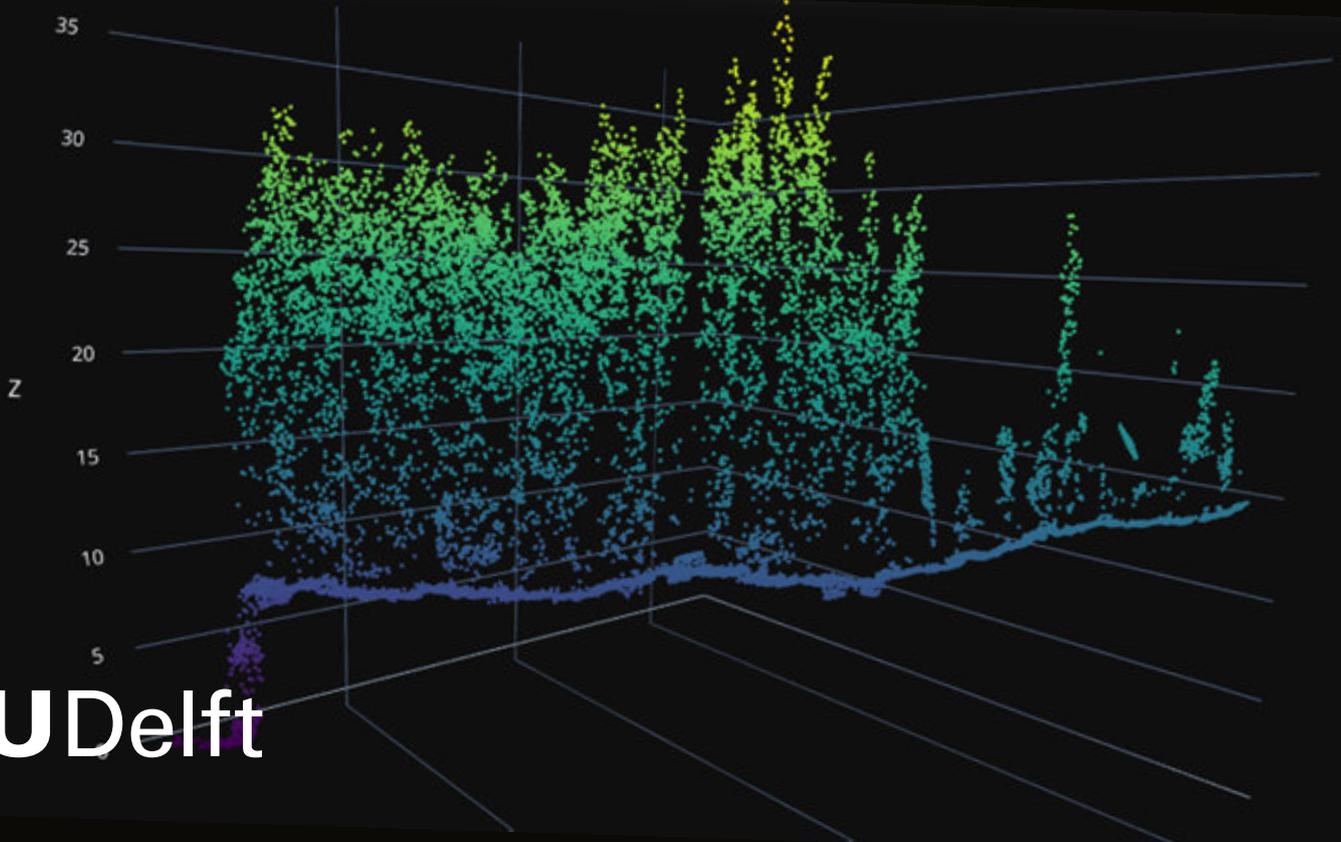


MSc thesis in Geomatics

# Canopy gap fraction estimation from ICESat-2 ATL08 product

Katrin Meschin

2023





MSc thesis in Geomatics

# **Canopy gap fraction estimation from ICESat-2 ATL08 product**

Katrin Meschin

April 2023

A thesis submitted to the Delft University of Technology in  
partial fulfillment of the requirements for the degree of Master  
of Science in Geomatics

Katrin Meschin: *Canopy gap fraction estimation from ICESat-2 ATL08 product* (2023)  
© ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The work in this thesis was carried out in the:



3D geoinformation group  
Delft University of Technology

Supervisors: Maarten Pronk  
Dr. Hugo Ledoux

# Abstract

With the increasing pressure on global forests due to deforestation and habitat loss, it is more important than ever to understand our forest ecosystems on a larger scale. Canopy gap fraction is an indicator used for estimating forest biomass and better understanding of ecosystem functioning. For decades there have been research on computing canopy gap fraction using ground measurements as well as Airborne Laser Scanning (ALS). These approaches, however, are limited by the data available. A global coverage of data on canopy gap fraction could be made available by space-based laser altimetry mission Ice, Cloud and land Elevation Satellite (ICESat-2). Although its main scientific objectives are focused on polar areas, it has already proven to facilitate broader scientific disciplines. The ATL08 data product is focused on land-vegetation and already provides global data on canopy heights. However, there is ongoing research for using ATL08 data for canopy gap fraction estimation. Although some approaches have been suggested in the literature, a tested workflow to achieve this goal has not been published.

This thesis tests two methods for estimating canopy gap fraction from ICESat-2 ATL08 data and evaluates the results against openly available ALS data. First, a simple method of using canopy to total photon ratio is used. Then, an alternative method that aims to correct for the surface reflectivity is tested. The results from both methods are similar, therefore the computationally less expensive method is recommended. Although this thesis does not achieve to present a sufficiently accurate approach for using ICESat-2 ATL08 data for canopy structure estimation, it is shown that further research is needed and the results are promising. Furthermore, it is demonstrated that annual trends in canopy gap fraction can be seen in ATL08 data. Considering the global coverage of ICESat-2 data, it is concluded that despite the accuracy not meeting the expectations, using ATL08 for studying canopy gap fraction on a global scale and through time has high value and great potential for environmental research.



# Acknowledgements

I would like to take a moment to show my gratitude towards people that supported me not only through the process of finishing this thesis, but since I started my journey in the Geomatics program in 2019.

I would first like to thank my supervisor Maarten Pronk for giving an excellent lecture on ICESat-2 in the Digital Terrain Modelling course which sparked my interest in this topic. I would not have made it far without his knowledge and support throughout the process from finding a topic to finally finishing my thesis, pushing me when I was slacking and encouraging me when I was doubting myself. I am also sincerely thankful for Hugo Ledoux who together with Maarten gave me constructive feedback to guide me to the right direction. I am lucky to be in a department that allowed me to pick a topic that interests me, even though it has little to do with the built environment. I realise that I made it more difficult for my supervisors and co-readers who now have to dive into the concepts about tree canopies and forest density. However, the skills needed for analysing the built and the natural environment are similar and personally I am unable to decide which one I like to study more.

I would like to extend my gratitude to the entire teaching staff in the MSc Geomatics program for the exceptional passion and dedication you have. The energy you put into teaching and the will you have to constantly improve the courses really is appreciated by the students, although it is known that students are often lousy at showing it. Even more so, I can say from my experience that the quality of teaching stayed high before, during and after the pandemic. I cannot imagine how difficult it was to switch lecture halls where you could meet the students to Zoom calls where everyone was muted with their cameras turned off, while your own personal lives were also most likely thrown upside down. Somehow through all this, you managed to keep the high standards and your passion was contagious even through the online learning.

I do not expect my friends and family to ever read my work on ICESat-2 data or even understand fully what I studied in Delft. I did, however, had a lot of support from my closest ones. If any of them ever reads these lines - thank you for listening me when I complained about my deadlines and for your patience when I had to cancel my plans. It meant the world for me to have people by my side throughout the journey.



# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Background and Motivation . . . . .	1
1.2. Research Objectives . . . . .	3
1.3. Research Scope and Challenges . . . . .	3
1.4. Thesis Outline . . . . .	3
<b>2. Related work</b>	<b>5</b>
2.1. The concept of canopy gap fraction . . . . .	5
2.1.1. Field measurements of canopy gap fraction . . . . .	5
2.1.2. Measuring canopy gap fraction from ALS . . . . .	6
2.2. Overview of ICESat-2 ATL08 data product . . . . .	6
2.3. Validating ATL08 data . . . . .	7
2.4. ATL08 for canopy structure . . . . .	8
2.4.1. Canopy gap fraction from ATL08 data . . . . .	9
<b>3. Methodology</b>	<b>13</b>
3.1. Study Area . . . . .	13
3.2. ICESat-2 data pre-processing . . . . .	13
3.2.1. ATL08 data filtering . . . . .	14
3.2.2. ICESat-2 reflectivity analysis through radiometric histograms . . . . .	16
3.3. ALS data pre-processing . . . . .	17
3.3.1. The ALS data available . . . . .	17
3.3.2. Retrieval and processing of ALS data . . . . .	17
3.4. Canopy gap from ALS . . . . .	18
3.5. Validating the ALS data . . . . .	19
3.5.1. Validating the canopy gap fraction estimates from ALS data against ground data . . . . .	19
3.5.2. Validating canopy height estimation from ALS and ATL08 . . . . .	19
3.6. Canopy gap from ATL08 . . . . .	20
3.6.1. Canopy to total photon ratio . . . . .	20
3.6.2. Accounting for the surface reflectance in canopy gap fraction computation . . . . .	21
3.6.3. Finding the trends in canopy gap fraction results . . . . .	23
3.6.4. Error estimation . . . . .	23
3.6.5. Implementation . . . . .	23
<b>4. Results</b>	<b>25</b>
4.1. Data validation . . . . .	25
4.1.1. Analysis of the radiometric histograms . . . . .	25
4.1.2. Validation of canopy gap fraction from ALS as reference data . . . . .	26
4.1.3. Validation of canopy height estimation from ALS and ATL08 . . . . .	28
4.2. ATL08 canopy gap fraction from canopy to total photon ratio . . . . .	29

*Contents*

4.3. ATL08 canopy gap fraction using radiometric profile . . . . .	30
4.4. Trends in canopy gap fraction estimated from ATL08 . . . . .	31
4.4.1. Canopy gap fraction in different forest types . . . . .	32
4.4.2. Canopy gap fraction throughout the year . . . . .	32
<b>5. Discussion</b>	<b>35</b>
<b>6. Conclusion and future work</b>	<b>37</b>
<b>A. Reproducibility self-assessment</b>	<b>39</b>
A.1. Marks for each of the criteria . . . . .	39
A.2. Self-reflection . . . . .	39

# List of Figures

1.1. Canopy gap fraction describes the amount of light reaching the forest floor. It is a unitless measure from 0 to 1 with 0 indicating full vegetation coverage with thick canopy and 1 indicating absence of vegetation. . . . .	2
2.1. Illustration comparing the ALS point cloud data with ICESat-2 photon data in ATL03 and ATL08 data products. . . . .	7
2.2. Beam configuration for ATLAS instrument on ICESat-2 with strong beams (1, 3, 5) and weak beams (2, 4, 6) which fourth of the energy. Spot 7 on the diagram is a virtual point representative of the sub-satellite point. Figure from <a href="#">Neuenschwander et al. 2022a</a> . . . . .	8
2.3. Various modalities of lidar detection. The photon counting probability distribution function (PDF) shown is illustrating a theoretical situation if hundreds or more photons were reflected from a target. Figure from <a href="#">Neuenschwander et al. 2022b</a> . . . . .	9
2.4. Illustration on how surface reflectance can alter the canopy cover estimation from ATL08. Taken from <a href="#">Neuenschwander et al. 2022a</a> . . . . .	10
2.5. Radiometric profiles by <a href="#">Neuenschwander et al. 2022a</a> for six study sites. The green lines were inserted manually by the authors to highlight the trend. Taken from <a href="#">Neuenschwander et al. 2022a</a> . . . . .	11
3.1. Overview of workflow used in this thesis. . . . .	13
3.2. Estonia in Europe and a map of Estonia with the ICESat-2 transect polygons used in this thesis . . . . .	14
3.3. Forest types in Estonia. . . . .	15
3.4. Workflow for creating polygon bounding box around each ATL08 segment and using it to clip ALS point cloud for each segment. . . . .	16
3.5. ALS data used in the thesis is open data by Estonian Land Board, in each year covering a quarter of Estonia. Figure adapted from Estonian Land Board. . . .	18
3.6. A map showing the location of Järvelja study site from where <a href="#">Kuusk et al. 2018</a> reports canopy gap fraction measurements in three forest stands which are 100 x 100 meters in size. This thesis computed the canopy gap fraction from ALS point clouds clipped to 12 x 100 meter plots inside these three study sites to validate the ALS results against values reported by <a href="#">Kuusk et al. 2018</a> . . . . .	20
4.1. Canopy and ground radiometric histograms for strong and weak beams. . . .	26
4.2. Comparison of canopy gap estimations computed for Järvelja test plots from summer and spring ALS. . . . .	27

*List of Figures*

4.3. Difference in canopy height estimation from ALS and ATL08 data. On the right all ATL08 segments are included while in the middle only segments from the night acquisition are added. The violin plot on the right shows the spread of data with only night acquisition and segments with canopy height difference of less than 10 meters. . . . .	28
4.4. Canopy to total photon ratio (Equation 3.3) plotted against canopy gap fraction (Equation 3.1) on the left and SCI (Eq 3.2) on the right for each segment. .	29
4.5. Radiometric profile of Estonia including all segments from clear August night	30
4.6. Canopy gap fraction computed from ATL08 using radiometric profile to solve Equation 3.4 plotted against canopy gap fraction from ALS (Equation 3.1) on the left and SCI (Eq 3.2) on the right . . . . .	31
4.7. Canopy gap fraction estimation from ATL08 in different tree species. The number under the tree species name indicates the number of segments included in the analysis. . . . .	32
4.8. Canopy gap estimation from ATL08 in different months by strong and weak beam. . . . .	33
4.9. Canopy gap fraction computed from ATL08 for the entire year in spruce (top) and birch (bottom) forest. . . . .	34

## List of Tables

4.1. Modes of ground and canopy radiometric histograms in different conditions. .	25
4.2. Canopy gap fraction estimation from summer ALS data using Equation 3.1 compared with the results shown in Kuusk et al. 2018. . . . .	28
4.3. Statistics on canopy height estimation from ALS dataset and ATL08 product. .	29
4.4. Statistics on canopy gap fraction estimation from ATL08 product compared to canopy gap fraction (CGF) computed from ALS (Eq 3.1) and SCI (Eq 3.2) . . . .	31



# Acronyms

ALS	Airborne Laser Scanning	v
TLS	Terrestrial Laser Scanning	6
LiDAR	light detection and ranging	1
ICESat-2	Ice, Cloud and land Elevation Satellite	v
ATLAS	Advanced Topographic Laser Altimetry System)	6
GEDI	Global Ecosystem Dynamics Investigation	1
QGIS	Quantum Geographic Information System	23
LPI	Light Penetration Index	18
SCI	Solberg's Cover Index	18
RANSAC	RANdom SAMpling Consensus	22
RMSE	Root Mean Square Error	23
MAD	Median Absolute Deviation	23
ATBD	Algorithm Theoretical Basis Document	7



# 1. Introduction

## 1.1. Background and Motivation

Forests are important habitat for global biodiversity and play a key role in the Earth's carbon cycle, a main regulator of the Anthropogenic climate change. Therefore, reliable data on the state of forests on a global scale is crucial for understanding and mitigating the climate change and biodiversity loss. This has led to a shift in forestry research from the traditional focus on timber resource monitoring to increased focus on ecological aspects of forests (Korhonen and Morsdorf 2014).

One of the key indicators needed for estimating the biomass and for better understanding of ecosystem functioning is the canopy structure of the forests (Ozanne et al. 2003). However, ground measurements of the canopy are laborious and expensive, therefore canopy structure is commonly measured by remote sensing (Korhonen and Morsdorf 2014). Specifically, light detection and ranging (LiDAR) technologies from ALS has proven to be superior over passive optical imagery as the quality of LiDAR data in forest environment does not depend on light and shadow conditions and the three-dimensional data produced allows better understanding of the canopy structure (Tang et al. 2019). Although countries and administrative units are increasingly providing more ALS data, the cost of such programs is still high, especially for large countries such as Brazil, Russia or Canada where large global forest reserves are located. Hence, a reliable data on forest structure on global scale is missing.

Space-based laser altimetry has potential to map vegetation structure at global scale that cannot be achieved with airborne lidar systems. Specifically designed for this task is NASA's mission Global Ecosystem Dynamics Investigation (GEDI) that has been collecting measurements since April 2019, using full waveform lidar. However, GEDI does not collect data above the latitudes of  $52^\circ$ , leaving most of the northern hemisphere and boreal forests uncovered. The higher latitudes are covered by the space-based laser altimetry mission ICESat-2 which was launched by NASA in 2018 with the main objective to record elevation time series in the cryosphere. In addition to the main scientific objective, ICESat-2 system has proven to facilitate a broad range of scientific disciplines such as the land-vegetation along track product (ATL08) which provides canopy height estimates. In addition, a product for gridded canopy height and canopy cover (ATL18) is still in production and there is scientific interest in developing a model for estimating canopy structure from ICESat-2 data. Contrary to GEDI, ICESat-2 does not use full waveform, but instead single photon lidar, which allows higher repetition rates and increases the spatial resolution (Neumann et al. 2019), but complicates the canopy structure estimation.

Forest canopy structure is a broad term and can be described through different traits, such as foliage height diversity, vertical distribution of plant material and many more. This the-

## 1. Introduction

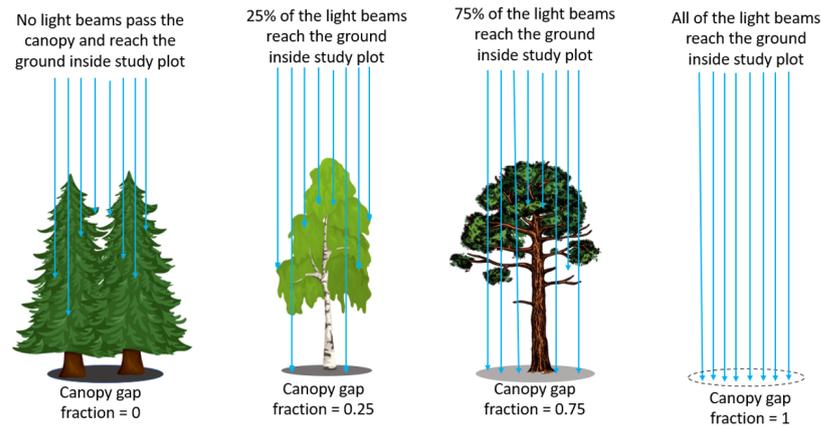


Figure 1.1.: Canopy gap fraction describes the amount of light reaching the forest floor. It is a unitless measure from 0 to 1 with 0 indicating full vegetation coverage with thick canopy and 1 indicating absence of vegetation.

sis focuses on canopy gap fraction which is a measure of forest canopy density and how much light reaches through the canopy to forest floor. Canopy gap fraction is defined as the probability of beam penetrating through the canopy and intercepting the ground given that the beam zenith angle is fixed at nadir (Fisher et al. 2020). It is a unitless measure within a range from 0 to 1. If the forest is dense, not much light from vertical angle can reach the forest floor, resulting in canopy gap fraction close to 0 while sparse forest has canopy gap fraction value close to 1 (Figure 1.1). Understanding canopy gap fraction allows estimation of the forest biomass, understanding of the habitat in the forest and it is a good indication of forest age and growth status. However, estimating canopy gap fraction from ICESat-2 data has not been accomplished yet, although some theories about a possible approach exist in the literature.

This thesis looks at two suggested methods for determining canopy gap fraction from ICESat-2 ATL08 product. The first is using the canopy to total photon ratio as described by Neuenschwander et al. 2022b. The other is described by Neuenschwander et al. 2022a proposing a method using ICESat-2 radiometric profiles for correcting for the surface reflectivity. This thesis aims to test both of those methods and compare the results with canopy gap fraction values retrieved from ALS data. The study is focused on Estonia, a country located in Northeastern Europe, between latitudes 57°N and 59°N. About half of Estonian territory is covered by forest which is a mix of deciduous and conifer species (Lang et al. 2018). There is ALS data freely available by Estonian Land Board. The main aim of this study is to test the methods for estimating canopy gap probability using ICESat-2 ATL08 data.

## 1.2. Research Objectives

The main objective of this thesis is test if it is possible to compute canopy gap fraction from ICESat-2 ATL08 product using two proposed approaches from the literature. Therefore, the key research question raised is:

*To what extent can canopy gap fraction be estimated from ICESat-2 ATL08 product?*

The results are compared with canopy gap fraction estimations derived from ALS point clouds. To achieve this goal, several relevant sub-questions are raised as follows:

- What are the optimal environmental conditions for ICESat-2 data acquisition that allow canopy structure estimation from ATL08? Answered in Chapters 4.1.1 and 4.1.3.
- Out of the two methods used in this thesis for estimating canopy gap fraction from ICESat-2 data, which performs better? Answered in Chapter 5.
- Does the canopy gap fraction derived from ATL08 reflect differences in different forest types? Answered in Chapter 4.4.1.
- To what extent does the canopy gap fraction derived from ATL08 reflect the changes in forest structure throughout the year? Answered in Chapter 4.4.2.
- To what extent is the ALS data provided by Estonian Land Board suitable for validating canopy gap fraction estimation from ATL08? Answered in Chapter 5.
- How could the methods for computing canopy gap fraction from ICESat-2 ATL08 data be further improved? Answered in Chapter 5 and 6.

## 1.3. Research Scope and Challenges

This thesis focuses on estimating canopy structure through canopy gap probability estimation and evaluating the results against the values determined using ALS point clouds.

This thesis uses only open data - no data was specifically collected for this project. Therefore the results of the canopy gap estimation from the ALS data are limited by the quality of the data available. The results retrieved from the ALS point clouds are compared and evaluated against the limited comparable results found in literature.

Although there are other ICESat-2 data products available, thesis focuses only on using the ICESat-2 ATL08 data product as it is suggested as the most promising approach for computing canopy gap fraction (Neuenschwander et al. 2022a, Neuenschwander et al. 2021).

## 1.4. Thesis Outline

This thesis is organised as follows:

- Chapter 2 reviews the related work first on the canopy gap fraction estimation from the ALS data and then on the canopy gap fraction estimation from ICESat-2 data.

## 1. Introduction

- Chapter 3 gives an overview of the entire pipeline used for data processing and analysis.
- Chapter 4 shows the results.
- Chapter 5 presents the discussion on the results.
- Chapter 6 gives a brief conclusion together with some suggestions for future work.

This thesis uses only open data and the code used to produce the results is available in Github ([https://github.com/Mschn-k/MSc\\_thesis](https://github.com/Mschn-k/MSc_thesis))

## 2. Related work

### 2.1. The concept of canopy gap fraction

Forest canopy gap fraction is a commonly used ecological indicator describing the amount of light reaching the forest floor and the density of forest canopy. The term has different definitions, mainly varying in terms of whether light reaching the forest floor from all angles or only from the zenith is considered. In this thesis the canopy gap fraction considers all the gaps in forest canopy when viewed from zenith (Armston et al. 2013). Canopy gap fraction is closely linked to the measure of canopy cover as shown in Equation 2.1. While high canopy cover value indicates dense forest, high canopy gap fraction value indicates sparse forest.

$$\text{Canopy gap fraction} = 1 - \text{Canopy cover} \quad (2.1)$$

Canopy gap fraction and canopy cover are both measured per study plot rather than per unit area. In this thesis, canopy gap fraction is computed per each segment of ICESat-2 ATL08 product. As shown in Figure 1.1, more dense forest would block more vertical light beams reaching the ground and result in a smaller canopy gap fraction value. Therefore, ATL08 segments with small canopy gap fraction value are assumed to have more dense forest while larger value indicates sparser forest. The canopy gap fraction changes throughout the year being higher in winter months when deciduous trees have lost their leaves, and lower during the summer months when the leaves are present and blocking more light from reaching the ground. This yearly fluctuation would be smaller in coniferous forests where most of the trees are evergreen throughout the year.

#### 2.1.1. Field measurements of canopy gap fraction

There are many methods for measuring canopy gap fraction on the field for ground data, five relevant ones are briefly described here. It can be measured manually using a sighting tube, often called Cajanus tube, which is a simple handheld tool with a mirror to look upward. Studies typically define small plots within the study area that are sampled by determining whether or not the point seen through the tube is covered by vegetation or not. Alternatively, the LAI-2000 Plant Canopy Analyzer is commonly used which is an electronic device with convex optical sensor measuring light interception at five angles instead of considering only light from zenith. TRAC (Tracing Radiation and Architecture of Canopies) is another optical instrument used by moving it around the study site while it records peaks in transmitted direct light which are interpreted as gaps in the canopy. Using hemispheric

## 2. Related work

or canopy photographs is another common method for estimating solar radiation by using photos taken by upward looking extreme wide-angle lens and often analysed by specific computer programs. Lastly, Terrestrial Laser Scanning (TLS) is used to measure canopy gap fraction from the ground throughout the angular profile. Studies have shown that the canopy gap fraction estimations from TLS is consistently lower than from passive methods such as hemispheric photographs and the LAI-2000 instruments as the latter tend to overestimate the gaps in the canopy (Korhonen et al. 2011). However, there is no agreement in literature which of those methods is the most precise. While canopy gap fraction measurements from different years yield similar results in mature forest, estimations obtained through using different methods can vary considerably (Kuusk et al. 2018).

### 2.1.2. Measuring canopy gap fraction from ALS

Airborne LiDAR has been increasingly used for quantitative estimates of vegetation structure for over two decades (Fisher et al. 2020) which has led to the development of variety of survey configurations and canopy structure indicators. Estimating canopy gap fraction from ALS data has shown to provide similar results to the field measurements while being less prone to sampling errors and allowing larger coverage. The methods using ALS can roughly be divided to two - those that use discrete returns and those using waveform data. The type of ALS data used affects the canopy gap fraction estimations and different techniques have been developed for different ALS data types. However, a single technique for estimating canopy gap fraction does not exist.

Most methods for estimating canopy gap fraction from discrete ALS data calculate the proportion of the canopy hits above a specified height threshold. The difference in methods is whether only the first returns or all returns are considered. Overall, if only first returns are used, only the larger gaps between the trees are captured. In addition to considering the canopy to total return ratio, different laser penetration metrics (LPMs) have been suggested, which can be used as proxies for canopy gap fraction estimation. Korhonen et al. 2011 shows strong correlation between the LPMs computed from ALS data and canopy gap fraction values from field measurements. However, using such LPMs as a prediction for gap probability, has shown to produce systematic bias for about 3-4% (Korhonen et al. 2011).

## 2.2. Overview of ICESat-2 ATL08 data product

Although the main objective of ICESat-2 mission is to continue the elevation time series in the cryosphere from its predecessor ICESat, it contributes to a broad spectrum of science disciplines beyond the primary scientific goal. Some of the examples include studying inland water bodies (Zhang et al. 2019), Amazon rainforest regrowth (Milenković et al. 2022) and for developing global lowland digital terrain model (Vernimmen et al. 2020). The measurements of ICESat-2 are done with the Advanced Topographic Laser Altimetry System (ATLAS) which uses a single 532 nm laser that is split into six beams that are organised as three beam pairs approximately 3 kilometers apart (Figure 2.2) (Magruder et al. 2020, Magruder et al. 2021). Each of the three beam pairs consist of one strong and one weak beam. The strong beams were designed to detect up to 16 photons per outgoing shot while the weak beams

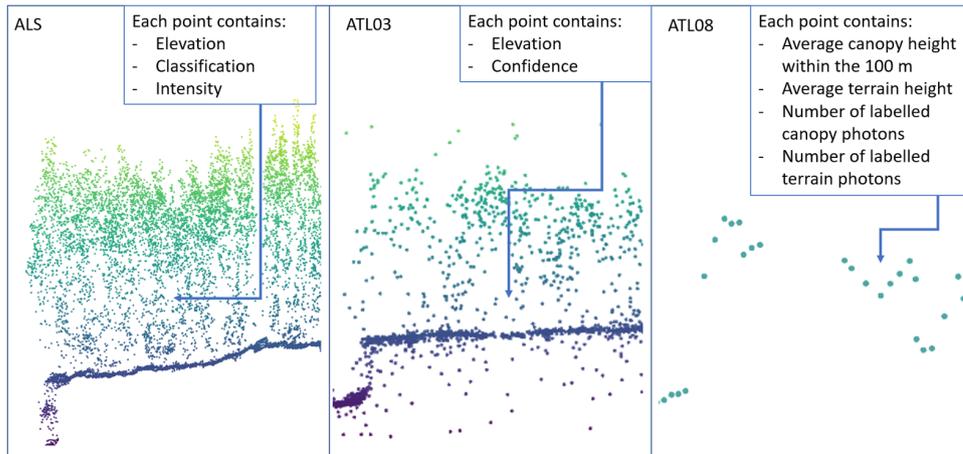


Figure 2.1.: Illustration comparing the ALS point cloud data with ICESat-2 photon data in ATL03 and ATL08 data products.

were designed to detect only one fourth of this - up to 4 photons per shot. However as shown by [Neuenschwander et al. 2022a](#), the ratio between the strong and weak beams is smaller in reality because the weak beams perform better than expected.

While the three pairs of *ICESat-2* tracks are separated by about 3 kilometers on landscape (Figure 2.2), the along-track resolution between shots is much higher. It is estimated that the shots are separated by about 70 centimeters along-track. Data with such high along-track resolution is stored in ATL03 product. The ATL08 data product, however, is designed for understanding Earth's biosphere and vegetation. The data is provided at a fixed segments of 100 meters along ground track, which essentially is a summary of the ATL03 data of photons recorded within each segment (Figure 2.1). As the footprint of *ICESat-2* has been estimated to range between 10-12 meters in diameter ([Magruder et al. 2021](#)), the footprint of the ATL08 segments could be visualized as 100 x 12 meter polygons. The *ICESat-2* ATL08 product contains heights for both terrain and canopy together with other descriptive parameters for each of these segments. The signal photons in ATL08 are classified as top of the canopy, mid-canopy and terrain.

## 2.3. Validating ATL08 data

According to the 2022 release of *ICESat-2* ATL08 Algorithm Theoretical Basis Document (ATBD) ([Neuenschwander et al. 2022b](#)), the preferred validation data for the *ICESat-2* mission is swath mapping airborne lidar, as it is widely available and the errors associated are well known. For validating canopy height, it is suggested that the difference between the *ALS* and ATL08 estimations differ by less than 2 meters for temperate forest. The suggested point density for the *ALS* data used for validation is 5 pts/m<sup>2</sup>, however it is noted that data with lower point density which still meets the vertical height accuracy given may be utilized. To validate terrain and canopy heights in the ATL08 transects, residuals should be computed between the respective terrain and canopy height values given in the 100 m segment and in

## 2. Related work

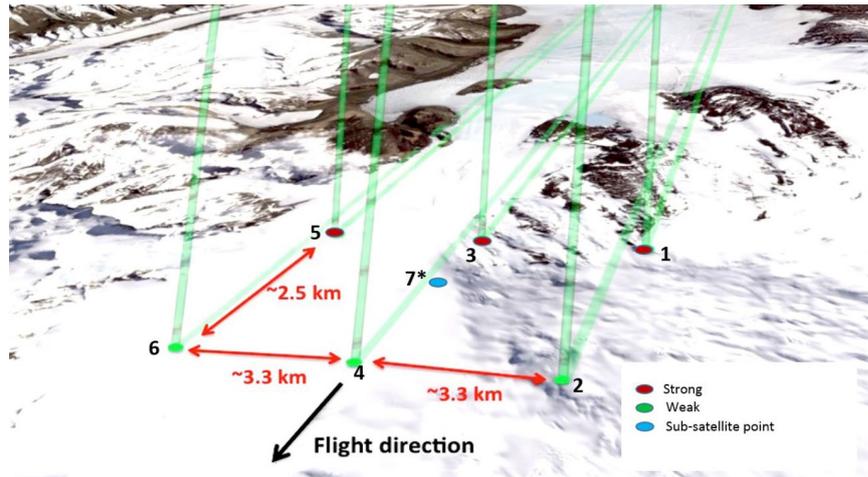


Figure 2.2.: Beam configuration for ATLAS instrument on ICESat-2 with strong beams (1, 3, 5) and weak beams (2, 4, 6) which fourth of the energy. Spot 7 on the diagram is a virtual point representative of the sub-satellite point. Figure from [Neuenschwander et al. 2022a](#)

validation data for that same representative distance.

It is also recommended by [Neuenschwander et al. 2022b](#) to use ancillary data sets, such as Landsat-derived annual forest change maps, to avoid comparing non-equivalent content between the validation and ATL08 datasets, for example in cases of forest clear cuts between the two acquisitions. Such dataset is developed by [Hansen et al. 2013](#) and updated yearly, using 30-meter resolution Landsat dataset to map the global forest disturbances. The dataset maps global mature forests as well as the areas that have experienced disturbance such as fire or clear cutting. However, smaller selective removals that do not leave the area non-forested, are not included.

The *ATBD* also gives some pointers for the canopy cover validation. Canopy cover values are not part of the ATL08 version 5 and while writing this thesis the gridded ATL18 data product with canopy features is not yet available. However, it is suggested that to validate the canopy cover of the ATL08 data product, the relative canopy cover for the same area should be computed from the validation *ALS* data. However, no suggestions for the desired threshold for agreement between the two datasets in canopy cover estimation is made.

## 2.4. ATL08 for canopy structure

In order to use the ATL08 data for quantifying canopy structure, it is important to understand from where within the canopy the photon is likely to be reflected. Different laser detector modalities behave differently as shown in Figure 2.3. Full waveform sensors digitize the entire profile of reflected energy as a function of time. The returns are later derived from the waveform during post-processing by threshold technique where a return is recorded

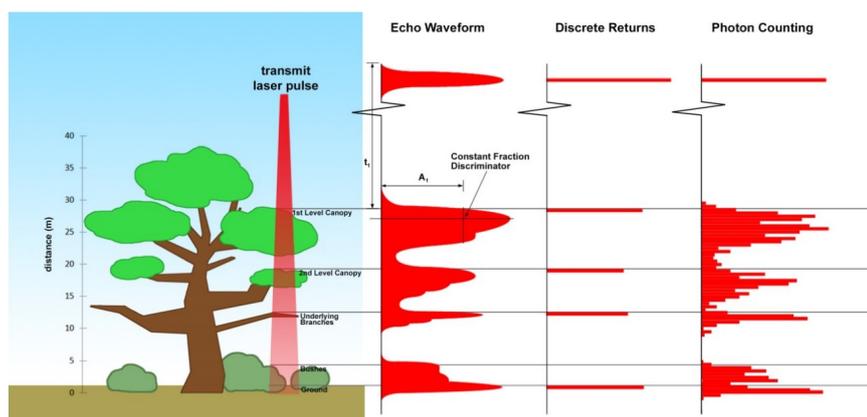


Figure 2.3.: Various modalities of lidar detection. The photon counting probability distribution function (PDF) shown is illustrating a theoretical situation if hundreds or more photons were reflected from a target. Figure from [Neuenschwander et al. 2022b](#).

whenever the power of the waveform exceeds a fixed threshold. Discrete return sensors however record discrete, time-stamped trigger pulses in real time and are typically limited to recording 1-6 returns per pulse. A photon counting system records the time when a single photon is detected which can occur anywhere within the vertical distribution of the reflected signal. However, the probability distribution function (PDF) of a single photon is the same as of the full waveform. Hence, if a significant number of shots would be recorded by photon counting lidar system over the same surface, the reflected photons would resemble a full waveform ([Neuenschwander et al. 2022b](#)).

### 2.4.1. Canopy gap fraction from ATL08 data

As most methods for estimating canopy gap fraction from airborne LiDAR involve finding the proportion of canopy hits in relation to total returns, a similar approach is suggested for finding canopy gap fraction using ATL08 data. For ICESat-2 data, the most simple approach would be to find the proportion of canopy photons to total signal photons from each ATL08 segment. However, ATLAS is a photon counting lidar system, meaning that it transmits a low power laser pulse while its detectors are sensitive at single photon level. The number of signal and background photons detected per pulse depends on the laser wavelength, solar conditions, surface reflectance, transmitted laser energy and scattering in the atmosphere. This means that any returned photon, either from reflected signal or solar background, can be detected by ATLAS. The returns from terrain and canopy class depend on the reflectance of ground and canopy ([Neuenschwander and Magruder 2016](#)) as well as on the vegetation structure ([Queinnec et al. 2021](#)).

Since the ATLAS sensor is sensitive to the surface reflectance, calculating a simple photon ratio for finding canopy gap fraction may not be sufficient. This can be illustrated by using the example by [Neuenschwander et al. 2022a](#) shown in Figure 2.4. If a forest would have snow on the ground, the ground reflectivity would increase, leading to higher ground

## 2. Related work

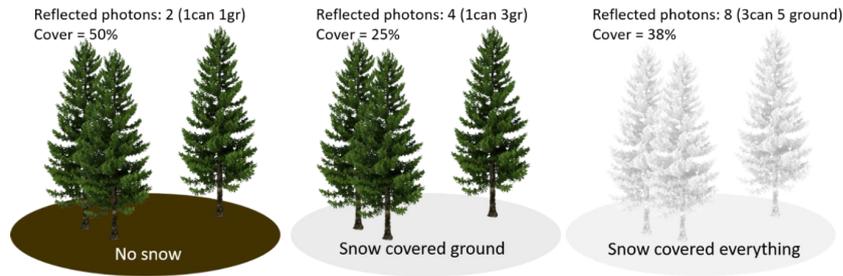


Figure 2.4.: Illustration on how surface reflectance can alter the canopy cover estimation from ATL08. Taken from [Neuenschwander et al. 2022a](#).

signal photons within the ATL08 segment and in return a higher canopy gap fraction estimation than if no snow was present. This change in the proportion of detected ground signal photons is caused by the changes in the reflectivity, not the actual canopy structure. The presence of snow is to emphasize the effect of reflectivity and in actual canopy gap fraction computation, the segments that have acquired during the presence of snow should be filtered out. However, aspects such as the color and water content of the canopy and the ground can change the surface reflectivity and therefore affect the results of the canopy gap fraction.

In order to account for the surface reflectivity, [Neuenschwander et al. 2022a](#) suggested using the ratio of canopy and ground reflectivity values that are found by using radiometric profiles (Figure 2.5). Radiometry for ICESat-2 is defined as the number of signal photons detected per outgoing laser shot. Radiometric profile is the relationship of ground radiometry against canopy radiometry for a given region. The profiles shown in Figure 2.5 are plotted per selected country and including only the ATL08 segments that were acquired during clear July night in order to only include the data from the most optimal conditions.

The most important aspect of the radiometric profiles for computing canopy gap fraction, are the intercepts of x and y axis. The points close to the x axis mark the ATL08 segments where most of the returned signal photons were from the ground while the points close to the y intercept mark segments where most of the returned signal photons were from canopy. In temperate forest, a relatively clear linear relation is formed between the two intercepts while in tropical forest there are fewer segments with more ground signal photons per shot. The theory presented by [Neuenschwander et al. 2022a](#) suggests that the ratio of the x and y intercepts is the ratio of vegetation and ground reflectivity and can be used to correct for the different reflectivity values that may affect the canopy gap fraction estimation.

It is important to note that the method proposed by [Neuenschwander et al. 2022a](#) for using the reflectivity ratio in canopy gap fraction computation is adapted from [Armston et al. 2013](#) who developed this for airborne LiDAR that allows computing the reflectivity ratio from the waveforms. It is assumed, that the data included is reflecting constant conditions, meaning that the reflectivity of ground and canopy does not change. While for studies using airborne LiDAR, the data is normally collected in a short time frame and in an area where the vegetation is relatively well known, it is very different when using ICESat-2 data. Namely, the

## 2.4. ATL08 for canopy structure

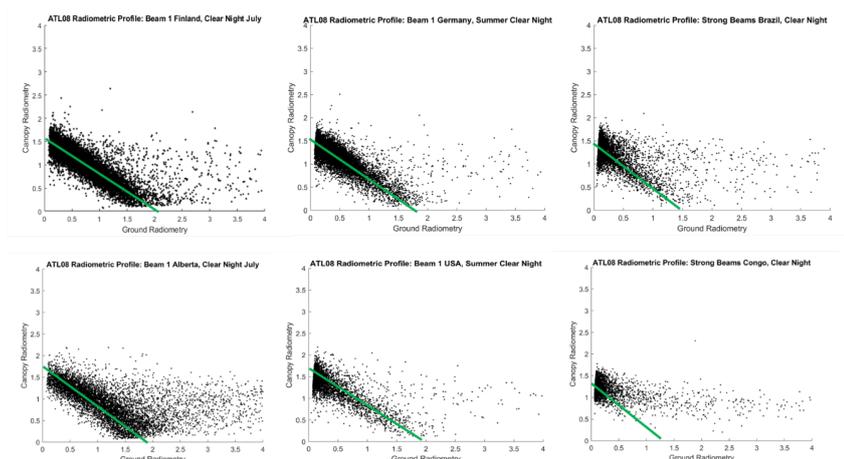


Figure 2.5.: Radiometric profiles by [Neuenschwander et al. 2022a](#) for six study sites. The green lines were inserted manually by the authors to highlight the trend. Taken from [Neuenschwander et al. 2022a](#).

radiometric profiles shown in Figure 2.5 include data from entire countries and therefore from different forest types. It has been shown that different types of vegetation can reflect light differently. For example ground covered with dry grasses has higher reflectance compared to green grasses ([Asner 1998](#)).

Overall, there are two methods proposed in literature for computing canopy gap fraction from ICESat-2 ATL08 data - one using simple photon ratio and other using in addition the radiometric profile to correct for reflectivity. None of the two methods, however, have been tested when this thesis was written. Furthermore, the second method that includes correcting for surface reflectivity was presented by [Neuenschwander et al. 2022a](#) as a theoretical concept but was not put into practice. Therefore, it is unknown if either of the methods will work. Furthermore, when presenting the radiometric profile, [Neuenschwander et al. 2022a](#) do not discuss different approaches of which data should be included in such profiles or how the x and y intersects should be determined as the green lines shown in Figure 2.5 were inserted manually by the authors to highlight the trend.



## 3. Methodology

In this chapter, a description of the study area and a workflow of the ATL08 and ALS data pre-processing and canopy gap fraction computation is given. The summary of the workflow is shown in Figure 3.1.

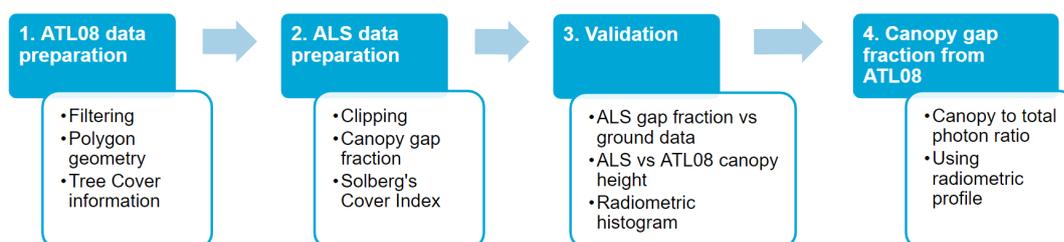


Figure 3.1.: Overview of workflow used in this thesis.

### 3.1. Study Area

To meet the research objectives of this thesis, full ICESat-2 time-series (version 5) over the entire mainland territory of Estonia (about 41 000 km<sup>2</sup>) was examined (Figure 3.2). According to the Estonian Forest 2020 Yearbook, about half (53.5%) of the Estonian territory is covered by forest which is a mix of deciduous and conifer species. The most common tree species (Figure 3.3) is Scots pine (*Pinus sylvestris*) which is the dominant tree species in about 30% of the forested area, followed by silver birch (*Betula pendula*) dominating 29% of the forest and Norway spruce (*Picea abies*) in 19% of the forest (Agency 2022). Estonian forests are part of the transition zone from broadleaf temperate forest typical to central Europe to boreal needleleaf forests that are characteristic for northern Scandinavia. Estonian topography is generally flat with the highest point just 317 m above the sea level.

### 3.2. ICESat-2 data pre-processing

The pre-processing of ICESat-2 data consisted of two steps. First, all the data available for the study area was downloaded and filtered. Then, the performance of the data under different environmental conditions was analysed using radiometric histograms. The findings from the histograms were used as input for later decisions on data analysis.

### 3. Methodology

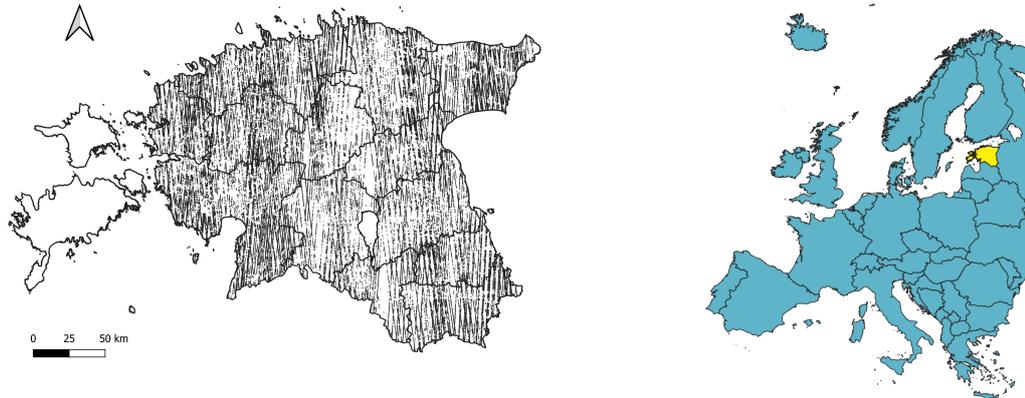


Figure 3.2.: Estonia in Europe and a map of Estonia with the *ICESat-2* transect polygons used in this thesis

#### 3.2.1. ATL08 data filtering

All the H5 files containing data of the *ICESat-2* transects intersecting the bounding box of the study area (Figure 3.2) were downloaded from National Snow and Ice Data Center's CMR Search API (Snow and Center 2022). The data covers the time period from October 2018 to December 2021. The downloaded data was filtered to only keep the segments that are inside the study area. Then, several additional filters were applied to remove segments with errors or unsuitable data.

Segments with canopy heights ( $h_{\text{canopy}}$ ) larger than 50 meters were removed as the tallest tree recorded in Estonia is 48,6 meters tall (Center 2015). Second, the segments with radiometric parameter values exceeding 16 photons per shot (sum of  $\text{photon\_rate\_can}$  and  $\text{photon\_rate\_te}$ ) were eliminated as the ATLAS detector can only detect 16 photons per outgoing shot. Lastly, using the parameter representing the height difference between the *ICESat-2* estimated ground surface and the reference DEM used by the *ICESat-2* ground systems ( $h_{\text{dif\_ref}}$ ), all segments with ground surface elevation difference larger than 30 meters were rejected. Lastly, as the signal return strength can vary due to snow (Neuenschwander et al. 2022a), the transects acquired in the presence of snow were removed by using ATL08 snow flag. The snow flag is derived from the daily NOAA Global Multi-sensor Snow/Ice Cover map (Palm et al. 2018). The remaining ATL08 segments were projected to the Estonian Coordinate System (EPSG: 3301).

The following information was stored from ATL08 product for each segment:

- ATLAS track id
- timestamp
- segment id
- beam number and type

- canopy height
- canopy photon count
- top of canopy photon count
- ground photon count
- canopy photon rate
- terrain photon rate
- snow flag
- solar elevation
- msq flag for scattering

Once the initial filtering of ATL08 data was completed, polygons for each ICESat-2 transects were created. ATL08 data is by nature point data with a pair of latitude and longitude coordinates marking the center of the ATL08 segment. However, each ATL08 segment has a footprint of 100 m along-track and 10-12 meters across (Magruder et al. 2021). Each ATL08 data point describes the canopy within the 100 x 12 meter polygon. Therefore, for each ATL08 segment, a polygon geometry surrounding the point was computed and stored. These polygon geometries were later used for clipping ALS point clouds for each segment (Figure 3.4).

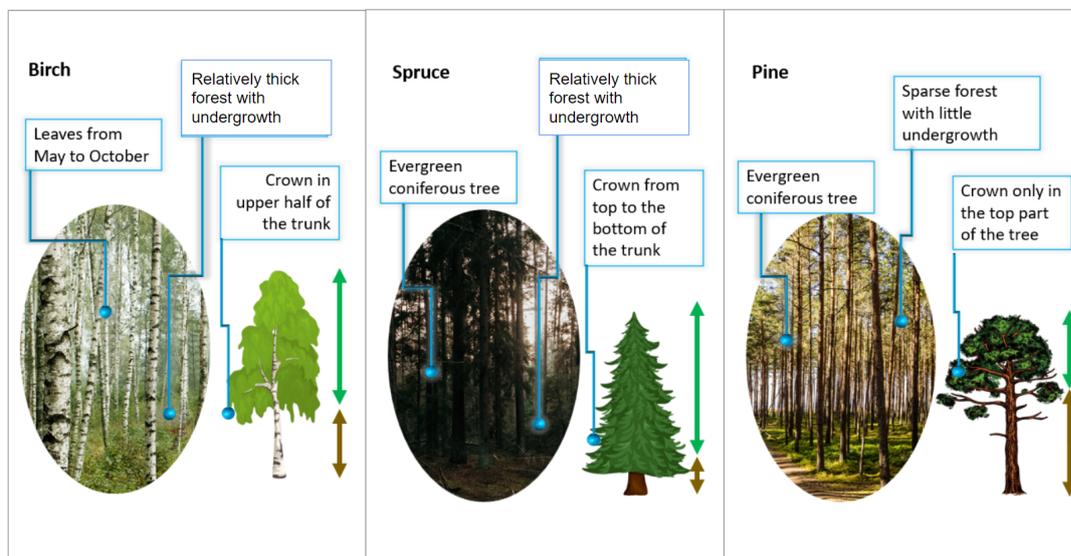


Figure 3.3.: Forest types in Estonia.

The last step in ATL08 data filtering was removing transects that did not contain any forest or where forest had been disturbed. This was necessary because ATL08 and ALS data used in this thesis had been collected on separate years and any forest disturbance would lead to incomparable setting. For this step Hansen Global Forest Change dataset version 1.9 which is a time-series analysis of Landsat images for characterizing global forest extent and change from the year 2000 to 2021 was used (Hansen et al. 2013). Only transects that contained undisturbed forest according to this dataset were kept. For the remaining transects, the tree species dominance and heterogeneity was characterised using the dataset by Lang et al. 2018. After completing this step, 105 328 ATL08 segments were stored.

### 3. Methodology

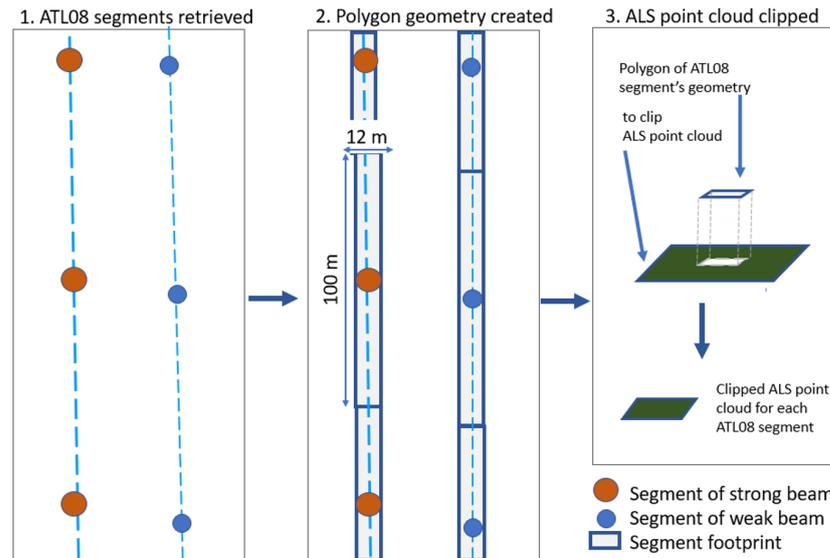


Figure 3.4.: Workflow for creating polygon bounding box around each ATL08 segment and using it to clip ALS point cloud for each segment.

Lastly, in order to match each transect with ALS a point cloud tile from Estonian Land board, a vector layer from the Estonian Geoportal was used. The layer of 1:2000 map tiles holds the IDs of each the respective tile together with information on the type and year of point cloud data available for the respective tile. For each ATL08 transect, an attribute was added to hold the ID of the respective point cloud tile.

#### 3.2.2. ICESat-2 reflectivity analysis through radiometric histograms

Methods for computing canopy gap fraction from ATL08 data are largely based on the ratio of canopy and ground photons per outgoing shot. Therefore, it is important to understand how the ground and canopy radiometry performs under different conditions and in strong and weak beams. If the canopy or ground radiometry are influenced by environmental conditions, rather than the observed vegetation, then it would affect the results of canopy gap fraction estimation from ATL08 data. Therefore, canopy and ground radiometry was studied using radiometric histograms which represent the frequency distribution of the number of labeled ground or canopy photons per outgoing shot. Using the data from ATL08 product indicating the solar elevation and scattering, histograms were plotted in five sets:

1. all segments;
2. segments with positive solar elevation and scattering value zero (clear day);
3. segments with positive solar elevation and scattering value above zero (cloudy day);
4. segments with negative solar elevation and scattering value zero (clear night);
5. segments with negative solar elevation and scattering value above zero (cloudy night);

Both the shapes of histograms as well as calculated modes were analysed. The histograms indicated, that the light condition in ICESat-2 daytime acquisition had large impact on the radiometry observed from the weak beams. Meanwhile, although by design the weak beams of ICESat-2 should have about four times smaller radiometry compared to the strong beams, in night conditions weak beams showed radiometry comparable to strong beams. Therefore, the radiometric histograms gave a reason to leave out segments that were acquired by the weak beams during the day. The results of the reflectivity analysis will be discussed further in Chapter 4.

## 3.3. ALS data pre-processing

### 3.3.1. The ALS data available

Each year Estonian Land Board collects three types of ALS data:

1. Regular aerial laser scanning of quarter of Estonia in spring (no-leaf season) at 2000 or 2600 m with 2.1 pts/m<sup>2</sup> (from here referred to as spring-time ALS).
2. Aerial laser scanning of quarter of Estonia in the summer (leaf-on season) at 3100 m and 0.8 pts/m<sup>2</sup> (from here referred to as summer-time ALS).
3. Annual aerial laser scanning of Estonian cities at 1200 m and 18 pts/m<sup>2</sup>, not used in this thesis.

The type of datasets (spring or summer acquisition) available are shown in Figure 3.5. This study used ALS datasets obtained in summer months to compare with ICESat-2 data acquired during leaf-on season from May to end of September, and ALS datasets from spring were used to compare with ICESat-2 data from leaf-off period from October to end of April.

The ALS data used in this study was acquired by Estonian Land Board using Riegl VQ-1560i which is a waveform LiDAR and scans at 1064 nm wavelength. The data available is already pre-processed from waveform to discrete point clouds. The points have already been classified to ground, vegetation, buildings, water, bridges and noise. The data is freely available to download from Estonian Geoportal in laz file format. Each file is marked by an ID that marks a square in a 1:2000 map grid covering Estonia.

### 3.3.2. Retrieval and processing of ALS data

As described earlier, during the pre-processing step of ATL08 data, for each ATL08 segment the ID of the intersecting ALS point cloud was stored. The necessary point cloud files were downloaded from Estonian Land Board's website through API. Then, the downloaded laz files were clipped according to the geometry of the polygon marking the ATL08 segment computed earlier (Figure 3.4). The clipped point clouds were extracted to las format and points that were classified as noise, bridges, buildings or an overlap bit, were filtered out. The 98th percentile of canopy height was calculated for each segment and lastly, the number of canopy, ground and total returns was computed and stored to be used in canopy gap fraction computation.

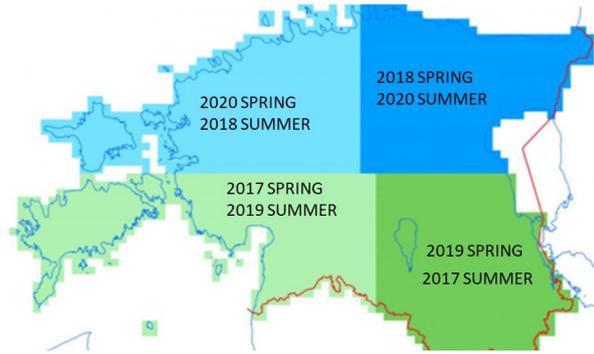


Figure 3.5.: ALS data used in the thesis is open data by Estonian Land Board, in each year covering a quarter of Estonia. Figure adapted from Estonian Land Board.

### 3.4. Canopy gap from ALS

As discussed in Chapter 2, methods to estimate canopy gap fraction from LiDAR data can be grouped to those using waveforms and those analysing the discrete returns. This thesis uses freely available point cloud data, which was acquired using full-waveform sensor, but has already been processed to discrete points. Therefore, estimating canopy gap fraction will follow the methods of using discrete returns as the waveform data was not available. As there is no single way to compute the canopy gap fraction, several methods were used and the results evaluated.

Two types of canopy gap fraction values were computed. First, canopy gap fraction was computed by finding the ratio of all canopy returns against all returns as shown in equation 3.1:

$$Canopy\ gap\ fraction = 1 - \frac{\sum All_{canopy}}{\sum All} \quad (3.1)$$

In addition, the Solberg's Cover Index (SCI) developed by Solberg et al. 2009 was computed for each segment. It is a Light Penetration Index (LPI) that is shown by Korhonen et al. 2011 to perform well with little bias for estimating canopy gap fraction. SCI is computed using the following equation:

$$SCI = 1 - \frac{\sum Single_{ground} + 0.5(\sum First_{ground} + \sum Last_{ground})}{\sum Single_{all} + 0.5(\sum First_{all} + \sum Last_{all})} \quad (3.2)$$

where coefficient 0.5 is a weight if both first and last echoes are produced, and it can be adjusted if necessary.

### 3.5. Validating the ALS data

This section first discusses the methodology to validate the results of canopy gap fraction estimation from ALS data against ground and ALS measurements reported by [Kuusk et al. 2018](#) from a test site in Estonia. Then, the method to evaluate agreement of the canopy height between ALS and ATL08 data is described.

#### 3.5.1. Validating the canopy gap fraction estimates from ALS data against ground data

There are not many studies that have published their canopy gap fraction estimations from ALS in Estonia. However, a comprehensive study by [Kuusk et al. 2018](#) measured canopy gap fraction in Järvselja Training and Experimental Forestry District in southern Estonia (58.30°N, 27.26°E). They compare the gap fraction estimates from ALS with ground measurements using described in Chapter 2, including Cajanus tube, plant-canopy analyser LAI-2000, optical TRAC instrument, hemispherical photos and terrestrial laser scanner. Three different forest stands that are 100 x 100 meter in size were measured - one dominated by pine trees, one by spruce and the last one by birch forest.

To test the methods used in this thesis for canopy gap fraction estimation from ALS, summer and spring ALS data was clipped according to the 100 x 100 meter study plots used by [Kuusk et al. 2018](#). In addition, similarly to the approach shown in Figure 3.4, additional five 12 x 100 meter plots were created inside each of the three forest stands to imitate the size of the ATL08 transect (Figure 3.6). This was done to test if the plot size affects the computed canopy gap fraction. For each of the six plots in three test sites, canopy gap fraction was calculated from the ALS data.

The results are shown in a scatter plot comparing the values derived from the spring and summer ALS data. The mean of the canopy gap fraction estimated for each of the three forest stands is shown in a table together with the results reported by [Kuusk et al. 2018](#). The results are shown in Chapter 4 together with discussion whether the ALS data used in this study is suitable substitution for ground data.

#### 3.5.2. Validating canopy height estimation from ALS and ATL08

In addition to validating if the computed canopy gap fraction results are comparable with values found in literature, it is necessary to ensure that the ALS data was comparable with the ATL08 data. According to [Neuenschwander et al. 2022b](#), ALS data can be used as a reference for ICESat-2 if the canopy height estimation from the two datasets differ by less than two meters. While ATL08 dataset contains the canopy height for each segment, it had to be computed for the ALS dataset by finding the 98th percentile height of all canopy returns. Then, the canopy height from ALS and ATL08 for each segment were plotted against each other. As this thesis uses two types of ALS data from spring and summer acquisition with different point density, both datasets were tested separately. The data was plotted on a scatter and a violin plot and RMSE, MAD and median height difference was computed. Also,

### 3. Methodology

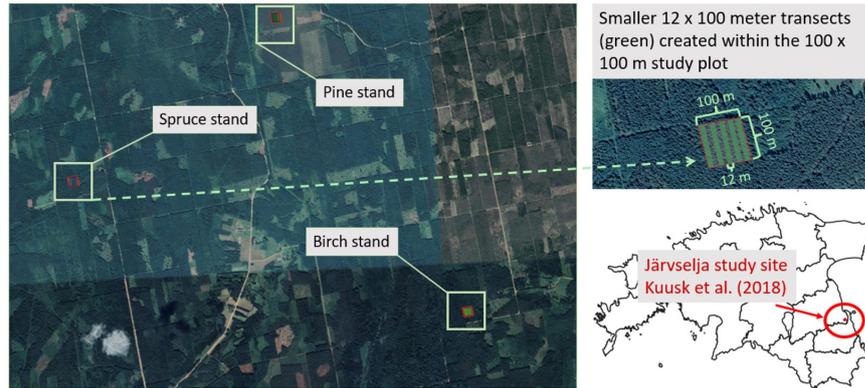


Figure 3.6.: A map showing the location of Järvelja study site from where [Kuusk et al. 2018](#) reports canopy gap fraction measurements in three forest stands which are 100 x 100 meters in size. This thesis computed the canopy gap fraction from ALS point clouds clipped to 12 x 100 meter plots inside these three study sites to validate the ALS results against values reported by [Kuusk et al. 2018](#).

the percentage of segments where the height difference was within 2 and 3 meter error was computed. In addition, the agreement of canopy height estimation was analysed in transects of strong and weak beams separately.

Overall, validating the canopy height estimation from ALS and ATL08 revealed that [ICESat-2](#) daytime acquisition resulted in a lot of noise in the data. This agrees with the findings from the radiometric histograms that showed daytime acquisition of [ICESat-2](#) to result in decreased radiometry for weak beams. Removing the segments of only weak beams from the daytime acquisition left still considerable noise. Removing segments of [ICESat-2](#) daytime acquisition of both strong and weak beams resulted in distinctively decreased noise. Therefore, from here on, canopy gap fraction was only computed for transects for which the [ICESat-2](#) data was acquired in night condition.

## 3.6. Canopy gap from ATL08

This section first describes the methodology for calculating canopy gap fraction using ATL08 canopy to ground ratio. Then, the method by [Neuenschwander et al. 2022a](#) adopted from [Armston et al. 2013](#) is introduced and an approach for computing a radiometric profile needed for its implementation is described.

### 3.6.1. Canopy to total photon ratio

The [ATBD](#) of ATL08 data product ([Neuenschwander et al. 2022b](#)) states that relative canopy cover can be estimated from ATL08 by dividing canopy returns by total returns. As shown in equation 2.1, canopy gap fraction can be calculated by subtracting canopy cover from one.

As the ATL08 product has signal photons labelled in three categories - canopy ( $R_{\text{canopy}}$ ), top of canopy ( $R_{\text{topcanopy}}$ ) and terrain ( $R_{\text{terrain}}$ ) - for each transect the sum of the canopy and top of canopy return rate is divided by the sum of all signal photons. The resulting equation for calculating canopy gap fraction from ICESat-2 ATL08 data is shown in Equation 3.3.

$$\text{Canopy gap fraction}_{ATL08} = 1 - \frac{\sum R_{\text{topcanopy}} + \sum R_{\text{canopy}}}{\sum R_{\text{topcanopy}} + \sum R_{\text{canopy}} + \sum R_{\text{terrain}}} \quad (3.3)$$

### 3.6.2. Accounting for the surface reflectance in canopy gap fraction computation

As shown in Figure 2.4 in Chapter 2, the number of canopy and ground signal photons recorded by ATLAS can be affected by the surface reflectivity of canopy and ground. In order to account for this Neuenschwander et al. 2022a proposed a methodology adapted from Armston et al. 2013 which is explained here through three steps. First the equation is explained, then the methods for finding the radiometric profile is shown and last, the approach to find the intersects from the radiometric profile is discussed.

#### Canopy gap fraction calculation

Armston et al. 2013 developed a method to retrieve canopy gap fraction from full waveform airborne lidar, proposing the following formula:

$$\text{Canopy gap fraction}_{ATL08} = 1 - \frac{\sum_{z_i}^{z_{\text{max}}} R_v}{1 + \frac{\rho_v}{\rho_g} \frac{R_g}{R_v}} \quad (3.4)$$

where  $R_v$  is the integrated vegetation backscatter component of the waveform from the top of the canopy down to elevation  $z_i$  which is the canopy cutoff point under which the vegetation is not considered as part of the canopy, and  $R_g$  is the ground backscatter component. As LiDAR returns are affected by the canopy and background reflectivity ratio, the  $\rho_g/\rho_v$  represent the ratio of the canopy volume backscattering coefficients ( $\rho_v$ ) and background reflectivity ( $\rho_g$ ). This ratio depends on canopy architecture as well as foliage spectral characteristics. The model assumes that the backscattering coefficients  $\rho_v$  and  $\rho_g$  are constant, hence the vegetation cover is homogeneous across the area for which the same reflectivity ratio is applied to.

As ICESat-2 is photon sampling system not a waveform LiDAR system, Neuenschwander et al. 2022a suggest substituting  $R_g$  and  $R_v$  with the number of detected photons for both the ground and canopy respectively. However, the values of the ground reflectance  $\rho_g$  and vegetation reflectance  $\rho_v$  for ICESat-2 are unknown. As a possible solution, Neuenschwander et al. 2022a propose a method of plotting the canopy photon rate against the terrain photon rate for each segment in the study area. This results in a linear trend called the radiometric profile (Figure 2.5), where the x and y-intercepts reflect the  $\rho_v$  and  $\rho_g$  values. Therefore,

### 3. Methodology

radiometric profiles must be plotted and the intercepts found before the equation 3.4 can be solved.

#### **ATL08 radiometric profiles**

The radiometry is defined as the number of signal photons detected per outgoing laser shot. Therefore, radiometric profile can be created by plotting the rate of and the relationship of the ground radiometry against the canopy radiometry creates the radiometric profile for a defined region which in this thesis is the mainland of Estonia. The radiometric profile of Estonia was created by plotting the photon rate for canopy photons within each segment (*photon\_rate\_can*) against photon rate for terrain (*photon\_rate\_te*). These photon rate values are calculated as the number of respective signal photons divided by the total number of laser shots within each ATL08 segment (Neuenschwander et al. 2022a) and are part of the ATL08 data product.

Equation 3.4 proposed by Armston et al. 2013 assumes a homogeneous vegetation as it is designed to be used for ALS data, which in general would have a smaller study area. Using this for ICESat-2 data poses a question how the reflectivity of the ground and terrain should be assessed. This thesis explores whether it would be sufficient to estimate the reflectivity per geographic region, or whether the time of the year should also be accounted for. To analyse this, two types of radiometric profiles were plotted.

First, a radiometric profile was plotted using all segments where the ATL08 data was acquired only during clear August night from all strong beams. Using data from only one month is adapted from Neuenschwander et al. 2022a and it allows enough data while also avoiding mixing data that reflects very different environmental condition. August was chosen as the most optimal month, because in the higher latitudes the nights are light in June and the first part of July. Also, the leaves on the deciduous trees start forming in the second part of April and are not fully developed in May. Using September or later months would mean including data from the time when the leaves start changing color in autumn.

#### **Finding the intersect in the radiometric profile**

Linear regression was used on the plotted profiles for finding the intersections with x- and y-axis. Since the data appeared to be noisy, a more robust solution for the linear fit was needed that would not be affected by the outliers. Therefore, RANdom SAMpling Consensus (RANSAC) regression was used which splits the data into inlier points and outliers and the line is fitted only by the identified inliers. RANSAC works by first randomly selecting two points to form a line and then checking for each point if they fit with the line by computing the distance from the point to line which is compared against user-defined threshold. However, due to the random nature of RANSAC, the output may differ with each run. Therefore, the model was ran 10 times for each radiometric profile. Each time the location of x- and y-intercept was noted. The final values used in the Equation 3.4 were the mean of the x and y-intercepts from the 10 runs.

### 3.6.3. Finding the trends in canopy gap fraction results

One way to evaluate the canopy gap fraction results from ICESat-2 data is to compare them against values computed from ALS data. However, in addition to knowing how close the values get to the reference data, there is also value in knowing which trends are picked up by the canopy gap fraction computed from ATL08. As described earlier, canopy gap fraction reflects the density of the canopy and should show a trend between different forest types. For example, more sparse pine forest should in general have lower canopy gap fraction than spruce forest (Figure 3.3). In addition, as the ICESat-2 acquisition continues throughout the year, it should pick up the annual changes in canopy gap fraction. To investigate this, the canopy gap fraction values computed from ATL08 data will be grouped by different months and different forest types, to see if any trends are present in the data.

### 3.6.4. Error estimation

To evaluate which of the two methods for finding canopy gap fraction from ATL08 had better agreement with the results computed from ALS, the following measures were used:

- Root Mean Square Error (RMSE) measuring the magnitude of error;
- Median Absolute Deviation (MAD) indicating the spread of the value, robust to outliers;
- Violin plots which are a type of box plots to visualize the distribution of data.

### 3.6.5. Implementation

To implement the methodology describe above, the following software was used:

- Python in Jupyter notebook environment;
- Lastools;
- Geopandas and Pandas library;
- Pyplot scatter plot;
- Seaborn violin plot;
- Quantum Geographic Information System (QGIS);
- Sklearn linear regression.

This thesis used the following datasets which all are available as open data:

- ATL08 version 5 data openly available from National Snow and Ice Data Center's CMR Search API;
- Grid of the airborne LiDAR file IDs from Estonian Land Board Geoportal;
- Airborne LiDAR data from spring and summer scanning from Estonian Land Board Geoportal;
- Hansen Global Forest Change dataset (Hansen et al. 2013) for including only intact forest segments;

### 3. *Methodology*

- Raster of forest types by dominant tree species in Estonia ([Lang et al. 2018](#)).

## 4. Results

This chapter first discusses the results from the data validation step. It then analyses the outcome of the canopy gap fraction estimation from ATL08 data using the two approaches and comparing the results with ALS data.

### 4.1. Data validation

The data validation was divided to three steps as shown in the workflow diagram in Figure 3.1. The results are first presented from the radiometric histograms, followed by validation using ground data found in literature and lastly comparing the canopy height estimation from airborne LiDAR and ICESat-2 data.

#### 4.1.1. Analysis of the radiometric histograms

The radiometric histograms, shown in Figure 4.1, were plotted in order to understand how the ground and canopy radiometry derived from ATL08 data performs under different conditions and in strong and weak beams. The canopy and ground radiometry should depend on vegetation in order to be used for the canopy gap fraction estimation, not on the cloud or light conditions. Also, more photons detected should improve the accuracy of the measurements. Therefore, each plotted histogram shows how the ground or canopy radiometry performs under cloudy day, cloudy night, clear day and clear night conditions. The modes calculated from the histograms are shown in Table 4.1.

The radiometry for day and night acquisition should be equivalent by the original design of ICESat-2, because although there is more background noise during the day, the amount of signal photons reflected from the surface should be the same. However, Neuenschwander et al. 2022a reported a mistake in ATL08 data product with the reported radiometric rates as they were not corrected for background noise. Nevertheless, even after the correction, higher signal radiometry for night acquisition was still reported by Neuenschwander et al.

	Ground mode		Canopy mode	
	Strong beam	Weak beam	Strong beam	Weak beam
<b>Clear day</b>	0.33	0.07	1.17	0.52
<b>Clear night</b>	0.31	1.01	0.98	0.38
<b>Cloudy day</b>	0.35	0.07	0.75	0.59
<b>Cloudy night</b>	0.66	1.17	0.99	0.48

Table 4.1.: Modes of ground and canopy radiometric histograms in different conditions.

## 4. Results

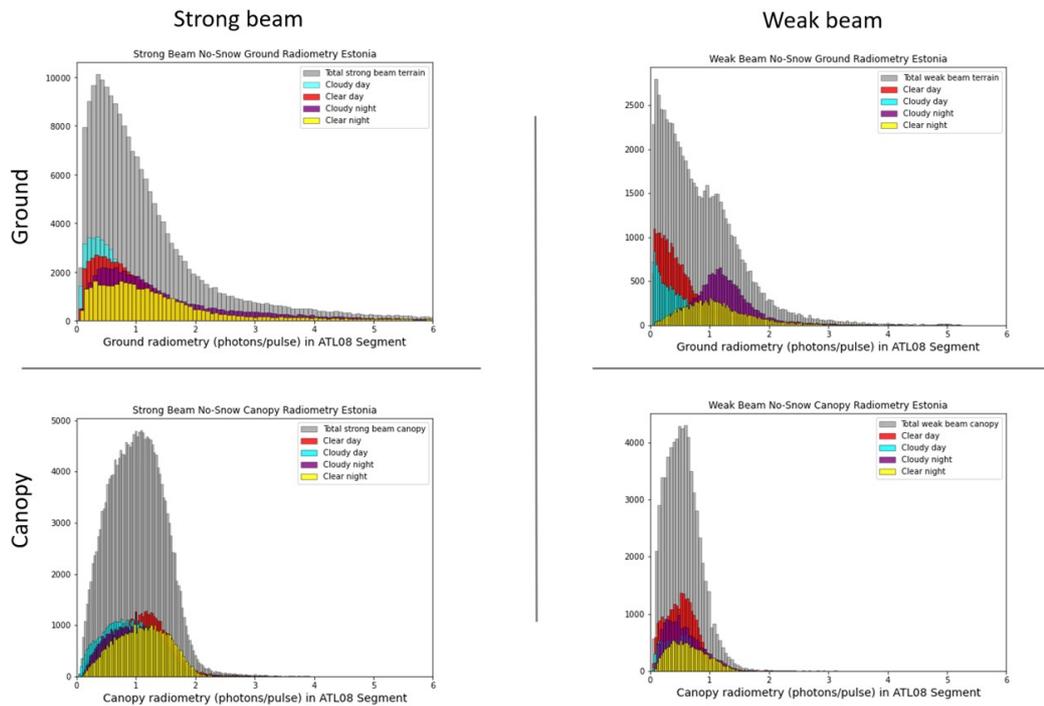


Figure 4.1.: Canopy and ground radiometric histograms for strong and weak beams.

2022a. As the data used in this thesis has not been corrected for the background noise, a clear difference between daytime and nighttime radiometry can be seen in the weak beam. This is evident from the top right histogram in Figure 4.1 where there is clear separability as the ground radiometry is much lower for daytime acquisition than for nighttime. The same is clear from the smaller daytime modes of weak beam shown in Table 4.1.

Based on these results, nighttime conditions are more optimal than daytime for using ATL08 data for canopy gap fraction estimation. In addition, it is noted that according to Table 4.1, the mode for weak beam in night acquisition is always higher for ground than for canopy, while the opposite is true for strong beam. This is unexpected, because although Neuen-schwander et al. 2022a reports that weak beam performs better than was expected from the design of ICESat-2, such pattern between strong and weak beam have not been reported.

### 4.1.2. Validation of canopy gap fraction from ALS as reference data

This study uses canopy gap fraction computed from ALS data as a replacement for ground truth. However, in order to check the feasibility of this, canopy gap fraction from the spring and summer ALS data was computed for the same study plots used by Kuusk et al. 2018 that reports their canopy gap fraction results from ALS and ground measurements. First, the values were computed from point clouds clipped according to the 100 x 100 meter study

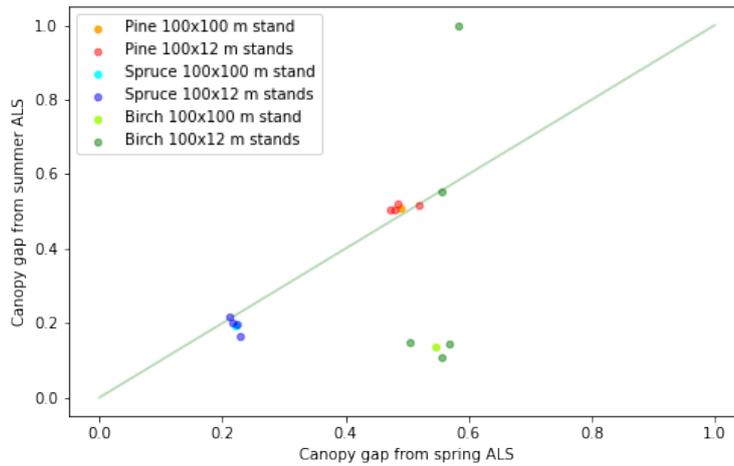


Figure 4.2.: Comparison of canopy gap estimations computed for Järvselja test plots from summer and spring ALS.

areas used by [Kuusk et al. 2018](#). Then, smaller 12 x 100 meter polygon-shaped study plots were created inside the same 100 x 100 meter plots, to see if the smaller area constraint from ATL08 data changes the canopy gap estimation (Figure 3.6).

Figure 4.2 shows the canopy gap fraction values computed from the spring [ALS](#) dataset on x-axis and results from summer dataset on y-axis. Points that are concentrated near the green line have similar canopy gap fraction estimation from the two datasets. It is evident that the area of the plot does not affect the estimated canopy gap fraction as the values from 100 x 100 meter and 100 x 12 meter plots are similar. Also, the estimations for coniferous pine and spruce stands are similar from summer and spring [LiDAR](#) data. However, there is some noise in the data reflecting the birch stand. First, while the canopy gap fraction values from the spring [ALS](#) are concentrated around 0.55, the estimation is below 0.2 from the summer data. This meets the expectation as for deciduous forest, canopy gap fraction should be lower during summer when leaves are on. However, there is some noise in summer [ALS](#) estimation. While the results from three plots are concentrated around 0.2, the estimation for two other plots is much higher. Reason for such noise was not determined and it might indicate some noise in the data or that the point density of the summer [ALS](#) is not sufficient to accurately estimate the canopy gap fraction of a dense forest.

Table 4.2 compares the results computed from the summer [ALS](#) data with canopy gap fraction results reported in [Kuusk et al. 2018](#). Their results are compared only with canopy gap fraction values computed from the summer [ALS](#) dataset, because [Kuusk et al. 2018](#) carried out their measurements in leaf-on season. [Kuusk et al. 2018](#) computed the canopy gap fraction from waveform [ALS](#) measured at 350 meters above ground, registering up to 15 targets per pulse and in total having the pulse density of 220 m<sup>2</sup>. In addition, ground measurements were made using five different methods, described in Chapter 2, but here only the averages of the five results are shown in Table 4.2. Overall, the results computed for this thesis have

#### 4. Results

	Reference data <a href="#">Kuusk et al. 2018</a>		Canopy gap fraction ALS	
	ALS	Ground measurements	100x100 m plot	segment-sized plots
<b>pine stand</b>	0.55	0.45	0.52	0.53
<b>birch stand</b>	0.3	0.25	0.11	0.34
<b>spruce stand</b>	0.35	0.3	0.18	0.19

Table 4.2.: Canopy gap fraction estimation from summer ALS data using Equation 3.1 compared with the results shown in [Kuusk et al. 2018](#).

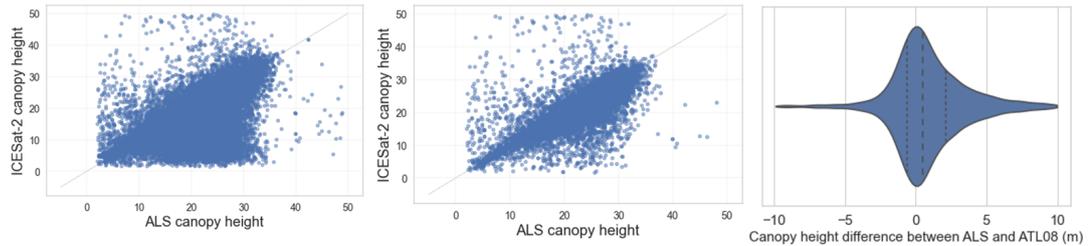


Figure 4.3.: Difference in canopy height estimation from ALS and ATL08 data. On the right all ATL08 segments are included while in the middle only segments from the night acquisition are added. The violin plot on the right shows the spread of data with only night acquisition and segments with canopy height difference of less than 10 meters.

the best agreement for the pine stand. Meanwhile the estimated canopy gap fraction for birch and spruce plots are underestimated compared to the values by [Kuusk et al. 2018](#). This may be due to the fact that the summer ALS data used in this thesis was collected at 3100 meters which means that the footprint of the laser beam is larger than the footprint of the laser flown by [Kuusk et al. 2018](#) at 350 meters. Therefore, the beams by [Kuusk et al. 2018](#) were able to penetrate smaller canopy gaps, leading to higher canopy gap fraction estimation. Hence, it must be considered that data used in this thesis may be underestimating the canopy gap fraction in more dense forest.

#### 4.1.3. Validation of canopy height estimation from ALS and ATL08

The canopy heights estimated from ALS and ATL08 data for each segment were compared using a scatter plot (Figure 4.3) in order to evaluate the suitability of the ALS data as reference. In addition, canopy height difference was calculated by subtracting the ATL08 value from ALS value together with other statistical values shown in Table 4.3. The desired outcome was to see the canopy height estimation from the two sources to differ by less than two meters. However, a lot of noise was present, especially in transects with lower canopy height (left on Figure 4.3). Based on the findings in the pre-processing step using the radiometric histograms, the segments for which the ICESat-2 data was acquired during the day were filtered out. This decreased the noise noticeably as shown in the middle of Figure 4.3.

#### 4.2. ATL08 canopy gap fraction from canopy to total photon ratio

	Day and night	Only night
<b>RMSE</b>	6.24	4.42
<b>MAD</b>	1.86	1.36
<b>Median difference (m)</b>	0.75	0.52
<b>Number of segments</b>	64048	31232
<b>% within 2 m</b>	64.3	72.2
<b>% within 3 m</b>	80.8	80.6

Table 4.3.: Statistics on canopy height estimation from ALS dataset and ATL08 product.

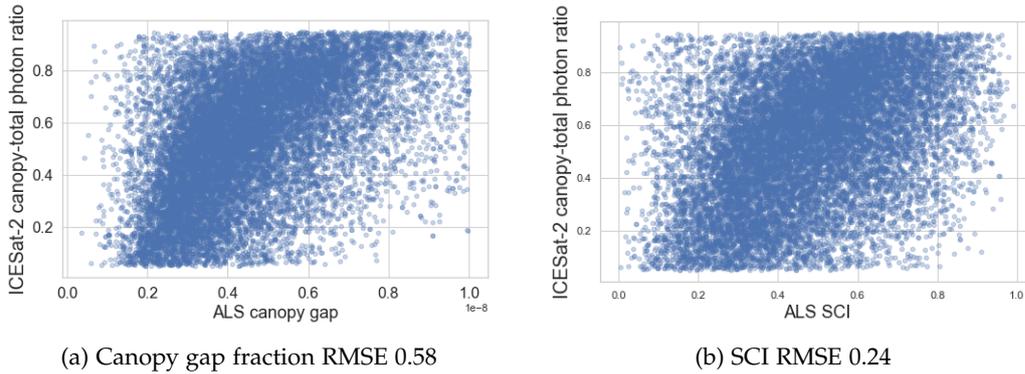


Figure 4.4.: Canopy to total photon ratio (Equation 3.3) plotted against canopy gap fraction (Equation 3.1) on the left and SCI (Eq 3.2) on the right for each segment.

Looking at the scatter plots as well as the violin plots in 4.3, on average ATL08 slightly underestimated the canopy height. Furthermore, some transects had height difference more than 10 meters which were filtered out for creating the violin plot. Table 4.3 shows that removing the transects from ICESat-2 daytime acquisition reduces the RMSE by a third. However, the number of segments left when the ATL08 daytime acquisition is removed is halved.

## 4.2. ATL08 canopy gap fraction from canopy to total photon ratio

The most straightforward approach for computing canopy gap fraction from ATL08 data would be to use Equation 3.3 which divides canopy photons by total photons in each segment. Figure 4.4 compares the results from this method with canopy gap fraction and SCI computed from ALS data. In both cases, the results from ATL08 tend to overestimate the canopy gap fraction, which however is more pronounced in Figure 4.4a than when compared to the SCI. Also, smaller RMSE values when compared to SCI indicate that there is better fit with the SCI than the actual canopy gap fraction value computed from ALS.

## 4. Results

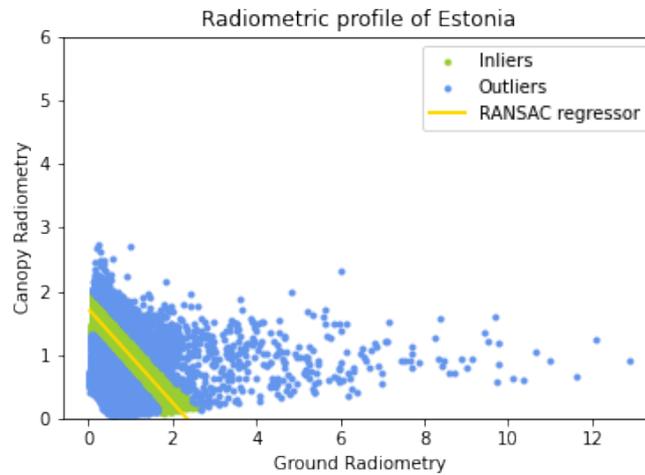


Figure 4.5.: Radiometric profile of Estonia including all segments from clear August night

### 4.3. ATL08 canopy gap fraction using radiometric profile

The second method tested in this thesis uses the equation 3.4 which uses radiometric profile to correct for the surface reflectance. The radiometric profile is used to find the ground and canopy reflectivity values which would be used in equation 3.4 to calculate canopy gap fraction. The clustering of radiometric values to a linear trend provides the canopy cover line where the y-intercept marks the ATL08 segments with total canopy cover and the x-intercept corresponds to transects with no canopy. The line fitted through is generated using RANSAC. In addition, this thesis looks at how much the radiometric profiles vary throughout the year.

Figure 4.5 shows the radiometric profile of Estonia when data acquired on a clear August night from all beams is included. It should be noted, that while this thesis only uses forested segments to compute the canopy gap fraction, all segments despite their forest cover are included in the radiometric profile to derive more complete line. The RANSAC regression estimated the y-intercept around 1.8 and x-intercept around 2.3. Therefore, based on this the  $\rho_v$  and  $\rho_g$  ratio needed for Equation 3.4 would be 0.72. This would suggest that the ground reflectivity is higher than the reflectivity of vegetation. Such condition would lead to more photons being reflected from the ground not because of the small canopy cover, but because of the nature of the ground surface. It would therefore indicate that without the correction, canopy gap fraction results would be over estimated as more photons would be reflected from the terrain due to its higher reflectivity.

Using these reflectivity values in Equation 3.4, canopy gap fraction for all the ATL08 segments from night acquisition was calculated. The results were plotted against the canopy gap fraction and the SCI computed from the ALS data and the results are shown in Figure 4.6. A lot of noise is evident and the results are similar to the first method (Figure 4.4). Compared to the canopy gap fraction from ALS (Figure 4.6a), there is still overestimation in the results computed from ATL08. This may suggest that the correction for reflectivity does not have too much effect on the results. As of the comparison with SCI (Figure 4.6b), there

#### 4.4. Trends in canopy gap fraction estimated from ATL08

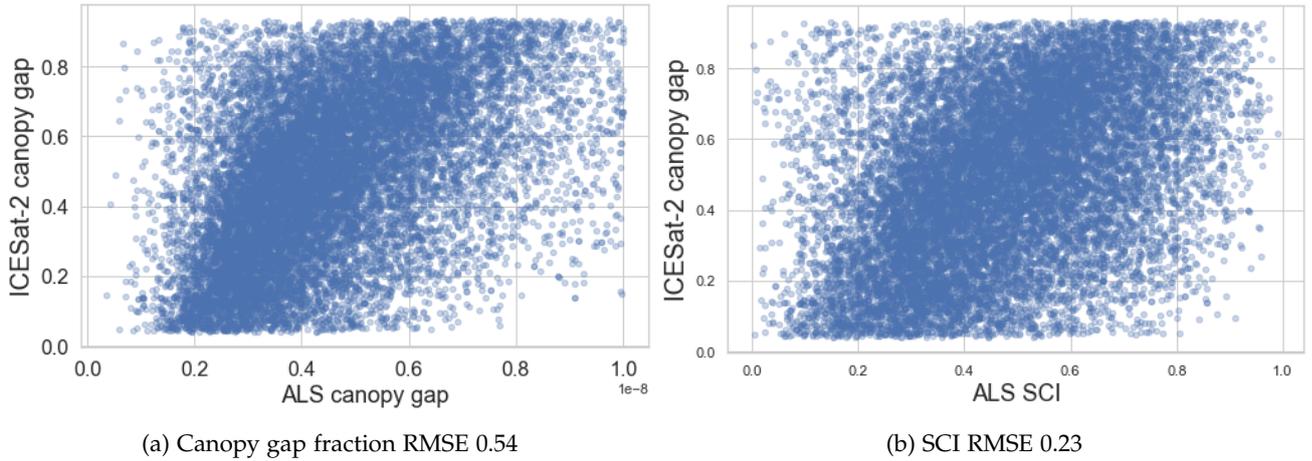


Figure 4.6.: Canopy gap fraction computed from ATL08 using radiometric profile to solve Equation 3.4 plotted against canopy gap fraction from ALS (Equation 3.1) on the left and SCI (Eq 3.2) on the right

	CGF using photon ratio (Eq 3.3)		CGF with reflectivity correction (Eq 3.4)	
	ALS CGF	ALS SCI	ALS CGF	ALS SCI
<b>RMSE</b>	0.58	0.24	0.54	0.23
<b>MAD</b>	0.17	0.17	0.19	0.19

Table 4.4.: Statistics on canopy gap fraction estimation from ATL08 product compared to canopy gap fraction (CGF) computed from ALS (Eq 3.1) and SCI (Eq 3.2)

is less of a trend.

Comparing the results of these two methods in Table 4.4, it is clear that the results are very similar. Both methods show very similar RMSE and the spread of the data, as well as MAD. Overall, the resulting canopy gap fraction estimation from ATL08 is not as accurate as was hoped. Although some linear trend is apparent, there is far too much noise to consider this approach sufficient for canopy gap fraction estimation. As using the second method involving radiometric profile is computationally more expensive, Equation 3.3 is considered as better approach of the two methods studied in the thesis.

#### 4.4. Trends in canopy gap fraction estimated from ATL08

In addition to comparing the results with data from ALS as done in previous section, this study also raised a question to what extent does the canopy gap fraction data computed from ATL08 pick up trends in vegetation. First, trends in different forest types are investigated, followed by looking for annual trend in canopy gap fraction.

#### 4. Results

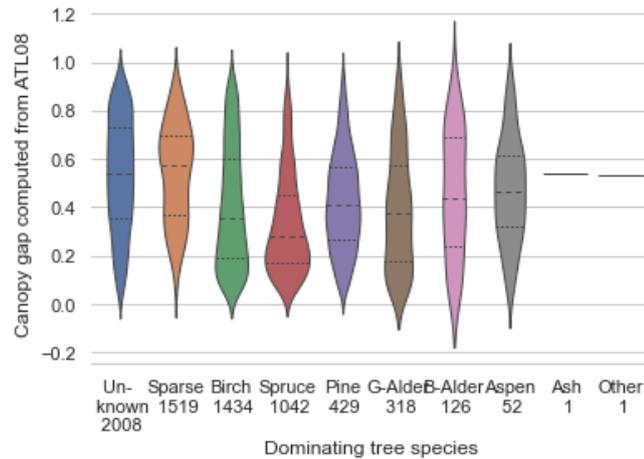


Figure 4.7.: Canopy gap fraction estimation from ATL08 in different tree species. The number under the tree species name indicates the number of segments included in the analysis.

##### 4.4.1. Canopy gap fraction in different forest types

As illustrated in Figure 3.3, forest dominated by different tree types is expected to have different canopy density. To answer one of the research questions on whether the canopy gap fraction computed from ATL08 differ in different forest types, the data was separated according to dominant tree types (Figure 4.7). Additional filters were used for computing this figure. Namely, only data from summer months was used and only segments where species variety according to the map used by Lang et al. 2018 was less than 2, meaning that the forest is assumed to be relatively homogeneous, were included.

Although the trend shown in Figure 4.7 is not very strong, it can be noted that the median canopy gap fraction value for spruce and birch forests are lower than for pine forest. Also, for forests that were marked as sparse by the species map (Lang et al. 2018), meaning that the forest was not dense enough for determining dominant tree species, the canopy gap fraction values are higher. However, it should be kept in mind that dominant tree species is not a strong indicator to be used as ground truth for estimating forest density. However, the trend shown in Figure 4.7 is indicating that ATL08 data does pick up some general trend between forest types.

##### 4.4.2. Canopy gap fraction throughout the year

Figure 4.8 shows the canopy gap fraction estimation computed from ATL08 data in different months throughout the year by strong beam and weak beam of ICESat-2. What can be noticed is that the canopy gap fraction values are higher from November to April and drop from May to September. This trend is more clear in the strong beam. The question that this result raises is that whether this is due to the higher canopy gap fraction in winter months when there are no leaves on the trees. Or, it could be due to snow cover that has not been highlighted by ATL08 snow flag. Presence of snow from December to February is very likely

#### 4.4. Trends in canopy gap fraction estimated from ATL08

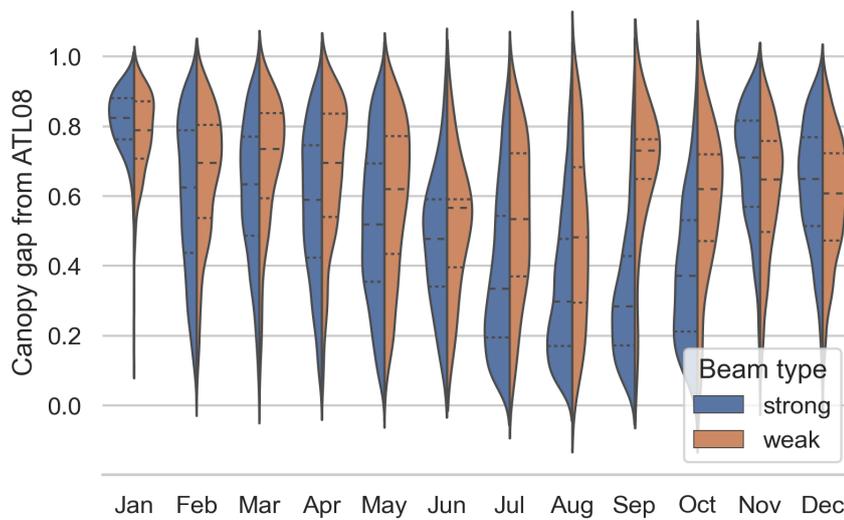
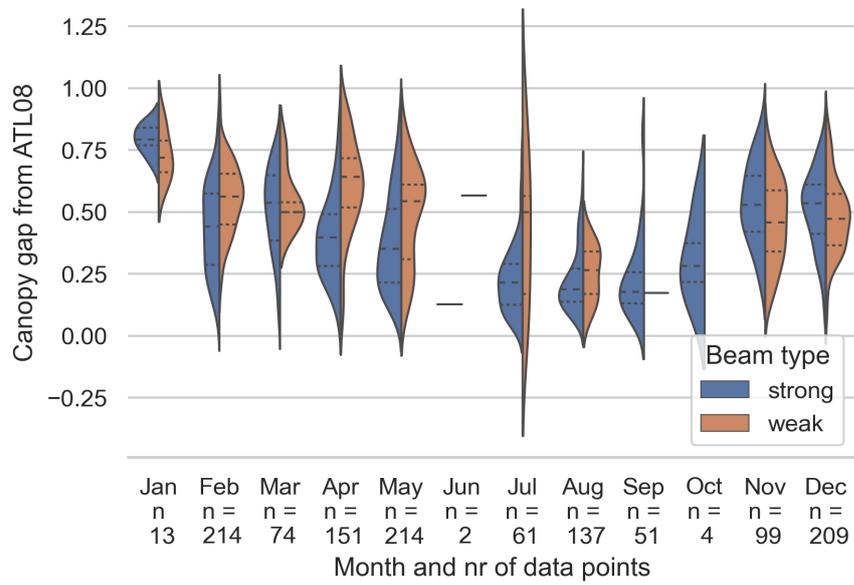


Figure 4.8.: Canopy gap estimation from ATL08 in different months by strong and weak beam.

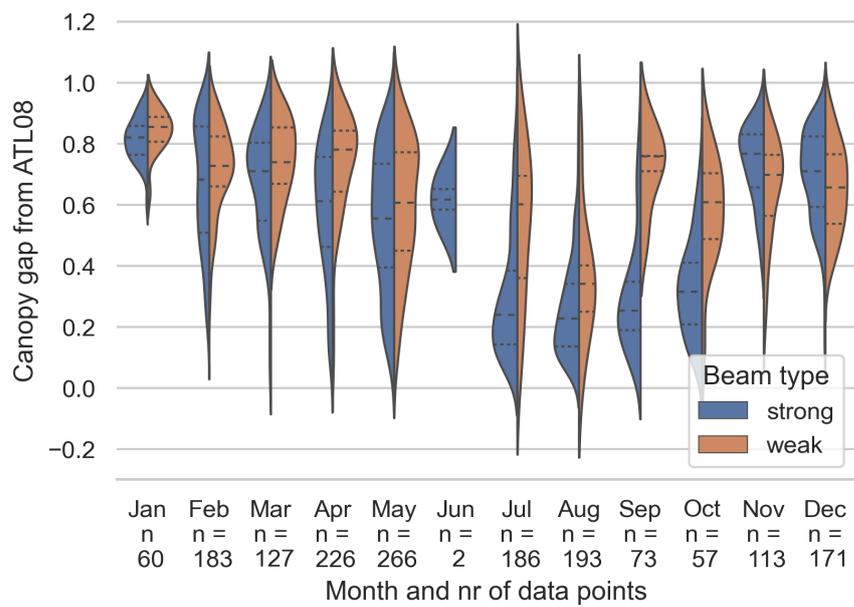
in Estonia and as explained earlier (Figure 2.4), it could lead to more photons detected from the ground and falsely low canopy gap fraction estimation.

To test whether the annual trend seen in Figure 4.8 is caused by errors from reflectivity or by changes in vegetation, canopy gap fraction through year was plotted separately for spruce (Figure 4.9a) and birch forest (Figure 4.9b). It should be noted that for producing these plots, only segments that according to the map by Lang et al. 2018 had low species variety and the respective dominant tree species were used. Therefore, for each month only very few segments can be used. Nevertheless, both plots show an annual trend of higher canopy gap fraction in the winter months than in the summer. However, while the values fluctuate from 0.5 to 0.2 for spruce, the annual fluctuation is more pronounced for birch going from 0.7 to 0.2. This is promising, because although as shown in Chapter 4.3, comparing the canopy gap fraction values derived from ICESat-2 data can be inaccurate, the results presented in this chapter indicate that the ATL08 data can pick up trends caused by different vegetation types.

4. Results



(a) Spruce



(b) Birch

Figure 4.9.: Canopy gap fraction computed from ATL08 for the entire year in spruce (top) and birch (bottom) forest.

## 5. Discussion

This chapter aims to discuss the results presented in Chapter 4, highlight the new knowledge produced by this thesis and make suggestions for to further improve the canopy gap fraction estimation from ICESat-2 ATL08 data.

This thesis used ALS data freely available by Estonian Land Board to evaluate the canopy gap fraction computed from ICESat-2 ATL08 data. In ideal circumstances, the ALS data would have been collected specifically for this study with parameters more suitable for the task. Namely, as suggested by Neuenschwander et al. 2022b, the preferred point density for the ALS data would be at least 5 pts/m<sup>2</sup> while the ALS data used in this study had point density of 2.1 pts/m<sup>2</sup> for the data retrieved in spring and only 0.8 pts/m<sup>2</sup> for the data retrieved in summer. This could be a reason why as shown in Chapter 4.1.2, the computed canopy gap fraction differed from results reported by Kuusk et al. 2018 who had much higher quality ALS data available. Furthermore, some anomalies were discovered in the canopy gap fraction results in the birch stand used in Kuusk et al. 2018 which may indicate some flaws in the data or possibility that the ALS data used in this thesis may fail in more dense forest. This could be the source of some of the noise in the results shown in Figure 4.4 and Figure 4.6, but most probably not all the noise. It could be said that the open ALS data used in this thesis was not of sufficient quality to have high confidence as validation data and therefore it is suggested that better results could be obtained in future research if higher quality ALS data together with more ground data would be used.

Two methods for computing canopy gap fraction from ATL08 data were used. While the first method using canopy to total photon ratio (Equation 3.3) is very simple to use, the second involving estimating the canopy to terrain reflectivity ratio (Equation 3.4) is computationally more expensive. Nevertheless, the results from the two methods shown in Chapter 4.3 did not differ much and therefore, the method of only using the canopy to total photon ratio is considered to perform better as it is also computationally less expensive. It was hypothesised that the canopy gap fraction computed from ATL08 will underestimate the results from ALS data. However, the opposite is shown in the results as overestimation of canopy gap fraction can be seen by ATL08 in Figure 4.4 and Figure 4.6. One theory to explain such overestimation is the higher reflectivity of terrain surface compared to the reflectivity of canopy. As shown in a theoretical example with snow in Figure 2.4, higher ground reflectivity could lead to more terrain signal photons to be detected by ATLAS. The radiometric profile shown on Figure 4.5 indicated slightly higher reflectivity from the ground. However, it should be considered that although clear linear trend is seen in the profile, there is still considerable amount of noise. This means that even when using RANSAC regression to fit a linear line through, there are many possibilities where the linear fit may indicate the x- and y-intercepts. If the difference in the surface reflectivity is the cause of the canopy gap fraction overestimation in ATL08 data, then better method for quantifying the ratio of vegetation to ground reflectivity is needed. Furthermore, the way of using the reflectivity

## 5. Discussion

ratio in Equation 3.4 may not have been sufficient to have affect on the outcome.

Overall, this thesis did not reach the desired outcome when validating the canopy gap fraction estimation with the ALS data. However, it was revealed that the ATL08 data can pick up on some trends in vegetation structure, which is a field not well studied in literature. Chapter 4.4.1 showed that different forest types overall do show a different canopy gap fraction estimated from ATL08 and it is possible also to observe annual trends as shown in Chapter 4.4.2. This proves that ATL08 can be valuable dataset for ecological studies. Whether the canopy structure data from ICESat-2 is sufficiently precise or not depends on the aim of the study. If very precise canopy gap fraction estimation is needed for a certain forest stand, then using ALS or methods is reasonable. However, there is no ALS dataset that has global coverage like ICESat-2 and periodic revisit times. Therefore, for large-scale estimation of canopy gap fraction, ATL08 has a lot of potential, but more research is needed.

## 6. Conclusion and future work

This thesis evaluated two methods for canopy gap fraction estimation from ICESat-2 ATL08 product. The results were compared to canopy gap fraction estimation and a SCI values computed from ALS data. Overall, the results from both methods had better agreement with canopy gap fraction predicted by SCI. However, a lot of noise was present in the data and the final results did not meet the initial expectation. Method for computing canopy gap fraction using canopy to total photon ratio from ATL08 was determined more preferable due to its lower computation time and similar results when compared to the alternative method.

This thesis suggests that ATL08 data reflects trends in forest structure. Both trends in different forest types and annual changes in canopy structure are reflected. That suggest that using ATL08 data for environmental studies at large scale has potential and should be researched further. Based on the lessons learned during the process of completing this thesis, the following suggestions are made for future studies on estimating canopy gap fraction from ATL08:

- Thanks to the potentially high quality and relatively large coverage, ALS data is suitable to be used as validation data for ATL08. However, high quality ALS data is needed to have high confidence in the canopy gap fraction results. The suggestion by Neuenchwander et al. 2022b of aiming for minimal quality of 5 pts/m<sup>2</sup> should be kept in mind;
- It seems likely that mechanism for correcting the difference in vegetation and terrain reflectivity is needed to improve the accuracy of canopy gap fraction estimation from ATL08 data. More research should be done on developing methods for estimating the reflectivity component.
- This thesis used the ATL08 version 5 which has errors in daytime data. Assuming that newer versions of ATL08 will fix this problem, using the version 5 is not recommended.
- This thesis did not investigate whether combining data of the same segment acquired at different times could improve the result of canopy gap fraction estimation. However, if the conditions in the canopy do not change between the acquisitions, such approach may lead to higher quality results.



# A. Reproducibility self-assessment

## A.1. Marks for each of the criteria

Grade/evaluate yourself for the 5 criteria (giving 0/1/2/3 for each):

1. input data: 3;
2. preprocessing: 2;
3. methods: 2;
4. computational environment: 2;
5. results: 1;

## A.2. Self-reflection

The results of this thesis are reproducible as only open data is used and the sources clearly referenced. Furthermore, the code used for preprocessing and analysis is made available on Github. As Jupyter notebooks were used for much of the data analysis, the author hopes that the code is easy to follow even without running it. A README file is shared to help setting up the computational environment with the required libraries. However, the generated results are not uploaded in tables as the data size used and generated in this thesis is relatively large. The results are only shared through figures and summary tables presented in this thesis.



# Bibliography

- Agency, E. E. (2022). Yearbook forest 2020. Online. <https://keskkonnaportaal.ee/sites/default/files/Teemad/Mets/Mets2020.pdf>.
- Armston, J., Disney, M., Lewis, P., Scarth, P., Phinn, S., Lucas, R., Bunting, P., and Goodwin, N. (2013). Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sensing of Environment*, 134:24–38.
- Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote sensing of Environment*, 64(3):234–253.
- Center, S. F. M. (2015). Estonia's new record-breaking trees. Online. <https://www.rmk.ee/organisation/press/news/news-2015>.
- Fisher, A., Armston, J., Goodwin, N., and Scarth, P. (2020). Modelling canopy gap probability, foliage projective cover and crown projective cover from airborne lidar metrics in australian forests and woodlands. *Remote Sensing of Environment*, 237:111520.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., et al. (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853.
- Korhonen, L., Korpela, I., Heiskanen, J., and Maltamo, M. (2011). Airborne discrete-return lidar data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, 115(4):1065–1080.
- Korhonen, L. and Morsdorf, F. (2014). Estimation of canopy cover, gap fraction and leaf area index with airborne laser scanning. In *Forestry applications of airborne laser scanning*, pages 397–417. Springer.
- Kuusik, A., Pisek, J., Lang, M., and Märdla, S. (2018). Estimation of gap fraction and foliage clumping in forest canopies. *Remote Sensing*, 10(7):1153.
- Lang, M., Kaha, M., Laarmann, D., and Sims, A. (2018). Construction of tree species composition map of estonia using multispectral satellite images, soil map and a random forest algorithm. *Forestry Studies*, 68(1):5–24.
- Magruder, L., Brunt, K., Neumann, T., Klotz, B., and Alonzo, M. (2021). Passive ground-based optical techniques for monitoring the on-orbit icesat-2 altimeter geolocation and footprint diameter. *Earth and Space Science*, 8(10):e2020EA001414.
- Magruder, L. A., Brunt, K. M., and Alonzo, M. (2020). Early icesat-2 on-orbit geolocation validation using ground-based corner cube retro-reflectors. *Remote Sensing*, 12(21):3653.
- Milenković, M., Reiche, J., Armston, J., Neuenschwander, A., De Keersmaecker, W., Herold, M., and Verbesselt, J. (2022). Assessing amazon rainforest regrowth with gedi and icesat-2 data. *Science of Remote Sensing*, page 100051.

## Bibliography

- Neuenschwander, A., Magruder, L., Guenther, E., Hancock, S., and Purslow, M. (2022a). Radiometric assessment of icesat-2 over vegetated surfaces. *Remote Sensing*, 14(3):787.
- Neuenschwander, A., Pitts, K., Jelley, B., Robbins, J., Klotz, B., Popescu, S., Nelson, R., Harding, D., Pederson, D., and Sheridan, R. (2021). Atlas/icesat-2 l3a land and vegetation height, version 5.
- Neuenschwander, A., Pitts, K., Jelley, B., Robbins, J., Markel, J., Popescu, S., Nelson, R., Harding, D., Pederson, D., Klotz, B., and Sheridan, R. (2022b). Ice, cloud, and land elevation satellite-2 (icesat-2) algorithm theoretical basis document (atbd) for land - vegetation along-track products (atl08). *National Aeronautics and Space Administration*.
- Neuenschwander, A. L. and Magruder, L. A. (2016). The potential impact of vertical sampling uncertainty on icesat-2/atlas terrain and canopy height retrievals for multiple ecosystems. *Remote Sensing*, 8(12):1039.
- Neumann, T. A., Martino, A. J., Markus, T., Bae, S., Bock, M. R., Brenner, A. C., Brunt, K. M., Cavanaugh, J., Fernandes, S. T., Hancock, D. W., et al. (2019). The ice, cloud, and land elevation satellite-2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. *Remote Sensing of Environment*, 233:111325.
- Ozanne, C. M., Anhof, D., Boulter, S. L., Keller, M., Kitching, R. L., Korner, C., Meinzer, F. C., Mitchell, A., Nakashizuka, T., Dias, P. S., et al. (2003). Biodiversity meets the atmosphere: a global view of forest canopies. *Science*, 301(5630):183–186.
- Palm, S., Yang, Y., and Herzfeld, U. (2018). Icesat-2 algorithm theoretical basis document for the atmosphere, part i: Level 2 and 3 data products. *National Aeronautics and Space Administration, Goddard Space Flight Center*.
- Queinnec, M., White, J. C., and Coops, N. C. (2021). Comparing airborne and spaceborne photon-counting lidar canopy structural estimates across different boreal forest types. *Remote Sensing of Environment*, 262:112510.
- Snow, N. and Center, I. D. (2022). The common metadata repository search api. Available online, <https://cmr.earthdata.nasa.gov>.
- Solberg, S., Brunner, A., Hanssen, K. H., Lange, H., Næsset, E., Rautiainen, M., and Stenberg, P. (2009). Mapping lai in a norway spruce forest using airborne laser scanning. *Remote Sensing of Environment*, 113(11):2317–2327.
- Tang, H., Song, X.-P., Zhao, F. A., Strahler, A. H., Schaaf, C. L., Goetz, S., Huang, C., Hansen, M. C., and Dubayah, R. (2019). Definition and measurement of tree cover: A comparative analysis of field-, lidar-and landsat-based tree cover estimations in the sierra national forests, usa. *Agricultural and forest meteorology*, 268:258–268.
- Vernimmen, R., Hooijer, A., and Pronk, M. (2020). New icesat-2 satellite lidar data allow first global lowland dtm suitable for accurate coastal flood risk assessment. *Remote Sensing*, 12(17):2827.
- Zhang, G., Chen, W., and Xie, H. (2019). Tibetan plateau's lake level and volume changes from nasa's icesat/icesat-2 and landsat missions. *Geophysical Research Letters*, 46(22):13107–13118.

## Colophon

This document was typeset using  $\text{\LaTeX}$ , using the KOMA-Script class `scrbook`. The main font is Palatino.



