CHARLES UNIVERSITY

Faculty of Science

Department of Applied Geoinformatics and Cartography

Study program: Geography

Curriculum: Cartography and Geoinformatics

Bc. Adéla ŠEDOVÁ

Forest monitoring using the GEDI sensor

Monitorování lesa pomocí GEDI senzoru

Master thesis

Supervisor: Ing. Markéta Potůčková, Ph.D.

Prague 2022

Vysoká škola: Univerzita Karlova **Fakulta:** Přírodovědecká

Katedra: Aplikované geoinformatiky a kartografie **Akademický rok:**

Zadání diplomové práce

pro Bc. Adélu Šedovou

obor Kartografie a geoinformatika

Název tématu: Využití senzoru GEDI pro monitorování lesních porostů

Zásady pro vypracování

Global Ecosystem Dynamics Investigation (GEDI) je projekt monitorování lesních porostů využívající full-waveform LiDARový systém nesený mezinárodní vesmírnou stanicí (ISS). Volně dostupné GEDI produkty obsahují mezi jinými nadmořskou výšku zemského povrchu, výšku vegetace, pokryvnost vegetace (canopy cover fraction). Diplomová práce má za cíl stanovit metodiku verifikace těchto parametrů včetně přesnosti prostorového určení stopy GEDI LiDARového paprsku na základě dostupných leteckých LiDARových dat (ALS). Dalším cílem je ověřit, jak lze GEDI data využít pro systematické monitorování výšky a zápoje porostu v kombinaci s daty ALS. Pro harmonizaci ALS a GEDI bude prozkoumán a popř. využít volně dostupný GEDI Simulátor.

Rozsah grafických prací: dle potřeby

Rozsah průvodní zprávy: cca 50 stran

Seznam odborné literatury:

Dubayah R. et al. (2020): The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1, 14.

Bouch.er P.B., Hancock S, Orwig D.A., Duncanson L., Armston J., Tang H., Krause K., Cook B., Payner I., Li Z., Elmes A., Schaaf C. (2020): Detecting Change in Forest Structure with Simulated GEDI Lidar Waveforms: A Case Study of the Hemlock Woolly Adelgid (HWA; Adelges tsugae) Infestation. *Remote Sensing*. 12, 8, 1304.

Potapov P., Xinyuan L., Hernandez-Serna A., Tyukavina A., Hansen M.C., Kommareddy A., Pickens A., Turubanova S., Tang H., Silva C. E, Armston J., Dubayah R., Blair J. B., Hofton M. (2021): Mapping global forest canopy height through integration of GEDI and Landsat data, *Remote Sensing of Environment*, 253.

Vedoucí diplomové práce: Ing. Markéta Potůčková, Ph.D.

Konzultant diplomové práce:

Datum zadání diplomové práce: 14.1.2021

Termín odevzdání diplomové práce: 20.7.2022

Platnost tohoto zadání je po dobu jednoho akademického roku.

…… .....……

Vedoucí diplomové práce Garant studijního oboru

V Praze dne 20.7.2022

Declaration

Hereby I declare that I have written this thesis myself and that I quoted all references accordingly. Neither this thesis nor its part was used to obtain other academic degree of the same or other levels. I hereby agree with using this thesis for academic purposes and I agree with its inclusion in the evidence of academic works.

In Prague on 20 July 2022

…….

Adéla Šedová

Acknowledgment

Hereby I would like to thank the thesis supervisor, Ing. Markéta Potůčková, Ph.D., for her patience, time, support, shared knowledge, and helpful advice throughout the time I have been working on this thesis. I would like to thank the research institute of CzechGlobe, namely Ing. Jan Hanuš, for kindly providing laser scanning dat. This work is based on use of Large Research Infrastructure CzeCOS supported by the Ministry of Education, Youth and Sports of CR within the CzeCOS program, grant number LM2018123. The beginning introduction into the data and problematic was supported by the kind and helpful advice from Markus Hollaus, Dipl.-Ing. Dr.techn. from Department of Geodesy and Geoinformation of Technical University in Vienna, who let me focus on the topic in his seminar and shared valuable knowledge. I would like to thank my friends, namely Alex Šrollerů for his valuable advice and keeping my spirits high, Jakub Dvořák for technical advice, and Daniela Kebertová for help with language corrections. Finally, I want to thank my family for their constant encouragement and support during my studies. Without you, I could not have finished.

Abstract

The overall objective of this thesis was to explore the use of GEDI and its integration with Airborne Laser Scanning (ALS) for large scale forest monitoring. The study was carried out using a sample of GEDI footprints that fell into the timeline of three available ALS datasets that were acquired during the same year. The study area, located in Southeast of Czechia, is covered with mature 121-year-old forest monoculture of Norway spruce (*Picea abies*), and due to frequent disturbances caused by infestation is closely monitored as a part of research on forest dynamic. As a result, the forest is highly fragmented, and due to its dynamic character, close dates of acquisitions were preferred to a larger dataset. Canopy gaps and low tree densities are known to pose a challenge for large-footprint full-waveform LiDARs. The specific of GEDI sensor, such as its footprint size, were specially designed to overcome these challenges. The options of optimising GEDI's geolocation accuracy were explored. A tool for integrating GEDI and ALS data, the GEDI Simulator, was used to standardise both data sources and derive elevation height, Relative Height (RH) and Canopy Cover Fraction (CCF). The metrics were derived from real GEDI waveforms, simulated GEDI-like waveforms, and calculated from the ALS point cloud, and used to validate the performance of GEDI and the Simulator. The results have shown that fragmented forest still poses a challenge for GEDI, on both the employed geolocation optimalisation approach called waveform matching and the derived vegetation metrics. Furthermore, dead-standing trees, that were frequently observed in selected study site, might be linked to the observed RMSEs. The author also suggests using the CCF metric with caution when integrating ALS and GEDI, as there were observed large biases between the values derived by GEDI and values derived by the GEDI Simulator.

Key words: GEDI, LiDAR, forest monitoring, GEDI Simulator, ALS, data fusion, geolocation

Abstrakt

Hlavním zaměřením této práce je využití GEDI a jeho integrace s daty leteckého laserového skenování (Airborne Laser Scanning – ALS) pro monitorování lesa ve velkém měřítku. Studie byla provedena na vzorku GEDI dat, které spadaly do časové osy tří dostupných souborů dat ALS nasnímaných během téhož roku. Oblast zájmu, ležící na jihovýchodě Česka, je charakteristická 121 let starou smrkovou monokulturou (*Picea abies*) a vzhledem k častým narušením lesní dynamiky (napadení stromů) je oblast bedlivě sledována. V důsledku častých změn je les vysoce fragmentovaný, a zejména proto bylo s ohledem na jeho dynamiku upřednostněno užší časové rozmezí nasnímaných dat před jejich kvantitou a plošným rozsahem. Je známo, že nízká hustota stromů a velké mezery mezi nimy představují výzvu pro velkostopé full-waveform LiDAR senzory. Specifika GEDI senzoru, jako například velikost stopy snímání, byla speciálně navržena pro překonání těchto problémů. V této práci byly nejprve zkoumány možnosti optimalizace geolokační přesnosti GEDI. Následně byl pro integraci dat GEDI a ALS využit dedikovaný nástroj GEDI Simulator, který byl použit ke standardizaci obou zdrojů a odvození metrik nadmořské výšky, relativní výšky (RH) a vegetačního pokryvu (CCF). Metriky byly odvozeny z dat GEDI, simulovaných GEDI signálů a manuálně vypočteny z mračna bodů ALS a byly použity ke zhodnocení přesnosti GEDI a simulátoru. Výsledky ukázaly, že fragmentovaný les pro sensor GEDI stále představuje výzvu, a to jak z hlediska použitého přístupu optimalizace geolokace (tzv. „waveform matching") tak odvozených vegetačních metrik. Kromě toho mohou být výsledné hodnoty RMSE spojeny s hojným výskytem stále stojících odumřelých stromů v oblasti zájmu. Autorka také doporučuje opatrnost s použitím metriky CCF při integraci ALS a GEDI, a to z důvodu vypozorovaných velkých odchylek mezi hodnotami odvozenými z GEDI a hodnotami odvozenými ze simulátoru GEDI.

Klíčová slova: GEDI, LiDAR, monitorování lesa, GEDI Simulator, LLS, integrace dat, geolokace

Table of contents

List of abbreviations

AGB - Aboveground Biomass
AGBD - Aboveground Biomass Density
ALS - Airborne Laser Scanning
ASCII - American Standard Code for Information Interchange
ATLAS - Advanced Topographic Laser Altimeter System
CCF - Canopy Cover Fraction
CHM - Canopy Height Model
CzechGlobe - Global Change Research Institute of the Czech Academy of Sciences
CzeCOS - Czech Carbon Observation System
DEM - Digital Elevation Model
DLR - German Aerospace Center
ESFRI - European Strategy Forum on Research Infrastructures
ETRS89 - European Terrestrial Reference System 1989
GEE - Google Earth Engine
GEDI - Global Ecosystem Dynamics Investigation
GLAS - Geoscience Laser Altimeter System
GMTED DEM - Global Multi-resolution Terrain Elevation Data Digital Elevation Model
HDF5 - Hierarchical Data Format version 5
ICESat - Ice, Cloud, and Land Elevation Satellite
ICESat-2 - Ice Cloud and Elevation Satellite-2
ICOS - Integrated Carbon Observation System
ISS - International Space Station
LAI - Leaf Area Index
LiDAR - Light Detection and Ranging
LPDAAC - Land Processes Distributed Active Archive Center
LTER - Long Term Ecological Research
LVIS - Land, Vegetation, and Ice Sensor
MODIS - Moderate Resolution Imaging Spectroradiometer
NASA - National Aeronautics and Space Administration
NISAR - NASA ISRO Synthetic Aperture Radar
ORNL DAAC - Oak Ridge National Laboratory Distributed Active Archive Center
RH - Relative Height

ROI - Region of Interest

SAR - Synthetic Aperture Radar

SRTM - Shuttle Radar Topography Mission

UTM33 - Universal Transverse Mercator Zone 33N

List of Figures:

List of Tables:

1 Introduction

Forests have an important role in the carbon cycle and climate regulation, they provide a variety of ecosystem services, and along with other ecosystems are a key component of habitat quality (Lang et al., 2021). Forest distribution and structure are changing fast as a result of human activities and natural processes (Rishmawi et al., 2021).

The height of a canopy is an important morphological attribute that aids in quantifying ecosystem response to external forces like climate and land use change (Dubayah et al., 2020; Silva et al., 2018; Lang et al. 2021). Furthermore, canopy height is an important biodiversity variable and a powerful predictor of species richness at all scales, from local to global (Lang et al., 2021). Forest canopy height is used to quantify aboveground biomass and wood volume in forests to monitor the effects of forest degradation, to assess the effectiveness of forest restoration, and to simulate other important ecological variables including primary production and biodiversity (Potapov et al., 2021).

The three primary sources of remotely sensed data used in forest monitoring are optical, radar, and Light Detection and Ranging (LiDAR). Current optical and radar sources are, however, limited by sensor saturation at particular biomass levels. LiDAR systems, whether airborne or spaceborne, can record the horizontal and vertical structure of canopy in great detail, allowing to estimate biomass with more precision than approaches based on radar or optical data (Fayad et al., 2021) and are therefore powerful instruments for acquiring ground elevation and canopy height information on vast scales (Adam et al., 2020). Airborne laser scanning (ALS) has been in long term use as a source of canopy data. However, the cost of airborne LiDAR data restricts repeated large-area monitoring, resulting in limited spatial coverage (Roy et al., 2021; Adam et al., 2020). Spaceborne lidar instruments, on the other side, can provide worldwide coverage of elevation height and vertical canopy structure (Adam et al., 2020). In the future, space-based LiDAR systems dedicated to vegetation assessment might be expected and required (Milenković et al., 2017).

Since the beginning of spaceborne LiDAR missions, such as the Geoscience Laser Altimeter System (GLAS) and Ice, Cloud, and Land Elevation Satellite (ICESat), ways to derive elevation height and vertical canopy structure information were explored and implemented. Currently, spaceborne lidar is widely acknowledged to be a strong instrument for global vegetation monitoring (Rishmawi et al., 2021). Hovewer, none of the mentioned instruments were not initially designed to extract vegetation structure, therefore the sensor characteristics such as a footprint size are not well suited for these applications.

The Global Ecosystem Dynamics Investigation (GEDI) LiDAR is the first spaceborne LiDAR created primarily to analyse ecosystem structure by providing vertical profiles of forest canopies (Dubayah et al., 2020). Given its high sampling rate and a footprint size designed to overcome blending of ground and canopy signals (Fayad et al., 2021), GEDI is meant to address the main challenges in biodiversity and carbon cycle monitoring (Guerra-Hernández & Pascual, 2021).

The potential and use of pre-launch simulated GEDI data were analysed by multiple studies (Boucher et al., 2020; Silva et al., 2018). However, the post-launch data might differ from the predictions, therefore validation studies are highly encouraged by the GEDI development team to improve the performance of next generations of GEDI data products and related tools (Guerra-Hernández & Pascual, 2021). A tool developed during pre-launch validations and calibrations, a GEDI Simulator, has shown a great potential for integrating data acquired by ALS and GEDI. The Simulator has been suggested as a standardisation tool for creation of a multimodal, multitemporal dataset consisting of GEDI and ALS data, possibly taken by different sensors, to monitor structural canopy changes (Huettermann et al., 2022 Roy et al., 2021; Boucher et al., 2020). Large-footprint full-waveform lidar data have proven valuable for monitoring structural changes and disturbances in forests (Tang & Dubayah, 2017), such as identifying developmental phases of forest plots (Marselis et al., 2018; Drake et al, 2002; Harding et al., 2001), estimating biomass change (Huang et al., 2013; Dubayah et al., 2010) or estimating historical hurricane damage (Weishampel et al., 2007). Fusing ALS and GEDI data with different time acquisitions could bring new opportunities for LiDAR use in operational forestry (Dorado-Roda et al., 2021), however, first it is necessary to understand and evaluate the new GEDI data and the potential of its integration with ALS data.

The objective of this thesis is to explore the feasibility of using GEDI for local (or large scale) forest monitoring in a temperate forest in the Southeast of Czechia, where the frequent forest disturbances caused high fragmentation of the forest. The overall goal consists of following tasks:

- explore the options of optimising GEDI's geolocation accuracy
- research the impact of various GEDI sensor parameters on the performance of the instrument, use a suitable process to filter out undesirable data
- standardise both data sources (GEDI and ALS) using GEDI Simulator

- find vegetation metrics that can be derived from standardised data, as well as calculated from ALS point cloud
- validate the performance of GEDI and the GEDI Simulator tool

2 Global Ecosystem Dynamics Investigation (GEDI)

The Global Ecosystem Dynamics Investigation (GEDI) LiDAR is the first spaceborne LiDAR mission specialized on monitoring vegetation structure, with a focus to improve forest aboveground biomass (AGB) quantification, to model land surface carbon budgets, and to understand the impact of vegetation structure on habitat quality and biodiversity (Dubayah et al., 2021). The mission's primary instrument is a full waveform LiDAR. The coverage of the Earth's surface is between 51.6° N and 51.6° S. The LiDAR instrument is made up of three full-waveform lasers that capture footprints with a diameter of 25 meters. This footprint size is suggested as appropriate for vegetation detection by Pang et al. (2011) and it is also similar to what is frequently used for field plot size in temperate climate forest surveys (Milenković et al., 2017). The lasers pulse at a rate of 242 Hz, emitting 1064nm light. One of the three lasers is divided into two beams (coverage beams), while the other two continue to operate at full power (power beams). The coverage beam is split into two transects, which are then dithered to reach the ground 600m apart, yielding four ground transects in total (Fig.1). The power beams are also dithered with the same spacing. Thus, the instrument produces eight beam ground transects, spaced approximately 60 m along the track and 600 meters apart in the cross-track direction (Luthcke et al., 2019).

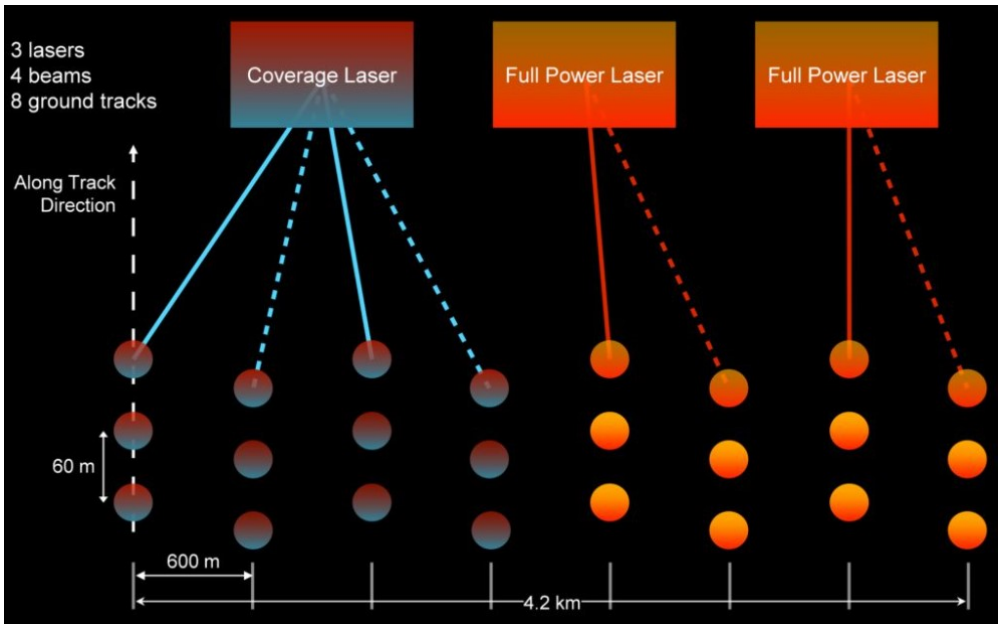

Fig. 1: GEDI's ground sampling pattern

Source: (Hofton et al., 2019)

The GEDI instrument was launched to the International Space Station (ISS) in December 2018 with on-orbit checkout in April 2019 (Dubayah et al., 2020). On January 21, 2020, the first set of GEDI data was released, followed by multiple derived products and versions. GEDI was initially planned to be a two-year mission, during which approximately 10 billion cloud-free observations have been expected to be obtained (Hofton et al., 2019). The data acquisition was prolonged until 2023, which is supposed to assure even denser sampling. An example of current coverage is shown in Fig. 2.

Mounted to the ISS, GEDI follows its orbit. The International Space Station (ISS) is a challenging platform for Earth observation. Due to a variety of reasons, it produces unique structural vibrations and orbit disturbances. The orbit is not constant because of the huge structural size of the station, relatively low flying altitude of 415 km, and orbit maintenance operations. Because the ISS is not in a recurring orbit, the GEDI measurements might not be periodically repeated. However, GEDI has a six-degree rotation capability, allowing the beams to be oriented up to 40 kilometres of either side of the ISS's ground track (Fayad et al., 2020).

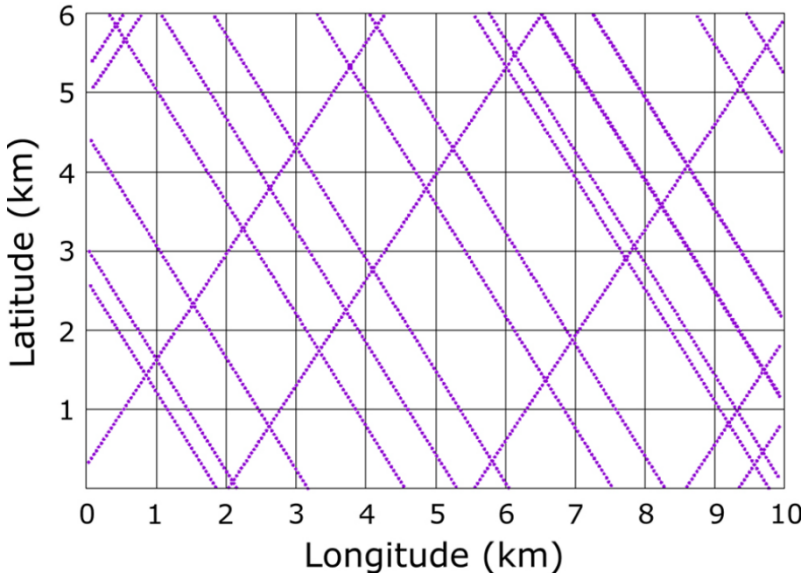

Fig. 2: An example of GEDI's track coverage over the equator

Source: GEDI Ecosystem Lidar (2022)

Relation to other NASA missions

The GEDI mission was implemented as a response to the goals of the National Academy of Sciences' and National Aeronautics and Space Administration (NASA) Science Mission Directorate's observational goals. The goals mention a raising necessity for global lidar vertical structure measurements. The main objective of the GEDI mission therefore

became to address major challenges in carbon cycle and biodiversity (GEDI Ecosystem Lidar, 2022). The mission lays the path for the incorporation of global, multitemporal, and full-waveform laser data into operational forestry (Duncanson et al., 2020). The GEDI data is supposed to support further improvement of upcoming NASA missions, such as the next generation of NASA ISRO Synthetic Aperture Radar (NISAR) and the Ice Cloud and Elevation Satellite-2 (ICESat-2) data products. In addition, GEDI and the German Aerospace Center (DLR) have a formal agreement to integrate GEDI's data with the DLR's TanDEM-X SAR interferometry mission to generate global maps of canopy heights (GEDI Ecosystem Lidar, 2022)

Calibration and validation

To construct relevant pre-launch calibration equations, GEDI relied on crowd-sourced in situ and airborne observations from around the world. Airborne laser scanning and ground plot field inventory datasets were gathered, standardized, and GEDI-like waveforms were simulated using the GEDI Simulator. The biomass equations were then calibrated with the field biomass values as reference data using the characteristics of simulated waveforms and derived metrics (Hofton et al., 2019). The calibration was complicated by the large strata of topographic, climatic, and edaphic conditions across the GEDI's observation zone, and also by using both high and low power beams, which introduces variations of noise levels (Lang et al., 2021).

2.1 GEDI products overview

The GEDI datasets are divided into multiple tiers (Levels), from the raw data to the more refined data obtained through waveform signal analysis and metrics extraction. As a result, there is a large number of derivatives available for each sensed footprint (Fayad et al., 2021).

There are several levels of products (Tab. 1) available for open-access download stored as single waveforms in Hierarchical Data Format version 5 (HDF5, .h5). The assigned levels indicate the amount of processing that the data has undergone after collection. The levels are assigned according to the following structure: radiometrically corrected raw waveforms (Level 1), waveforms with added geophysical variables (Level 2) and value-added data products mapped on uniform space-time grids (Level 3 and 4).

Product name	Description	Resolution	Availability
L1A	raw waveforms	25 m	unavailable
L1B	geolocated waveform, geolocation information	25 m	LPDAAC
L2A	ground elevation, canopy top height, RH metrics	25 m	LPDAAC
L2B	LAI, LAI profile, CCF, CCF profile	25 m	LPDAAC
L3	gridded L2 metrics	1 km	ORNL DAAC
L4A	footprint estimations of AGBD	25 m	ORNL DAAC
L4B	gridded AGBD	1 km	ORNL DAAC

Tab. 1: GEDI data product characteristics (RH – relative height, LAI – leaf area index, CCF – canopy cover fraction, AGBD – aboveground biomass density) and centres providing data download; LPDAAC – Land Processes Distributed Active Archive Center, ORNL DAAC – Oak Ridge National Laboratory Distributed Active Archive Center
Source: (Hofton et al., 2019)

The Level 1B data products are intended for users who want to analyse the entire waveform return and derive new geolocated waveforms by own approach. It is suggested to start with Level 2 data product (Blair et al., 2021), if the interest is output metrics, as the Level 2 data products are already geolocated using an algorithm described in (Luthcke et al., 2019). However, as the calibration analysis used for computing geolocation of the early products was preliminary and limited, the geolocation accuracy remains advised to be considered depending on intended use (Blair et al., 2021).

The first release of GEDI products - Version 1 was made publicly available in early 2020. Due to its relatively high geolocation error, it was suggested to use the products with caution (Adam et al., 2020, Lang et al., 2021, Potapov et al., 2021, Rishmawi et al., 2021, Roy et al., 2021), especially when using data collected over spatially varied canopies and for fusing GEDI with multiple data sources. After further reprocessing, Version 2 was released in April 2021. The changes in Version 2 include improved geolocation accuracy, inclusion of spatial metadata to allow visualizing orbit tracks and spatial querying in NASA Earthdata Search. The granule size was cut from one complete ISS orbit to four segments of each orbit. The filenames for Version 2 have been changed to contain the segment number and version

number. Parameters for quality filtering, such as degrade flag (described more in detail in section 4) were transformed from binary values to include specific values for each degraded condition. (Blair et al., 2021)

Elevation

All elevation values included in GEDI data products are relative to World Geodetic System 1984 (WGS84) ellipsoid. The elevation height is determined only for on-land areas, no bathymetric mapping is currently included in the GEDI data products. The GEDI datasets also contain Tan-DEM-X digital elevation model on footprint/gridded level, obtained by X-band Synthetic Aperture Radar (SAR). The Tan-DEM-X elevation values are reported to show high correlation to GEDI's elevation values (Spracklen & Spracklen, 2021).

The GEDI elevation metrics were evaluated in several studies (Liu et al., 2021; Spracklen & Spracklen, 2021; Adam et al., 2020). The elevation accuracy is influenced by slope, canopy density, solar elevation (Spracklen & Spracklen, 2021). Steep terrain seems to highly influence the inaccuracy of elevation height measurements, as shown by multiple studies (Huettermann et al., 2022; Adam et al., 2020). The study of Huettermann et al. (2022), observed almost linear relationship between the increasing slope and the increasing RMSE of height metrics derived from both real GEDI waveforms and simulated GEDI-like waveforms. Especially in case of large-footprint waveforms, the error can be increased if the terrain is sloped and forested; the returns from canopy and slope can occur at the same height, and cannot be distinguishable, as the ground returns in one waveform would be spread out over different heights (Adam et al., 2020). If a footprint is misplaced on a sloped area in the same direction as the slope, even larger height error is introduced.

The study of Huettermann et al. (2022) showed that a higher canopy cover caused lower elevation height accuracy. Misreading of elevation was reported to have strong correlation with misreading of canopy height (Spracklen & Spracklen, 2021). A factor that is more sensor-specific, that is beam sensitivity, was reported to have minor correlation (Spracklen & Spracklen, 2021), where higher sensitivity seemed to lead to higher error, or no correlation at all (Adam et al., 2020). The accuracy for GEDI elevation metrics is reported to be higher than of existing SRTM and The Global Multi-resolution Terrain Elevation Data (GMTED) DEM products (Liu et al. 2021).

Relative height (RH)

The vertical distribution of plant material in the canopy is described by relative height (RH) of returned LiDAR energy (Dubayah et al., 2010). The values indicate the height below which specified proportion of the LiDAR's energy is returned. The RH100, that is computed from the first and last recorded returns of the waveform, is meant to represent a top of the canopy. However, due to a frequent influence of noise on the first return, there is usually minimal leaf area to capture at the very top of the canopy. Thus, RH100 was reported as an unstable metric (Blair and Hofton, 1999 in Duncanson et al., 2022; Potapov et al., 2021) and either RH98 (Duncanson et al., 2022) or RH95 (Bauer et al., 2021; Potapov et al., 2021) is suggested instead as a top of the canopy indicator. The RH50 stands for median energy height and has been applied for AGBD modeling (Duncanson et al., 2022). In spots with minimal canopy cover, the lower GEDI RH (e.g., RH10 and lower) will often have sub-zero values (Spracklen & Spracklen, 2021). The reason for that is that a relatively high fraction of the waveform energy comes either from the ground or below the *elev_lowestmode* (the mean of elevation height within the footprint). To set an example, if the ground return contains 30% of the energy, RH1 through RH15 are likely to be negative, since half of the energy required to compute the footprint's mean ground elevation (*elev_lowestmode*) falls below the centre of the ground return (Blair et al., 2021).

2.2 Data fusion

The concept of fusing full waveform data from a spaceborne LiDAR with other data sources to derive vertical canopy metrics has been demonstrated in numerous studies. Canopy metrics derived from GLAS were interpolated nationally, continentally, and globally (Chi et al., 2015; Simard et al., 2011; Baccini et al. 2008 in Potapov et al., 2021) using Moderate Resolution Imaging Spectroradiometer (MODIS) optical data. At local and biome scales GLAS-based forest canopy height and biomass were interpolated using Landsat imagery (Hansen et al., 2016; Tyukavina et al., 2015; Duncanson et al., 2010; Hudak et al., 2002 in Potapov et al., 2021). Several studies have used GLAS waveforms to increase the quantity of training data available for height or biomass models (Potapov et al., 2019; Li et al., 2107; Dubayah et al., 2012 in Healey et al., 2020). However, the major goal of GLAS was not to investigate canopy structural characteristics but to provide global altimetry. Its approximate footprint diameter of 70 m was rather large for vegetation monitoring; moreover, the footprints were spaced by around 170 meters along tracks, often separated by more than 10 kilometres in cross-track direction (Montgomery et al., 2021). The GEDI is aimed to provide

significantly denser sampling, using an acquisition strategy (footprint size, acquisition rate and beam coverage) designed to provide vegetation information on tropical and temperate forests (Healey et al., 2020).

Multiple studies (Potapov et al., 2021; Healey et al., 2020) explored the integration of GEDI with Landsat. The diameter of GEDI's footprint roughly matches Landsat's 30 m ground sampling distance, enabling a relatively straightforward footprint-to-pixel matching (Healey et al., 2020). Fusion of spaceborne optical and spaceborne LiDAR data might provide periodic global forest monitoring (Potapov et al. 2021). A raster-based tool for Google Earth Engine (GEE) was already published (Healey et al., 2020) to enable flexible integration of Landsat and GEDI datasets. An integration of GEDI and radar (Sentinel-1) has also already been shown feasible (Chen et al., 2021), increasing the accuracy of forest stand volume mapping while requiring fewer field measurements. The fusion of GEDI and TanDEM-X DEM for predicting sub-canopy topography was introduced in Tan et al. (2020). Fusing GEDI with another remote sensing data source aids to overcome the sampling limitations of the inconstant ISS orbit and discrete footprint acquisition (Liu et al., 2022).

2.3 Quality filtering

2.3.1 Power/coverage beams

The use of only power beam is recommended especially for areas with dense canopy cover (Liu et al., 2021). The reason is that the strong beam penetrates deeper and extracts ground data more precisely. Power beams have 98% penetration threshold, while the threshold for coverage beams is 95%. That means, as explained in (Blair et al., 2021), that coverage beams were only designed to penetrate canopies of up to 95% canopy cover under "average" conditions, for which, though, they yield highly useful information (Liu et al., 2021).

2.3.2 Quality flags

Both the L2A and L2B products include a set of quality flags and metrics that can be used to filter shots with poor geolocation performance, signal quality, and waveforms influenced by cloud and/or other land surface conditions. The quality parameters are in detail described in (Blair et al., 2021) and briefly introduced in following paragraphs.

A quality flag parameter is a binary value that is dependent on energy, sensitivity, amplitude, real-time surface tracking quality, and difference to a DEM. If the pre-set standards for all mentioned variables are met, quality flag is set to 1. In L2B products, with the help of land cover datasets, quality flag is set to 0 for all shots over open water or urban infrastructure. Quality flag is part of most quality filtering combinations (Huettermann et al., 2022; Milenković et al., 2022; Musthafa & Singh, 2022).

A sensitivity parameter is a binary value that determines the ability of a waveform to penetrate canopies and detect the ground level. The threshold over land is set to be *sensitivity >0,9* in order for a waveform quality flag to be 1. When the value is set to 0, the detected ground level within a shot doesn't represent the true ground level due to weak returns caused by meteorological conditions and/or dense canopy cover. Blair et al. (2021) suggest that for conditions such as dense forest, setting higher sensitivity threshold (e.g. 0.95) might produce better results, however, Adam et al. (2020) found in their study no direct connection between the sensitivity parameter and vegetation metrics bias.

The geolocation degrade flag indicates that the geolocation was influenced by factors that might cause its deterioration (Huettermann et al., 2022; Milenković et al., 2022). For example, if the recorded waveform contains saturated intensities, the value is set to a non-zero indicator. There are specific values for each degraded condition. Single digits represent a deteriorated trajectory, whereas tens digits represent a deteriorated attitude. An example of the resulting influence on geolocation error was given in (Lang et al., 2021), who were working with the Version 1 of GEDI data products. For these data the reported systematic geolocation error is up to 19.7±10.7 m (mean±1σ), but in the case of orbits with degraded geolocation flags offsets may reach up to 60 m.

The solar elevation parameter represents the solar altitude angle in degrees. Sub-zero values indicate nigh-time data. Based on a comprehensive assessment of the terrain and canopy height accuracy of ICESat-2 data products (Neuenschwander et al., 2020 in Chen et al., 2021) the RMSE value for terrain height and canopy height was lowest when only using night-time data. The solar noise causes large number of outliers, that can be eliminated by only using night-time data, that are less impacted by background solar noise (Liu et al., 2021). The terrain accuracy of the GEDI data is not as affected by solar noise, but for canopy height retrieval, the use of only night-time data is recommended (Liu et al., 2021; Potapov et al., 2021; Adam et al., 2020; Duncanson et al., 2020).

2.4 Geolocation

One of the most difficult aspects of using a space-based LiDAR is processing its waveforms. To calculate canopy height and other structure variables like canopy cover and vertical plant area profiles, features like the location of the ground and the top of the canopy are crucial for accurate calculations (Dubayah et al., 2020).

The first releases of the data products were expected not to fulfil the geolocation expectations (Beck et al. 2021). The mean 1-sigma horizontal geolocation error of the Version 1 of GEDI is 23.8m, while the horizontal geolocation error for each 25 m footprint center was originally required to be within 10 m (1σ), supposing normally distributed geolocation errors (Roy et al. 2021). The ranging accuracy is reported to be ~56 cm (Adam et el. 2020).

The horizontal geolocation accuracy In the Version 2 data products has been significantly improved. The mean 1-sigma horizontal geolocation error is reported to be 10.2m (Blair et al., 2021). Several methods were used to evaluate the range/elevation performance, such a direct altimetry to known surfaces, comparisons to independent data, and crossover analysis. The ranging precision is stated to be ~3 cm.

2.4.1 General GEDI geolocation algorithm overview

The geolocation for any point within the waveform is computed using the ranges to the first and last bins of the received waveform. The geolocation for any point inside the waveform may be estimated using simple linear interpolation between the two ranging points. The geolocation to the first and last bin range locations is included in the L1A and L1B data products. (Beck et al 2021). The intention for the GEDI mission geolocation is to be in the same frame as the one of the ICESat-2 and to use consistent geophysical corrections in order to facilitate the direct comparison and use of these laser altimeter mission data (Beck et al 2021).

2.5 Theory of Waveform Matching

The match between waveforms simulated from a discrete-return point cloud and real waveforms recorded by a full-waveform sensor, so called “waveform matching”, is commonly recognized as an efficient approach to improve the waveforms' geolocation (Ni et al., 2021, Tan et al., 2020, Hancock et al., 2012). The approach was performed on GLAS

(Harding et al., 2005 in Tan et al., 2020) where they used correlation between simulated and shifted waveforms with normalized total amplitudes, and on GEDI (Bauer et al., 2021; Lang et al., 2021; A. Liu et al., 2021; Tan et al., 2020). The efficiency is highlighted multiple times (Hancock et al., 2012). Hancock (2019) simulator code, which was designed for pre-launch validations, is frequently used for waveform matching of GEDI waveforms (Bauer et al., 2021; Lang et al., 2021; Liu et al., 2021; Tan et al., 2020). Comparing the simulated waveform to a real GEDI waveform to find the real offset is supposed to reduce geolocation errors to a few meters on average (Lang et al., 2021; Ni et al., 2021). However, the accuracies of GEDI waveforms might vary from the accuracies of the ALS-simulated GEDI data due to multiple factors, such as different geolocation errors or atmospheric conditions during the acquisition of the data (Rishmawi et al., 2021).

Usually the GEDI-like waveforms are simulated with a range of potential horizontal geolocation offsets and the location of the simulated waveform with the greatest correlation (Liu et al., 2021; Lang et al., 2021) to the real waveform or with the best shape match (Bauer et al., 2021; Ni et al., 2021) is treated as a real offset. This method, referred to as “moving windows” (Fig. 3) is based on prior research of geolocation errors correction for GLAS (Neuenschwander et al. 2008 in Liu et al 2021). Bruening et al. (2021) define waveform matching as quantifying the similarities in shapes between two waveforms, hence computing the relative overlapping area between two waveforms. The potential horizontal offset can be either specified based on randomly generated position errors using for example a Monte Carlo simulation (Roy et al., 2021), or within a specified range with several-meter steps (A. Liu et al., 2021; Lang et al., 2021; Ni et al., 2021).

The study is not giving a clear answer on the efficiency of the Monte Carlo simulation for this application; however, the approach has been previously used in relation to ATLAS, with the methodology being utilised for a photon counting lidar.

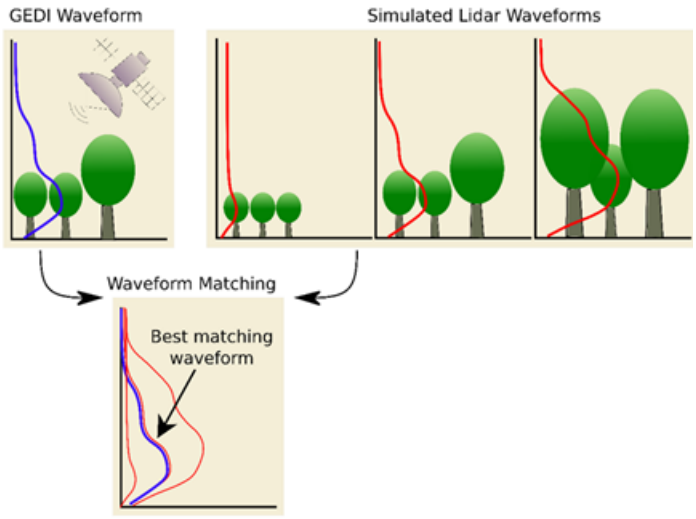

Fig. 3: The workflow of waveform matching

Source: Bauer et al. (2021)

According to Roy et al. (2021), majority of the shifted footprints fell closely around the GEDI footprint location, but a minority of footprints had their centres further than reported maximum error. In Ni et al. (2021), certain shifts were up to the size of GEDI footprint, which also exceeds the pre-set expectations on the geolocation accuracy. However, there wasn't observed a good shape match between the GEDI and simulated waveforms.

2.6 Topography and canopy cover influence on the return waveform

In the separation of canopy and ground signals there are lots of impacts that are either known (e.g., pulse shape, digitizer noise) and unknown (e.g., slope, multiple scattering). The ground return, for example, can be a weak signal that may be poorly detected against a strong background noise such as dense canopy cover (Dubayah et al., 2020). Hancock et al. (2012) describe that for distinguishing canopy and ground return in a spaceborne large footprint LiDAR, there needs to be a feature present in the return waveform which allows the separation of canopy and ground return, preferably a gap, as shown in Fig. 4 (top). If the waveform is decomposed into Gaussians, and either the last one or the mean of the last two is used to calculate the ground elevation, as done in multiple previous studies (Chen, 2010; Brenner et al. 2003; Hofton et al. 2000 in Hancock et al. 2012), a clear inflection point between the decomposed Gaussian functions is necessary. Multiple phenomena happen to hinder the legibility of the waveform return, namely topography (Fig. 4 middle) and low canopy (Fig. 4 bottom). Slope is well recognized to pose issues with LiDAR ground elevation

readings, especially for slopes steeper than 15–20° (Guerra-Hernández & Pascual, 2021; A. Liu et al., 2021; Silva et al., 2021; Spracklen & Spracklen, 2021). In the case of low canopy, the supposed ground return might contain low canopy return signal, leading to inaccurate ground measurements. Canopies that are taller than 20 m or have higher density than 90% also have a reported influence on accuracy deterioration (A. Liu et al., 2021, Spracklen & Spracklen, 2021).

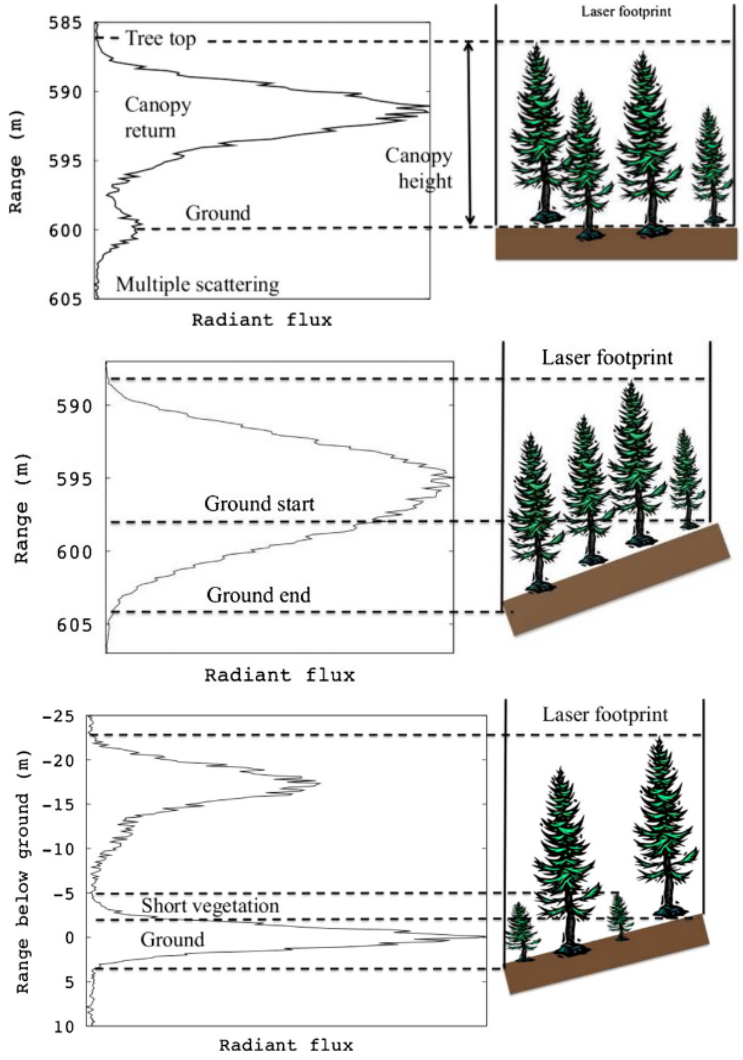

Fig. 4: Illustration of a 30m footprint waveform return in different topographic and canopy related scenarios. A simple waveform with a clear gap between the ground return and the canopy return (top). The less legible waveform impacted by a 30° slope in the footprint (middle). A waveform with a supposedly clear gap between the canopy return and the ground return, but the ground return being influenced by low canopy return (bottom). The features in the waveform are identified using the Gaussian decomposition method.

Source: (Hancock et al., 2012)

Shadows can also introduce a subtle error, as the shadow caused by branches and tree crowns reduce the intensity of the ground return, as illustrated on Fig. 5.

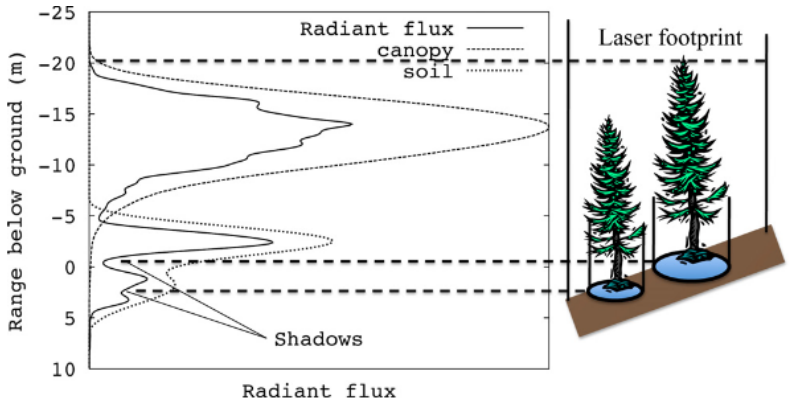

Fig. 5: The noise introduced by shadow to a waveform return.

Source: (Hancock et al., 2012)

2.7 GEDI Simulator

The GEDI Simulator was developed by Hancock et al., (2019) for the pre-launch validation and calibration. Airborne lidar returns are grouped vertically and weighted using a Gaussian distribution depending on their position in relation to the centroid of the footprint, with a Gaussian width equalling sigma of 5.5 m to match the width of a GEDI footprint. White Gaussian noise is added as suggested in (Hancock et al., 2011). The waveforms (Fig. 6) are imposed with the GEDI pulse shape (full width at half maximum, FWHM = 15.6 ns). Any ALS data may be used to generate GEDI-like signals, with simulations having the same characteristics and expected measurement errors as GEDI (Duncanson et al., 2020). In particular for high-density LiDAR point clouds, the simulator maintains consistency in point cloud flight characteristics (Leite et al., 2022). Hancock et al. (2019) provide detailed description of the GEDI simulator's development and validation.

The waveform simulator's vegetation metrics were tested against LVIS data and found to have mean bias < 0.22 and root mean square error (RMSE) < 5.7 m (Hancock et al., 2019 in Duncanson et al., 2022). The simulated data were found to be highly comparable with real

large-footprint data (Boucher et al., 2020). The GEDI Simulator is capable of simulating the waveforms with different noise conditions with different ground finding algorithms.

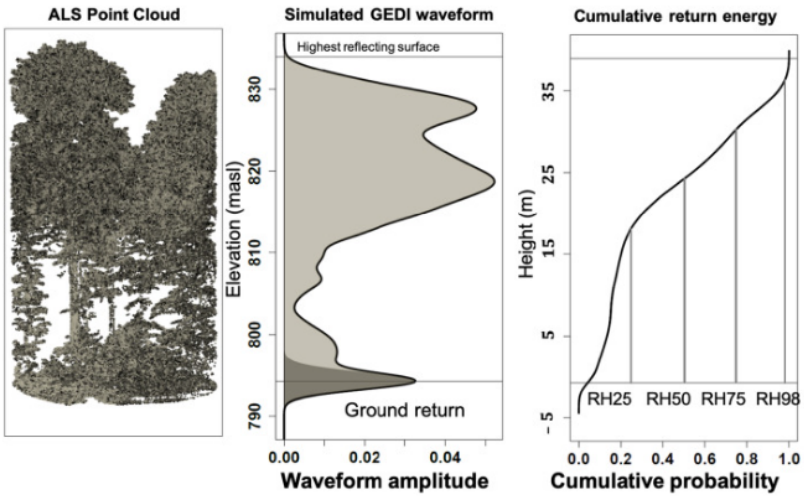

Fig. 6: Simulated waveform from an ALS point cloud with illustrated relative return energy
Source: (Duncanson et al., 2022)

Ground detection algorithm

The accuracy of used ground detection algorithm is acknowledged for having an erroneous influence on not only elevation readings, but also calculated RH metrics in full-waveform laser scanning (Leite et al., 2022; Guerra-Hernández & Pascual, 2021). It determines the ground classification, so along sensor properties such as beam sensitivity, or environmental characteristics such as atmospheric conditions and terrain and canopy related characteristics, it is an important factor.

The quality of the metrics relies on the ability to detect the ground signal which is expected to vary based on factors such as canopy cover, GEDI beam energy, weather conditions and topography. However, apart from the environmental characteristics and sensor properties, what determines the ground classification is the implemented algorithm. Hancock et al. (2019) describes in detail the three ground finding algorithms that the GEDI Simulator offers. A Gaussian fitting, that is based on decomposing the signal into separate Gaussian functions and identifying the ground return. The Lowest maximum method considers the lowest peak to be the ground signal. The Inflection point algorithm computes the second derivative of the waveform and computes the ground elevation.

Simulating the GEDI noise

The white Gaussian noise that is being added in the simulator is supposed to achieve the same signal-to-noise ratio (SNR) anticipated by pre-launch study of GEDI's performance for mean atmospheric transmission, solar noise depending on the time of acquisition (day/night), and expected detector response (Dubayah et al., 2020; Duncanson et al., 2020; Hancock et al., 2019). Boucher et al. (2020) suggested settings for simulating the noise in case of specifying the time of acquisition or the sensor parameters such as used beams.

Deliverables from simulation, GEDI and ALS

Leite et al. (2022) highlighted vegetation metrics, that are possible to be computed from simulated waveforms and that are also included in the GEDI L2A and L2B products. Selected metrics are elevation height, RH, and CCF. Those metrics are also possible to be derived directly from ALS point clouds.

Elevation height within a footprint is widely calculated as the mean height of ground-classified returns (Moudrý et al., 2022). The ground elevation provided in GEDI data products also represents the mean elevation within the footprint (B. Blair et al., 2021). In Adam et al. (2020), though, who compared the derived elevation height from GEDI and ALS, was recommended the use of median instead of mean to reduce the influence of potential outliers.

The RH is defined as the percentage of energy returned below certain threshold. All returns are being used for this calculation, and a thresholder percentile is used for deriving the RH metric from an ALS point cloud (Guerra-Hernández & Pascual, 2021).

The CCF is defined as the percent of the ground covered by the vertical projection of canopy material (Dorado-Roda et al., 2021). The methods for calculating CCF are described and compared in detail in (Ma et al., 2017). Methods based on LiDAR point cloud employ either only first returns or all returns. In both cases, the CCF is calculated as a ratio of LiDAR returns that represent canopy. The case of first returns assumes that later returns provide no more useful information on canopy, however, this method was found less accurate than calculations based on all returns. Another LiDAR-based method that is quite widely used is Canopy Height Model (CHM) where the cover is calculated as the percentage of pixels with a CHM value exceeding a set threshold for tree height, that has been set as 2 m in Ma et al. (2017) and also in Moudrý et al. (2022), who used this approach in relation with space-borne ICESat-2 in a temperate forest in Czechia. Other approaches employ landcover classification derived from optical data to support the CCF calculations.

2.8 Monitoring of forest change

Because of its capacity to penetrate gaps in the top canopy and capture structural change in the lower canopy, GEDI is particularly well adapted to monitoring wood infestation effects (Boucher et al., 2020). However, due to the nuances in the orbit of ISS, it is not likely that one track will be repeatedly scanned at the same geolocations. The comparison between GEDI footprints and coincident field measures such as field plots might be inaccurate due to the geolocation accuracy of GEDI.

It was proposed in Huettermann et al., (2022) to use the GEDI Simulator as a standardisation tool for LiDAR data sources acquired by different sensors. That would allow for analysis of forest structure based on multimodal, multitemporal dataset consisting of ALS and GEDI data. If sufficient ALS data are available for the area of study (recommended density over 2 pulses per square meter), a multitemporal analysis, where both GEDI waveforms and GEDI-like simulated waveforms are treated as the same data source, is possible (Boucher et al., 2020). Authors Huettermann et al. (2022) suggest the use of the Simulator even as a tool for standardising ALS-only datasets, as it minimises the influence of flight altitude and point density in the output simulated waveform.

3 Study area and data

3.1 Study area

The study site belongs to the national network CzeCOS (Czech Carbon Observation System) and the international network ICOS (Integrated Carbon Observation System). It is a national complement of significant infrastructures with-in ESFRI (European Strategy Forum on Research Infrastructures). The site is also a long-term research station LTER (Long Term Ecological Research) (Chawla et al., 2018). The site belongs to a network of ecosystem stations of the Global Change Research Institute – CzechGlobe, that focus on environmental research, particularly the challenge of global climate change. The ALS monitoring was done by CzechGlobe as a part of their research on the forest dynamics.

The study site is situated in Czech-Moravian highlands near a town Rájec-Jestřebí (Fig. 8). The coordinates of its centroid are 49°26'37"N, 16°41'47"E. Significant part of the area is covered with mature 121-year-old forest monoculture of Norway spruce (*Picea abies)* (Chawla et al., 2018). The trees are up to 30m of height and the forest contains occasional low canopy level, as seen on a cross section from an ALS point cloud (Fig.7).

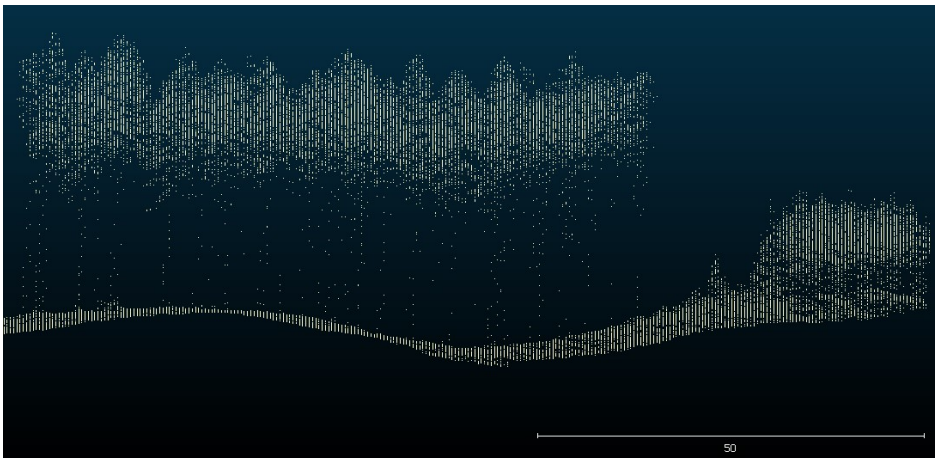

Fig. 7: A vertical profile of the study area

Source: Author

The study area, that is approximately 10,5 km^2 large, is of a very dynamical character. Fig. 8 shows the decrease of forested area in between years 2018-2020. The decrease of forest happens mostly due to logging.

Fig. 8: Comparison of aerial photos of the study area between the years 2018 (left) and 2020 (right)

The subdominant landcover categories that are slowly expanding are shrub, grassland, and agriculture. The topography is predominantly flat with mild slopes and occasional steep slopes ranging up to 80 degrees, with altitudes ranging between approximately 550 - 700 meters above sea level. The average annual temperature is 7 degrees Celsius, with mean summer maximums of 28 degrees Celsius and mean winter minimums of -3 degrees Celsius. Winters are characterised by prevailing snow cover. The average yearly rainfall is 600 mm, with highs during the summer period from June to August (Chawla et al., 2018).

3.2 GEDI and Airborne Laser Scanning data

In this study were analysed the available footprint-level GEDI data products L1B, L2A, L2B datasets. The L4A data product are as well available at footprint level, but an error has been noticed in the current Version 2 data with granule production version 01, resulting in inaccurate AGBD estimations for shots with algorithm setting group 10 (Dubayah et al., 2022). The data were downloaded over the area of interest for the entire up-to-date mission sensing period, from January 2019 to April 2022. Due to a small size of the study area and irregular ISS orbit, which prevents repeated data acquisition over the same tracks, only four acquisition dates were available for the area of interest, containing 1353 shots.

ALS data from a public research institution Global Change Research Institute of the Czech Academy of Sciences (CzechGlobe) were used as ancillary. The ALS data are stored in las format. The data were collected with a RIEGL LMS-Q780 scanner in three different acquisition dates within one year – April, August and September 2020. Mean point densities

are ranging around 9.8 pulses/m2. The suggested lower threshold for unbiased simulation of GEDI relative height measures is 2 pulses/m2 (Hancock et al., 2019).

The point cloud is horizontally referenced to the European Terrestrial Reference System 1989 (ETRS89) datum and projected to the Universal Transverse Mercator Zone 33N (UTM33) mapping frame. The vertical datum of ALS is relative to geoid ETRS89, while GEDI is referenced to geoid WGS84, however, the resulting horizontal difference between those two systems is so negligible, that it was deemed irrelevant to this study.

Detector type	Discrete return
Laser wavelength	near infrared
Laser repetition rate	Programmable, up to 400kHz
Laser beam divergence	≤ 0.25 mrad
Wavelength	Near infrared
Minimum range	50m
Accuracy	20mm
Precision	20mm
Vertical datum	ETRS89

Tab. 2: Riegl LMS Q780 airborne lidar instrument and product specifications

Source: (RIEGL Laser Measurement Systems GmbH, 2015)

4 Methodology

4.1 GEDI data pre-processing

4.1.1 GEDI Subsetter

The GEDI_Subsetter.py script, built in Python version 3.7., converts GEDI data products (HDF5, .h5) into GeoJSON files, to allow for its visualization. For the execution of this script, a desired region of interest (ROI) is specified either by bounding box coordinates or loaded from an external file in a GeoJSON or shapefile format. As default, output includes shot from all beams, however, if needed, there is an optional argument for subsetting by specific beams, e.g., only power beams (BEAM0101, BEAM0110, BEAM1000, BEAM1011), as shown in Fig. 9 on a script that was used in this study. All inputs are georeferenced in the WGS84 (EPSG:4326) coordinate system. According to the product of interest, each output includes default datasets; the full list is in the LP DAAC repository (LP DAAC, 2022). Due to the size of the output files, additional datasets, such as beam azimuth or metrics related to geophysical correction, shall be added manually through –sds parameter during execution of the script. All datasets are stored as columns in output GeoJSON files, where each row represents a GEDI shot.

```
conda create -n gedi -c conda-forge --yes python=3.7 h5py shapely geopandas pandas
conda activate gedi
python GEDI_Subsetter.py --dir d:\Path\ --roi study_area.geojson --beams BEAM0101,BEAM0110,BEAM1000,BEAM1011
```

Fig. 9: A script used for subsetting GEDI data and filtering power beams

Source: LP DAAC (2022)

Once the data is in a spatial data format that allows visualisation of its location, it can be assessed in Geographic Information System or a Remote Sensing software. Fig. 10 shows the location of L1B and L2A data centroids acquired on the same date and track.

Fig. 10: The spatial comparison of L1B (blue) and L2A (red) footprint centroids from the same track.

Source: Author

4.1.2 Shots filtering

First filtering that was done was temporal, aimed to keep only the shots that fell with their acquisition dates into the timeline between the acquisitions of the available ALS data. The ALS data were collected on 7th April, 12th August and 15th September of 2020. Two GEDI acquisitions were available within this timeline, one from 13th August and second one from 17th August 2020.

Based on suggestions and recommendations for quality filtering of Version 2 GEDI data products, that were mentioned in previous studies, a following filtering was applied to keep only high-quality shots relevant for selected study area:

1. Only shots acquired by power beams during night-time (*solar_elevation* < 0) were used for this study. In the case of vertically heterogenous forest, the shadows caused by branches and tree crowns are prone to influence the legibility of large-footprint waveform (Liu et al., 2021; Potapov et al., 2021; Adam et al., 2020; Duncanson et al., 2020; Hancock et al., 2012).
2. The shots had to fulfil the condition of *quality flag* being set to 1 and *degrade flag* being set to 0.

After quality filtering, only 67 shots were left for use in the remainder of the study (Fig. 11).

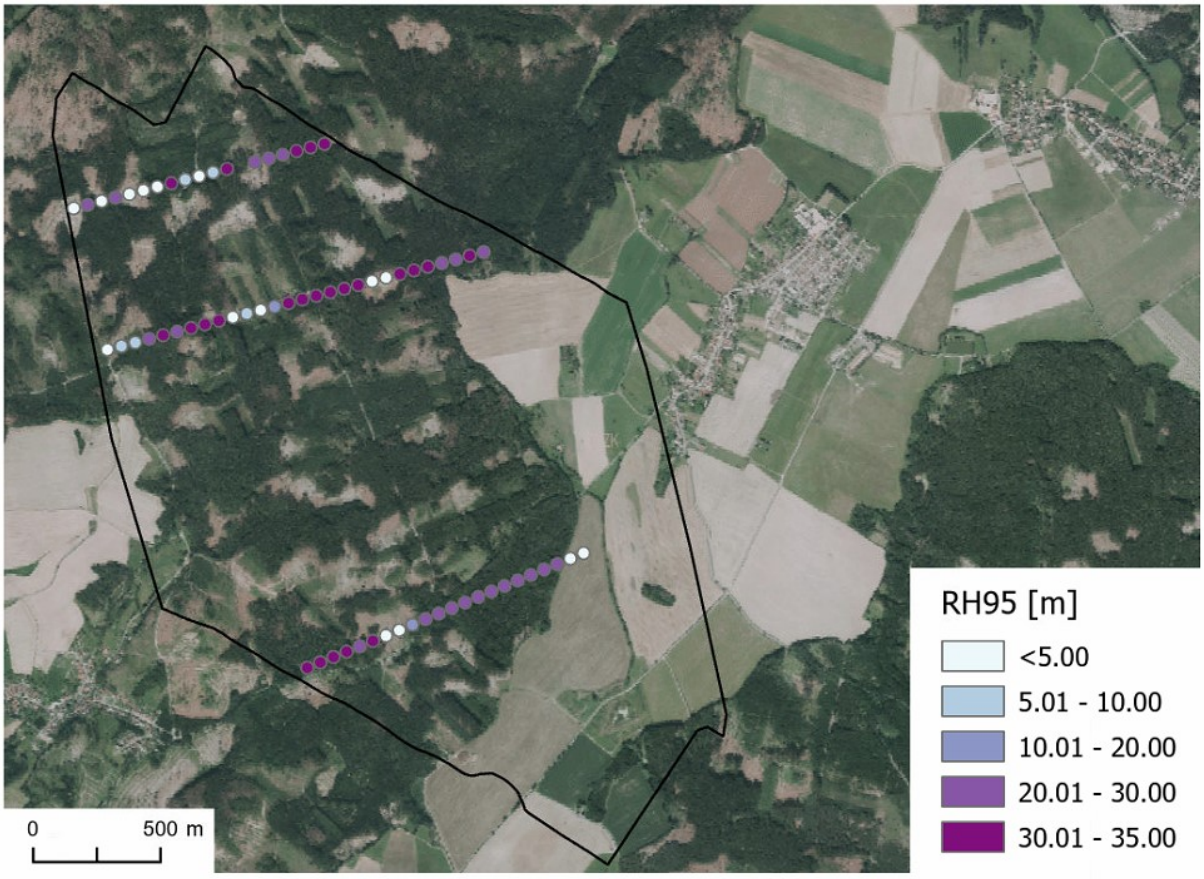

Fig 11: The remaining shots in the study with visualised RH95 metric available in the L2A data.

Source: Author

4.1.3 Retrieval of the GEDI waveform amplitude

A package for handling GEDI data in R and C programming languages was developed by Silva (et al., 2020). The rGEDI package (version 0.1.11) that was built on the version 3.6 of R allows processing and visualising GEDI data and simulating GEDI-like waveforms. With the use of functions *readLevel1B* and *getLevel1BWF* data tables with columns containing amplitude values and corresponding elevation values were derived for each shot. The data tables were exported as *.csv* files. Fig. 12 illustrates a shape of a return waveform with added RH information from L2A data for the specific footprint using the function *gedilevel2a.*

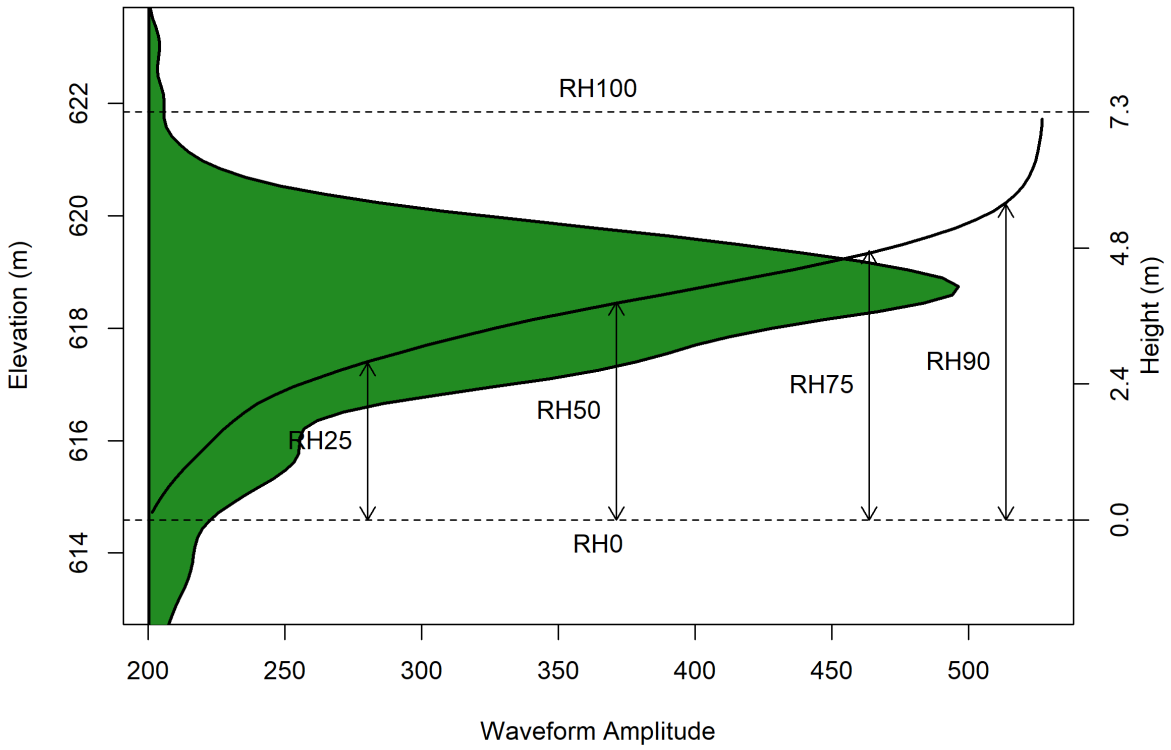

Fig. 12: Illustration of a return signal delivered from L1B data with added RH information from L2A data

4.2 Waveform matching

4.2.1 Simulating GEDI-like waveforms

The package rGEDI contains the function *GediWFSimulator*, that takes a *.las* point cloud as an input, and produces a *.h5* waveform file. The input point cloud has to be clipped to match the diameter of the GEDI fooprint, that is a circle with a diameter of 25 metres. Therefore, a process for extracting footprint-sized point cloud is necessary.

Footprint-sized point clouds

The QGIS software offers a *LasTools* plugin, which allows to combine spatial functions and LasTools functions for building a model that can be used as a batch processing tool.

The GeoJSON file derived by the GEDI Subsetter algorithm containing L2A shots that remained after quality filtering was loaded into QGIS. The function *Split Vector* was used to create a folder with a separate GeoJSON layer for each shot with the name containing the shot number. The following model (Fig.13) built in QGIS takes a GeoJSON shot and outputs a dataset of footprint-sized point clouds around the geolocation of the shot (Fig.14).

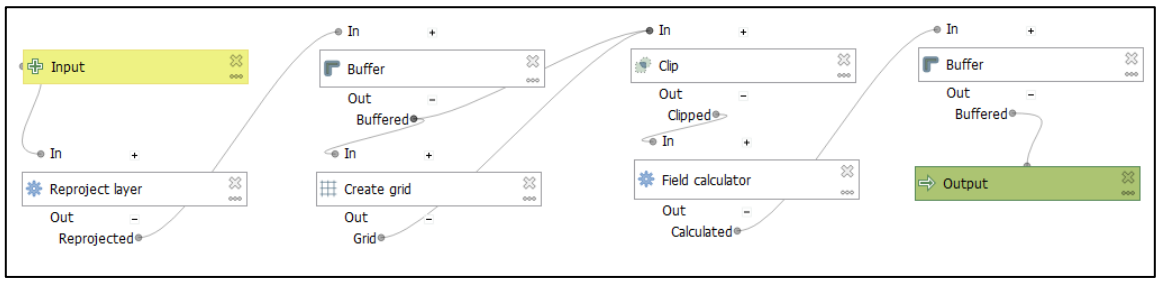

Fig. 13: The process of deriving a footprint-sized point cloud in QGIS

The shot numbers contained in GeoJSON layer derived from GEDI Subsetter are stored in a string format. In the description of rGEDI package a set of functions to convert the field "shot number" from string to character format is offered and suggested. The entire layer can be then exported as a shapefile. The author of this thesis recommends using that approach as it is simple and straightforward; however, the rGEDI package was not yet available when the work on this research begun, so the following process was developed prior to rGEDI's release.

The input GeoJSON point layer represents the centroid of a L2A footprint with name containing the shot number. The GeoJSON waveforms generated by the GEDI Subsetter script, alike GEDI h5 waveforms, are stored with assigned WGS84 coordination system. To generate a buffer using meter values, the data are projected into UTM, specifically ETRS89 UTM Zone 33, which is the coordinate system assigned to used ALS dataset. The spatial difference between the reference system is negligible regarding the scale of the assessed data and purposes of this study (ČÚZK, 2022). Following few steps are specific to a waveform matching approach used in this study. A buffer with diameter 50m defines the area where the real centroid of the waveform is being looked for. The diameter value was set as the GEDI footprint diameter plus the reported geolocation error. Within the extent of the footprint a grid of points with 5 x 5 m spacing is generated. The reason for using 5 x 5m spacing, and not 1 x 1 m as for example in (Liu et al., 2021), where they searched within a 40m diameter, is the size of generated data, although it might influence the accuracy. Around each point of the new grid, a footprint with 25m diameter is calculated, to simulate the GEDI footprints (Fig. 14). Finally, the *lasclip* function of *LasTools* is used to cut a footprint area out of the ALS point cloud. The function *lassclip* allows input clipping layer to be only in a few formats, shapefile being one of them, GeoJSON not. Therefore, in the step of creating a buffer around each point of the grid, the buffer layer is stored into a separate folder as a shapefile. Since this step is in

the model executed sooner than the next function, e.g. *lasclip*, the process executes without difficulties. Using and additional command *-split* in the *lasclip* console, each clipped ALS footprint is saved into a separate *.las* file.

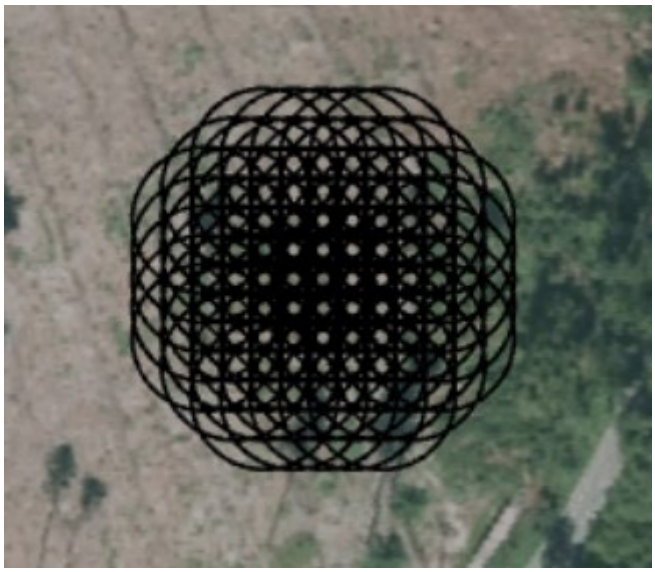

Fig. 14: The possible locations of a footprint within a selected range.

Source: Author

Simulation

The function *GediWFSimulator* in the rGEDI package was used to simulate GEDI-like waveforms from the *.las* footprints (Fig. 15).

```
dir<-"c:/path/to/folder/"
outdir<-dir
files<-list.files(path=dir, pattern = .las, full.names = TRUE)
for (file in files){
     las_1b<-readLAS(file)
     xcenter_1b = mean(las_1b@bbox[1,])
     ycenter_1b = mean(las_1b@bbox[2,])
        wf_1b<-gediWFSimulator(input=file,output=paste0(file,".h5"),coords = c(xcenter_1b, ycenter_1b))
        }
```

Fig. 15: Script for a loop simulating the GEDI-like waveform from a point cloud

Source: Author

The output *.h5* file contains the simulated waveform interpolated into 1022 regularly distributed points. The file contains 24 datasets, and the structure of the file can be examined for example through an open-source tool HDF5 Viewer. Instead of reference elevation values the file contains separate datasets for the elevation of the lowest return (Z0), ground peak

(ZG), and highest return (ZN). The rGEDI package contains functions for visualisation of the output .h5 file (Fig. 16).

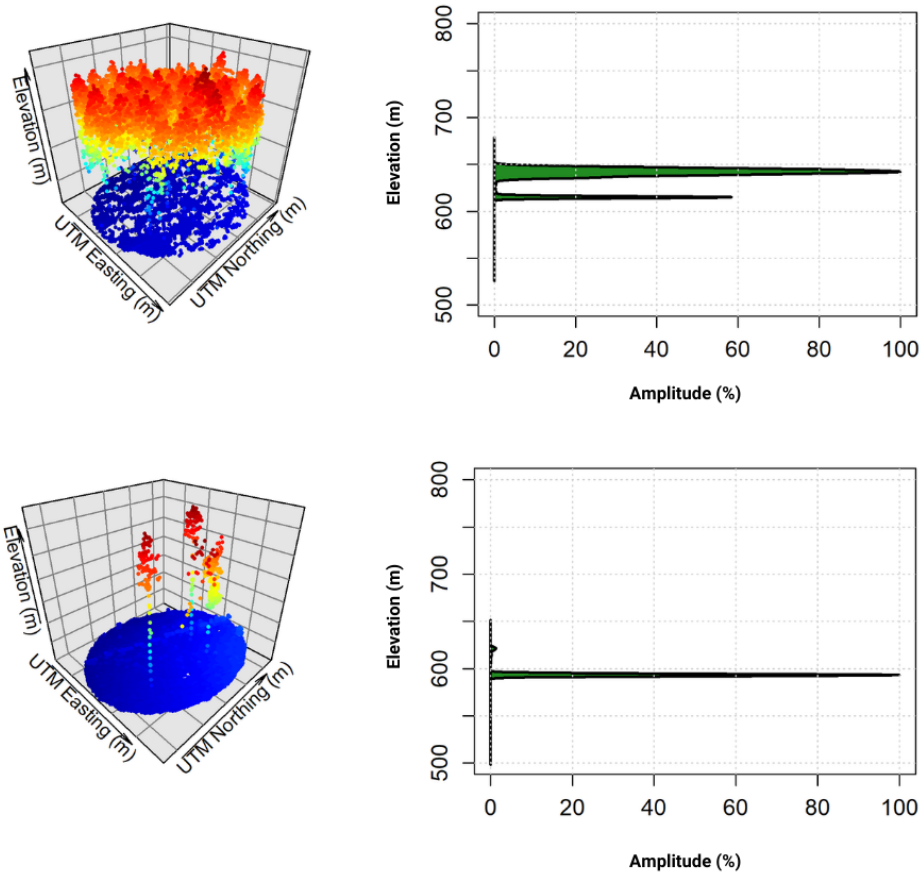

Fig. 16: Examples of simulated waveforms (right) from respective ALS point clouds (left). *Source: Author*

4.2.2 Matching based on waveforms' correlation and overlap of waveforms

For identifying the best match between the real and the simulated waveforms two approaches were tested out. The first one was based on the correlation of the waveform values (Liu et al., 2021, Bruening et al., 2021) and the second one on finding highest overlap of areas under the waveform curves (Lang et al., 2021). MATLAB scripts were developed for computation of both approaches Fig. 17 and Fig. 18. The initial process was identical in both cases – real GEDI waveforms and simulated waveforms were normalised by the amplitude, and the simulated waveform was interpolated to the same elevation range as GEDI waveforms, using the same interpolation method as GEDI uses, i.e. a spline function.

```
%a cycle that takes each .h5 file in the list
for i=1:length(fileList)
    sim=fileList(i).name;

        %load simulated waveforms, set specific datasets at variables
        %RX=the waveform, ZN=lowest return, Z0=highest return
        RX= h5read(sim,'/RXWAVEFRAC');
        ZN = h5read(sim,'/ZN');
        Z0 = h5read(sim,'/Z0');


        %create a new column of reference values corresponding
        %to the length of the simulated waveform
        v = [ZN:(Z0-ZN)/1022:Z0];

        %linear interpolation
        vq = interp1(v,RXf,Elev,"spline");

        %normalize interpolated amplitude values
        vqn=normalize(vq,'range');


        % set the variables
        x=Elevation;
        y=AmplitudeGEDI;
        y2=vqn;

        %computing a correlation coeficient
        CC = corrcoef(y,y2);

        %filling the correlation coeficient into the prepared matrix
        R(i)=CC(1,2);
        RC(i,1)=sim;
        RC(i,2)=CC(1,2);
        RCD = str2double(RC(:,2));

        %select rows with the lowest difference
        [XRsorted,I] = sort(RCD,"descend");
        YRsorted = RC(I);
        correlation=[XRsorted(1,1),YRsorted(1,1)];
```

Fig. 17: The script that interpolates the input simulated waveforms on an elevation axis and chooses the one that has the highest correlation coefficient with the GEDI waveform

For each shot, a loop goes though all simulated waveforms generated within the given offset from the GEDI waveform position. The correlation is determined by computing a correlation coefficient between the real waveform and each simulated waveform. A simulated footprint with highest correlation to the real waveform is chosen (see Fig. 16 and Fig. 19).

```
%define lower curve and higher curve
mx=double([y,y2]);
m=max(mx,[],2);
l=min(mx,[],2);


%integrate the area under both curves
maxcurve=trapz(m);
mincurve=trapz(l);

%calculate percentual overlap
ovl=mincurve*100/maxcurve;

%fill in the overlap into an array
overlapC(i,1)=sim;
overlapC(i,2)=ovl;
```

Fig. 18: Script for computing the highest overlapping area under the curves between the real GEDI and simulated waveforms

The area under the curve was computed using an integral. The overlap was defined as the ratio between the maximum y-value at a given point on an x-axis to the minimum y-value at the respective point on an x-axis.

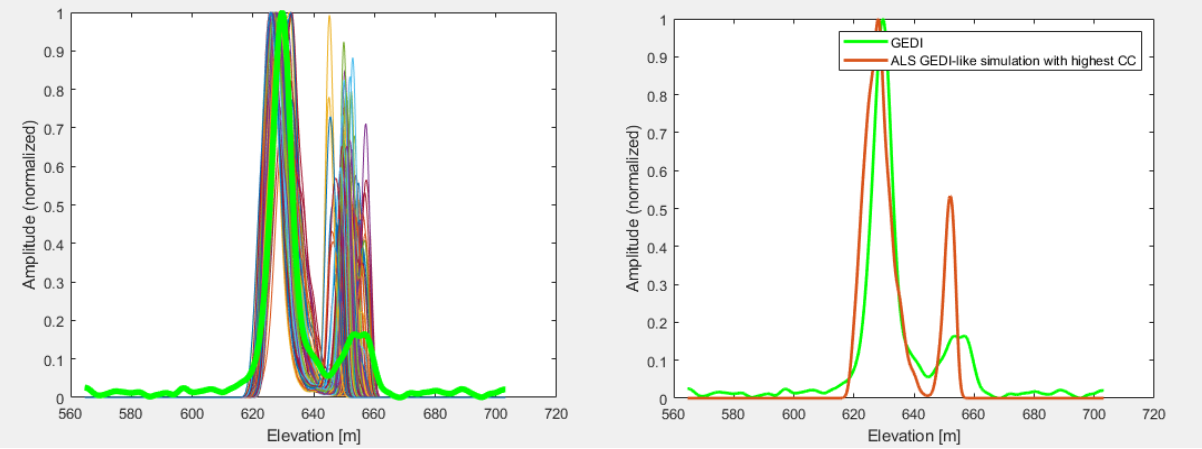

Fig. 19: Finding the waveform with the highest correlation. On the left are plotted waveforms from all possible footprints against the real GEDI waveform (green) and on the right is the real GEDI waveform plotted against a waveform with the highest correlation.

4.3 Metrics derived from an ALS point cloud

The ALS data was used for two purposes: i) simulating GEDI observations with the GEDI simulator, ii) computing vertical vegetation metrics to validate the values of the GEDI

products and the results derived from the GEDI Simulator. Tab. 3 gives the summary of the vegetation metrics that were used for the validation including their origin or the method of calculation.

derivates from ALS point cloud		**corresponding GEDI metrics**			
		GEDI simulator		**GEDI data products**	
variable	**description**	**variable**	**ground finding algorithm**	**variable**	**GEDI product**
ground elevation	ground elevation [m] relative to reference ellipsoid	*gHeight*	Gaussian fitting	*elev_lowestmode*	L2A
RH95	height in [m] below which was reflected 95% of the returned energy	*rhGauss 95*		*rh95*	L2A
CCF	the percent of the ground [m2/m2] covered by the vertical projection of canopy material	*cover*		*cover*	L2B

Tab. 3: The vegetation metrics used in this study

Gaussian fitting was used as the ground finding algorithm (Boucher et al., 2020) for all three metrics derived from the GEDI simulator

The ALS point cloud was pre-processed using *LasTools*. The point cloud was classified to ground and non-ground points with the *lasground* function and the above-ground height of ALS points was computed using *lasheight*. Due to the frequent occurrence of low vegetation, the last discrete LiDAR return was often reflected from the low canopy, not reaching the ground. Therefore, based on empirical assessment of individual footprints and (Liu et al. 2017), a threshold was set to 3 m of height, and all points that fell into the lowest 3 metres of the point cloud were classified as ground. The point cloud was transformed into an ASCII file with *las2txt* and the following metrics were computed in MS Excel.

The ground elevation metric was calculated as a median value of the points that are classified as ground, so an occasional occurrence of higher bushes in the same category did not considerably influence the calculated metrics. The ground elevation provided in GEDI data products represents the mean ground elevation within the footprint; but in Adam et al. (2020) it was suggested to use median instead to reduce the influence of potential outliers.

The RH95 metric has been calculated as the 95th percentile of all points in the point cloud.

The CCF is related to the vertical projection of the canopy on the ground and describes the percentage of covered ground. In other words, the metric reflects the compactness or gaps of the forest cover (Ma et al., 2017). Hence, the point cloud was rasterised with a cell size equal to the point spacing in the point cloud, that is 0.308 m. The pixels in the raster were reclassified into ground and nonground according to assigned height values. The threshold value for classifying ground pixels was set to be 2 metres above the median ground elevation of the point cloud, as in (Moudrý et al., 2022; Ma et al., 2017). This step was necessary due to slope – if the ground was completely flat, it would be possible to directly divide all the points above median ground elevation as non-ground, however, with slope, a threshold was necessary, according to empirical findings. The proportion of pixels classified as non-ground to all pixels gave the percentage of ground vertically covered by vertical canopy material.

4.4 Metrics derived from simulated waveforms

The metrics in Tab. 3 were derived from the simulated waveforms using the rGEDI function *gediWFMetrics.* A parameter adding noise to the simulated waveforms was included in the options of *gediWFMetrics* (Fig. 20). The white Gaussian noise that is being added is supposed to achieve the same signal-to-noise ratio (SNR) anticipated by pre-launch study of GEDI's performance. It is based on anticipated mean atmospheric transmission, solar noise depending on the time of acquisition (day/night) and expected detector response. Since only shots acquired by power beams at night are used in this study, the link margin was set to 4.956, as recommended for the use of power beams and night-time data (Boucher et al., 2020). The beam sensitivity was set to 0.95, which was the lowest beam sensitivity in the dataset of footprints. The beam width was left at 0.5 and all other parameters were left at default.

```
wf_1b_n<-gediWFMetrics(input=wf_1b,
                       outRoot=file.path(getwd(), "wf_1b_n"),
                       linkNoise= c(4.956,0.95),
                       maxDN= 4096,
                       sWidth= 0.5,
                       varScale= 3)


metrics_noise<-rbind(wf_1b_n)
rownames(metrics_noise)<-c("rajec")
head(metrics_noise[,6])
head(metrics_noise[,33])
head(metrics_noise[,11])
```

Fig. 20: Script for simulating the noise and deriving selected waveform metrics from the simulated waveforms

Source: rGEDI, author

5 Results

5.1 Waveform matching

The results of waveform matching reflect the challenging character of the fragmented forest. Considering ALS the ground truth and employing the approach of highest correlation, the RMSE of the GEDI ground elevation height increased of 0.35 m (from RMSE of 2.28 m of nominal, i.e a centroid of a Version 2 GEDI footprint to RMSE 2.63 of the optimised, i.e. a centroid of a footprint with an offset defined by waveform matching, data. The RMSE of GEDI RH95 increased by 0.68 m from 8.60 m for the nominal data to 9.28 m for the optimised data. The approach of highest overlap of areas under the curves yielded even more significant rise of RMSE, the RMSE of GEDI ground elevation height rose of 0.42 m (from RMSE of 2.28 m of the nominal data to RMSE 2.70 m of the optimised data), and the RMSE of GEDI RH95 rose by 0.71 m from 8.6 m for the nominal data to 9.31 m for the optimised data. The maximum horizontal shift was 25 m for both approaches, which equals an entire size of the GEDI footprint. Mean slope in the study area is 7.37°, that means for a footprint located on a slope, where the horizontal shift was calculated in the direction of the slope, the geolocation accuracy makes for a 3.23 m of ground elevation difference. More so, the correlation coefficients values in number of footprints did not reach the threshold $c > 0.95$ set by Lang et al. (2021), neither did the overlap threshold of >70% set by Bauer et al. (2021). Former studies worked with significantly larger GEDI datasets and dropped the footprints that did not have clear enough match (Bruening et al., 2021). With 65 footprints remaining in the study, this approach was decided unfeasible. No trend of shifting towards a certain angle was observed, neither within one recorded track nor among all footprints with corresponding acquisition date. Due to many observed erroneous readings by GEDI that are caused by the character of the canopy and explained and discussed in following chapters, and no way to validate the actual position of GEDI (Ni et al., 2021), the optimalisation of the geolocation was in the end found to possibly introduce more error than solve, so the original version of GEDI Version 2 data products was used for the validations. The original data products still bring in the geolocation accuracy of 10 m, however, it is not making a difference in quantifying the accuracy of the simulator.

5.2 Ground elevation

Very strong correlations were found between the ground elevation derived from GEDI, simulator and point cloud (Tab. 5). The distribution (Fig. 21) shows minimum outliers, and there is a slight trend of overestimating the ground elevation by both GEDI and the Simulator, according to the median of ALS ground height. The intercomparison of median difference, maximum absolute difference and RMSE between all three datasets is shown in Tab. 4. All three datasets show strong positive correlations as seen in Tab. 5, the GEDI ground elevation having 0.998 correlation coefficient with the simulator derived ground elevation.

	GEDI x Simulator	GEDI x ALS	Simulator x ALS
Median difference [m]	1.09	0.12	0.8
Maximum absolute difference [m]	4.85	10.11	14.23
RMSE [m]	1.64	2.36	2.28

Tab. 4: The intercomparison of median difference, maximum absolute difference and RMSE of ground elevation among datasets

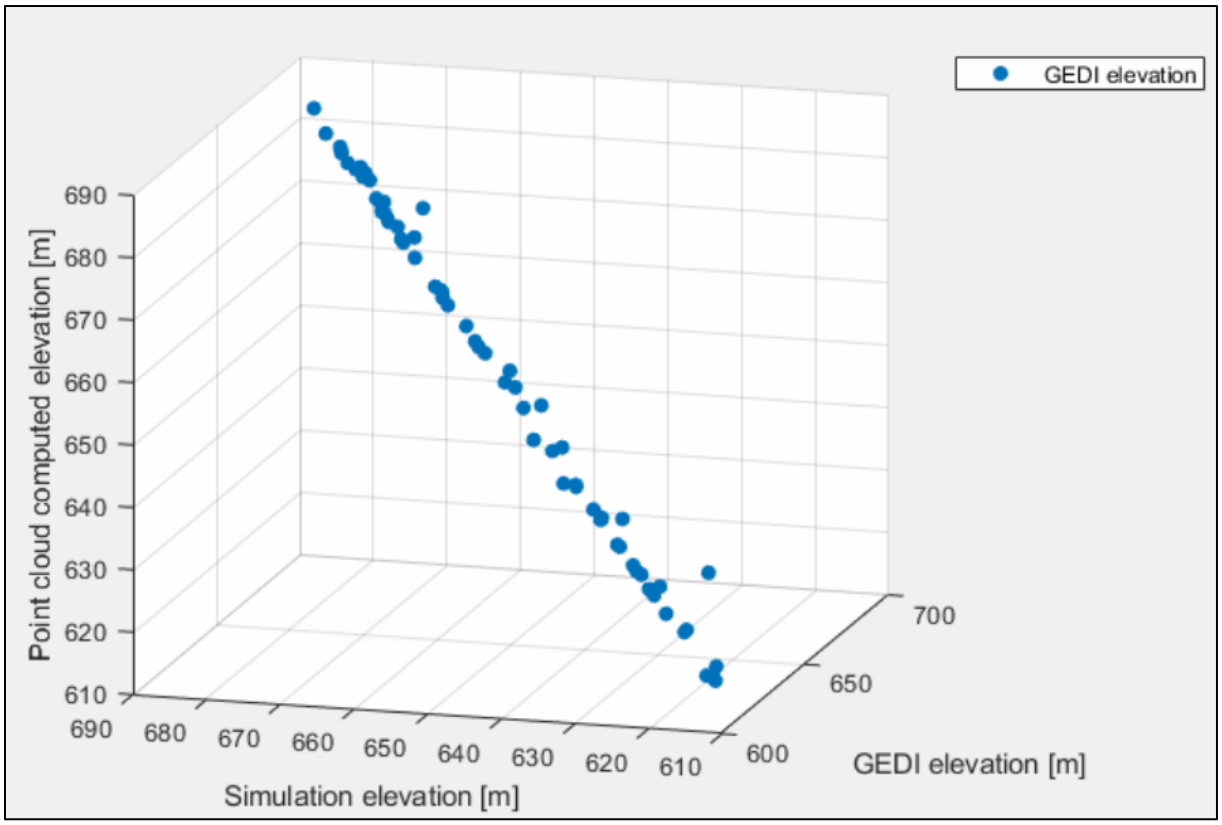

Fig. 21: A scatterplot illustrating the distribution of GEDI derived ground elevation in relation to the ground elevation estimates from ALS simulated waveforms and the median of the ALS ground height of each footprint

ground elevation derived from	GEDI	Simulation	Point cloud
GEDI	1.0000	0.9981	0.9943
Simulation	0.9981	1.0000	0.9947
Point cloud	0.9943	0.9947	1.0000

Tab. 5: Correlation coefficient of ground elevation height among datasets

There was found a correlation 0.24 between the canopy cover and ground elevation deviations, representing a weak positive relationship between the two variables (Fig. 22), indicating that dense canopy cover has only weak influence on GEDI's tendency to overestimate the terrain height. A correlation 0.31 was found between the ground elevation deviations and RH95 indicating a weak positive relationship between the overestimation of ground elevation and canopy height (Fig. 23).

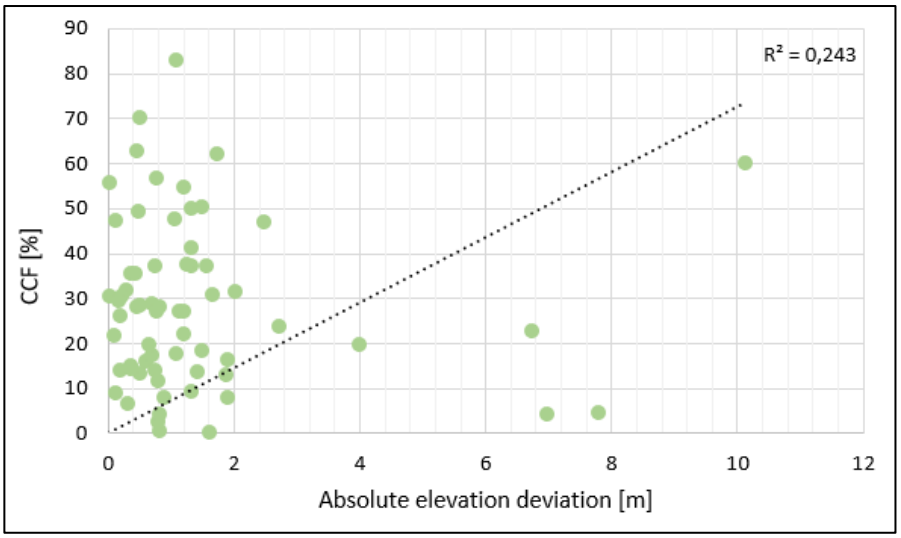

Fig. 22: A scatterplot illustrating the relationship between GEDI derived CCF and the deviation of ground elevation height between GEDI and the ALS benchmark

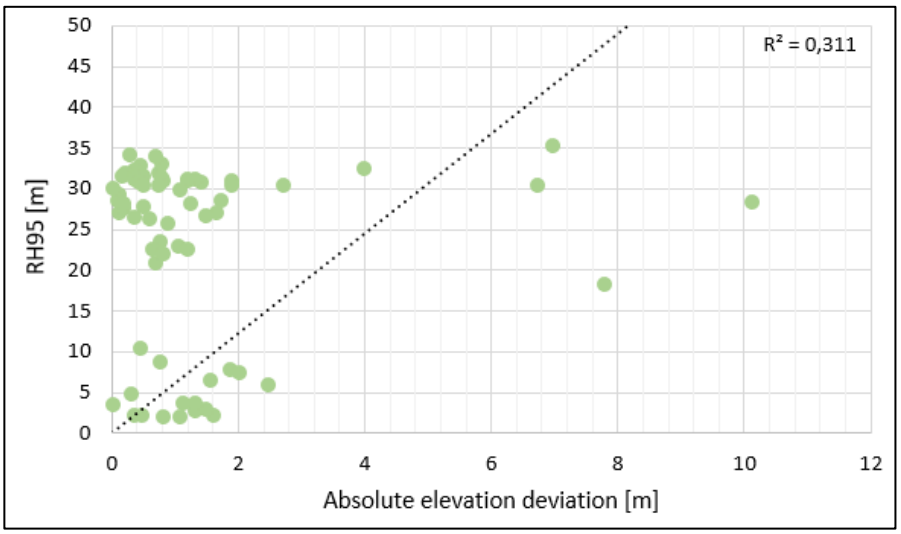

Fig. 23: A scatterplot illustrating the relationship between GEDI derived RH95 and the deviation of ground elevation height between GEDI and the ALS benchmark

5.3 Canopy height

The distribution of derived canopy height metrics (Fig. 24) shows high positive relationship between all the datasets. In Tab. 6, the median difference, maximum absolute difference and RMSE of derived RH95 are compared among the datasets. The maximum absolute height differences represent in the observed monoculture forest the height difference between forested area and non-forested area.

	GEDI x Simulator	GEDI x ALS	Simulator x ALS
Median difference [m]	0.94	1.13	0.92
Maximum absolute difference [m]	28.25	28.1	30.69
RMSE [m]	6.95	8.6	6.8

Tab. 6: The intercomparison of median difference, maximum absolute difference and RMSE of derived RH95 among datasets

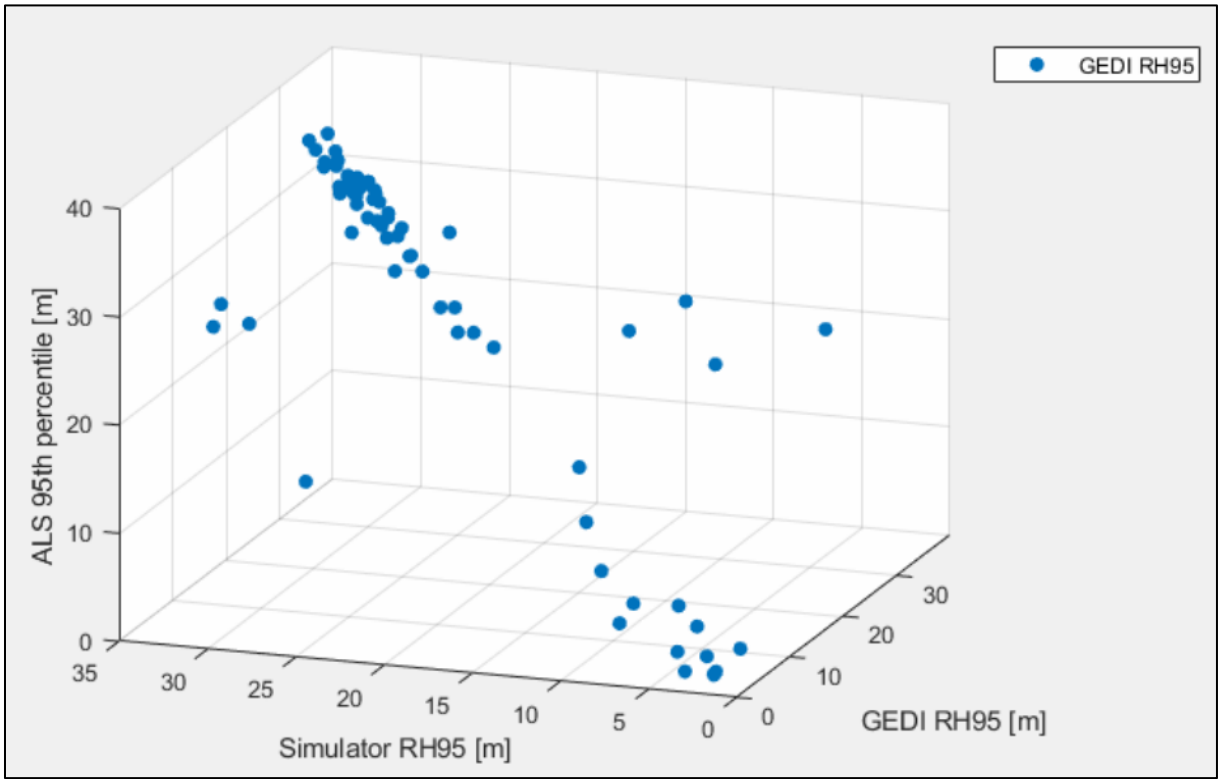

Fig. 24: A scatterplot illustrating the distribution of GEDI derived RH95 in relation to the RH95 estimates from ALS simulated waveforms and the 95th height percentile of the ALS point cloud of each footprint

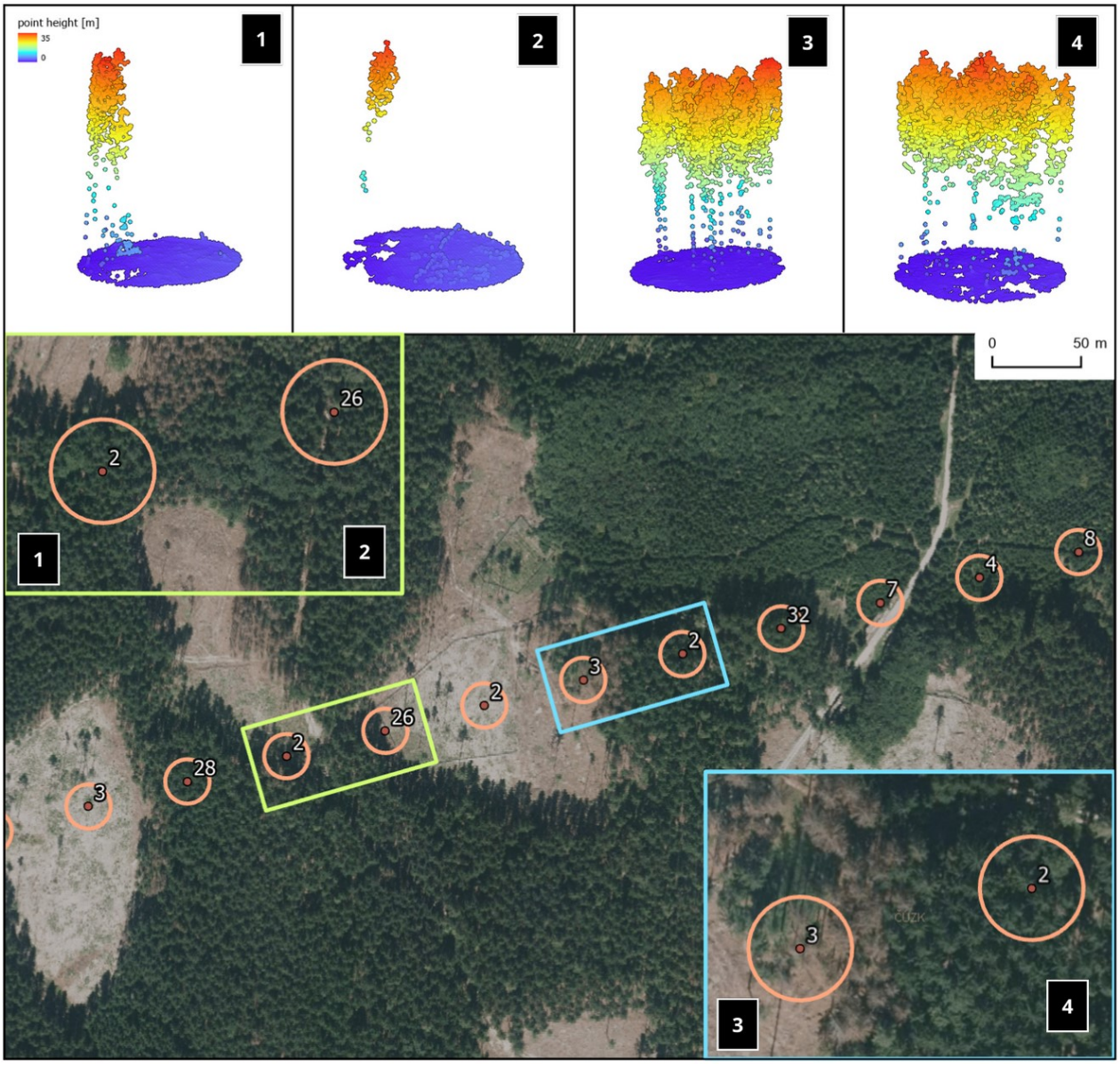

Fig. 25: Illustration of the GEDI footprint with respective values representing the GEDI RH95 metrics. Four close-ups show the footprints with the corresponding ALS point clouds.

Fig. 25 shows a closer look at individual footprints with large RH95 differences among the datasets. The values for simulated RH95 and ALS 95th percentile, respectively, are [1] 28.3 m and 29.5 m, [2] 2.95 m and 23.69 m, [3] 30.23m and 30.61 m, [4] 30.29 and 28.99 m. It is visible that the 95th percentile computed from ALS data reflects the presence of a single tree in the footprint, as 95% of the discrete returns are located below the height of the tree, while the simulator and GEDI reflect the challenges that are specific to large footprint

LiDAR sensing over a sparse canopy, that are more discussed in following section. In the close-ups it is also visible that some of the footprints [3], [4] are located in areas with snags.

As seen in Tab. 7, the GEDI derived RH95 has strong positive correlation with the RH95 derived from the simulator. The correlation between the ALS 95th percentile and GEDI is lower, however, the correlation between simulator derived RH95 and the ALS 95th percentile is higher than the correlation between simulator derived RH95 and GEDI derived RH95.

	GEDI RH95	Simulation RH95	Point cloud 90th percentile
GEDI RH95	1.0000	0.8107	0.7126
Simulation RH95	0.8107	1.0000	0.8109
Point cloud 90th percentile	0.7126	0.8109	1.0000

Tab. 7: Correlation coefficient of RH95 among datasets

Low correlation of 0.19 was found between the GEDI RH95 deviations and canopy cover (Fig. 26). That indicates low or no effect of canopy cover on the RH95 accuracy.

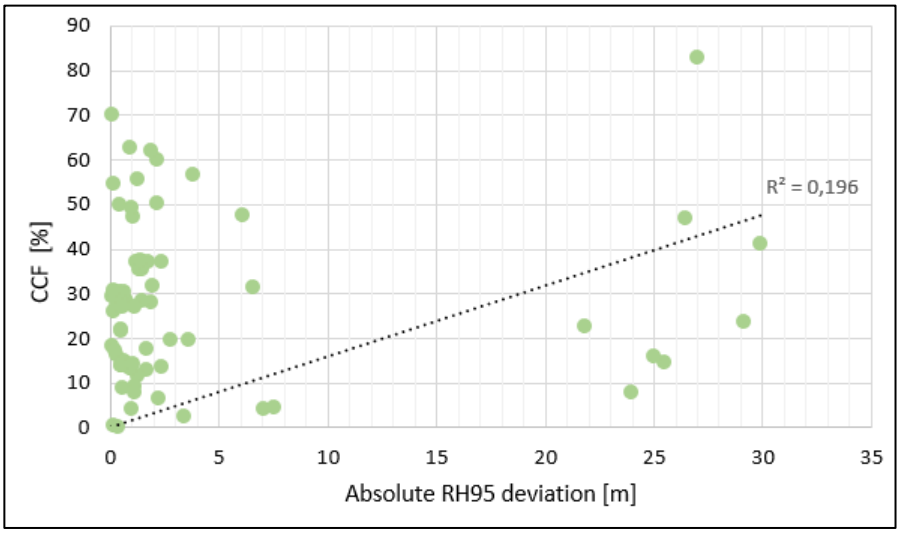

Fig. 26: A scatterplot illustrating the relationship between GEDI derived CCF and the deviation of RH95 height between GEDI and the ALS benchmark

5.4 Canopy cover

Fig. 27 shows that GEDI derived canopy cover pronounce quite different distribution compared to the simulator derived canopy cover. In Tab. 9 it is shown that the correlation between the canopy cover derived from the simulator and the canopy cover GEDI estimates is relatively low. The median difference, maximum absolute difference and RMSE comparison among the datasets is shown in Tab. 8. The is trend generally showing underestimation of the CCF by GEDI. The simulator has the highest correlation with the manually computed values from ALS point cloud, which suggest that the variables in the simulator algorithm that are set to a constant value to simulate the sensor parameters of GEDI still produce results that show stronger relationship with discrete-return calculated data.

	GEDI x Simulator	GEDI x ALS	Simulator x ALS
Median difference [percent points]	22	26	8
Maximum absolute difference [percent points]	96	83	79
RMSE [percent points]	38	33	21

Tab. 8: The intercomparison of median difference, maximum absolute difference and RMSE of CCF among datasets

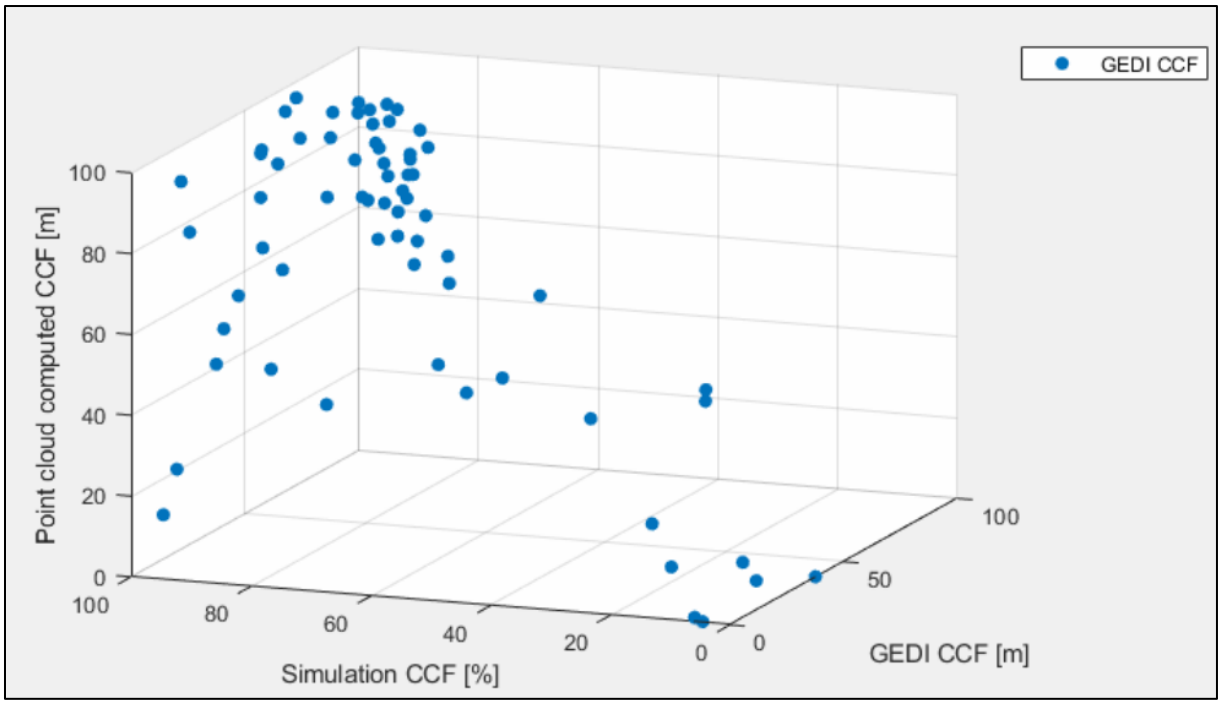

Fig. 27: A scatterplot illustrating the distribution of GEDI derived RH95 in relation to the CCF estimates from ALS simulated waveforms and the CCF of the ALS point cloud of each footprint

CCF derived from	GEDI	Simulation	Point cloud
GEDI	1.0000	0.5426	0.6446
Simulation	0.5426	1.0000	0.7294
Point cloud	0.6446	0.7294	1.0000

Tab. 6: Correlation coefficient of CCF among the datasets

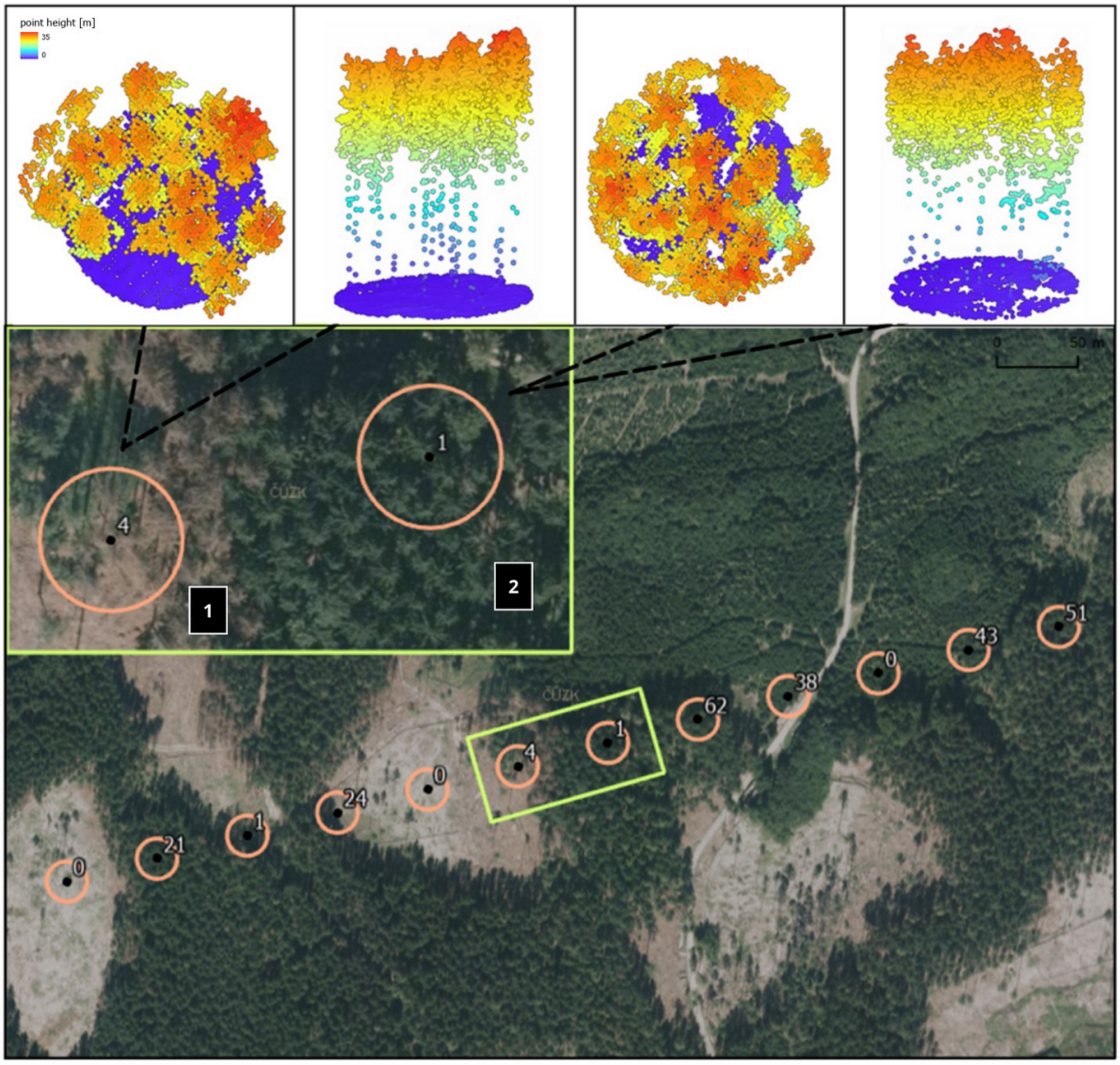

Fig. 28: Illustration of the GEDI footprint with respective values representing the GEDI CCF metrics. Two close-ups show the footprints with the corresponding ALS point clouds.

The Fig. 28 shows two footprints where the values of CCF derived by GEDI and by simulator vary significantly (by 64% and 77% for [1] and [2], respectively). The tree crowns are shown also from above, as the canopy gaps reflect significantly into the CCF value. The ALS data were acquired 5 days in advance of the GEDI acquisition, so there is only a slim chance of deforestation in the meantime. As seen in the close-up in the Fig. 26, the footprint with GEDI-reported cover value of 4% is in majority covered in dead standing trees, that are in the aerial image barely distinguishable from the ground. The GEDI geolocation error could also provide an explanation, however, it cannot explain the case of the neighbouring footprint, which even when shifted within the maximum reported geolocation error, would be still

located in a forested area. In said footprint with GEDI-estimated cover of 1% is also in the close-up visible a dead-standing tree. From another case of footprints, where the aerial image showed forested area but both GEDI and point cloud reported not forested area is presumed that the aerial imagery was acquired in a span of months prior to the GEDI and ALS acquisition. It is possible that the condition of the trees further deteriorated since the acquisition of the photo imagery and was not captured by the discrete lidar sensing.

5.4.1 Use of GEDI in time series of ALS datasets

Fig. 29 and Fig. 30 show examples of using GEDI among ALS data after standardising them to the same format and deriving described metrics. The graphs show metrics derived from simulated waveforms and real GEDI waveforms in 4 acquisition dates within one year. The ground elevation shows relatively small deviation among all the data, which is in accordance with the results of section 5.1.2. After visual comparison of the graphs with the locations of the respective footprints shown on Fig. 29, it is shown that a horizontal structure of the forest can be observed from the RH95 metric. However, there are some behavioural phenomena specific to large-footprint lidar that are caused by inaccuracies of the simulated metrics and could be interpreted wrongly. For example, in footprint [4], based only on RH95, the most logical interpretation would be a deforestation; however, on the ALS point cloud it is visible that no significant change other than a slight decrease of tree cover happened in the specific footprint between August and September. When looking at the graph of ground elevation, the deviation of ground elevation for [4] can supposedly mean overestimation of the ground elevation that results in underestimation of RH, which is described in multiple studies (Dorado-Roda et al., 2021; Guerra-Hernández and Pascual, 2021; Potapov et al., 2021). The CCF derived from GEDI generally underestimates the cover in comparison to the simulator, which is in accordance with the results of 5.1.4.

Fig 29. Location of footprints corresponding to graphs on Fig. 30

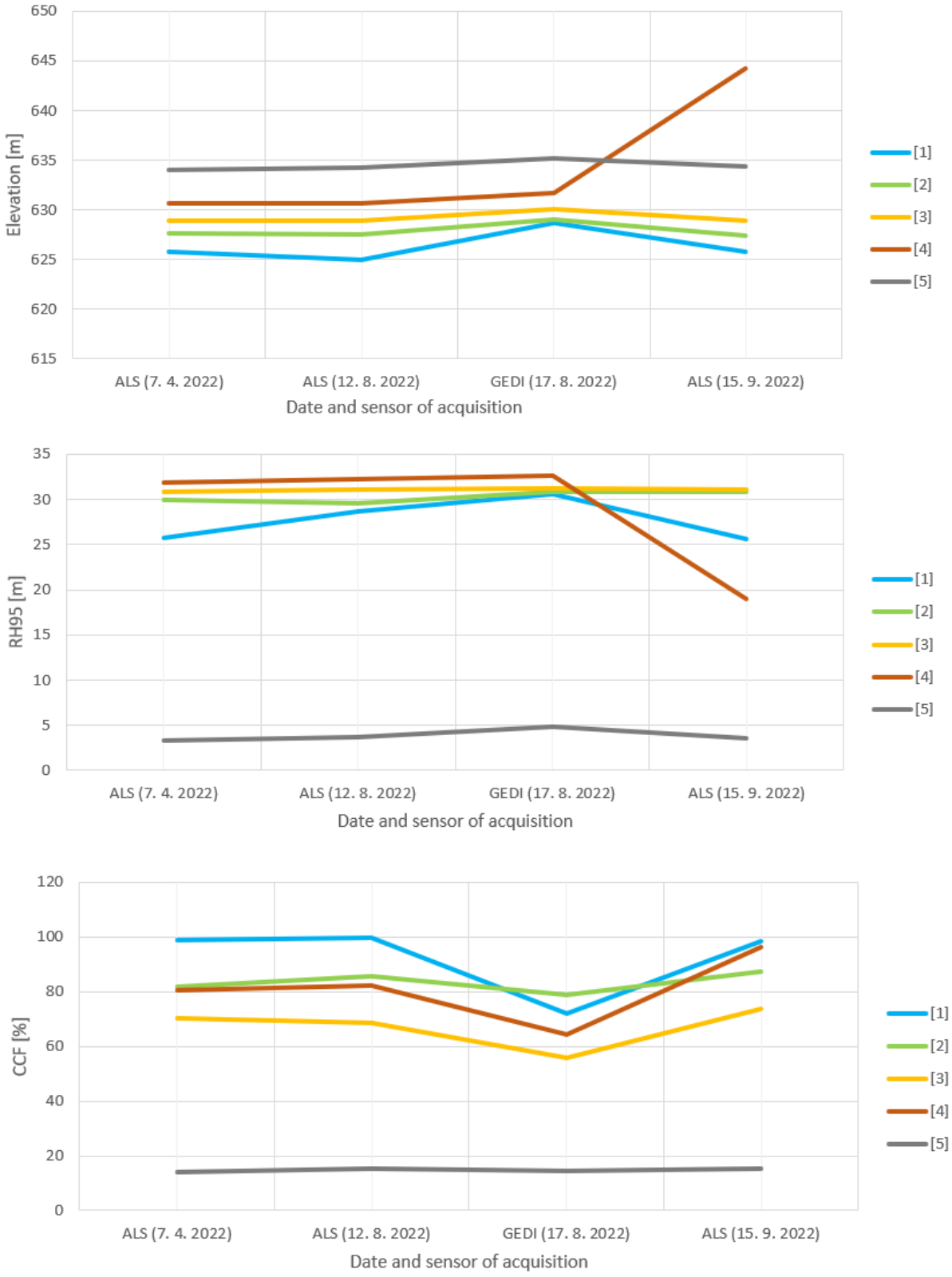

Fig. 30: The use of GEDI data among ALS data, showcase on selected metrics

6 Discussion

A limited number of studies (Guerra-Hernández & Pascual, 2021) published prior to this study focused on possibility of incorporating GEDI data products (i.e. not simulations) as a part of a multitemporal LiDAR dataset. With recurrent GEDI acquisitions and state-of-art LiDAR products, the GEDI development team encourages carrying out validation studies to support a further improvement of coregistration of GEDI data products and the performance of related tools (Guerra-Hernández & Pascual, 2021).

First, the geolocation accuracy of the GEDI data was tried to be optimised by a waveform matching approach, and ground elevation and RH95 metrics were derived from both nominal, i.e a centroid of a Version 2 GEDI footprint and optimised, i.e. a centroid of a footprint with an offset defined by waveform matching, data. The RMSE of ground elevation increased by waveform matching based on highest correlation by 0.35 m, and the RMSE of RH95 rose by 0.68 m. In the case of waveform matching based on highest overlap of areas under waveform curves, the RMSE rose by 0.42 m and 0.71 m. These results go in contrary with (A. Liu et al., 2021; Roy et al., 2021), where the RMSE generally lowered. However, all mentioned studies highlighted the challenge of waveform matching in areas that include forest edges and large canopy gaps. As the area of focus in this study is small with highly fragmented forest, and almost half of the footprints are located on the edges of larger gaps, forest edges, or spots with uneven tree density, the author suggests the character of the area as a cause for such results, based on empirical assessment of the individual footprints. The results might also be influenced by relatively small number of footprints in the study, or the sampling of 5x5 metres in the grid of possible footprint locations that was preferred to denser sampling due to computational complexity. The lowest 5 metres of GEDI waveforms were reported to often have blurred ground peaks by Bauer et al. (2021), possibly influenced by understory vegetation and uneven terrain, and were therefore discarded in their study. That approach was considered not feasible for the number of shots in this study, as well as for the character of the study site, since when the focus is on the change of forested area in a highly dynamic study site, eliminating lower parts of the waveforms might cause a loss of information on the horizontal structure of the forest, e.g. gaps. The author suggests caution with using waveform matching to optimise GEDI's geolocation based on the characteristics (e.g., terrain complexity, variability in vegetation height, structure and density) of the studied area, in accordance with the drawbacks of large footprint full-waveform lidar, that are discussed in following text. As the geolocation optimisation method is more of a guess (Ni et al., 2021)

and there is no way to validate the actual position of GEDI, the author was concerned that in case of this study, using the results of waveform matching as opposed to the original data might introduce additional error, as also mentioned in (Lang et al., 2021; Potapov et al., 2021), since a shift of few horizontal metres can introduce an error of several metres in height. Therefore, the original data were used as discussed below.

A sample of GEDI footprints that were acquired over an area covered in a forest with high fragmentation, large canopy gaps and uneven tree density was assessed and compared to the performance of GEDI simulator. Metrics that were derived from both real and simulated GEDI waveforms were validated against ALS benchmark data, and their accuracy is discussed in following paragraphs. The computed metrics vary among the footprints and between GEDI, simulated and computed metric.

The vertical accuracy of Version 2 of GEDI data product is reported to be close to 50 cm (Dubayah et al. 2020). In this study, the accuracy of GEDI ground elevation measurements with described data filtering criteria, when the median of ALS ground classified points was considered as ground truth, was 2.29 m in terms of RMSE. That is lower than in Guerra-Hernández et al. (2021) and Hancock et al. (2019) with RMSEs 4.48 m and 5.7 m, respectively, possibly due to more elaborate quality filtering settings and absence of significantly sloped terrain. The maximum absolute ground elevation error was 10.11 m, which is significantly lower than in the study of (Spracklen and Spracklen, 2021), where they associated the highest absolute ground elevation errors (up to 43.8 m) with high solar elevation, i. e. day-time shots. The thorough quality filtering that was used in this study yielded relatively clean data that showed 0.998 correlation coefficient with validation dataset. The high GEDI's performance in terms of ground elevation accuracy was highlighted by the frequent occurrence of understory canopy that is often found to challenge LiDAR ground detection abilities. The understory canopy might have led to the trend that the terrain heights in this study were generally overestimated (with median deviation 0.8 m), as opposed to (Guerra-Hernández et al., 2021) whose study showed underestimation in ground elevation.

Only a weak relationship was found between the ground elevation deviations and canopy cover. A weak relationship was also found between the ground elevation deviations and RH95 metric, indicating that neither canopy cover nor canopy height have significant influence on GEDI's ground elevation reading error. Both results align with previous study conducted over temperate coniferous forest (Adam et al., 2020). Stronger relationship between those phenomena were found in studies performed over different biomes (Huettermann et al., 2022;

Musthafa and Singh, 2022; Liu et al., 2021; Silva et al., 2021; Spracklen and Spracklen, 2021).

The canopy top was represented by an RH95 metric in this study. The median, an indicator chosen for being robust to outliers, showed 1.22 m of deviation between GEDI RH95 and ALS validation dataset. The mean difference, though, was 4.56 m, which is notably higher than in Potapov et al. (2021) (mean difference 0.7 m). The higher number was mostly caused by frequent canopy gaps, that resulted in a high number of large outliers. The maximum deviation between GEDI and the ALS reference dataset in terms of canopy height is 29.89 m, which in the forest monoculture of *picea abies* (height up to 35 m) means the difference between forested and non-forested area. In this study a relationship between the RH95 deviations and GEDI derived canopy cover was not proven, which is likely due to relatively high inaccuracies of the GEDI canopy cover metric for this study site. However, in Neuenschwander et al. (2020) and Dorado-Roda et al., (2021) it is documented that in the areas with sparse canopy cover, the energy is more likely to be reflected from the ground surface than from the canopy for both full-waveform GEDI and photon counting ICESat-2, which might often lead to an error. From visual analysis of the aerial images, the study site contains significant number of dead-standing trees that are spread across the study site, which is another phenomenon influencing the accuracy of derived RH metrics as highlighted also in (Huettermann et al., 2022). Their findings suggest that the spread-out distribution of return energy caused by dead-standing trees might present a challenge for GEDI's detection of RH95; this observation is supported by the findings in this study. The characteristics of the RH metrics makes the indicator sensitive to even a single tree occurrence within a footprint, as in that case, the specified (e.g. 95%) of energy is supposed to be returned below the tree's height. That brings geolocation inaccuracy of GEDI as another possible cause of large height differences between GEDI RH metrics and real canopy height, specifically in case of fragmented forest with large canopy gaps, lower tree density and frequent forest edges, as mentioned also in other studies (Huettermann et al., 2022; Dorado-Roda et al., 2021; Guerra-Hernández & Pascual, 2021; Roy et al., 2021). Several of the outliers might also be caused by artificial objects such as power lines, as GEDI does not discriminate between natural and artificial objects, and the study site is located close to urban areas. In Potapov et al. (2021) it is also indicated that sensor-specific parameters of GEDI such as the pulse with might also have a notable influence on large errors between GEDI RH metrics and true canopy height; however, further research in that area hasn't been carried out yet.

The CCF GEDI values have shown only medium positive correlation (0.64) with the validation ALS dataset. The maximum absolute difference between the CCF derived from GEDI and the manually computed values from ALS point cloud is 83 percent points, with median deviation 26 percent points. The causes of wrong canopy cover readings were so far found hard to explain (Spracklen and Spracklen, 2021). Based on an assessment of individual footprints, the results of this study suggest to further examine the effect of dead-standing trees on the waveform return, as it might be linked to the cover metrics error in similar way that it is linked to the RH95 metrics error. The reason for overall high differences between the CCF values among all the datasets, that were observed in this study, might be a slightly different approach of calculating the metric for each dataset. From ALS point cloud, the CCF is calculated with a CHM approach as in (Moudrý et al., 2022; Ma et al., 2017). In the calculations of CCF that are included in GEDI datasets are as variables the specific return signal parameters, such as the irradiance of emitted laser pulse, that is based on derived atmospheric transmittance (Tang et al., 2019). In the GEDI Simulator, constants are set instead of the varying GEDI pulse parameters.

The simulator showed correlation with ALS benchmark data of 0.99, 0.81 and 0.72 for ground elevation, RH95 and CCF respectively. The differences are caused by the added noise and characteristics specific to a large-footprint, full waveform LiDAR. The maximum absolute differences between simulated metrics and ALS data are 14.23 m in ground elevation height, 28.1 m in RH95 and 79 percent points in canopy cover. In relation to used study site, the result of maximum deviation for RH95 show that the simulator reflects the challenge that large canopy gaps present for a large-footprint, full waveform lidar. The simulator underestimated the terrain in relation to ALS benchmark with a RMSE 2.36 m, which is in accordance with the trend observed in GEDI data for the study site.

In comparison with ALS benchmark data, the simulator overestimates the RH95 metrics with RMSE 5.9 m, which is a higher bias than in Huettermann et al., (2022) (RMSE 4.2 m), even though both results were derived from datasets of similar point densities (over 9 pulses/m^2). The higher RMSE value in this study is likely because of adding noise to the simulations. The study of Huetterman et al., (2022) implies that the RMSE difference between GEDI RH95 and simulated RH95, that were 8.46 m for trees of the same height category as in this study, could be potentially minimised by adding the noise to the simulated waveforms. That is confirmed in this study, where the noise setting that was specific to employed quality filtering (night-time sensing, power beams) resulted in lower RMSE of 6.95 m.

The simulator was suggested as a standardisation tool by (Huettermann et al., 2022) for not only multimodal datasets containing a large-footprint full waveform lidar, but also for datasets containing ALS from different sources. The simulator is reported to minimise the influence of different acquisition height and point density. However, based on the findings of this study, it is recommnded to experiment with the selected footprint size, as the footprint size of GEDI, although designed to reduce the influence of canopy density and fragmentation, still possesses several drawbacks that are characteristic for large footprint lidars.

The ground elevation derived from simulator showed nearly 100% correlation (0.998) with the ground elevation derived by GEDI, with median deviation 1.09 m. The maximum absolute difference between GEDI derived ground elevation and ground elevation from simulator is 4.85 m. In comparison with (Huettermann et al., 2022), it shows that adding the noise has effect on improving the correlation between GEDI and simulated metrics also in the terms of ground elevation. Different ground-finding algorithm was used in both studies. In Huettermann et al. (2022) the authors favoured Lowest maximum that was found to result in higher accuracy, while in this study was used Gaussian fitting as in Hancock et el. (2019) and Bauer et al. (2021). However, as it is shown on the results of this study, that the correlation between GEDI and simulated data clearly benefited from adding GEDI-like noise.

The maximum absolute difference between GEDI estimated RH95 and RH95 derived from simulator is 28.25 m. The median difference between those two variables is 0.94 m with RMSE 6.95 m. For the RH metrics, the quantified measure of the simulator's accuracy is a RMSE of 5.7 when compared to large-footprint, full-waveform LVIS waveforms during GEDI's pre-launch validations (Hancock et al. 2019). As the simulators accuracy was averaged over a wide range of forest types, following studies (Guerra-Hernández and Pascual, 2021; A. Liu et al., 2021; Potapov et al., 2021; Roy et al., 2021) that employed real GEDI data products and more local scales showed generally higher RMSEs, which is in accordance with the result of this study. The required pulse density set by Hancock et al. (2019) is 1.5 pulse/m2, while this study shows RMSE 6.95 m for pulse density 9,8 pulse/m2, which goes in accordance with Huettermann et al. (2022), who as well shows higher RMSEs for densities over 9 pulses/m2.

The comparison between canopy cover derived from the simulator and from GEDI has shown median difference 22 percent points and maximum absolute difference 96 percent points. Those are notably high numbers for using both sources in a multitemporal dataset. The difference might be caused by slightly different ways of computing the variable by GEDI and

the simulator. The suggestion for future research is to use a different noise setting for computing the CCF or a different ground-finding algorithm.

7 Conclusion

This research comprised two parts, both carried out with the focus on using GEDI for local forest monitoring. The first part was focused on optimising the geolocation of GEDI, as few metres of geolocation inaccuracy might introduce erroneous reading of canopy height in forest edges and spaces with large canopy gaps. However, it was shown that fragmented forest causes a challenge for the geolocation optimalisation approach adopted by GLAS, and as selected study site is characterised by highly dynamic fragmented forest, the tried approaches have shown to not be beneficial. There might be a more suitable geolocation optimalisation approach for such areas, which shall be an object of further research.

In the second part, the study followed on previous research on integration of GEDI and ALS data for forest change monitoring purpose. The GEDI Simulator has shown to be a valuable tool for integrating ALS and GEDI, and the ground elevation height derived by both sources has shown very high correlation. Canopy gaps and low tree densities generally cause erroneous readings in large-footprint full-waveform LiDARs. That was supposed to be overcome by GEDI's footprint size. The results of this study and assessment of individual footprints have shown that fragmented forest still presents a challenge for GEDI. As a result, trends that can be observed in a timeline created from LiDAR sensing of a forested area, such as changes in canopy height or canopy cover, might be caused by either wrong GEDI reading or inaccuracies of the simulated metrics and could lead to wrong interpretation of the horizontal forest structure. Apart from the fragmentation of the forest, a frequent occurrence of dead-standing trees might be linked to the RMSEs observed in this study, and the author suggests further research into the influence of dead-standing trees on the GEDI's return waveforms. The author also suggests using the CCF metric with caution when integrating ALS and GEDI, as there are large biases between the values derived by GEDI and values derived by the GEDI Simulator.

It is acknowledged that comparing data derived from the same data source is not an ideal validation approach, as the validation data should preferably come from an independent dataset. However, the intercomparison between real GEDI data, simulated waveforms and ALS derived metrics is a widely adopted approach (Huettermann et al., 2022). It is also acknowledged that the used GEDI dataset was considerably small and can be hardly taken as representative, however, GEDI is supposed to provide a sample of the forest structure rather than a census. The GEDI quality filtering applied in this study yielded relatively clean dataset, however, only 5% of footprints acquired over the study site fulfilled the required conditions.

With the mission still ongoing, a denser sampling is expected. Based on the findings in this thesis, GEDI might be considered for use alongside other LiDAR datasets as an valuable data source for long-time research on vegetation and forest monitoring.

8 References

Scientific Publications

Adam, M., Urbazaev, M., Dubois, C., & Schmullius, C. (2020). Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters. *Remote Sensing*, *12*(23), 1–28. https://doi.org/10.3390/rs12233948

Bauer, L., Knapp, N., & Fischer, R. (2021). Mapping amazon forest productivity by fusing GEDI lidar waveforms with an individual-based forest model. *Remote Sensing*, *13*(22), 4540. https://doi.org/10.3390/RS13224540/S1

Blair, B.; Wirt, B.; Armston, J.; Hofton, M.; Luthcke, S.; & Tang, H.(2021). Global Ecosystem Dynamics Investigation (GEDI) Level 02 User Guide. (Document version 2.0). Available online: https://lpdaac.usgs.gov/documents/986/GEDI02_UserGuide_V2.pdf (accessed on 15 March 2022).

Boucher, P. B., Hancock, S., Orwig, D. A., Duncanson, L., Armston, J., Tang, H., Krause, K., Cook, B., Paynter, I., Li, Z., Elmes, A., & Schaaf, C. (2020). Detecting change in forest structure with simulated GEDI lidar waveforms: A case study of the hemlock woolly adelgid (HWA; adelges tsugae) infestation. *Remote Sensing*, *12*(8), 1304. https://doi.org/10.3390/RS12081304

Bruening, J. M., Fischer, R., Bohn, F. J., Armston, J., Armstrong, A. H., Knapp, N., Tang, H., Huth, A., & Dubayah, R. (2021). Challenges to aboveground biomass prediction from waveform lidar. *Environmental Research Letters*, *16*(12), 125013. https://doi.org/10.1088/1748-9326/ac3cec

Chen, L., Ren, C., Zhang, B., Wang, Z., Liu, M., Man, W., & Liu, J. (2021). Improved estimation of forest stand volume by the integration of GEDI LiDAR data and multi-sensor imagery in the Changbai Mountains Mixed forests Ecoregion (CMMFE), northeast China. *International Journal of Applied Earth Observation and Geoinformation*, *100*, 102326. https://doi.org/10.1016/j.jag.2021.102326

Chen. Q., (2010). Retrieving vegetation height of forests and woodlands over mountainous areas in the pacific coast region using satellite laser altimetry. *Remote Sensing of Environment*, *114*(7), 1610-1627. https://doi.org/10.1016/j.rse.2010.02.016

Dorado-Roda, I., Pascual, A., Godinho, S., Silva, C. A., Botequim, B., Rodríguez-Gonzálvez, P., González-Ferreiro, E., & Guerra-Hernández, J. (2021). Assessing the accuracy of gedi data

for canopy height and aboveground biomass estimates in mediterranean forests. *Remote Sensing*, *13*(12), 2279 . https://doi.org/10.3390/rs13122279

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., & Prince, S.D. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, *79*(2-3), 305–319. https://doi.org/10.1016/S0034-4257(01)00281-4

Dubayah, R., Bryan, B.,Goetz, S.,Fatoyinbo, L.,Hansen, M.,Healey, S.,Hofton, M.,Hurtt, G.,Kellner, J.,Luthcke, S.,Armston, J.,Tang, H.,Duncanson, L.,Hancock, S.,Jantz, P., Marselis, S.,Patterson, P., Qi, W.,Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., & Chazdon, R.L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica.J. *Journal of Geophysical Research Atmospheres*. *115*(G2). https://doi.org/10.1029/2009JG000933

Duncanson, L., Kellner, J. R., Armston, J., Dubayah, R., Minor, D. M., Hancock, S., Healey, S. P., Patterson, P. L., Saarela, S., Marselis, S., Silva, C. E., Bruening, J., Goetz, S. J., Tang, H., Hofton, M., Blair, B., Luthcke, S., Fatoyinbo, L., Abernethy, K., & Zgraggen, C. (2022). Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment*, *270,* 112845. https://doi.org/10.1016/j.rse.2021.112845

Fayad, I., Baghdadi, N. N., Alvares, C. A., Stape, J. L., Bailly, J. S., Scolforo, H. F., Zribi, M., & Maire, G. le. (2021). Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 7095–7110. https://doi.org/10.1109/JSTARS.2021.3092836

Guerra-Hernández, J., & Pascual, A. (2021). Using GEDI lidar data and airborne laser scanning to assess height growth dynamics in fast-growing species: a showcase in Spain. *Forest Ecosystems*, *8*(1), 14. https://doi.org/10.1186/s40663-021-00291-2

Hancock, S., Lewis, P., Foster, M., Disney, M., & Muller, J. P. (2012). Measuring forests with dual wavelength lidar: A simulation study over topography. *Agricultural and Forest Meteorology*, *161*, 123–133. https://doi.org/10.1016/j.agrformet.2012.03.014

Harding, D.; Lefsky, M.; Parker, G.; & Blair, J. (2001). Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. Remote Sensing of Environment ,*76*(3), 283–297. https://doi.org/10.1016/S0034-4257(00)00210-8

Hofton, M., & Blair, J., (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products. https://lpdaac. usgs.gov/documents/581/GEDI_WF_ATBD_v1.0.pdf. (Accessed 13 March 2022).

Huang, W., Sun, G., Dubayah, R., Cook, B., Montesano, P., Ni, W., & Zhang, Z. (2013). Mapping biomass change after forest disturbance: Applying LiDAR footprint-derived models at key map scales. *Remote Sensing of Environment*. *134*, 319–332. https://doi.org/10.1016/j.rse.2013.03.017

Huettermann, S., Jones, S., Soto-Berelov, M., & Hislop, S. (2022). Intercomparison of Real and Simulated GEDI Observations across Sclerophyll Forests. *Remote Sensing*, *14*(9), 2096. https://doi.org/10.3390/rs14092096

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., & Wegner, J. D. (2021). Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sensing of Environment* , *268*, 112760. https://doi.org/10.1016/j.rse.2021.112760

Leite, R. V., Silva, C. A., Broadbent, E. N., Amaral, C. H. do, Liesenberg, V., Almeida, D. R. A. de, Mohan, M., Godinho, S., Cardil, A., Hamamura, C., Faria, B. L. de, Brancalion, P. H. S., Hirsch, A., Marcatti, G. E., Dalla Corte, A. P., Zambrano, A. M. A., Costa, M. B. T. da, Matricardi, E. A. T., Silva, A. L. da, & Klauberg, C. (2022). Large scale multi-layer fuel load characterization in tropical savanna using GEDI spaceborne lidar data. *Remote Sensing of Environment*, *268*, 112764. https://doi.org/10.1016/J.RSE.2021.112764

Liu, A., Cheng, X., & Chen, Z. (2021). Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, *264,* 112571. https://doi.org/10.1016/j.rse.2021.112571

Liu, L., Coops, C. N., Aven, W. N., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, *200*, 170-182. https://doi.org/10.1016/j.rse.2017.08.010

Liu, X., Su, Y., Hu, T., Yang, Q., Liu, B., Deng, Y., Tang, H., Tang, Z., Fang, J., & Guo, Q. (2022). Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2 data. *Remote Sensing of Environment*, *269,* 112844. https://doi.org/10.1016/j.rse.2021.112844

Luthcke, S. B., Rebold, T., Thomas, T., & Pennington, T. (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI Waveform Geolocation for L1 and L2 Products. lpdaac.usgs.gov/documents/579/GEDI__WFGEO_ATBD_v1.0.pdf (Accessed 18 December 2021).

Ma, Q., Su, Y., & Guo, Q. (2017). Comparison of Canopy Cover Estimations from Airborne LiDAR, Aerial Imagery, and Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(9), 4225–4236. https://doi.org/10.1109/JSTARS.2017.2711482

Marselis, S.M., Tang, H., Armston, J.D., Calders, K., Labrière, N. & Dubayah, R. (2018). Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lope National Park, Gabon. *Remote Sensing of Environment*, 216, 626–634. https://doi.org/10.1016/j.rse.2018.07.023

Milenković, M., Reiche, J., Armston, J., Neuenschwander, A., de Keersmaecker, W., Herold, M., & Verbesselt, J. (2022). Assessing Amazon rainforest regrowth with GEDI and ICESat-2 data. *Science of Remote Sensing*, *5*, 100051. https://doi.org/10.1016/j.srs.2022.100051

Milenković, M., Schnell, S., Holmgren, J., Ressl, C., Lindberg, E., Hollaus, M., Pfeifer, N., & Olsson, H. (2017). Influence of footprint size and geolocation error on the precision of forest biomass estimates from space-borne waveform LiDAR. *Remote Sensing of Environment*, *200*, 74–88. https://doi.org/10.1016/j.rse.2017.08.014

Moudrý, V., Gdulová, K., Gábor, L., Šárovcová, E., Barták, V., Leroy, F., Špatenková, O., Rocchini, D., & Prošek, J. (2022). Effects of environmental conditions on ICESat-2 terrain and canopy heights retrievals in Central European mountains. *Remote Sensing of Environment*, *279,* 113112. https://doi.org/10.1016/j.rse.2022.113112

Musthafa, M., & Singh, G. (2022). Forest above-ground woody biomass estimation using multi-temporal space-borne LiDAR data in a managed forest at Haldwani, India. *Advances in Space Research*, *69*(9), 3245–3257. https://doi.org/10.1016/j.asr.2022.02.002

Ni, W., Zhang, Z., & Sun, G. (2021). Assessment of Slope-Adaptive Metrics of GEDI Waveforms for Estimations of Forest Aboveground Biomass over Mountainous Areas. *Journal of Remote Sensing*, 1–17. https://doi.org/10.34133/2021/9805364

Pang,Y., Lefsky, M., Sun, G., & Ranson, J., (2011). Impact of footprint diameter and off-nadirpointing on the precision of canopy height estimates from spaceborne lidar. *Remote Sensing of Environment, 115*, 2798–2809. https://doi.org/10.1016/j.rse.2010.08.025

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., &

Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/J.RSE.2020.112165

Riegl (2014). RIEGL LMS-Q780 Full Waveform Digitizing Airborne Laser Scanner for Wide Area Mapping. www.riegl.com (Accessed 25 January 2022).

Rishmawi, K., Huang, C., & Zhan, X. (2021). Monitoring Key Forest Structure Attributes across the Conterminous United States by Integrating GEDI LiDAR Measurements and VIIRS Data. *Remote Sensing*, *13*(3), 442. https://doi.org/10.3390/rs13030442

Roy, D. P., Kashongwe, H. B., & Armston, J. (2021). The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. *Science of Remote Sensing*, *4*, 100024. https://doi.org/10.1016/j.srs.2021.100024

Silva, C. A., Duncanson, L., Hancock, S., Neuenschwander, A., Thomas, N., Hofton, M., Fatoyinbo, L., Simard, M., Marshak, C. Z., Armston, J., Lutchke, S., & Dubayah, R. (2021). Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. *Remote Sensing of Environment*, *253*, 112234. https://doi.org/10.1016/J.RSE.2020.112234

Silva, C.A., Saatchi, S., Garcia, M., Labriere, N., Klauberg, C., Ferraz, A., Meyer, V., Jeffery, K.J., Abernethy, K., & White, L. (2018). Comparison of Small-and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study from Central Gabon. IEEE J. Sel. Top. *Applied Earth Observation Remote Sensing*, *11*, 3512–3526.

Spracklen, B., & Spracklen, D., (2021). Determination of Structural Characteristics of Old-Growth Forest in Ukraine Using Spaceborne LiDAR. *Remote Sensing, 13*(7), 1233. https://doi.org/10.3390/RS13071233

Tan, P., Zhu, J., Fu, H., Wang, C., Liu, Z., & Zhang, C. (2020). Sub-Canopy Topography Estimation from TanDEM-X DEM by Fusing ALOS-2 PARSAR-2 InSAR Coherence and GEDI Data. *Sensors*, *20*(24), 7304. https://doi.org/10.3390/S20247304

Tang, H., & Armston, J., (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics. https://lpdaac.usgs.gov/documents/588/GEDI_FCCVPM_ATBD_v1.0.pdf (Accessed 25 June 2022).

Tang, H., & Dubayah, R., (2017). Light-driven growth in Amazon evergreen forests explained by seasonal variations ofvertical canopy structure. *The Proceedings of the National Academy of Sciences*, *114*, 2640–2644. https://doi.org/10.1073/pnas.1616943114

Wang, C., Wang, C., Wang, C., Zhu, X., Zhu, X., Zhu, X., Nie, S., Nie, S., Nie, S., Xi, X., Li, D., Zheng, W., & Chen, S. (2019). Ground elevation accuracy verification of ICESat-2 data: a case study in Alaska, USA. *Optics Express, 27*(26), 38168–38179. https://doi.org/10.1364/OE.27.038168

Weishampel, J.F., Drake, J.B., Cooper, A., Blair, J.B., & Hofton, M. (2007) Forest canopy recovery from the 1938 hurricane and subsequent salvage damage measured with airborne LiDAR. *Remote Sensing of Environment*, 109, 142–153. https://doi.org/10.1016/j.rse.2006.12.016

Other sources

ČÚZK (2022). INSPIRE služba transformace souřadnic. https://geoportal.cuzk.cz/(S(y1c1tebwlpqq3xguhoyobwj1))/Default.aspx?mode=TextMeta&text=sit.trans.uvod&side=sit.trans&head_tab=sekce-03-gp&menu=34 (Accessed 5 April 2021)

GEDI Ecosystem Lidar (2022). Instrument Overview. https://gedi.umd.edu/instrument/specifications/ (Accessed 5 December 2021)

Krehbiel, C. (2021). GEDI Spatial and Band/Layer Subsetting and Export to GeoJSON Script. https://git.earthdata.nasa.gov/projects/LPDUR/repos/gedi-subsetter/browse (Accessed 10 May 2021)

Silva, C.A., Hamamura,C., Valbuena, R., Hancock,S., Cardil,A., Broadbent, E. N., Almeida, D.R.A., Silva Junior, C.H.L, & Klauberg, C., (2020). rGEDI: NASA's Global Ecosystem Dynamics Investigation (GEDI) Data Visualization and Processing. version 0.1.9. https://CRAN.R-project.org/package=rGEDI (Accessed 20 February 2022)

Used data and software

Core Team R (2020). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/. (Accessed 5 December 2021)

Dubayah, R., Hofton, M., Blair. J.,Armston, J.,Tang, H., & Luthcke, S. (2021). GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002 [Dataset]. https://doi.org/10.5067/GEDI/GEDI02_A.002 (Accessed 6 June 2021)

Dubayah, R., Luthcke, S., Blair, J., Hofton, M., Armston, J., & Tang, H. (2020). GEDI L1B Geolocated Waveform Data Global Footprint Level V002 [Dataset]. https://doi.org/10.5067/GEDI/GEDI01_B.002. (Accessed 4 June 2021)

Isenburg, M. (2020). LAStools-efficient tools for LiDAR processing. Version 200216. http://lastools.org. (Accessed 10 December 2021)