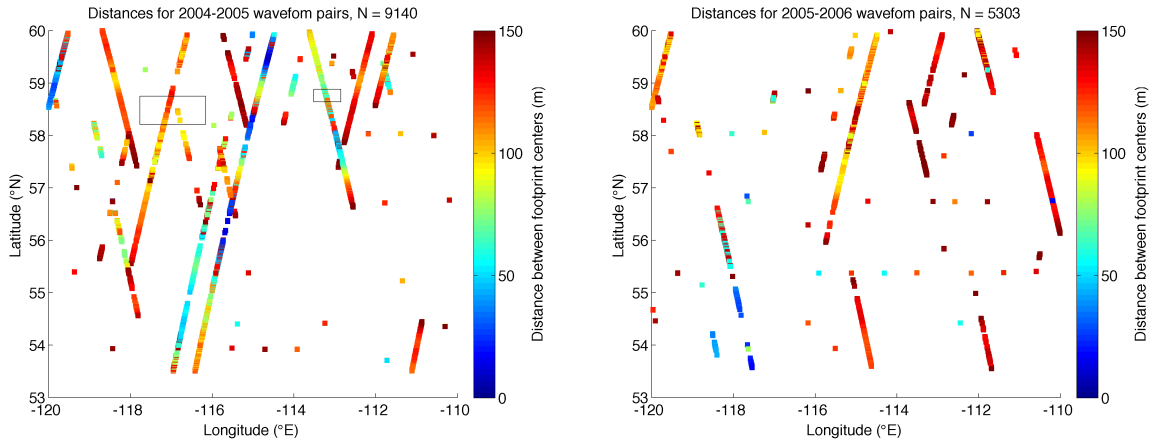
In this chapter we present the results of applying the methods described in the previous chapter. In advance on the following chapter, interpretations of the results are discussed intermittently.

4.1 Spatial distribution of high-quality waveform pairs

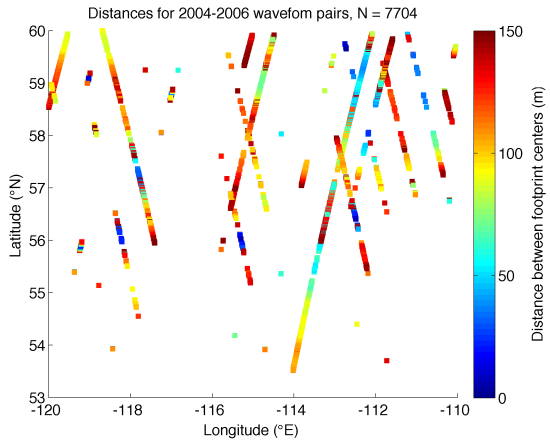
From the three summer campaigns (L2C, L3C and L3F), a total of 22309 waveform pairs of high-quality measurements with a distance smaller than 150 m between the footprint centers were identified for northern Alberta. Here, high-quality measurements refer to the fulfillment of the requirements regarding signal-to-noise ratio, saturation, terrain index and agreement with SRTM elevations described in the previous chapter. Figure 4.1 shows the spatial distribution of pairs for different measurement time intervals and the respective total numbers of measurement pairs. The distances between footprint centers are indicated by color, except for measurement intervals smaller than 27 days (figure 4.1d). For these pairs, the first and the second measurement were performed during the same campaign. Many of the squares represent multiple pairs (typically three) which are indistinguishable at this scale. These same-campaign measurement pairs with a footprint center distance smaller than 150 m are relatively rare.

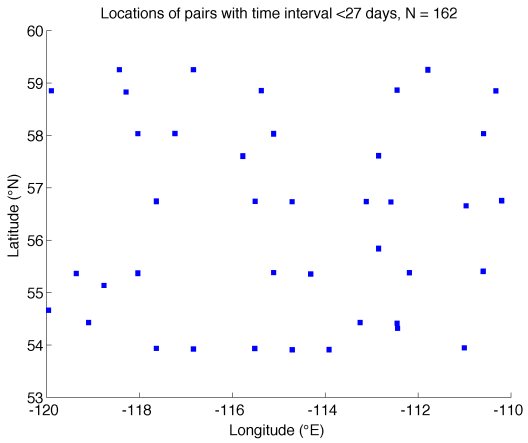
The footprint center distances for interannual measurement pairs are spatially correlated. At some parts of the tracks, these distances alternate between two distinct values as a footprint from one campaign may be within 150 m reach of two footprints from another campaign. Despite the activation of reference track pointing for the latitudes considered in this study during the 2005 and the 2006 campaigns (table 3.3), the combination of these latter two summer campaigns yielded fewer high-quality waveform pairs than the other two interannual combinations (2004-2005 and 2004-2006).

Figure 4.1: Spatial distribution of measurement pairs with a maximum footprint center distance of 150 m. Footprint center distances are indicated in color for graphs (a)-(c). *(Continued on the next page.)*



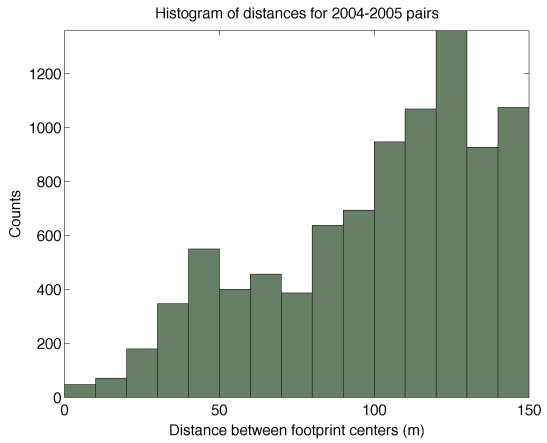
(a) First shot in 2004 (L2C), second shot in 2005 (L3C). (b) First shot in 2005 (L3C), second shot in 2006 (L3F).

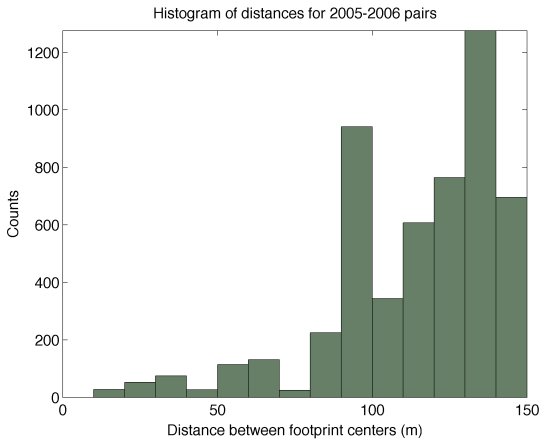




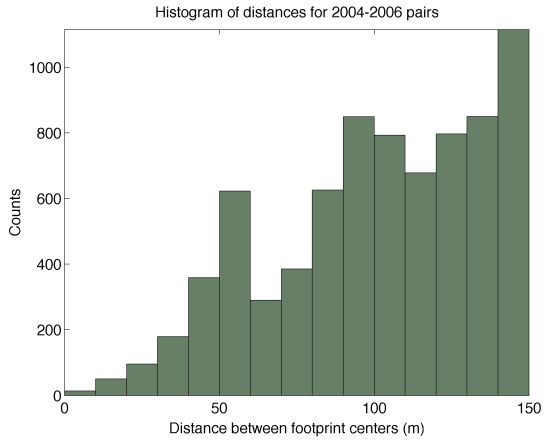
(c) First shot in 2004 (L2C), second shot in 2006 (L3F). (d) First and second shot in the same year/campaign.

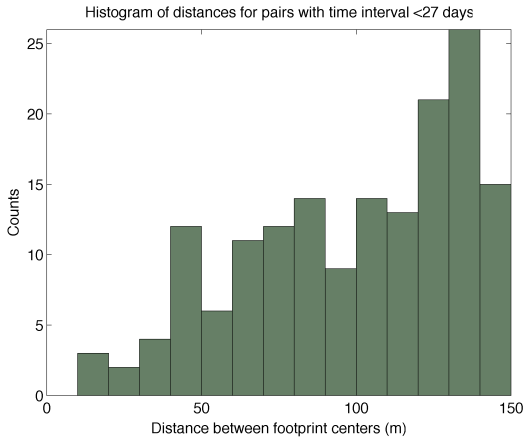
Upon inspection of the histograms of footprint center distances, shown in figure 4.2, we find that the 2005-2006 measurement interval also yields a remarkably low fraction of pairs with a footprint center distance below 90 m, compared to the other measurement time intervals. The abundance of measurement pairs generally increases with increasing footprint center distance.





(a) First shot in 2004 (L2C), second shot in 2005 (L3C). (b) First shot in 2005 (L3C), second shot in 2006 (L3F).





(c) First shot in 2004 (L2C), second shot in 2006 (L3F). (d) First and second shot in the same year/campaign.

Figure 4.2: Histograms for the distance between the footprint center of the first and the second shot within a measurement pair.

Figure 4.3 shows the distribution of high-quality waveform pairs over latitudes between 53.5°N and 60°N, at a resolution of 0.5 degrees latitude. The number of pairs over the full 10 degrees longitude, per 0.5 degrees latitude, have been divided by 5 to arrive at the average number of pairs per square degree. For the interannual measurement pairs, the number of pairs per square degree generally increases with increasing latitude. This is not observed for the distribution of same-campaign waveform pairs across the same latitudes. We point out that the sets of waveform pairs corresponding to different time intervals differ also in spatial sampling.

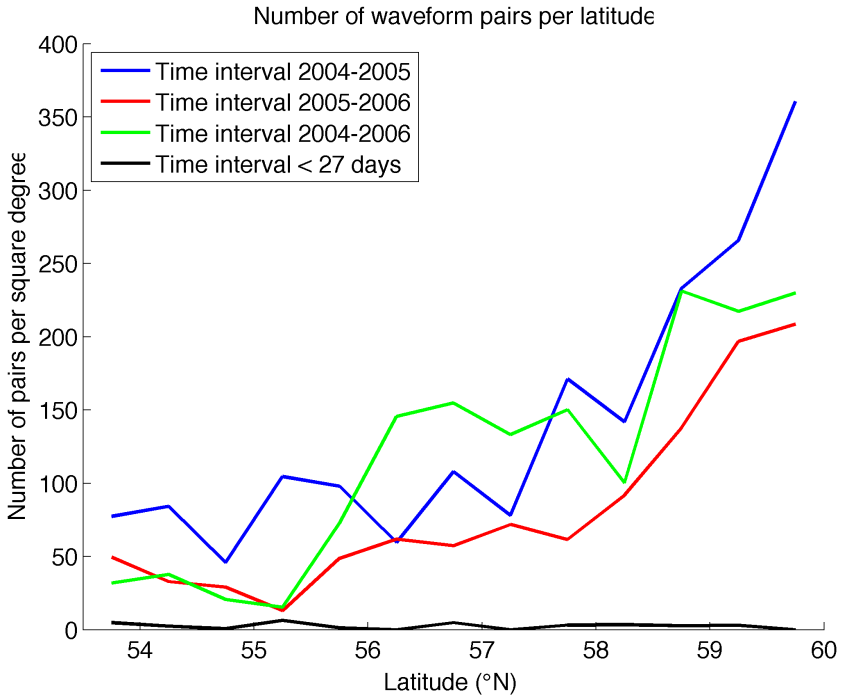


Figure 4.3: Distribution of the selected waveform pairs over latitudes, in pairs per square degree, averaged over 10 degrees longitude (110 – 120°W).

Author's interpretation

Interruption of the tracks of interannual measurement pairs may be due to the maximum allowed footprint center distance, but may also be a result of individual shots not fulfilling the requirements for high-quality measurements. When the footprint center distance steadily increases along the track until it reaches 150 m just before an interruption, the former cause is especially plausible. Where the footprint center distances are well below 150 m just before an interruption of the track, the latter case may be more likely. Some tracks may only yield crossover track footprint pairs if the repeat track footprint pairs do not fulfill the requirements concerning measurement quality or footprint center distance. Due to the 91-day ground repeat track for the summer campaigns, pairs with a time interval smaller than 27 days (figure 4.1d) are all found at track crossover locations. Each of the summer campaigns covers approximately a third of the ground repeat track (table 3.3).
Since the distributions of pairs over latitudes (figure 4.3) are expressed in pairs per square degree, the observed increases in measurement pair abundance towards higher latitudes (for interannual pairs) are *not* due to the convergence of ground tracks but a result of the high-quality data selection process. Since the surface area (in km^2 for instance) of a square degree decreases from the equator towards the poles, the relative abundance of high-latitude pairs in units of pairs per km^2 would be higher than it is here in units of pairs per square degree, and *would* include the effect of decreasing ground track spacing towards the poles.

4.2 Reproducibility of waveform extent measurements

For the two approaches (described in chapters 2 and 3) to delineating the waveform extent, the correlation between the first and second measurements within the pairs are compared in figure 4.4. Once more, a distinction between different measurement time intervals. A number of observations can be made from the graph:

- The correlation coefficients are all above 0.7
- The correlation is highest for the smallest measurement time interval
- The correlation is lowest for the largest (two-year) measurement time interval
- The correlation is lower for the one-year interval from 2005 to 2006 than for the one-year interval from 2004-2005
- The zero-crossing approach yields higher correlations than the threshold-crossing approach for all time intervals
- The zero-crossing approach yields correlation coefficients above 0.75 for all time intervals.

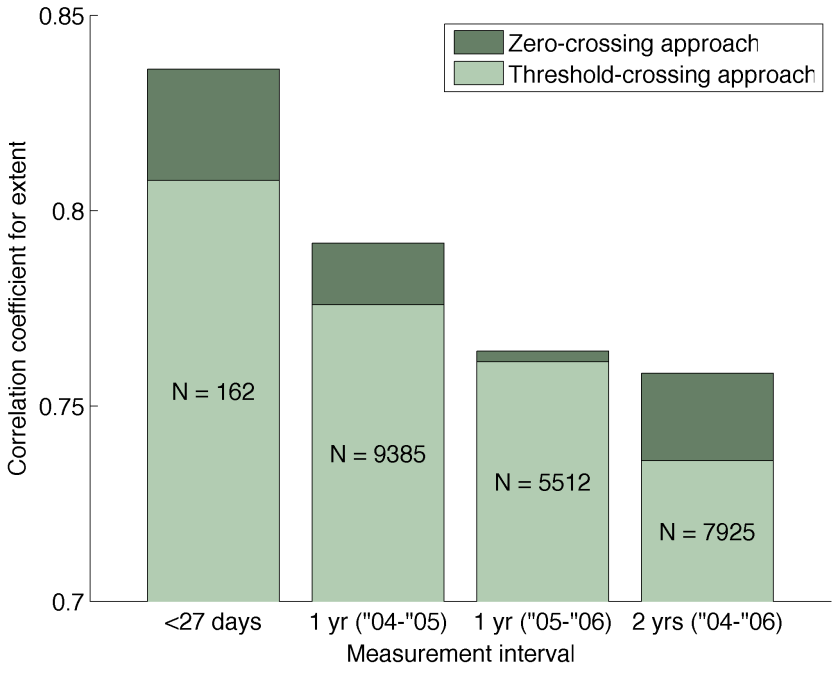
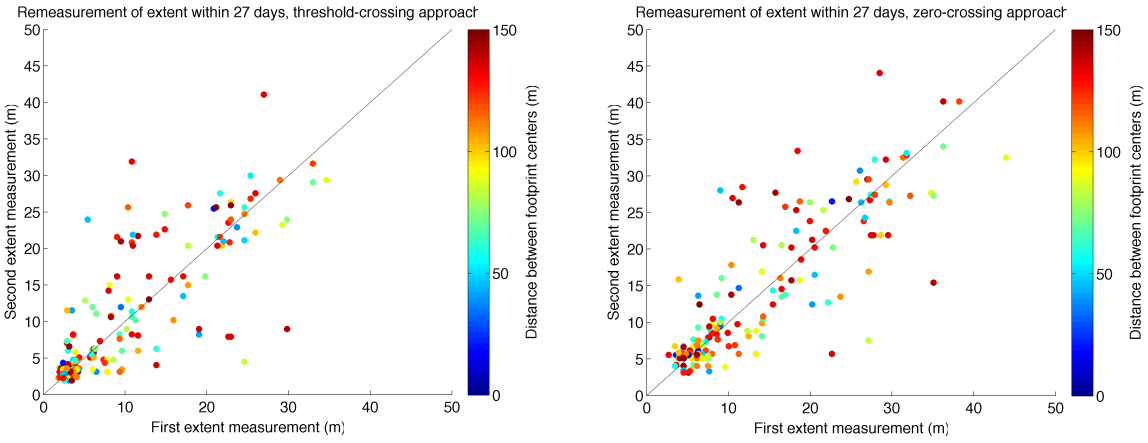


Figure 4.4: Correlation between the first and second waveform extent measurements of the selected waveform pairs, for different time intervals. Two approaches to delineating the waveform extent are compared.

Author's interpretation

The higher correlations for the zero-crossing approach compared to the threshold-crossing approach suggest that waveform extents estimated with the former approach are affected less by changes in laser power and the horizontal distribution of laser energy (including a shift of the footprint center). The correlations for time intervals smaller than 27 days are especially strong evidence for this, as terrain and vegetation canopy height are generally unlikely to change much within 27 days making differences in laser pulse power and horizontal distribution the main error source in reproducing the measurement.
Lower correlations for larger time intervals (figure 4.4) are expected due to canopy height changes in between measurements, but we cannot exclude the possibility that changes in the measurement instrument (e.g. laser power and width of the transmit pulse) have led to (or contributed to) these differences in correlation. If we can assume there were no significant instrumental changes during individual campaigns (a stable laser power and transmit pulse shape on time scales smaller than 27 days), interannual instrumental changes cause biases which may be nearly constant for each combination of two campaigns. For instance, if the transmit pulse FWHM was 4 ns during the entire summer-2004 campaign and 8 ns during the entire summer-2005 campaign, the waveform extent measurement from 2005 will be biased with respect to those from 2004 and the bias will be approximately constant (depending somewhat on the shape of the impulse response). When the bias is constant, it does not affect correlation.
The lower correlations for 2005-to-2006 pairs compared to 2004-to-2005 pairs are likely a result of the generally higher footprint center distances for the former (compare figures 4.2a and 4.2b).

The reproducibility of waveform extent measurements within 27 days is inspected more closely in figure 4.5. The colors indicate the distances between footprint centers. The zero-crossing approach yields higher values for waveform extent than the threshold-approach does. For the threshold-approach, we find a cluster of points at the lower-end of the range of extent values. No particularly striking improvement in reproducibility — the agreement between the measurement and remeasurement — is seen when comparing small-distance remeasurements with larger-distance remeasurements.



(a) Threshold-crossing approach.

(b) Zero-crossing approach.

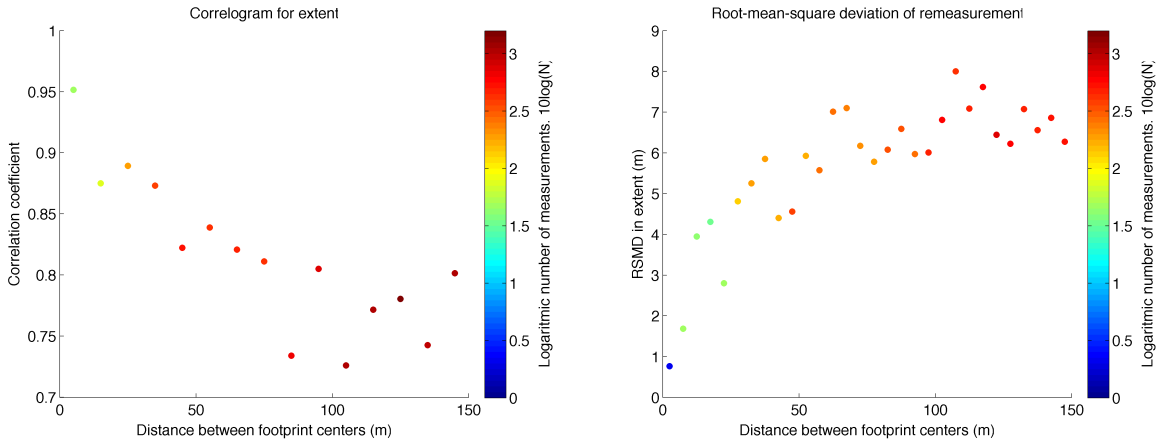
Figure 4.5: The waveform extent of the second shot within a pair plotted against that of the first shot for all same-campaign pairs with a maximum footprint center distance of 150 m. Two approaches to delineating the waveform extent are compared. The dashed lines indicate the 1:1 ratios.

Author's interpretation

There are significant errors in remeasurement (figure 4.5). This is to be expected since pairs have a footprint center distance of up to 150 m. According to the simplified description of a footprint being circular with a diameter of about 70 m, some of the footprint pairs have no overlap. Still the first and second measurement are correlated and the errors of remeasurement are random errors, not systematic biases.

We continue with the results from the zero-crossing approach (unless specifically indicated otherwise). To inspect the relation between correlation in waveform extent and the footprint center distance, we include pairs with a time interval of summer 2004 to summer 2005 to increase the amount of pairs with respect to the set of same-campaign pairs. The resulting correlogram is shown in figure 4.6a. The correlation coefficient is above 0.72 for all distances smaller than 150 m. Overall, the correlation coefficient increases as distance decreases. This is particularly clear for distances smaller than 80 m. For distances between 0 and 10 m, the correlation coefficient exceeds 0.95. For distances between 80 and 150 m the coefficient fluctuates between 0.73 and 0.81. The colors signify the amount of measurements per 10-meter-wide bin on a logarithmic scale. These range from 47 pairs (with distances between 0 and 10 m) to 1417 pairs.

For the same set of pairs, the root-mean-square deviation (RMSD) of remeasurement is shown as a function of footprint center distance in figure 4.6b (equal to the square-root of the variogram.) The distance-bin width (equal to twice the lag tolerance) is set lower (bins of 5 m width) than for the correlogram so that the RMSD of pairs with a small footprint center distance can be inspected more specifically. The RMSD generally increases with footprint center distance, reaching about 7 m for distances between 100 and 150 m. The smallest RMSD values, 0.8 m and 1.7 m, are found for distances from 0 to 5 m and 5 to 10 m, respectively, although the former is based on only two pairs. The latter is based on 45 pairs.



(a) Correlogram.

(b) Square-root of the variogram.

Figure 4.6: The correlogram (lag tolerance 5 m) and square-root of the variogram (lag tolerance 2.5 m) for waveform extent measurements. One-year interval pairs (2004-2005) have been included to increase the number of pairs.

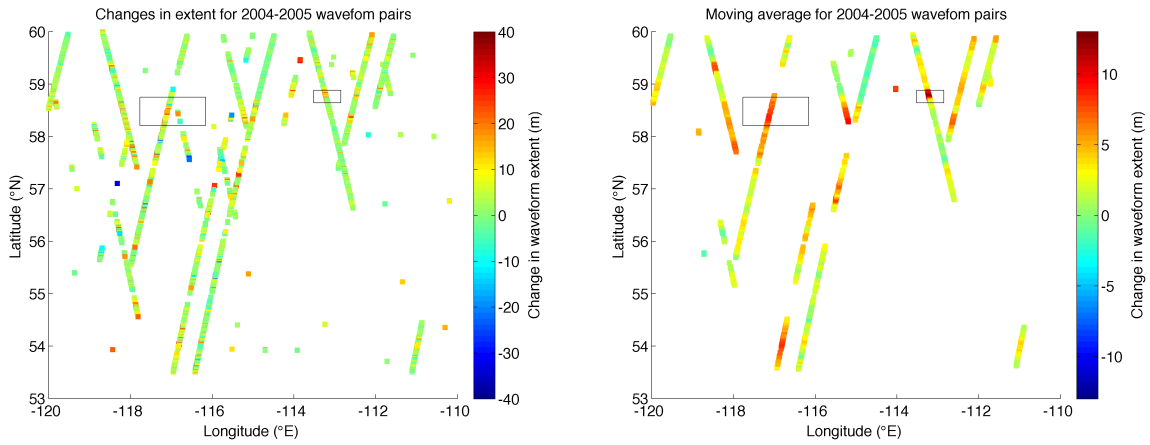
Author's interpretation

The correlation coefficient and RMSD are dependent on the distance between footprint centers (figures 4.6a and 4.6b). The fact that these do not respectively decrease and increase *monotonically* with increasing distance implies that there are factors other than footprint center distance that contribute to errors in remeasurement. One is the horizontal distribution of laser energy around the footprint center. Another is the shape of the transmitted waveform. It is also important to remember that the majority of the pairs span a measurement interval of about one year, so any changes in the vertical distribution of reflective surfaces in the footprints (for instance due to vegetation growth or construction activities) will contribute to the errors in remeasurement. Despite these other contributing factors, the correlogram and square-root of the variogram suggest that a horizontal shift of the footprint center is main cause of remeasurement errors. Since the correlation coefficient nearly approaches unity and the RMSD nearly approaches zero as the footprint center distance approaches zero, other factors contributing to remeasurement errors have a negligible effect for at least the pairs in the smallest-distance bins.

4.3 Changes in maximum canopy height

4.3.1 Application of a moving-average filter

Figure 4.7a displays the changes in waveform extent[1] — the proposed estimator for changes in maximum canopy height —between summer 2004 (campaign L2C) and summer 2005 (campaign L3C). Although changes between -10 m and +10 m clearly dominate, extreme changes of roughly ±40 m are also observed. The result of the application of a moving average filter to the data, with a window-size of 101 pairs, is presented in figure 4.7b. The filtered data has lower extreme values; the range of changes in waveform extent is about -3 m to +13 m. Small groups of isolated pairs (less than 101 pairs) are removed during the filtering process, as are the first and last 50 pairs of partial tracks.



(a) Unfiltered changes in waveform extent.

(b) Moving average of changes in waveform extent.

Figure 4.7: Waveform extent changes for 2004-to-2005 waveform pairs, before and after the application of a moving average filter. Satellite images for the regions indicated by the black rectangles are presented in figures 4.8a (western rectangle) and 4.8b (eastern rectangle).

[1] In this report, a *"change in waveform extent"* refers to the waveform extent of the second shot in a pair minus that of the first shot and when preceded by *"rate of"*, this will indicate a subsequent division by the time interval between first and second shot in the pair.

Two regions where the filtered values reach their positive extreme are indicated with black rectangles. Note that the rectangles are also indicated in figure 4.1a, showing that the footprint center distances within the western rectangle are in the 100-to-150 m range and those in the eastern rectangle within the 50-to-100 m range. True-color satellite imagery of the region within the western rectangle is shown in figure 4.8a, with an overlay of footprint pair locations (including those removed during filtering). The track crosses lands that are used for timber production and/or agriculture. The region indicated by the eastern rectangle is given a closer look in figure 4.8b. The satellite image shows a river with surrounding wetlands.

(a) Western rectangle.

(b) Eastern rectangle.

Figure 4.8: Satellite images for the regions indicated by the rectangles in figures 4.1a and 4.7. The grey dots indicate locations of unfiltered 2004-to-2005 waveform pairs (figure 4.7a).

Author's interpretation

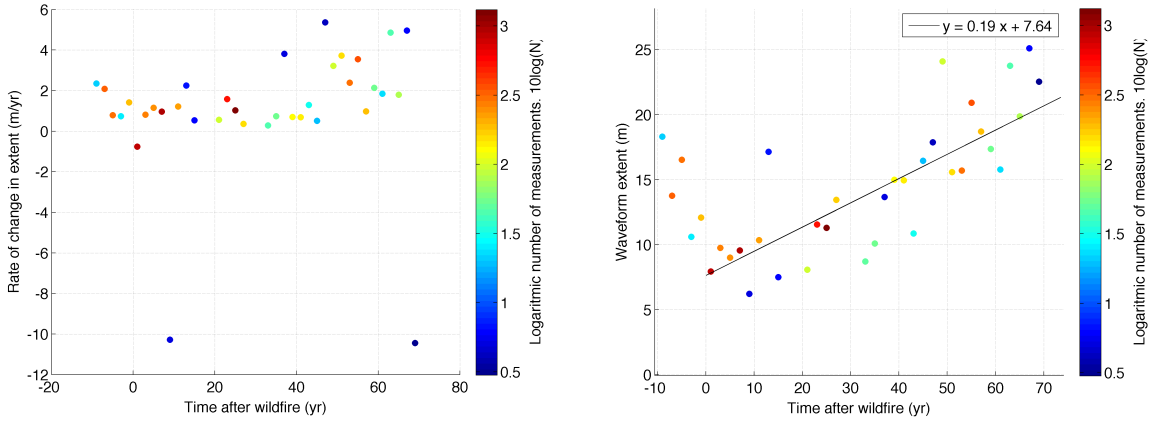
Due to footprint misalignments, differences in waveform extent from individual pairs (figure 4.7a) are not readily interpretable as changes in maximum canopy height. Taking differences in waveform extent due to misalignment as errors in the estimates of changes in maximum canopy height and assuming this is a random error (expectation value of zero), averaged differences in waveform extent for a set of pairs *can* be interpreted as average changes in maximum canopy height. The estimate of average changes in maximum canopy height then contains an error $\sigma/\sqrt{N}$, where σ is the sample standard deviation and N is the number of observations. Thus, by averaging over 101 pairs instead of taking a single pair, the random error is decreased with approximately a factor 10. From the square-root of the variogram (figure 4.6b) we conclude that the sample standard deviation of estimated changes in maximum canopy height (equal to the RMSD of waveform extent) are typically no higher than 8 m. Therefore, the standard deviation of the mean of 101 observations of changes in maximum canopy height is no higher than 0.8 m.
Now this reasoning poses a problem with interpreting filtered changes in canopy height on the order of 10 m in one year (figure 4.7b). Therefore we note some aspects that were not taking into account:

- The assumption that footprint misalignment causes *random* errors with zero expectation value may not be valid locally (for 101 pairs), as the distance but also direction of footprint center shifts are spatially correlated and maximum canopy height may be very heterogeneous on the scale of 150 m. For instance, when the second pass of a specific part of the repeat track samples a forest just along the edge of a water body while the footprints of the first pass were all located over the water body (or the other way around), the estimated mean change in maximum canopy height will be severely affected.
- New constructions and previously absent large vehicles within footprint locations will be misinterpreted as canopy height changes.
- Not all canopy height changes may be natural: there are croplands and timber production sites within the study region.

The satellite images for the examples extreme values indeed suggest a highly heterogeneous canopy height along river (zero canopy height) boundaries and wetlands (figure 4.8b) and indicate the presence of croplands or timber production sites (figure 4.8a). Also note the East-West boundaries of the croplands and (production) forests: as the track is oriented approximately along such boundaries, heterogeneity in canopy height may again cause severe errors in estimates of mean change in maximum canopy height. This demonstrates some of the limitations to estimating changes in maximum canopy height. Land cover products may be used to exclude problem areas such as waterbodies and croplands from the study site or satellite imagery may be used to test for general homogeneity in the region covered by (and in between) the two footprints.

4.3.2 Diminishment of maximum canopy height due to wildfires

In chapter 2, differences in waveform extent were proposed as an estimator for changes in maximum canopy height. The mean rate of change in waveform extent (see footnote on page 25) is evaluated for two-year wide bins of time-after-wildfire. All interannual waveform pairs are used; measurement time intervals are either one year or two years, approximately. The result is shown in figure 4.9a. The logarithmic number of pairs per bin is indicated by the marker color. Among the bins with less than 10 pairs (dark-blue dots) are the extreme values of roughly -10 and +5 m/year. Apart from these bins, the lowest value of -0.8 m/year occurs at 0-to-2 years after a wildfire. The number of pairs within the bin is 826. All other bins with 10 pairs or more yield positive averages no higher than 5 m/year. Between 2 and 46 years after a wildfire, these averages are roughly 1 m/year and no higher than 2 m/year. After the 46-years mark, these well-sampled bins yield average rates of approximately 1 to 5 m/year.



(a) Mean rate of change in waveform extent.

(b) Mean extents of the second waveforms in the pairs.

Figure 4.9: Mean waveform extents and their rates of change determined from waveform pairs with one-year and two-year intervals, as functions of the time elapsed since the most recent wildfire. The transmit time of the second shot in a pair is taken as the measurement time. For negative values, the time indicates the time remaining until the first upcoming wildfire. Samples with a negative time have no recorded wildfire history before the measurement. The black line is the least-squares linear fit to the waveform extents before averaging over 2-year bins.

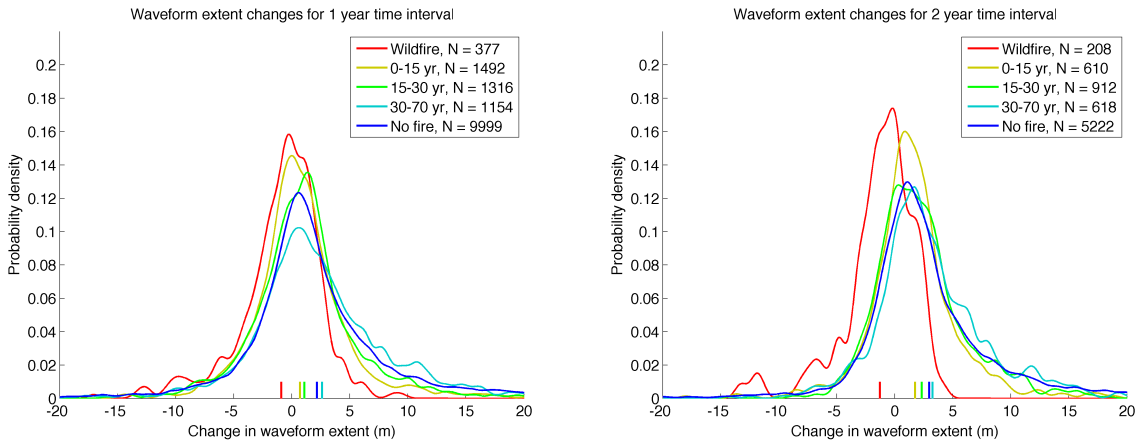
For the same set of waveform pairs and the same time-after-wildfire bins, the average waveform extents of the latter measurements within the pairs are determined. The result is presented in figure 4.9b. The values are roughly between 6 and 25 m. Along the positive part of the horizontal axis, a trend is observed in the waveform extent. The linear fit that minimizes the squared residuals of the original data (before averaging over 2-year bins) has an offset of about 8 m and a slope of approximately 0.2 m/year. Average waveform extents for negative values of time-after-wildfire (indicating the interval between the latter measurement within the pair and the first upcoming wildfire in case of no known wildfires before the measurement) are higher than those between 0 and 10 years after a wildfire.

Author's interpretation

Since the measurement intervals of the data inspected here are either one or two year long, approximately, most change measurements that are directly affected by wildfires are expected to fall within the time-after-fire bin of 0-to-2 years (figure 4.9a). This bin yields a lowest and only negative-valued estimate of rate of change in maximum canopy height than any other bin, excluding two outliers that are based on less than 10 measurements each and are therefore likely to be severely affected by random errors due to footprint misalignment. The results therefore suggest that the diminishment of maximum canopy height due to wildfires is detectable with the GLAS.
The mean values for other bins suggest a bias in the rates of change in canopy height as estimated from differences in waveform extent, since height-growth of roughly 4 meters per year seems unrealistic. Actual height growth is expected to be on the order of decimeters per year [17]. The mean extents in figure 4.9b do show an increase with time-after-fire which is realistic, about 0.2 m/year, but it must be clear that this is not an actual change observed over one region but a comparison between forest stands that differ in the years elapsed since the last wildfire.

The distributions of change in waveform extent for pairs with a one-year measurement time interval are shown in figure 4.10a in the form of empirical kernel-smoothed probability density functions (PDF's). The kernel function is a Gaussian peak of width 0.5 m. The width is purposely kept low so that some fluctuations in the PDF's of fewer samples are still preserved, giving a sense of the quality of the estimated PDF. The interannual pairs are divided into five categories based on the start date (or, in some cases, date of discovery) of the most recent recorded wildfire at the pair location before the latter of the two measurements in the pair:

1. The first shot was before the most-recent wildfire start date, but the second shot more than 60 days after.
2. The first shot was between 60 days and 15 years after the most-recent wildfire start date.
3. The first shot was between 15 years and 30 years after the most-recent wildfire start date.
4. The first shot was between 30 years and 70 years after the most-recent wildfire start date.
5. No wildfires have been recorded at the location of the measurement pair.



(a) Waveform pairs with a one-year interval.

(b) Waveform pairs with a two-year interval.

Figure 4.10: Kernel-smoothed probability density functions for waveform extent changes over one-year and two-year intervals. The five measurement categories (see legends) are defined on page 28 based on the time elapsed since the most-recent wildfire.

The mean changes in waveform extent for each category is indicated with a colored vertical bar at the horizontal axis. The first category yields the lowest and only negative mean value, about -1 m, as well as the only PDF that skewed to the left. The other four categories yield positive averages between 0.5 and 3 m and right-skewed PDF's. In the order of first, second, third, fifth, and then the fourth category, the mean change in waveform extent increases, the maximum probability density decreases and the full-width-at-half-maximum (FWHM) and skewness increase. The PDFs for first four categories, having fewer samples than the fifth, show some fluctuations, but all PDFs approximate unimodal distributions.

The same graph has been produced for waveform extent changes over two-year time intervals, between summer 2004 and 2006 (figure 4.10b). The average values for the five categories increase in the same order as in they did for one-year intervals, but their magnitudes are larger: -1.3 m for the first category and between 1.7 and 3.3 m for the other four. The maxima of the PDFs have increased for all categories except the third (15-30 years). The PDF of the first category is again skewed to the left and the other four are still right-skewed. Differences between FWHM and skewness for the latter four categories are not as clear as for the one-year interval PDFs. Once more, the first four categories have fewer samples than the fifth and yield multi-modal PDF's, yet categories 2, 3 and 4 yield only one mode on scales of 5 m or larger. The PDF for the first category has an extra small peak for changes between -15 and -10 m.

The two PDFs of each category may be combined by dividing the changes in waveform extent by the measurement time interval, resulting in average rates of change in extent during the intervals. The resulting PDFs are displayed in figure 4.11. These are very similar to the PDFs for one-year measurement intervals, except that the maximum probability densities are higher and the PDFs have lower FWHM-values. The same statements can be made concerning the order in which the means, maximum probability densities and skewness increase. The average rates of change for the first category is -0.8 m/year and ranges from 0.7 to 2.2 m/year for the other four categories.

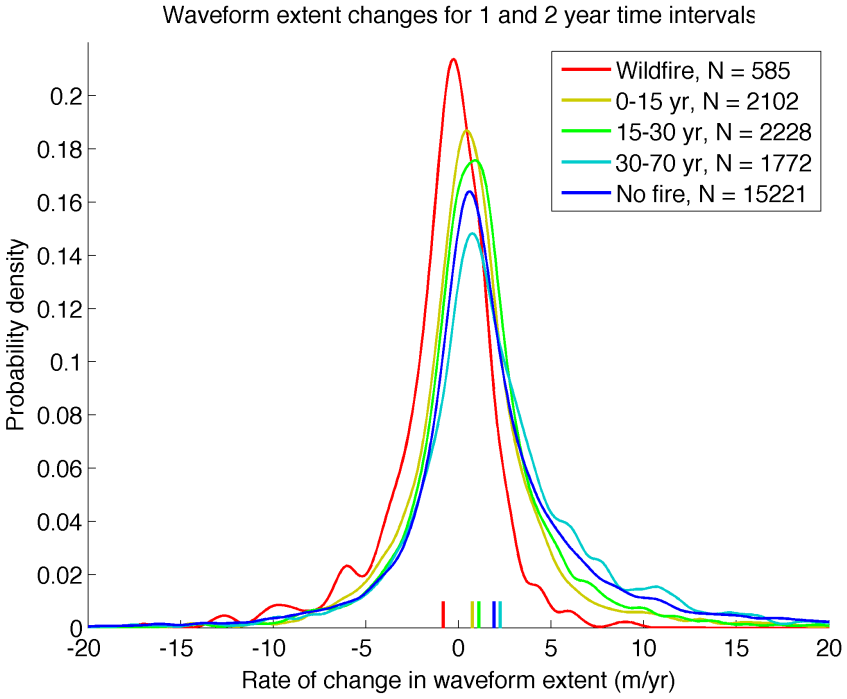
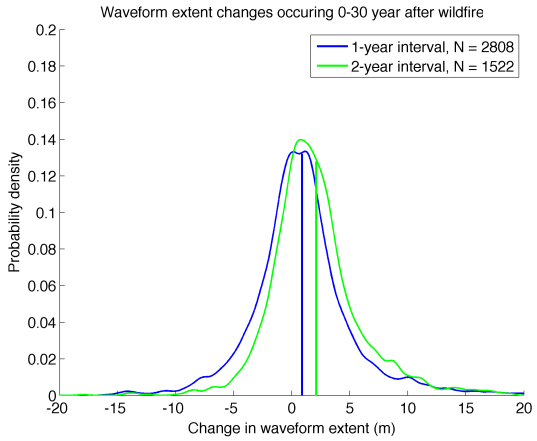


Figure 4.11: Kernel-smoothed probability density functions for rate of change in waveform extent from one-year and two-year interval pairs. The five measurement categories (see legend) are defined on page 28 based on the time elapsed since the most-recent wildfire.

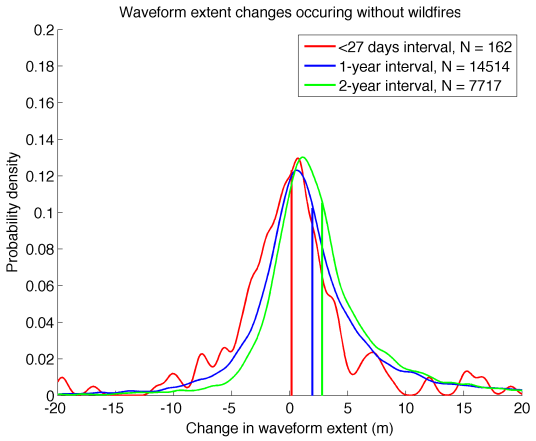
Author's interpretation

The mean rate of change in waveform extent is interpreted as an estimate of the mean rate of change in maximum canopy height. Category 1 is defined so that it contains measurements that are likely to be directly affected by wildfires. The other categories are defined so that it contains measurements that are unlikely to be directly affected by wildfires. Maximum canopy height then decreases with one meter on average during wildfires (figures 4.10a and 4.10b). As on page 26, the maximum error magnitude is estimated as $8\text{m}/\sqrt{N}$ and the number of observations $N = 585$ (both one-year and two-year intervals) then yields an error of ± 0.3 m. Therefore, the one-meter decrease encountered over wildfires is thought to be significant. However, it should be noted that instrumental changes (e.g. laser power or transmit pulse FWHM) between campaigns may have introduced a bias in this result. The PDFs for categories 2-5 are right-skewed and yield mean changes from 0.5 m to 3 m in one year (figure 4.10a). The higher of these increases in waveform extent cannot be entirely ascribed to actual changes in canopy height as an average height growth of 3 m in a year seems unrealistic. Therefore, we expect that the estimated changes in maximum canopy height are positively biased. As mentioned before, instrumental changes are a possible cause of bias in the estimation. Construction activities may also cause a positive bias due to misinterpretation of constructed elevated surfaces (e.g. buildings, bridges and street lamps) as canopies. Both may explain that the bias seems positively correlated to the time elapsed since the most-recent wildfire: an instrumental bias may be positively correlated to canopy height, which in turn is positively correlated to the time-after-wildfire, and construction activities may be less likely to take place in regions that have recently been burned. Since results suggest positive biases and the decreases in maximum canopy height estimated over wildfires are significant with respect to the random errors due to footprint misalignment despite a possible positive bias, we conclude that the diminishment of maximum canopy height due to wildfires is successfully detected.

Figure 4.12a illustrates the differences between changes during one-year and two-year intervals for union of category 2 and category 3 data. The mean changes in extent are 0.9 m for one-year intervals and 2.1 m for two-year intervals, differing by approximately a factor 2. The two-year interval PDF has a higher skewness than the one-year interval PDF, but both are skewed to the right. The width of the PDFs at half maximum are about the same, as are the maximum probability densities.



(a) The first 30 years after wildfires; union of category 2 and 3.



(b) No wildfire during measurement interval; complement of category 1.

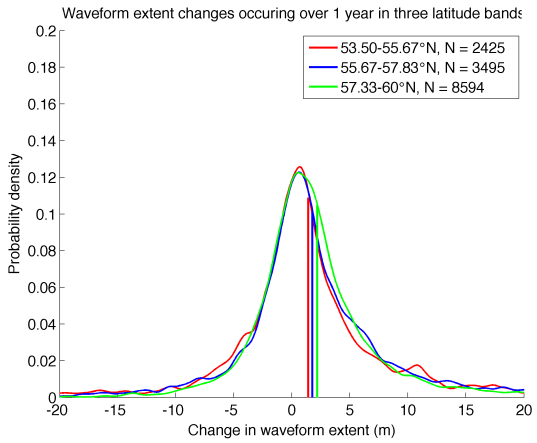
Figure 4.12: Kernel-smoothed probability density functions for waveform extent changes from one-year interval pairs, two-year interval pairs and same-campaign pairs. The measurement categories are defined on page 28.

Repeating the graph for the complement of category 1 (figure 4.12b), a PDF of same-campaign changes (an interval of less than 27 days) can be included, although the number of same-campaign pairs is still low (162 pairs). The mean value of the PDF is nearly zero. For one-year and two-year intervals, the mean values have increased slightly with respect to those in figure 4.12a.

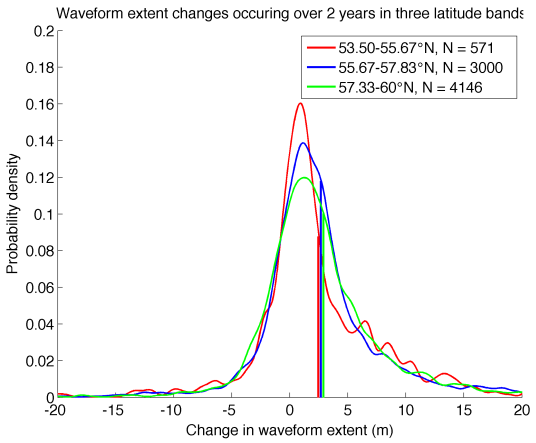
Author's interpretation

The finding that the mean changes in maximum canopy height (estimated from waveform extent differences) occurring the first 30 years after wildfires are approximately twice as high for two-year interval pairs as for one-year interval pairs (figure 4.12a) makes it plausible that these changes are in fact due to actual height-growth. Still some other possible causes (most notably the effect of instrumental changes) contributing to the averaged changes in waveform extent should be excluded before drawing such a conclusion. Annual height-growth of roughly one meter may still be plausible; in literature [17] we find rates up to about 0.7 m/year for a boreal forest study area ($53°34' - 54°24'$N and $105°25' - 107°52'$W) in Saskatchewan. Furthermore, saplings may have been planted after a wildfire for faster reforestation.
The nearly zero-mean for same-campaign pairs (figure 4.12b) shows that changes in maximum canopy occurring within 27 days are negligible with respect to the changes estimated from interannual pairs. The positive bias that is suspected due to unrealistically high increases in maximum canopy height within one year (figure 4.10a) appears not to affect same-campaign measurements. Therefore, if the bias is due to instrumental changes, these changes did not occur during the summer campaigns but (mostly) in between campaigns.

Taking again the complement of category 1, only one-year interval pairs, the distributions of waveform extent changes are inspected for three different latitudes bands of equal width within the study region. The PDFs are presented in figures 4.13a and 4.13b. The mean values differ slightly and increase with increasing latitude. The PDFs are nearly identical and approximate slightly right-skewed unimodal distributions.



(a) One-year interval pairs.



(b) Two-year interval pairs.

Figure 4.13: Kernel-smoothed probability density functions for waveform extent changes per latitude band, from one-year and two-year interval pairs that do not sample wildfires (complement of category 1, see page 28).

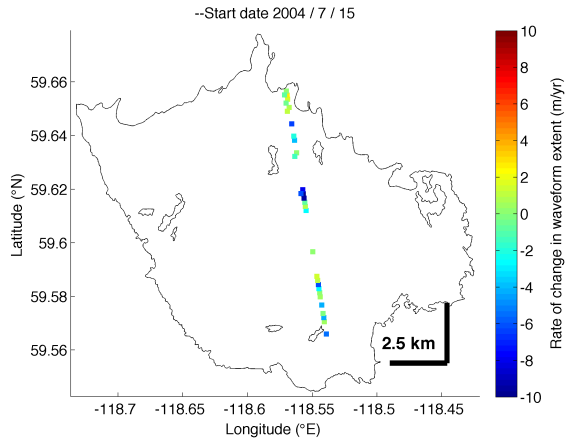
Author's interpretation

The PDFs for estimated changes in maximum canopy height within different latitude bands (figure 4.13) are nearly identical (apart from one being of lower quality due to a significantly fewer measurements, namely the red PDF in figure 4.13b). This suggests that the suspected bias is due to instrumental changes as opposed to construction activities which' impact on the estimated changes in maximum canopy height is unlikely to be spread evenly across these latitude bands.

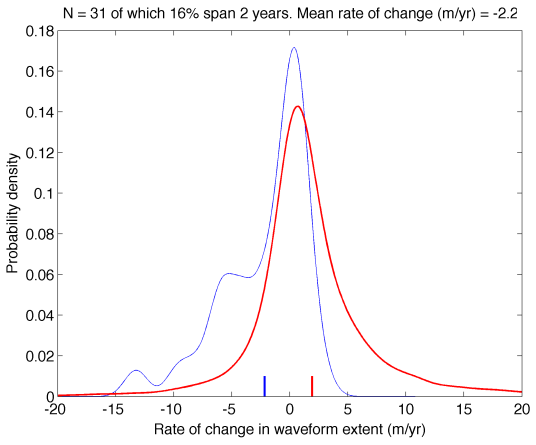
4.3.3 Individual analyses of four wildfires

Four wildfires sampled by more than 30 pairs are analyzed individually. This accounts for 545 of the 585 pairs (93%) in category 1. In agreement with the criteria for category 1, a pair that *samples a wildfire* refers to a pair of which the first shot was before the most-recent wildfire start date at the pair location and the second shot more than 60 days after. Thus, the pair location (halfway between the first and second footprint center) must be within the wildfire perimeters. Both one-year and two-year interval pairs are considered. Figures 4.14a, 4.14c, 4.14e and 4.14g show the rates of change in waveform extent and locations of the pairs within the wildfire perimeters. Figures 4.14b, 4.14d, 4.14f and 4.14h present the corresponding PDFs (blue lines; Gaussian kernel of width 1 m/year). For reference, the PDF for category-5 pairs (one-year and two-year intervals) has been added (red line). The number of pairs per wildfire ranges from 31 to 267 and the fraction of pairs with two-year measurement intervals ranges from 0 to 50%. For all four wildfires, the mean rate of change in waveform extent is negative, ranging from -2.2 to -0.1 m/year. Each PDF has its highest-amplitude peak at approximately 0 m/year and one or more peaks of smaller amplitude between -20 and 0 m/year. For the wildfire called "Lake Nine" (figures 4.14g and 4.14h), the PDF is spread over a significantly wider range of rates (taking for instance the full width at a probability density of 0.02), both in the positive and negative direction. Unlike the other three PDFs, it shows that a considerable fraction of the pairs have yielded rates between 5 and 10 m/year.

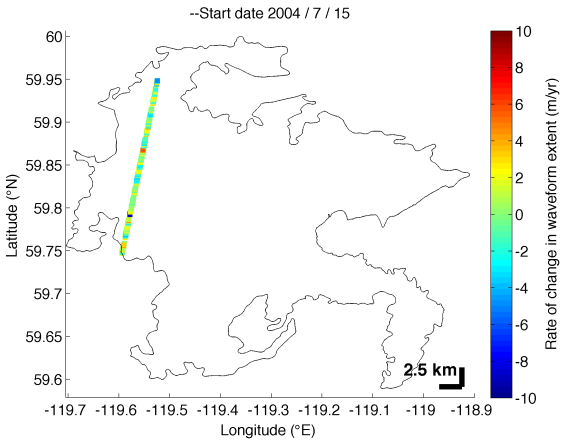
Figure 4.14: Rates of change in waveform extent for one-year and two-year interval pairs within wildfire perimeters (shown in black), first shot before the wildfire start/discovery date, second shot more than 60 days after. The corresponding kernel-smoothed probability density functions (PDFs) are shown in blue. For reference, the PDF of category-5 pairs (no wildfire, see page 28) has been added in red. *(Continued on the next page.)*



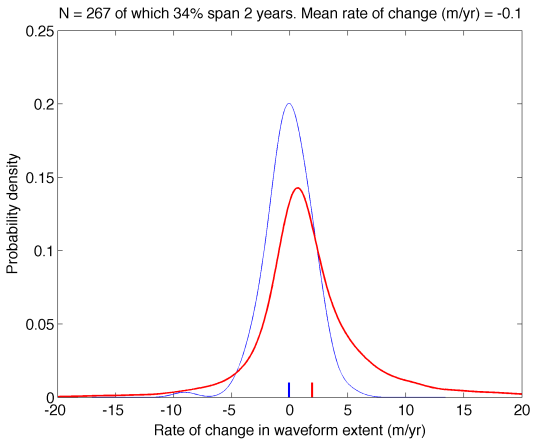
(a) Nameless wildfire 1: measurement locations.



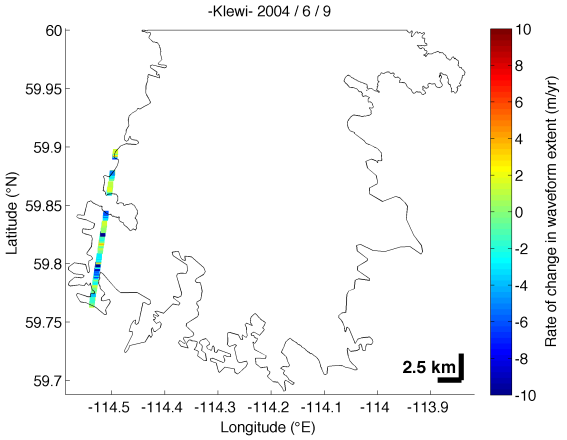
(b) Nameless wildfire 1: kernel-smoothed PDF (blue).



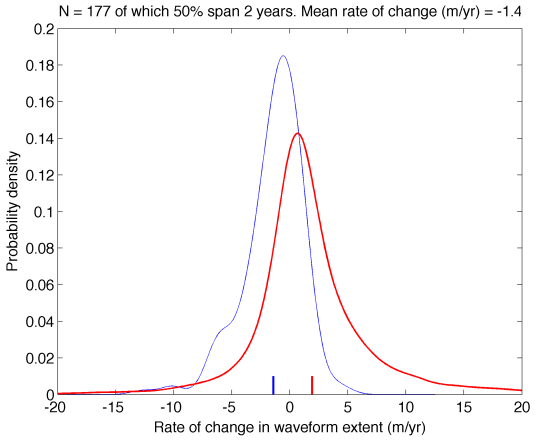
(c) Nameless wildfire 2: measurement locations.



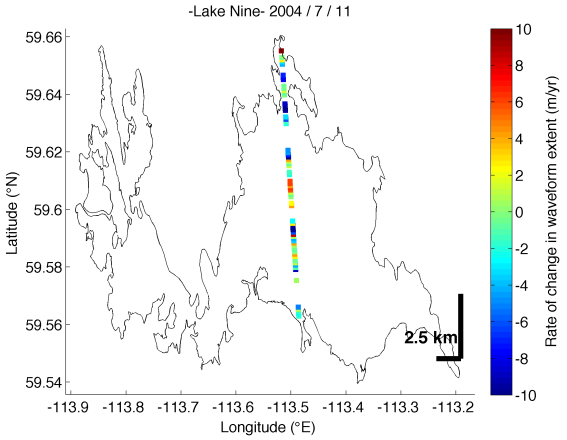
(d) Nameless wildfire 2: kernel-smoothed PDF (blue).



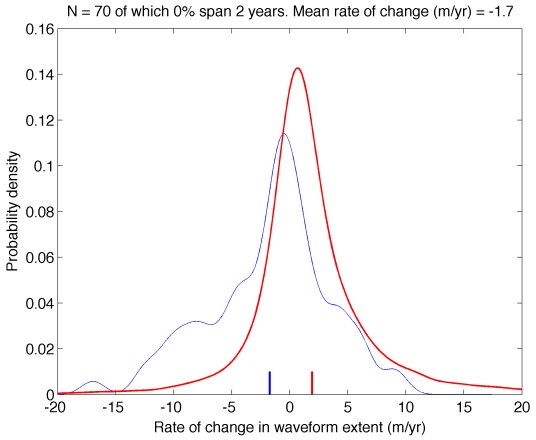
(e) "Klewi" wildfire: measurement locations.



(f) "Klewi" wildfire: kernel-smoothed PDF (blue).



(g) "Lake Nine" wildfire: measurement locations.



(h) "Lake Nine" wildfire: kernel-smoothed PDF (blue).

Author's interpretation

The analysis per wildfire is added for extra confirmation and illustration of the results presented earlier. The PDFs for estimated rates of change in maximum canopy height for individual wildfires (figure 4.14) have their highest-amplitude peaks located near zero and a FWHM of roughly 5 m/year. These are thought to be a result of pairs located over regions within the perimeters where the maximum canopy height was not affected by the wildfire. The FWHM-values of these peaks are then mostly caused by estimation errors due to footprint misalignments, but the peaks may be widened by the smoothing with a Gaussian kernel of width 1 m/year (in terms of standard deviation, equal to a FWHM of 2.35 m/year). The peaks to the left of these maximum-amplitude peaks are ascribed to actual changes in maximum canopy height due to wildfires. Since such peaks are not found to the right of the maximum-amplitude peaks, it is unlikely that they are all a result of footprint misalignments. One exception to this is found for the "Lake Nine" wildfire (figure 4.14h). The wider PDF suggests that maximum canopy height is more heterogeneous than for the other wildfires or the footprint center distances are larger, causing larger errors in the estimation of changes in maximum canopy height. Each of these four individual wildfires yields a negative estimate for the mean change in maximum canopy height. One of the mean values is significantly lower than the other three (figure 4.14d), which may indicate that the sampled region was virtually unaffected by the wildfire; vegetation may have been absent both before and after the wildfire.
The maximum-amplitude peaks seem to lack the suspected bias in the estimation of maximum canopy height changes. It was previously considered that the estimates might be biased due to instrumental changes. Perhaps a better explanation for the seemingly unrealistically high estimates for mean changes in maximum canopy height is horizontal growth of the crowns of high trees, which can increase maximum canopy height over land previously not covered by (high) vegetation with the entire height of the tree crown. This would explain that estimated changes between +15 and +20 m are encountered (e.g. the dark-blue PDFs in figure 4.10b) but estimates between -20 and -15 m/year are virtually non-existent. Changes in canopy height are not necessarily a result of trees growing in height. The expected contribution of horizontal crown growth to the mean of a set of changes in maximum canopy height is dependent on the widening of the crowns, the spacing between individual crowns, the crown heights at which horizontal growth occurs (say between h_1 and h_2) and the footprint dimensions. These affect the probability that footprint with a random location over vegetated land contains plant surfaces at heights above ground between h_1 and h_2 (assuming level ground for simplicity). With a geometric model for forest growth, the order of magnitude of the expected contribution may be estimated, but instrumental changes should preferably be investigated first to rule out a bias.

For wildfire "Klewi" (figures 4.14e and 4.14f), two LandsatLook Natural Color images are acquired [37]. One was recorded the day before the wildfire start date, the other approximately a year later. The images are displayed in figures 4.16 and 4.17 with an overlay of the rates of change in waveform extent for one-year and two-year interval pairs, including those located outside of the wildfire perimeters but within the boundaries of the map. Comparing the two LandSat images, one can clearly observe changes within the wildfire perimeters. Small parts of the land within the perimeters show no noticeable change. Apart from a small cloud and the shadow it casts (near the southwestern corner of the map) the two images are nearly identical outside of the wildfire perimeters.

The PDF for one-year and two-year interval pairs within the perimeters the Klewi wildfire is repeated in figure 4.15 (Gaussian kernel of width 0.5 m/year). The PDF for the set of pairs that cover the same time intervals (either summer 2004 to summer 2005, or summer 2004 to summer 2006) and are within the map boundaries, but *outside* of the wildfire perimeters, is added in red. While the pairs within the perimeters yield a mean rate of -1.4 m/year, the pairs located outside of the perimeters average to -0.1 m/year. The distribution of pairs within the perimeters yields a main peak that is shifted slightly in the negative direction with respect to the other PDF in the figure. The most notable differences occur between rates between rates of -15 and -5 m/year. These values are found within the perimeters, but outside of the perimeters, they make up for a negligible fraction of the measurements.

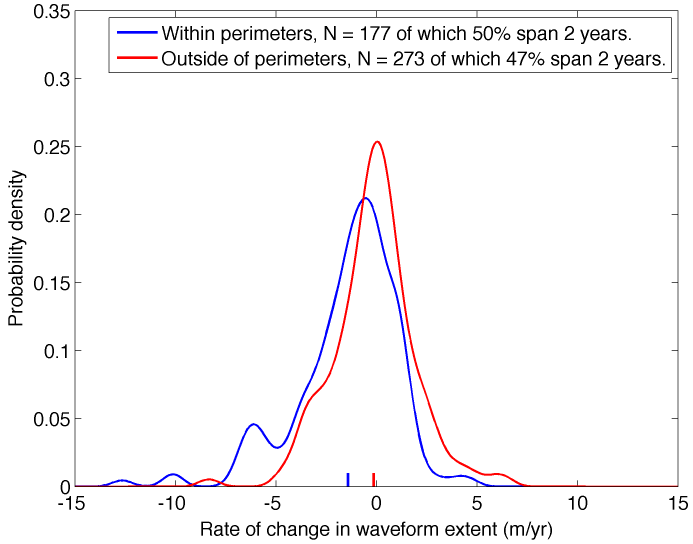


Figure 4.15: Probability density functions of the rates of change in waveform extent shown in figures 4.16 and 4.17 (first shot before the start/discovery date, second shot more than 60 days after): one for measurements within the wildfire perimeters (blue) and one for measurements outside of the perimeters (red).

Author's interpretation

The most extreme negative values for estimated change in maximum canopy height are located over some of the most pronounced burn scars in the LandSat image for the "Klewi" wildfire (where the track reaches latitudes 59.795°N and 59.842 in figure 4.17). These estimates suggest that these burn scars correspond to stand-replacing patches within a wildfire of mixed severity[a] The missing pairs where the track passes over a lake at 59.845°N are likely due to the upper limit of the allowed signal-to-noise ratio (set at 40, see chapter 3), since the fraction of near-infrared radiation reflected by still water is negligible [33].
The PDFs for pairs within the wildfire perimeters and for those outside of the perimeters, and the corresponding averages, show that the waveform-derived changes in maximum canopy height confirm that (1) canopy height diminishment has occurred within the wildfire perimeters, and (2) any change is very limited or non-existent outside of the perimeters (figure 4.15).

[a]Spatial variability in the severity of a wildfire is an interesting study subject. Collins & Stephens [8] address this aspect of wildfires.

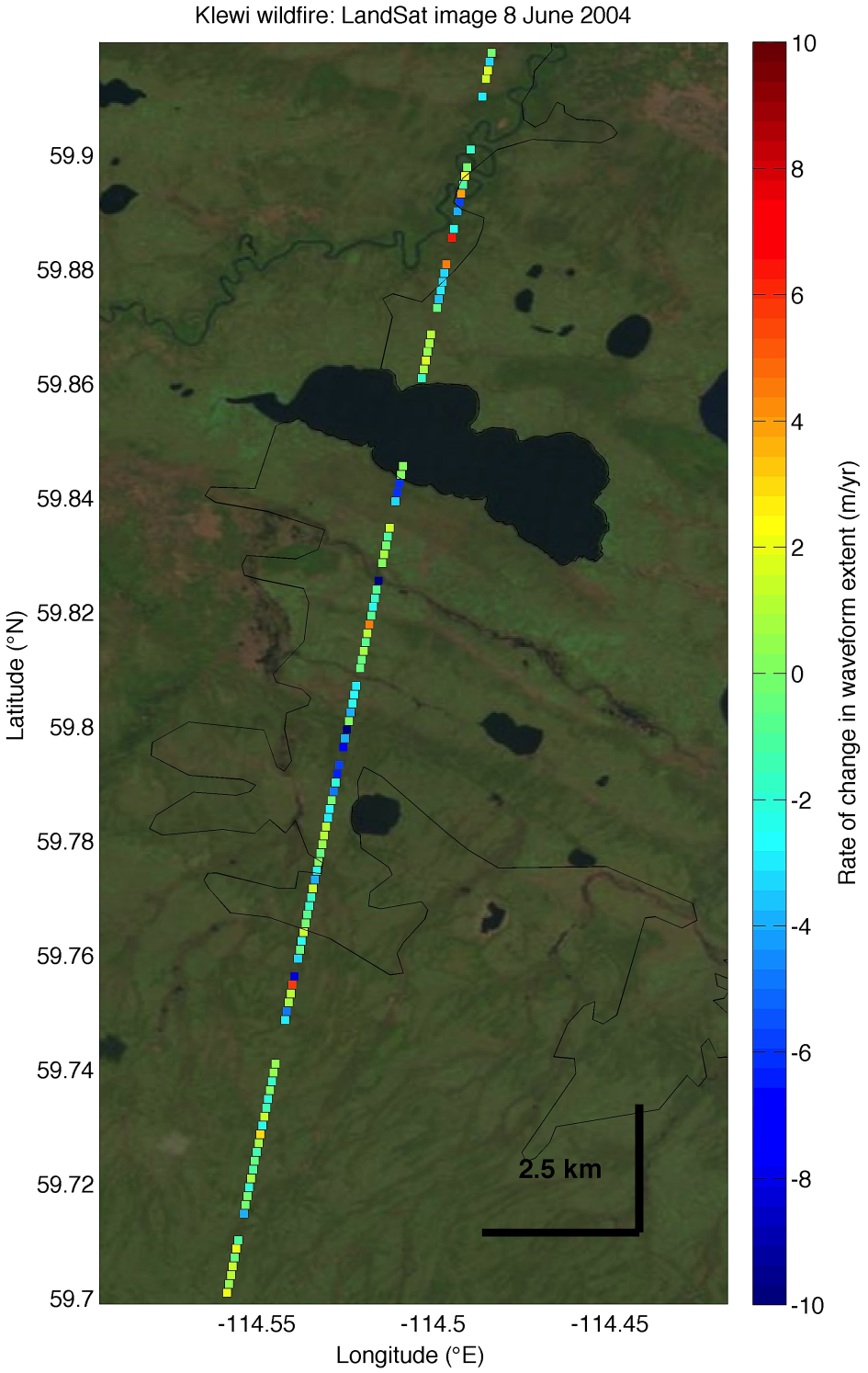


Figure 4.16: A LandsatLook Natural Color image recorded one day before the start/discovery date of "Klewi" wildfire [37]. Overlays of "Klewi" wildfire perimeters (black) and rates of change in waveform extent for one-year and two-year interval pairs (colored squares).

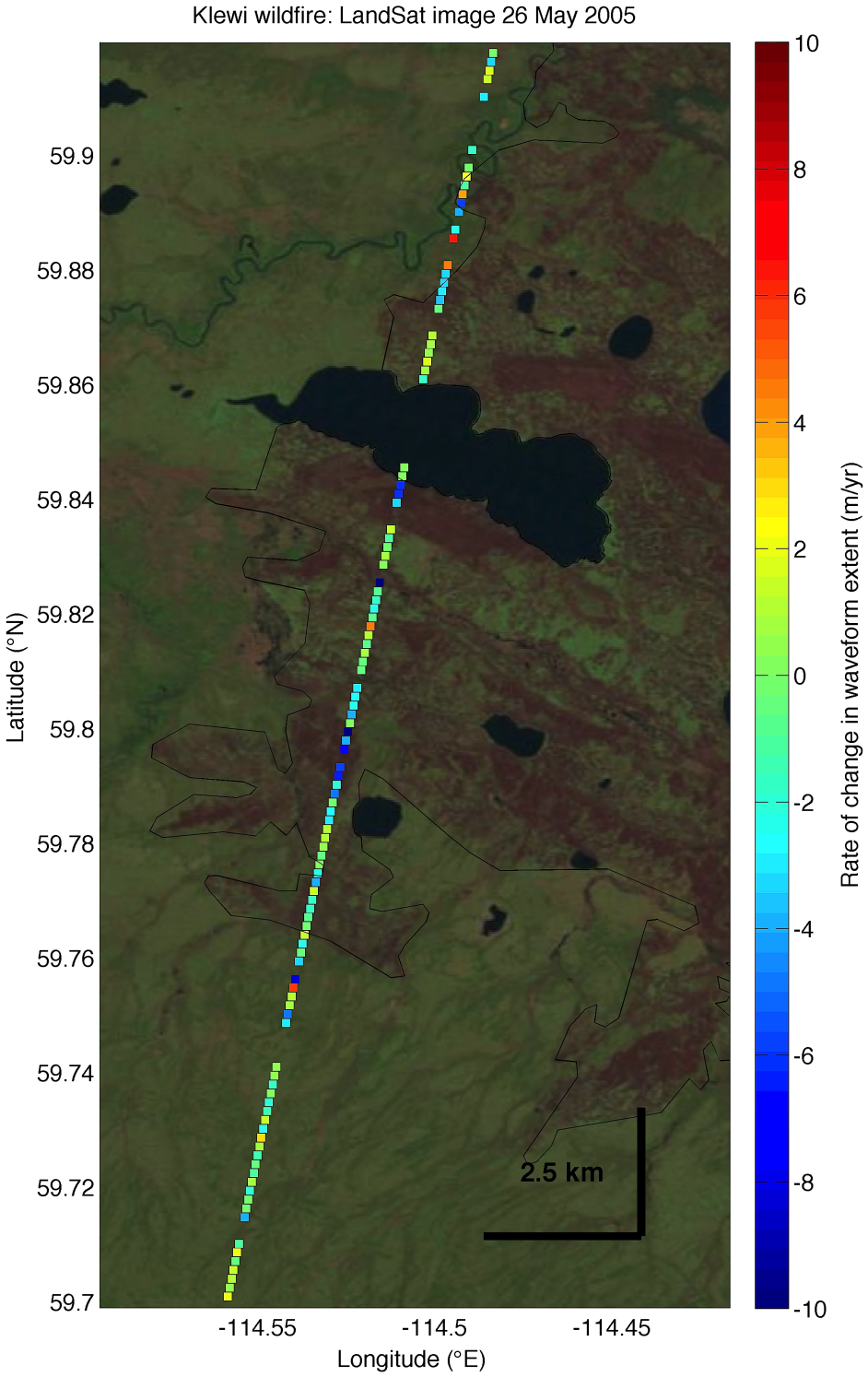


Figure 4.17: A LandsatLook Natural Color image recorded about a year after the start/discovery date of "Klewi" wildfire [37]. Overlays of "Klewi" wildfire perimeters (black) and rates of change in waveform extent for one-year and two-year interval pairs (colored squares).

4.4 Case studies of individual pairs

The conclude the chapter, we will look into some of the measurement pairs with the smallest footprint center distances more closely. For each of the selected pairs, the following is presented:

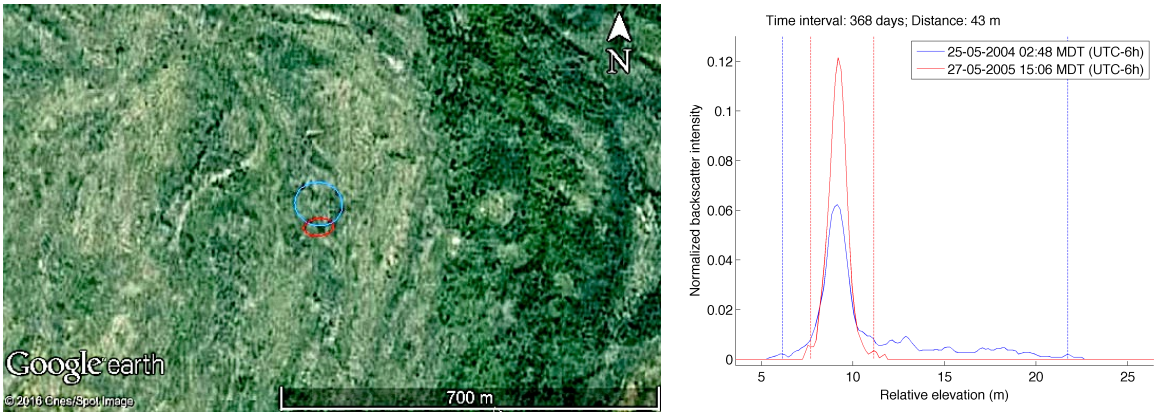
- Google Earth satellite imagery with an overlay of the boundaries of two elliptical footprints
- Both received waveforms, normalized and manually aligned along the horizontal axis, with noise removed according to the zero-crossing approach and the first and last threshold-crossings indicated

The first case is one of the pairs sampling the Klewi wildfire. For this case, additional smaller-scale maps with the LandSat images recorded on 10 July 2004 and 26 May 2005 are included. These show clearly observable changes at the pair location. Although the start/discovery date of the wildfire is 9 June 2004, the satellite image from July does not show any signs of burning near the case study location. In this case, the footprints overlap partially (figure 4.19a). The footprint of summer 2005 has a smaller major axis and eccentricity than the 2004-footprint. The distance between the footprint centers is 43 m. The waveform recorded in 2005 consists of a single peak (figure 4.19b. The 2004-waveform has a larger extent than the 2005-waveform for both the zero-crossing and threshold-crossing approach. It consist of a peak followed by a weaker signal which is overall declining (in the positive direction of the horizontal axis).

(a) LandSat image recorded 10 July 2004.

(b) LandSat image recorded 26 May 2005.

Figure 4.18: Location of the waveform pair (blue cross) for the first case study within the "Klewi" wildfire perimeters (pale yellow). LandsatLook Natural Color images are shown in the background [37].



(a) Case 1 ("Klewi" wildfire): footprints.

(b) Case 1 ("Klewi" wildfire): waveforms.

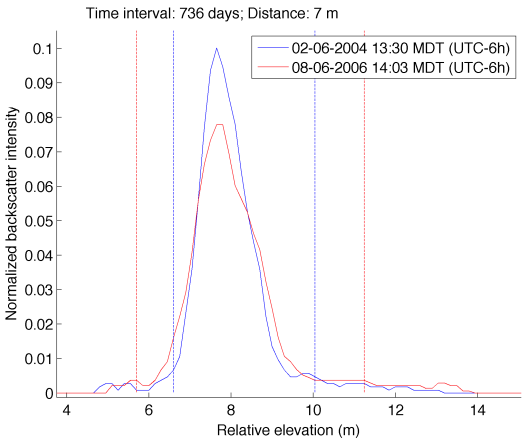
Figure 4.19: Footprints and waveforms of the first (blue) and second (red) measurements within the pairs that are selected for case studies. The zero-crossing approach has been used to remove noise from the waveforms. The first and last crossing of the signal threshold (five standard deviations of the background noise) are indicated with dashed vertical lines (same color as the corresponding waveform).

Author's interpretation

The LandSat image in figure 4.18a recorded on 10 July 2004 illustrates that the region of wildfire sampled by waveform pairs may not yet be affected by fire at the start/discovery date of the fire (9 June 2004 in this case). For this reason, the definitions of the first two wildfire categories (sampling a wildfire and within 15 years after a fire, see page 28) involved a 60-day buffer. The six pairs thereby excluded from category 1 were less certain to be affected by wildfires than the remaining pairs. Pairs excluded from category 2 were possibly affected by that were wildfires still burning during the measurement interval.
The LandSat image in figure 4.18b shows that the footprint pair for the first case study is located within a severely burned patch. Within the received waveforms shown in figure 4.19b we identify the maximum-amplitude peaks at relative elevations of about 9 m as the ground returns. The waveform of the second shot lacks the signal received before (at higher elevations than) the ground return in the waveform of the first shot. This signal is thought to be a result of laser reflections from vegetation surfaces and judging from the elevation difference between the highest detected surface (at about 23 m relative elevation for the zero-crossing approach) and the centroid of the ground return (at about 9 m relative elevation), the maximum canopy height within the footprint was about 14 m before the wildfire. In the same way, we may estimate maximum canopy height within other the footprint, after the wildfire, at approximately 3 m, although the true maximum canopy height may have been zero; the estimate may be biased by the non-zero duration of the transmit pulse (that is, the first detected signal may be part of the ground return).
Furthermore, we note that the end of the signal (lowest elevations) has not been reproduced accurately with respect to the centroid of the ground return. This may be due to a lower signal-to-noise ratio, which in turn may be a result of a decrease in laser power (laser degradation) or a decrease in reflectivity (e.g. due to carbonized materials and ashes), or it may be a result of differences between the laser pulse characteristics (transmitted waveforms). Surface topography combined with footprint misalignment may also contribute. Finally, we would like to mention that multiply scattered photons may be delayed due to an inclined travel direction (e.g. between a first scattering event in the canopy and a second one at the ground) and thus be misinterpreted as photons reflected from a lower elevation than is the case. Regardless of what caused the end of the signal not to be reproduced with respect to the ground-return centroid, taking the ground return centroids as the reference elevations for changes in maximum canopy height is expected to yield more accurate change estimates, provided that the ground returns can be accurately identified. Another advantage of this method would be the addition of a retrieval of maximum canopy height (as opposed to changes only).

The other three cases are from the fifth category of pairs (no recorded wildfires at the pair location). The first of these three has a measurement interval of approximately two years and a footprint center distance of 7 m (figures 4.19c and 4.19d). The 2006-footprint has a slightly larger major axis but a smaller eccentricity than the 2004-footprint. From 2004 to 2006, the waveform extent has increased slightly. The waveform extent according to the threshold-crossing approach has increased considerably. Both waveforms consist of a peak followed by a weaker, overall declining signal (again in the positive direction of the horizontal axis), but the main peak of the 2006-waveform is wider.

(c) Case 2 (59.746°N, 112.752°W): footprints.

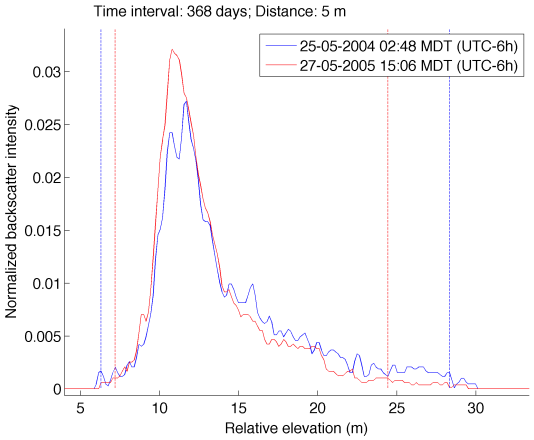


(d) Case 2 (59.746°N, 112.752°W): waveforms.

Figure 4.19: *Continued from page 38.*

For the next case (figures 4.19e and 4.19f), the second footprint falls completely within the first; it has a smaller major axis and eccentricity. The measurement time interval is approximately one year long and the footprint center distance is 5 m. Once more, both waveforms can, to good approximation, be described as a large peak followed by a weaker, overall declining signal (positive direction of the horizontal axis). However, in this case, the quite steadily declining signal contains a considerably larger fraction of the total integrated signal than in the previous cases. While the extent determined by threshold-crossings has decreased with about 5 m from 2004 to 2005, the zero-crossing approach gives good agreement between the waveform extents. Relative to the total integrated signal, the 2005-waveform has a larger peak amplitude and less signal from higher elevations.

(e) Case 3 (59.550°N, 114.611°W): footprints.



(f) Case 3 (59.550°N, 114.611°W): waveforms.

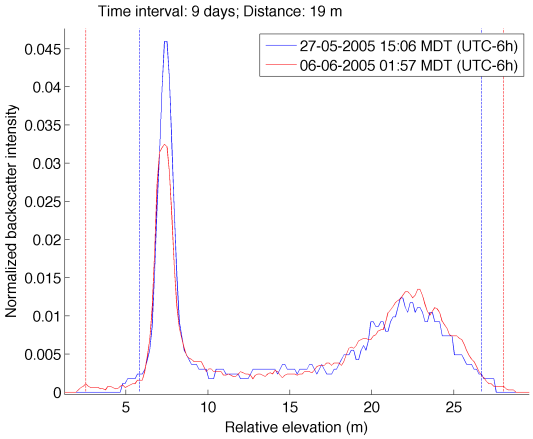
Figure 4.19: *Continued from page 38.*

The final case (figures 4.19g and 4.19h) is of a pair with a nine day measurement interval and a footprint center distance of 19 m. The footprints overlap considerably and their dimensions are nearly the same. The satellite image in figure 4.19g shows a dense forest at the footprint locations. Both waveforms (May 27th and June 6th) show two peaks of which the one at a lower elevation has a larger amplitude but smaller width than the other peak. In between the two peaks, the waveforms show some signal above zero. From May to June, one observed the following changes:

- The lower-elevation peak has decreased in normalized amplitude.
- The normalized amplitude of the higher-elevation peak has increased slightly.
- A tail is formed at low elevations.

Comparing the signal beginnings (largest elevations) of the aligned waveforms, one observes that it is placed at a lower elevation for the May-waveform than for the June-waveform, regardless of the choice between the zero-crossing approach and the threshold-crossing approach and the difference is about a meter for both. Comparing the signal ends (lowest elevations), one observes that it is placed at a higher elevation for the May-waveform than for the June-waveform. This is again regardless of the choice between the zero-crossing approach and the threshold-crossing approach and the difference is about 3 m for both.

(g) Case 4 (58.035°N, 115.110°W): footprints.



(h) Case 4 (58.035°N, 115.110°W): waveforms.

Figure 4.19: *Continued from page 38.*

Author's interpretation

For the latter three waveform pairs examined individually, we would especially like to point out that the zero-crossing approach to delineating the waveform extent yields a higher reproducibility than the threshold-crossing approach in each case, both in terms of extent and in terms of the entire cropped waveform (gates before/after signal beginning/end set to zero). Any biophysical variables (e.g. canopy cover as explained in chapter 2) retrieved from the waveform are very likely to yield an improved reproducibility and accuracy by using the zero-crossing approach instead of the threshold-crossing approach to delineate the beginning and end of the signal.

Chapter 5

Discussion & conclusions

In the previous chapter, interpretations of the results have already been discussed, so here we will address the research questions (section 1.2.1), summarize findings deemed most important and give recommendations for further research.

Main research question: "To what extent is it possible to detect canopy height changes with the Geoscience Laser Altimeter System?"

It is possible to detect changes in maximum canopy height from repeated waveform measurements, but since footprints rarely coincide, change estimates from individual pairs usually contain a large error (on the order of meters) due to footprint misalignment. Therefore, only the *mean* change in maximum canopy height from many waveform pairs is usually interpretable. The following answers to the subquestions highlight different aspects of the problem.

Subquestion i: "How can we efficiently identify and extract useful GLAS waveforms from the data sources?"

Based on parameters in the GLAH14 data products, useful GLAS shots can be identified. A selection was made based on:

- Signal-to-noise ratio: upper limit to the background noise standard deviation set at 1/40 of the maximum signal amplitude.
- Saturation: upper limit of the number of gates crossing the detector saturation threshold set to 2.
- In-footprint relief: a terrain index is derived from SRTM elevations to represent the elevation differences within the footprint, upper limit of 30 m.
- Agreement with SRTM elevations: if the difference between the SRTM elevation at the footprint center and the GLAS elevation is more than 100 m, the measurement is omitted (the laser pulse was probably intercepted by a cloud).

Furthermore, only GLAS data from the late-May to late-June campaigns (L2C, L3C and L3F; for brevity referred to as summer campaigns) have been included in this study to minimize the presence of snow cover in the footprints. This has resulted in a maximum time interval of approximately two years for the change measurements. Further research into the functioning of the described selection method may yield improvements; some useful measurements may have been wrongly excluded and some low-quality measurements may have been wrongly included in the analysis. This includes the selection of summer campaigns, as some of the other campaigns may still have resulted in measurements that are useful for estimating changes in canopy height. The spatial distribution of snow cover, its effect on change estimates and possible corrections for the effect are interesting topics for further research.

Once the waveforms that are deemed useful were identified, they were extracted from binary GLA01 data with a reading-program that is included in appendix .

Subquestion ii: "How can we estimate changes in maximum canopy height from repeated waveform measurements?"

Changes in maximum canopy height can be estimated from waveform pairs as the difference between waveform extents, assuming the characteristics of the transmitted laser pulses are the same in terms of horizontal energy distribution, transmitted waveform and pointing angle. A study of the variability in the characteristics of the transmitted pulses and possible corrections to the estimates is recommended for further research. Furthermore, footprint misalignments cause random errors on the order of meters in the change estimates, but these are thought to cancel out when averaging over a sufficiently large set of measurements. The number of measurements required depends on the accuracy needed and as the sample density is limited (along-track one every 172 m), only change phenomena occurring on large spatial scales can be detected with the GLAS. Furthermore, strong gradients or discontinuities in canopy height (e.g. at a forest-river boundary) can prevent the misalignment errors from canceling out when averaging over a local set of measurements (e.g. 101 pairs covering about 17.4 km along-track).

A revised method of delineating the waveform extent is proposed. The waveform extent is widened to include not only the signal between the first and last threshold-crossings, but also the signal between the first/last threshold-crossing and the last/first zero-crossing before/after it. This revision is thought to improve the accuracy of estimated change in maximum canopy height, since it yields significant improvements in reproducibility of the waveform extent. Retrievals of canopy structure variables from the signal between the estimated signal beginning and end are likely to be more accurate when the revised method is employed.

Subquestion iii: "Given a study site and a tolerance for measurement alignment, how are the repeated measurements spatially and temporally distributed?"

For norther Alberta (latitudes higher than 53.5°N) we found 22309 high-quality waveform pairs with a footprint center distance smaller than 150 m. Measurement time intervals smaller than 27 days (same-campaign pairs) are very rare (162 pairs) and found only at track crossover locations. The other waveform pairs span time intervals of either one or two years, approximately. We found 9140 pairs for the 2004-to-2005 interval, 5303 pairs for 2005-to-2006 and 7740 pairs for 2004-to-2006. For each of these interannual intervals, the number of pairs per square degree ranges from roughly 50 at a latitude of 53.5°N to roughly 250 at a latitude of 60°N.

Subquestion iv: "To what extent are the measurements reproducible?"

Reproducibility of the waveform extent depends on the method used to delineate the waveform extent, the time interval of remeasurement and the footprint center distance.

- The revision of the waveform extent delineation method (subquestion ii) improves reproducibility of waveform extent measurements. Any subsequently derived parameters are expected to yield improved reproducibility as a result of the revision.
- As the length of the time interval increases, the reproducibility decreases.
- As the footprint center distance increases, the reproducibility decreases.

The highest correlation coefficient was found for same-campaign pairs: about 0.84 (time interval of less than 27 days), where the unrevised method yielded a coefficient of 0.81. We assumed no changes in maximum canopy height during campaigns and between the 2004 and 2005 campaigns to find an upper limit for the expected errors due to footprint misalignments. As footprint center distance increases from 2.5 m to 150 m, RMSD increases from about 1 m to 7 m. Footprint misalignment is thus shown to be the strongest limiting factor for reproducibility of the waveform extent and the largest error source for the estimates of change in maximum canopy height from individual pairs.

Subquestion v: "Is diminishment of maximum canopy height as a result of wildfires detectable with the GLAS?"

Waveform pairs that are likely to be affected by wildfires have been identified based on wildfire perimeters and start/discovery dates. The 585 pairs in the set (either one-year or two-year time intervals) yielded a mean change in maximum canopy height of about -1 m. The estimated standard error of the mean was ± 0.3 m. Therefore, the one meter decrease is thought to be significant. Four wildfires have been analyzed individually and as the distributions of change estimates show both a spread due to footprint misalignments (symmetrical with zero-mean) and extra peaks for negative values, these distributions confirm that the diminishment of maximum canopy height due to wildfires was successfully detected. The estimated rates of change in maximum canopy height just outside the perimeters of the "Klewi" wildfire were shown to average to zero, while the pairs located within the perimeters yielded a mean of -1.4 m/year. The most extreme negative values were shown to be located over some of the most pronounced burn scars in LandsatLook Natural Color images for the "Klewi" wildfire. Changes in near-infrared reflectance of materials at the Earth surface due to wildfires are expected to have a minor effect on the change estimates, but this should be confirmed by further research.

Subquestion vi: "Do change measurement statistics suggest tree growth?"

The finding that the mean changes in maximum canopy height (estimated from waveform extent differences) occurring the first 30 years after wildfires were approximately twice as high for two-year interval pairs (2.1 m) as for one-year interval pairs (0.9 m) makes it plausible that these changes are in fact due to actual growth. The estimated changes occurring during a campaign (time interval smaller than 27 days) average to zero. Since average changes of roughly $+3$ m/year were also encountered for other measurement sets of about 10^4 pairs, factors other than height-growth are thought to contribute to these increases. The following can be investigated to shed further light on the problem:

- Horizontal growth of tree crowns. Given parameters describing crown widening, the spacing between individual crowns, the crown heights at which horizontal growth occurs and the footprint dimensions, how much does horizontal growth contribute to mean changes in maximum canopy height estimated from waveform pairs?
- Instrumental changes. How have the total energy, shape (transmitted waveform) and horizontal energy distribution of the transmitted laser pulses changed significantly over time, particularly in between campaigns? How may these changes have influenced the estimated changes in maximum canopy height and how can these effects be corrected for?
- Anthropogenic impact. Will exclusion of croplands, production forests and built-up environments from the study region affect the mean estimated changes significantly?

These factors will have to be studied before any conclusions concerning tree growth can be drawn from the results presented in this report.

Furthermore, this study has covered only changes in maximum canopy height, estimated from changes in waveform extent. It will be valuable to also involve the following biophysical variables in further research:

- Maximum canopy height itself, as opposed to only changes
- Canopy cover
- The vertical distribution of horizontally projected plant area

These can be retrieved from the waveform. If the ground return can be successfully identified and fitted with a Gaussian (assuming flat land and a Gaussian transmit pulse), the elevation difference corresponding to the interval between signal beginning and the ground return centroid gives the maximum canopy height. The canopy cover can be estimated as the integrated signal of the waveform with ground return removed, divided by the integrated original waveform. Lefsky et al. [22] describe how the vertical distribution of horizontally projected plant area (the canopy height profile) can be derived after removal of the ground return. By incorporating these variables in further research, a more complete picture of changes in canopy structure may be formed.

Bibliography

[1] Abshire, J.B.; Sun, X.; Riris, H.; Sirota, J.M.; McGarry, J.F.; Palm, S.; Yi, D.; Liiva, P. (2005) Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance. *Geophysical Research Letters 31* (21): November 2005

[2] Brandt, J.P.; Flannigan, M.D.; Maynard, D.G.; Thompson, I.D.; Volney, W.J.A. (2013) An introduction to Canada's boreal zone: ecosystem processes, health, sustainability, and environmental issues. *Environmental Reviews 21* (4): 207-226.

[3] Brenner, A.C., Zwally, H.J., Bentley, C.R., Csathó, B.M., Harding, D.J., Hofton, M.A., Minster, J., Roberts ,L., Saba, J.L., Thomas, R.H., & Yi, D. (2011) Geoscience Laser Altimeter System (GLAS) algorithm theoretical basis document: derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights. *Algorithm Theoretical Basis Documents (ATBD)*, 92pp.

[4] Brenner, A. C., DiMarzio, J. P., & Zwally, H. J. (2007) Precision and accuracy of satellite radar and laser altimeter data over the continental ice sheets. *Geoscience and Remote Sensing, IEEE Transactions on, 45*(2), 321-331. doi:10.1109/TGRS.2006.887172 0196- 2892

[5] Canadian Forest Service - Natural Resources Canada (2015) Canadian Wildland Fire Information System - Canadian National Fire Database. *Retrieved from http://cwfis.cfs.nrcan.gc.ca/ha/nfdb*

[6] Canadian Soil Information System - Agriculture and Agri-Food Canada National Soil DataBase (NSDB) - A National Ecological Framework for Canada. *Retrieved from http://sis.agr.gc.ca/cansis/nsdb/ecostrat/gis_data.html*

[7] Carroll, M., Townshend, J., DiMiceli, C., Noojipady, & P., Sohlberg, R. (2009). A New Global Raster Water Mask at 250 Meter Resolution. *International Journal of Digital Earth 2*(4).

[8] Collins, B.M., Stephens, S.L. (2010) Stand-replacing patches within a mixed severity fire regime: quantitative characterization using recent fires in a long-established natural fire area. *Landscape Ecologie 25*, 927-939 DOI 10.1007/s10980-010-9470-5

[9] Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., & Blair, J. B. (2002) Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment, 81*(2), 378-392.

[10] Duong, H.V. (2010) Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data. PhD Thesis, Delft University of Technology

[11] Duryea, M. L., & Malavasi, M. M. (1998) How Trees Grow in the Urban Environment. *Retrieved from https://edis.ifas.ufl.edu/fr002*

[12] Ecological Stratification Working Group (1996) A national ecological framework for Canada. *Agriculture and Agri-Food Canada and Environment Canada, Ottawa.*

[13] Flannigan, M. D., Amiro, B. D., Logan, K. A., Stocks, B. J., & Wotton, B. M. (2006) Forest fires and climate change in the 21st century. *Mitigation and adaptation strategies for global change, 11* (4), 847-859.

[14] Gamache, I. & Payette, S. (2004) Height growth response of tree line black spruce to recent climate warming across the forest-tundra of eastern Canada. *Journal of Ecology, 92*, 835-845. doi:10.1111/j.0022-0477.2004.00913.x

[15] Geospatial Data Abstraction Library (GDAL) GDAL Tool "ogr2ogr". *Retrieved from http://www.gdal.org/ogr2ogr.html*

[16] Hall, Dorothy K., George A. Riggs, & Vincent V. Salomonson. (2007), updated weekly. MODIS/Aqua Snow Cover 8-Day L3 Global 500m Grid V005, February 2003 to October 2009. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

[17] Gutsell, S. L., & Johnson, E. A. (2002) Accurately ageing trees and examining their height-growth rates: implications for interpreting forest dynamics. *Journal of Ecology, 90* (1), 153-166.

[18] Harpold, R., Urban, T., Webb, C., & Schutz, B. (2007) Assessment of ICESat repeat track estimation techniques for polar elevation change detection. *AGU 650 Fall Meeting 2007* Abstract #C23A-0943.

[19] Heinselman, M.L. (1981) Fire intensity and frequency as factors in the distribution and structure of northern ecosystems. *p. 7-57 in Proceedings of the Conference: Fire Regimes in Ecosystem Properties*, Dec. 1978, Honolulu, Hawaii. USDA, For. Serv., Washington DC, Gen. Tech. Rep. WO-26.

[20] Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E. 2008. Hole-filled SRTM for the globe Version 4, available from the *CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org*

[21] Lee, J.E. (2013) The GLAS Science Algorithm Software (GSAS) User's Guide Version 7. *ICESat (GLAS) Science Processing Software Document Series NASA/TM 2013-208641/Vol 17*

[22] Lefsky, M. A. (2010), A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Resears Letters 37*, L15401, doi:10.1029/2010GL043622.

[23] Mitchard, E. T., Feldpausch, T. R., Brienen, R. J., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., ... & Steege, H. (2014) Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Global Ecology and Biogeography, 23* (8), 935-946.

[24] Moffet, M.W. () What's "Up"? A Critical Look at the Basic Terms of Canopy Biology. *Biotropica 32* (4a): 569-596 2000

[25] Molijn, R.A., Lindenbergh, R.C., & Gunter, B.C. (2010) ICESat laser full waveform analysis for the classification of land cover types over the cryosphere. *International Journal of Remote Sensing 32* (23), 8799-8822.

[26] NASA Distributed Active Archive Center (DAAC) at NSIDC - ICESat / GLAS Data Laser Operational Periods. *Retrieved from http://nsidc.org/data/icesat/laser_op_periods.html*

[27] NASA Distributed Active Archive Center (DAAC) at NSIDC - ICESat / GLAS Data Order Data. *Retrieved from https://nsidc.org/data/icesat/order.html*

[28] NASA EOSDIS Reverb ECHO - The Next Generation Earth Science Discovery Tool. *Retrieved from reverb.echo.nasa.gov*

[29] NSICDC GLAS Science Algorithm Software (GSAS) Development Team NSICDC Fortran reader. *Retrieved from ftp://sidads.colorado.edu/pub/DATASETS/icesat/tools/f90/reader/GLA01_wf_mod.f90*

[30] Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., ... & Ciais, P. (2011) A large and persistent carbon sink in the world's forests. *Science, 333* (6045), 988-993.

[31] Phan, V.H., Lindenbergh, R.C., & Menenti, M. (2011) ICESat derived elevation changes of Tibetan lakes between 2003 and 2009. *International Journal of Applied Earth Observation and Geoinformation 17*, 12-22.

[32] Selkowitz, D.J., Green, G., Peterson, B., Wylie, B. (2012) A multi-sensor lidar, multi-spectral and multi-angular approach for mapping canopy height in boreal forest regions. *Remote Sensing of Environment 121*, 458-471

[33] Siegmund, A., Menz, Z. (2005) Spectral signatures of soil, vegetation and water, from: Fernes nah gebracht Satelliten- und Luftbildeinsatz zur Analyse von Umweltveränderungen im Geographieunterricht, *Geographie und Schule*, 154: 7, *Retrieved from http://www.seos-project.eu/modules/remotesensing/remotesensing-c09-p01.html#Siegmund2005*

[34] Slobbe, D.C., Lindenbergh, R.C., & Ditmar, P. (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.

[35] Smith, B. E., Bentley, C. R., & Raymond, C. F. (2005) Recent elevation changes on the ice streams and ridges of the Ross Embayment from ICESat crossovers. *Geophysical Research Letters, 32*, doi:10.1029/20050L024365

[36] Ter-Mikaelian, M. T., & Korzukhin, M. D. (1997) Biomass equations for sixty-five North American tree species. *Forest Ecology and Management, 97* (1), 1-24.

[37] United States Geological Survey (USGS) (2016) LandsatLook Viewer *Retrieved from https://landsatlook.usgs.gov/*

[38] Yang, W., Ni-Meister, W., Lee, S. (2011) Assessment of the impacts of surface topography, off-nadir pointing and vegetation structure on vegetation lidar waveforms using an extended geometric optical and radiative transfer model. *Remote Sensing of Environment 115*, 2810-2822

[39] Zwally, H. J., R. Schutz, C. Bentley, J. Bufton, T. Herring, J. Minster, J. Spinhirne, & R. Thomas. 2011. GLAS/ICESat L1A Global Altimetry Data, Version 33. Feb 2003 to Oct 2009, bounding coordinates (53.5 °N, 120 °W); (60 °N, 110 °W). Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. Retrieved from http://dx.doi.org/10.5067/ICESAT/GLAS/DATA121

[40] Zwally, H. J., R. Schutz, C. Bentley, J. Bufton, T. Herring, J. Minster, J. Spinhirne, & R. Thomas. 2014. GLAS/ICESat L2 Global Land Surface Altimetry Data, Version 34. Feb 2003 to Oct 2009, bounding coordinates (53.5 °N, 120 °W); (60 °N, 110 °W). Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. *Retrieved from http://dx.doi.org/10.5067/ICESAT/GLAS/DATA227*

Appendices

Appendix I: Matlab script "readGLA01.m"

```
function [WF, extent] = readGLA01(Rec,shot)

% The function readGLA01 reads binary GLA01 files. Given a list of record indices —Rec—
% and a list of shot number —shot—, the outputs are a cell—array of waveforms —WF—
% (decompressed, converted to volts and noise removed) and a list of waveform extent
% values for the given record indices and shot numbers. The reading function works
% through all files in the current directory that start with 'sGLA01_'.

filepath = ls('sGLA01_*');
filepath = strread(filepath,'%s','delimiter',' ');


[cnt2volt1 cnt2volt2] = cnt2volttables(); % load the counts—to—volt conversion tables


WF = cell(length(Rec),1);
extent = nan(length(Rec),1);
searchrec = unique(Rec);

for fcurrent = filepath'

    file = fopen(fcurrent{1},'r','b');
    fread(file,4660,'char=>char')'; % header record
    strread(ans, '%s', 'delimiter', '\n'); % lines to cells
    strread(ans{2},'NUMHEAD= %d;'); % read second line/cell, number of headers
    fseek(file,4660*(ans—1),'cof'); % skip remaining headers headers

    while ~feof(file)
        recndx = fread(file,1,'int=>int'); % recndx
        if isempty(recndx); break; end; % last lines may be empty, so quit
        if ~ismember(recndx,searchrec) % if the record is not wanted
            fseek(file,4660—4,'cof'); % skip to next record
            continue % and restart record—reading process
        end
        recid = Rec == recndx;

        %UTCtime = fread(file,2,'int=>int'); % read UTC seconds and fraction of seconds
        fseek(file,2*4,'cof');  % skip UTC time, skip to record type
        rtype = fread(file,1,'int16=>int'); % record type

        if rtype==2 % if record type is "LONG"
            subrec = subrec+1;
            %fprintf(' L') % notice for long record encountered
            fseek(file,120—14,'cof'); % skip to offset 120
            BgMean = fread(file,8,'int16=>double'); % and there read background noise mean
            BgSDEV = fread(file,8,'int16=>double'); % and background noise standard deviation
            fseek(file,16,'cof'); % skip to offset 168
            comptype = fread(file,1,'int8=>int8',7); % and there read the compression type,
            %0 for Npq—type, 1 for R—type
            wf = fread(file,[544 8],'uint8=>double'); % and then 8 long waveforms

            % WAVEFORM PROCESSING
            if comptype==0 & (P~=1 | Q~=1)
                 wf = [kron(wf(1:N,:), ones(P,1)); kron(wf(N+1:end,:), ones(Q,1))];
```

```
        % decrompress, Npq-type compression
    end
    if comptype==1 & R~=1
        %fprintf(['\nR = ' num2str(R)])
        wf = kron(wf, ones(R,1));% decrompess R-type compression
        if size(wf,1)>1000
            wf = wf(1:1000,:); % max 1000 bins per waveform
        end
    end
    % wf(544:-1:1,:) % reverse the waveform if required
    if instate(28)==1; % if primary digitizer was used
        wf = cnt2volt1(wf+1);
        % convert these to volts using the count-to-volt table for digitizer 1
        BgMean = 0.006675*BgMean*0.01 - 0.19528;
        % as well as the noise parameters. eq. from GSAS users guide
        BgSDEV = 0.006675*BgSDEV*0.01;
        % - 0.19528; % std is an interval, so no offset in conversion to volt
    else
        fprintf('Digitizer 2 used\n')
        wf = cnt2volt2(wf+1);
        % convert these to volts using the count-to-volt table for digitizer 2
        BgMean = 0.006625*BgMean*0.01 - 0.19383; % as well as the noise parameters
        BgSDEV = 0.006625*BgSDEV*0.01;
    end

    for i = 1:8
        wf(:,i) = wf(:,i)-BgMean(i); % adjust wf so that mean noise is now zero
        SigBeg = find( wf(:,i)>5*BgSDEV(i) ,1,'last' ); %  first threshold crossing
        SigEnd = find( wf(:,i)>5*BgSDEV(i) ,1,'first' ); %  last threshold crossing
        if isempty(SigBeg)
            SigBeg = 0;
        else
            ToZero = find( wf(SigBeg+1:end,i)<=0,1,'first' );
            % number of gates to first zero crossing from SigBeg
            if ~isempty(ToZero) % if a zero crossing is found
                SigBeg = SigBeg + ToZero - 1 ;
                % first zero crossing above threshold crossing is signal begin
            end
        end
        if isempty(SigEnd)
            SigEnd = 0;
        else
            SigEnd = find( wf(1:SigEnd-1,i)<=0,1,'last' );
            % SigEnd adjusted to last zero crossing after noise crossing
        end
        wf(SigBeg+1:end,i) = 0;
        % gates higher than SigBeg are set to zero; if no crossing is found,
        % all is set to zero.
        if ~isempty(SigEnd) % if a zero crossing is found
            wf(1:SigEnd,i) = 0;
            % gates SigEnd and those after SigEnd are set to zero;
            % if no crossing is found, none are set to zero.
        end
    end

    t = [0:size(wf,1)-1]';
    d = t * 299792458 / 1e9 / 2;

    shotid = find(recid & shot<=8*subrec & shot>8*(subrec-1));
    for i = 1:length(shotid)
        WF{shotid(i)} = wf(:,shot(shotid(i))-(subrec-1)*8);
        if sum(WF{shotid(i)}>0)>0
            extent(shotid(i)) = d(find(WF{shotid(i)}>0,1,'last'))
            - d(find(WF{shotid(i)}>0,1,'first'));
        else
            extent(shotid(i)) = 0;
        end
    end
    fseek(file,4660-4528,'cof'); % skip to next record
    continue % and restart record-reading process
end

if rtype==1 % if record type is "MAIN"
```

```
    %fprintf('\n HIT')
    subrec = 0;
    fseek(file,2624-14,'cof'); % skip to instrument state bytes
    inststate = fread(file,32,'ubit1=>int'); %IS(14)=primary detector;
    % IS(13)=secondary detector; IS(28)=primary digitizer; IS(27)=secondary digitizer
    fseek(file,2700-2628,'cof'); % skip to compression variables
    P = fread(file,1,'int16=>int16'); % read compression ratio P
    Q = fread(file,1,'int16=>int16'); % compression ratio Q
    N = fread(file,1,'int16=>int16'); % N value
    R = fread(file,1,'int16=>int16'); % and R value
    fseek(file,4660-2708,'cof'); % skip to next record
    continue % and restart record-reading process
end



if rtype==3 % if record type is "SHORT"
    subrec = subrec+1;
    fprintf(' Short record type encountered')

    fseek(file,276-14,'cof'); % skip to offset 120
    BgMean = fread(file,20,'int16=>double'); % and there read background noise mean
    BgSDEV = fread(file,20,'int16=>double'); % and background noise standard deviation

    fseek(file,40,'cof'); % skip to offset 396
    comptype = fread(file,1,'int8=>int8',19); % and there read the compression type
    wf = fread(file,[200 20],'uint8=>double'); % and then 20 short waveforms

    % WAVEFORM PROCESSING
    if comptype==0 & (P~=1 | Q~=1)
         wf = [kron(wf(1:N,:), ones(P,1)); kron(wf(N+1:end,:), ones(Q,1))];
         % decrompress, Npq-type compression
    end
    if comptype==1 & R~=1
         wf = kron(wf, ones(R,1));% decrompess R-type compression
         if size(wf,1)>1000
            wf = wf(1:1000,:); % max 1000 bins per waveform
         end
    end
    % wf(544:-1:1,:) % reverse the waveform
    if inststate(28)==1; % if primary digitizer was used
        wf = cnt2volt1(wf+1);
        % convert these to volts using the count-to-volt table for digitizer 1
        BgMean = 0.006675*BgMean*0.01 - 0.19528;
        % as well as the noise parameters. eq. from GSAS users guide
        BgSDEV = 0.006675*BgSDEV*0.01;% - 0.19528;
        % std is an interval, so no offset in conversion to volt
    else
        fprintf('Digitizer 2 used\n')
        wf = cnt2volt2(wf+1);
        % convert these to volts using the count-to-volt table for digitizer 2
        BgMean = 0.006625*BgMean*0.01 - 0.19383; % as well as the noise parameters
        BgSDEV = 0.006625*BgSDEV*0.01;
        % - 0.19383; % std is an interval, so no offset in conversion to volt
    end


    for i = 1:20
        wf(:,i) = wf(:,i)-BgMean(i); % adjust wf so that mean noise is now zero
        SigBeg = find( wf(:,i)>5*BgSDEV(i) ,1,'last' ); %  first threshold crossing
        SigEnd = find( wf(:,i)>5*BgSDEV(i) ,1,'first' ); %  last threshold crossing
        if isempty(SigBeg)
            SigBeg = 0;
        else
            ToZero = find( wf(SigBeg+1:end,i)<=0,1,'first' );
            % number of gates to first zero crossing from SigBeg
            if ~isempty(ToZero) % if a zero crossing is found
                SigBeg = SigBeg + ToZero - 1 ;
                % first zero crossing above threshold crossing is signal begin
            end
        end
        if isempty(SigEnd)
            SigEnd = 0;
```

```
                else
                    SigEnd = find( wf(1:SigEnd-1,i)<=0,1,'last' );
                    % SigEnd adjusted to last zero crossing after noise crossing
                end
                wf(SigBeg+1:end,i) = 0;
                % gates higher than SigBeg are set to zero; if no crossing is found,
                % all is set to zero.
                if ~isempty(SigEnd) % if a zero crossing is found
                    wf(1:SigEnd,i) = 0;
                    % gates SigEnd and those after SigEnd are set to zero;
                    % if no crossing is found, none are set to zero.
                end
            end

            t = [0:size(wf,1)-1]';
            d = t * 299792458 / 1e9 / 2;

            shotid = find(recid & shot<=20*subrec & shot>20*(subrec-1));
            for i = 1:length(shotid)
                WF{shotid(i)} = wf(:,shot(shotid(i))-(subrec-1)*20);
                if sum(WF{shotid(i)}>0)>0
                    extent(shotid(i)) = d(find(WF{shotid(i)}>0,1,'last')) -
                    d(find(WF{shotid(i)}>0,1,'first'));
                else
                    extent(shotid(i)) = 0;
                end
            end
            fseek(file,4660-4416,'cof'); % skip to next record
            continue % and restart record-reading process
        end
        fprintf('Unknown record type encountered')
    end
    fclose(file);

end


end
```