



University of
Nottingham
UK | CHINA | MALAYSIA

Assessing Mangrove Canopy Heights in Myanmar using GEDI & Sentinel-2 for Effective Monitoring

Submitted September 2023, in partial fulfilment of
the conditions for the award of the degree **MRes Geospatial Studies**.

Yashvini Shukla 20489718

Supervised by Professor Doreen Boyd

School of Engineering
University of Nottingham

Yashvini Shukla
20489718

ENG4014
Dissertation in Geospatial Science
September 23

CENTRE FOR DOCTORAL TRAINING IN GEOSPATIAL SYSTEMS

FACULTY OF ENGINEERING

Abstract

Mangrove forests are crucial ecosystems that store over three times as much carbon per hectare as terrestrial tropical forests (Donato et al., 2011) and host an important role in regulating global and local climate systems (Estoque et al. 2018). Alarming, the global mangrove area shrank by 50% between 1997 and 2016, with the most significant losses occurring in Southeast Asia (Estoque et al., 2018; Románach et al., 2018). Projections suggest that mangroves could vanish entirely within the next century (Polidoro et al., 2010), highlighting the urgent need for accurate mapping of their structural and spatial characteristics to aid conservation and restoration efforts. While readily available multispectral satellite data like Sentinel-2 provide insights into mangrove coverage, they offer limited information on three-dimensional (3D) structural characteristics such as Canopy Height (CH). Traditional methods for 3D mapping, such as Airborne Lidar Surveying (ALS) and Synthetic Aperture Radar (SAR), are expensive and geographically unscalable. Spaceborne Lidar missions, and specifically the launch of the Global Ecosystem Dynamics Investigation (GEDI), offer a new opportunity to obtain 3D mangrove canopy data. As GEDI samples only 4% of the Earth, it is often fused with contiguous imagery like Landsat and Sentinel-2 to produce Global Canopy Height Maps (CHMs), albeit with limitations at local scales and for non-standard forest structures like mangroves (Potapov, 2021; Lang et al., 2022). This study leverages a Random Forest (RF) algorithm to combine Sentinel-2 and GEDI data for producing a contiguous Mangrove CHM for a restored local region in Myanmar for the year 2019. Field measurements were obtained from Worldview International, the project facilitators in 32 sample field plots (Vanniarachchy and Jayakody, 2020). Three models, trained on data sets from 2019 and 2020, were tested against a GEDI validation set and field height measurements. The Relative Height (Rh) at the 60th Percentile Waveform Energy Return (Rh60) from GEDI's Level 2A product (that provides elevation and height metrics) was identified as the best predictor of field-measured heights out of other Rh metrics, yielding an R^2 value of 0.24, a Mean Error (ME) of 0.28 m, and a Root Mean Squared Error (RMSE) of 0.37 m. These results were compared with 3 baseline Global CHMs from Potapov et al. (2021), Lang et al. (2022), and Simard et al. (2019). These comparisons revealed that the model by Lang et al. (2022) had the highest ME and RMSE, followed by the models from Simard et al. (2019) and Potapov et al. (2021), respectively. Further, vertical structural analysis using GEDI's L2B (that provides biophysical metrics) product indicated that the mangroves studied are vertically uniform, unlike typical forests. This could explain why the mean Rh metrics provide a better approximation of true heights as opposed to the commonly used Rh90+ metrics in the baseline Global CHMs. This research suggests that CH estimates are overestimated by over 50%, and as a consequence Above Ground Biomass (AGB) estimates could be grossly inaccurate in short stature (<3m) mangroves in current Global CHMs. This emphasizes the wider need for their local calibration. These findings have a direct impact on estimating National Carbon Stocks, contribute to the accreditation of community-based forest conservation and afforestation projects, and aid wider efforts in understanding and mitigating climate change.

Acknowledgements

I hereby thank Prof. Doreen Boyd – especially for her first principles approach to complex problems that guided me through this research. I also thank Prof. Keshav Srinivasan for his intuitive insight and experience with GEDI. I appreciate Amelia Halcomb’s generosity in replying to my questions, Dr. Sujit Ghosh for his incredible patience and explanations, Stenver Jenku, and Rez from Solid World for the opportunity.

Thank you to all the brilliant people at the NGI,

...family and friends.

To protecting our home.

1. Table of Contents

Abstract.....	18
Acknowledgements	19
List of Figures	22
List of Tables.....	24
List of Abbreviations	25
1. Introduction	27
1.1. Motivation and Research Purpose	30
2. Literature Review.....	31
2.1. Introduction to mangrove monitoring.....	31
2.2. Use of Space-borne lidar in CH measurements.....	33
2.2.1. ICESat	33
2.2.2. GEDI	34
2.3. A Summary of Global CHMs	37
2.4. Mangrove monitoring in VCS1764 - Ayeyarwady, Myanmar.....	40
2.5. Research Gap.....	42
3. Aims and Objectives.....	43
3.1. Research Aim.....	43
3.2. Research Objectives.....	43
4. Research Design Methodology	44
4.1. Project site	44
5. Data and Pre-processing	46
5.1. Field data.....	46
5.2. GEDI	48
5.2.1. GEDI L1B, L2A & L2B (Objective 1: Explore GEDI products over VS1764)	48
5.2.1. GEDI L2A (Objective 2: Create a Mangrove CHM)	49
5.3. Sentinel-2	52
5.4. Baseline data (Objective 3: Comparison with Baseline CHMs).....	55
6. Training the model.....	56
6.1.1. Random Forest Model	56
7. Results	60
7.1. Objective 1: Explore GEDI products over VS1764	60
7.1.1. GEDI L1B.....	60
7.1.2. GEDI L2A.....	65
7.1.3. GEDI L2B.....	68
7.2. Objective 2: Create a Mangrove CHM	69
7.2.1. GEDI L2A Rh Exploration with Sentinel-2 bands	69
7.2.2. Random Forest Model – Training the model	73

7.2.3.	Final results	73
7.2.4.	Examining Rh60.....	78
7.3.	Objective 3: Comparison with Baseline CHMs	79
7.3.1.	Comparison with local data	81
8.	Discussion	83
8.1.	Methods and Model	83
8.2.	Results.....	85
8.2.1.	Objective 1: Explore GEDI products over VS1764	85
8.2.2.	Objective 2: Create a Mangrove CHM	86
8.2.3.	Objective 3: Comparison with Baseline CHMs	88
8.3.	Future work and suggestions	90
9.	Conclusion.....	91
10.	Data availability.....	92
11.	References.....	93
12.	Appendix.....	98
12.1.	Field Heights.....	98
a)	Distribution of Heights.....	98
12.2.	Model Performance Scores	100
a)	Model 2019_s	100
a)	Model 2019.....	102
b)	Model 2020.....	104
12.3.	Alternative visualization of model results	106
12.4.	Performance Scores	107
a)	Test data.....	107
b)	Field data.....	108
12.5.	Summary of GEDI transects used by model	109
12.6.	Data Management Plan	110
12.7.	Gantt Chart.....	113

List of Figures

Figure 1.1: Location of VCS1764 in Myanmar.	29
Figure 2.1: GEDI’s sampling methodology (NASA, 2020).	35
Figure 2.2: VCS1764 Project location in Myanmar.	41
Figure 2.3: Visible degradation from 1985 imagery in 2010 and restoration in 2020 in the southern part of the site on Google Earth in 2010.	42
Figure 4.1: Location of sample field plots by year of planting VCS1764 Project location in Myanmar.	45
Figure 5.1: Field photographs of mangroves at an unspecified location in VCS1764 in 2023. (e) The height of a mangrove looks about $\frac{3}{4}$ of the average height (160cm) of a person– amounting to 120 cm. (Vanniarachchy and Jayakody, 2020).	47
Figure 5.2: The image acquisition process of GEDI (NASA, 2021).	48
Figure 5.3: Map of GEDI shots with Quality flag 1 and 0 passing over site in 2019.	50
Figure 5.4: Tide conditions in Goyangyi Kyun (Tidechecker, 2023).	52
Figure 6.1: Summary of Data Sources, filtration procedures, RF model parameters and outputs of the study.	56
Figure 7.1: GEDI Transect on 22 May 2019 over VCS1764.	60
Figure 7.2: Examining a GEDI transect Shot 2494110020003598, closest to a Sample Field plot 12.	61
Figure 7.3: a) Comparing the GEDI L1B waveforms of a shot near VCS1764 with that of b) in Redwood National Park.	61
Figure 7.4: a) Comparing different algorithms for Shot 2494110020003598. b) σ represents the standard deviation of the background noise level (Tang and Armston, 2019).	62
Figure 7.5: The Full waveform and relative height metric over mangrove and bare fields in May 2019.	63
Figure 7.6: Waveform returns over water and coast.	64
Figure 7.7: a) Rh covering 50 shots to the approximate right and left of the Sample Field Plot 12 in the GEDI in the along-track direction, on 22 May 2019. b) Rh across Redwood National Park.	65
Figure 7.8: Comparison of a) Mangroves on-site in VCS1764 and b) Californian Redwood trees.	66
Figure 7.9: a) GEDI Rh100 Canopy heights over VCS1764. b) Distribution of Rh100.	67
Figure 7.10: 3D Visualisation of Rh100 on 22 May 2019 (Shots in Yellow are those over mangrove in VCS1764) Raster values in Green are Potapov et al., (2021)’s Global CHM.	67
Figure 7.11: GEDI Energy profiles of 2 shots within 30m of their respective Field Plots in 2019.	68
Figure 7.12: GEDI L2B Beam 1011 PAVD over VCS1764 on 22 May 2019.	69
Figure 7.13: GEDI L2B BEAM1011 PAVD over Redwood National Park on June 19, 2019.	69
Figure 7.14: The canopy height metrics superposed with the Field heights, 22 May 2019.	70
Figure 7.15: GEDI L2A Rh Distribution in the Bounding Region of VCS1764 in model- 2019.	70
Figure 7.16: Correlation Heatmap for GEDI shots in the VCS1764 region - 2019.	71
Figure 7.17: Correlation between some Sentinel-2 bands and Rh products.	72
Figure 7.18: Correlations between the Sentinel-2 bands and GEDI-derived Rh values, and sample field heights.	72
Figure 7.19: Optimal mtry for each Rh metric model.	73
Figure 7.20: Performance metric across the testing set.	74
Figure 7.21: Performance metrics of each model on field data.	74
Figure 7.22: Filtering model_2020 for night-time shots.	75
Figure 7.23: GEDI validation set metrics.	76
Figure 7.24: Field validation set metrics.	77
Figure 7.25: The Importance Scores of some Rh variables in the trained RF model.	78

Figure 7.26: The importance score vs. correlation for the Rh60 model.	78
Figure 7.27: Sample values of CH (m) from Rh60, 3 baseline CHMs & Field Heights.	79
Figure 7.28: Distribution of CH amongst all models; Rh60-2019 & the 3 global CHM baselines... ..	80
Figure 7.29: Comparison of Field heights with predictions from baseline CHM models and newly created Rh60-2019.	80
Figure 7.30: Histogram of Field heights measured in 2019 (Vanniarachchy and Jayakody, 2020).. ..	81
Figure 7.31: Final Model Rh60-2019 Predicted Mangrove Height Map at 25 m resolution.....	82
Figure 7.32: Distribution of mangrove heights in VCS1764 using the best-performing model from Rh60-2019.....	82
Figure 12.1: Sample distributions of Rh metrics in model-2019.....	98
Figure 12.2: Sample distributions of (Rh metrics) ^{0.5} in model-2019.	99
Figure 12.3: Model-2019_s' performance across GEDI testing set.....	100
Figure 12.4: Model-2019_s' performance across Field data.....	101
Figure 12.5: Model 2019's performance across GEDI test data.....	102
Figure 12.6: Model-2019's performance across Field data.	103
Figure 12.7: Model-2020's performance across GEDI test data.	104
Figure 12.8: Model-2020's performance across Field data.	105
Figure 12.9: Visualisation of metric values across test data.	106
Figure 12.10: Visualisation of metric values across Field data.	106
Figure 12.11: Composite scores across test data.....	107
Figure 12.12: Composite scores across field data.....	108

List of Tables

Table 2.1: Summary of available GEDI products.....	35
Table 2.2: Summary of global attempts at producing Global Canopy Height Maps using Spaceborne lidar.	37
Table 5.1: Table with the description of each Sentinel band variable extracted from GEE.	53
Table 5.2: Summary of Data sources used.....	55
Table 6.1: Summary of the models and the shots used in them.....	58
Table 2: Summary of L2A beams with date of acquisition used in each model.....	109

List of Abbreviations

AGB: Above-Ground Biomass

ALS: Airborne Laser Scanning / Airborne Lidar Survey

CH: Canopy Height

CHM: Canopy Height Model

CNN: Convolutional Neural Network

ESA: European Space Agency

Field data: Data obtained from the Worldview Foundation (Vanniarachchy and Jayakody, 2020)

GBM: Gradient Boosting Machine

GEDI: Global Ecosystem Dynamics Investigation

GEDI L1A: Global Ecosystem Dynamics Investigation Level 1A Data

GEDI L1B: Global Ecosystem Dynamics Investigation Level 1B Data

ICESat: Ice, Cloud, and land Elevation Satellite

ICESat-GLAS: Ice, Cloud, and land Elevation Satellite - Geospatial Laser Altimeter System

InSAR: Interferometric Synthetic Aperture Radar

LAI: Leaf Area Index

Landsat-ARD: Landsat - Analysis Ready Data

Lang-2020: Lang et al., (2022)'s CHM Model

LiDAR: Light Detection and Ranging

Local data: Any non-global third-party datasets other than from the Worldview Foundation

(Vanniarachchy and Jayakody, 2020), e.g. papers published studying a specific region like Australia, Malaysia, etc.)

MAE: Mean Absolute Error

ME: Mean Error

MODIS: Moderate Resolution Imaging Spectroradiometer

NIR: Near-Infrared

Potapov-2019: Potapov et al., (2021)'s CHM Model

R: Pearson Correlation Coefficient

REDD+: Reducing Emissions from Deforestation and forest Degradation

RF: Random Forest Algorithm

Rh: Relative Height

Rh60: Relative Height at 60th Percentile

R²: Coefficient of Determination

RMSE: Root Mean Square Error

SAR: Synthetic Aperture Radar

Simard-2000: Simard et al., (2019)'s CHM model

SRTM: Shuttle Radar Topography Mission

TanDEM-X: TerraSAR-X Add-On for Digital Elevation Measurement

UAV: Unmanned Aerial Vehicle

UN: United Nations

1. Introduction

Mangroves are a unique type of plant cover that are a foundational species in both marine and above-ground ecosystems along the world's tropical and subtropical coastlines, estuaries, lagoons and rivers (Chatting et al., 2022). They have the highest total carbon density of any forest on Earth, 49–98% of which is stored in the first 3 m of soil (as opposed to 20% for typical forests) and have carbon burial rates per year 25 times that of tropical terrestrial forests (Stovall et al., 2021). Though they only cover 0.1% of the earth's continental surface (see Figure 1.1), they account for 11% of the total input of terrestrial carbon into the ocean (Giri et al., 2011; Jennerjahn and Ittekkot, 2002) and 3-4% of global carbon sequestration by the total tropical forest area (Alongi, 2014). Quite concerning, the rate of disappearance of mangroves worldwide is 1% to 2% a year from baseline, which could mean that 100% of mangrove forests could be lost in the next century (Polidoro et al., 2010) impacting immediate microclimates and worsening the impacts of global climate change. With regard to global climate change, increasing saltwater incursion from rising sea levels can have negative effects on mangrove health, which could cause negative feedback loops in ecosystems that are supported by it (Ahmed et al., 2022; Bhowmik et al., 2022; Song et al., 2011). However, the most salient threat to mangroves is direct human activity in the form of aquaculture and agriculture (Bhowmik et al., 2022).

According to Bhowmik et al. (2022), there has been an 8,600 km² decline in global mangrove coverage between 1990 and 2020, with the highest decline in South and Southeast Asia. According to Global Mangrove Watch, Indonesia faced the largest net change in mangrove habitat extent between 1996 and 2002, followed by Australia, Mexico and Myanmar. Myanmar has the highest annual rates (~2%) of mangrove deforestation in the world and may have the third-highest potential annual CO₂ emissions from mangrove deforestation (Friess et al., 2020). Due to deforestation, Myanmar also suffered a net loss of 2,397 million US\$/year in its mangrove ecosystem service value (i.e. 28.7% decrease from 2000), in which maintenance of fisheries nursery populations and habitat and coastal protection were among the services that were greatly affected (Estoque et al., 2018). The primary driver for the deterioration of mangroves has been shrimp farming, settlement areas and public constructions, which has caused fragmentation of the mangroves - a trend also seen in other places like China (Hai et al., 2022).



Figure 2.1: Global distribution of mangrove forests in 2000 (Giri C., 2011).

It is important to conserve and restore mangrove projects, as they have been shown to have a role in global climate regulation (Estoque et al., 2018), have positive social benefits on the community (Su and Gastaparatos, 2023), have been shown to store 5 times the carbon of rainforests in soil (Aye et al., 2022). This has been recognized and considered by the United Nations (UN), in the form of a declaration as the UN Decade on Ecosystem Restoration (2021 – 2030), “increasing the need for efforts to monitor and enable mangrove restoration” (Su and Gastaparatos, 2023). Initiatives to restore mangrove ecosystems worldwide are fortunately underway. The UN Framework Convention on Climate Change (UNFCCC) introduced a framework in 2007 called Reducing Emissions from Deforestation and Forest Degradation (REDD) to enhance conservation and restoration (Mitchard et al., 2013). This framework evolved into REDD+, a voluntary climate change mitigation strategy playing a significant role in the global cap and trade carbon market. Key concepts within this framework include additionality, permanence, and leakage, which all rely on precise carbon stock estimations.

Though national-level funding constraints hindered the materialization of this market (Duchelle et al., 2018), the Verified Carbon Standard (VCS), an independent quality standard, has leveraged the developed framework to certify carbon emissions reductions in the Voluntary Carbon Markets (VCM). One notable conservation project under the VCS, VCS1764, has been implemented in the Northern part of Ayeyarwady Division of Myanmar, covering 2265.47 ha of degraded lands (see Figure 1.1). This project, which falls under the Afforestation, Reforestation, and Revegetation (ARR) segment of the VCS, showcases private efforts to mitigate environmental impacts through transactions involving carbon credits awarded based on sequestration. Administered by Verra, founded in 2005, the VCS has become the world's most widely used carbon crediting program that directs finance towards projects that reduce emissions, enhance livelihoods, and protect nature (Verra, 2023).



Figure 1.1: Location of VCS1764 in Myanmar.

To continue receiving funding for the carbon sequestered by VCS1764, regular measurement and verification of the health of mangroves in terms of Above Ground Biomass (AGB) is necessary. Measuring AGB takes the form of mangrove height and diameter at breast height (DBH) and requires an understanding of ecosystem structure and change, both horizontally (land cover and land cover change) and vertically (forest height and biomass) (Campbell et al., 2022). Traditionally, low-tech and highly unreliable or unscalable methods have been employed for mangrove monitoring, lacking the full capabilities of current remote sensing data and tools. An investigation in 2021 by The Guardian revealed that although many forest projects were doing valuable conservation work, the credits that they generated appeared to be inconsistent with previous levels of deforestation in the areas, “*with room to generate credits that had no impact on climate whatsoever*” (Greenfield, 2023). Many projects additionally award credits to places with reducing forest cover to help the community dependent on the additional income from carbon credits – sometimes losing the main plot of carbon sequestration (Ayrey, 2022). This erosion of trust due to inaccurate/non-existent AGB estimates in the carbon markets could have detrimental consequences on climate mitigation efforts by reducing finance towards quality carbon sequestering projects in the long term (Macquarie, 2022).

1.1. Motivation and Research Purpose

While the horizontal structure of mangroves has been thoroughly investigated through the extensive utilization of multispectral and radiometric satellite data, the understanding of vertical mangrove structure in terms of Canopy Height (CH), essential for accurate growth and AGB estimation remains relatively uncharted (Campbell et al., 2022). This knowledge gap is largely attributed to the challenges associated with geographically scaling vertical geospatial data sources across various forest types.

A study into this topic aligns with the Geospatial CDT research theme of Big Data and Spatial Analytics as a raster file will be made publicly available with predicted heights in the VCS1764 region, contributing to the Spatial Analysis & Modelling, and Visualisation and Decision Support theme. These themes collectively contribute to a broader understanding that facilitates conservation, afforestation, and other ecologically sustainable practices.

The research conducted also falls within the Building a Green Future, and Energy and Decarbonization priorities of the Engineering and Physical Sciences Research Council (EPSRC). Additionally, it aligns with the global objectives outlined in the Sustainable Development Goals (SDG), particularly those related to Goal 13 - Climate Action, Goal 14 - Life on Land and Goal 15 - Life below water. These focus areas highlight the interdisciplinary nature of this work and underscore its relevance in contributing to a sustainable and technologically advanced future, consistent with international efforts to foster responsible growth and environmental stewardship.

2. Literature Review

2.1. Introduction to mangrove monitoring

Mangrove monitoring initially relied upon the extrapolation of field data, environmental conditions and partial extent maps. Measuring can take the form of investigating vegetation structure metrics such as CH and leaf area index (LAI), which are crucial for understanding environmental shifts and biodiversity (Tang and Armston, 2019). Both passive and active remote sensing technologies have been used to measure mangrove extent on mainly local and since 2011, global scales. L-Band SAR (Synthetic Aperture Radar), which uses radar returns to map the earth's surface, and Landsat have been used by the Global Mangrove Watch to provide a historical global mangrove extent change map, the first baseline for the year 2010 (Bunting et al., 2016), and subsequent ones that mapped changes in Global Mangrove extent since, one from 1996-2016 (Bunting et al., 2019) and the other from 1996 -2020 (Bunting et al., 2022). Multispectral instruments like the Moderate Resolution Imaging Spectrometer (MODIS) and Landsat have been also used to provide valuable data on the horizontal structure of mangroves since the 1980s. For example, the first attempt that made use of Landsat imagery to create a global mangrove coverage map for the year 2000 was made by Giri et al., (2011).

While it is mostly passive optical platforms that have been used to map mangrove extent, change and extrapolate carbon storage with field data, they face three principle challenges; a) spectral signal saturation in dense forests, b) lack of vertical canopy structure measurements and, c) have inconsistencies between sensors (Campbell et al., 2022).

To address these limitations, researchers have begun to adopt active sensors such as Airborne LiDAR Scanning (ALS), SAR and Interferometric Synthetic Aperture Radar (InSAR). These sensors have demonstrated their effectiveness in measuring structural attributes of mangroves, such as height, thus complementing the initial extent maps. This approach was first showcased by Simard et al., (2006), who utilized radar data from the Shuttle Radar Topographic Mission (SRTM), harmonized with ALS data, to create a detailed mangrove Canopy Height Map (CHM) (Campbell et al., 2022). The SRTM radar, Remarkably, Simard et al., (2006) achieved a height error of 2m over a 30m pixel resolution in the Everglades National Park, USA. To enhance the precision of CH

explorations, SAR data is often paired with ALS data—the latter which is known for its proficiency in mapping the vertical structure of vegetation.

The lidar technology used in ALS is a method that involves recording the time it takes for infrared light to return from a surface. In the case of full waveform lidar, this time is then translated into a Relative height metric (Rh) that ranges from 1-100 representing the percentage energy return. For example, an Rh of 50%, Rh50 represents the height at which 50% of the lidar energy returns to the sensor. This instrument can be on board Unmanned Ariel Vehicles (UAVs) like drones, Airplanes, and satellites. UAV or ALS is the most accurate instrument for measuring mangrove height, providing estimates correct to within 2 cm vertically and 1 cm horizontally (Ganz et al., 2019). A study by Q. Li et al., (2023) combined multi-temporal ALS and Sentinel-2 to assess the disturbances and recovery of mangrove forests in China after a typhoon and insect damage. CHMs made using Lidar were assessed in the process, and crucial insights into how different species of mangroves had differing layers of canopy structure i.e. single, double or multi-storeyed that could be differentiated (Q. Li et al., 2023).

A notable study that brings all the different ways of obtaining Mangrove CHM discussed so far was carried out by Lagomasino et al., (2016) when conducting a study into the mangroves of the Zambezi River Delta. The relative accuracies of Very-High Resolution (VHR) stereophotogrammetry (3D information extracted from photographs), TerraSAR-X add-on for Digital Elevation Measurement, and Shuttle Radar Topography Mission (TanDEM-X), and ALS were compared with each other (Lagomasino et al., 2016). Canopies over 10m were found to be accurately predicted by all CHMs. Notably, mean CHs were not statistically different between field measured (10.1 m), lidar (10.76 m) and VHR (10.95 m) technologies. It is interesting to note however, that ALS (15.26 m), which used Rh100 had the highest deviation of - 4.1 m from field data (10.1 m), followed by VHR (12.26 m), Tandem-X (11.67 m) and SRTM (10.72 m). This comparison is revealing because AGB estimates need good height accuracies to be true representations of carbon storage.

Studies show that ALS (Ganz et al., 2019; Guerra-Hernández and Pascual, 2021), and UAVSAR (Unmanned Ariel Vehicle SAR) (Stovall et al., 2021) have the lowest calibration error in both mangrove and forest CH measurements. However, the main limitation of a lot of these accurate methods is cost; ALS and SAR start at about 100 dollars per 1km² which increases depending on the resolution (Donato et al., 2011; Hopkins, 2014). Additionally, field surveys like lidar that

document forest carbon stocks, deforestation, and forest degradation can be costly, time-consuming, and dangerous, increasing the impetus to find methods that can geographically scale through cheap remote sensing (Lagomasino et al., 2016). Whereas there is abundant literature available on the usage of ALS and SAR to measure mangrove heights in localised areas (Ganz et al., 2019; Woodroffe et al., 2020), the application of lidar to produce global datasets was unexplored till 2003 with the launch of ICESat (Ice, Cloud and Land Elevation Satellite)– a Spaceborne Lidar instrument.

2.2. Use of Space-borne lidar in CH measurements

2.2.1. ICESat

ICESat-1 - GLAS – Geoscience Lidar Altimetry System) was the first attempt at spaceborne lidar technology designed to measure ice-sheet mass balance, cloud and aerosol heights as well as land topography and vegetation characteristics. It utilizes photon-counting Lidar, a technology that counts individual photons returning to the sensor which is particularly sensitive to changes in elevation and can be used to detect subtle variations in CH. The sampling of ICESat-2 is non-continuous – it just had a single laser beam with a footprint diameter of 60 m and a spacing of 172 m between the footprints (Li et al., 2020).

Though mainly designed for monitoring in the polar regions, its use has been demonstrated in mapping vertical forest structure (Duncanson et al., 2020). The Landsat mangrove coverage map by Giri et al. (2011) described in Section 2.1 was used in conjunction with ICESat to produce a *Global Forest Canopy Height Map* (CHM) by Simard et al., (2011). ICESat-2/ATLAS (Advanced Topographic Laser Altimeter System) was later used by the same author to produce a *Mangrove Canopy Height Map* for the year 2000 in conjunction with SRTM (Simard et al., 2019), to become a widely used map in many scientific publications as a baseline (Lang et al., 2022a). SRTM and Landsat were the contiguous datasets that were used to overcome the small sampling limitations of lidar. This process, known as *data fusion* involves first integrating different spatial and spectral resolution data sources to produce fused data that contains more detailed information than each of the sources (Zhang, 2010). Data fusion is often implemented when using lidar because of sampling limitations and the expense required to get wall-to-wall coverage. This is followed by regression of the lidar and the contiguous dataset, like Landsat to produce a contiguous lidar dataset.

Apart from Simard. et al. (2019)'s global attempts at producing CHM, several local studies too have explored the use of ICESat in measuring CHs through fusion with other contiguous data sources. For example, Li et al., (2020) used ICESat-2 Lidar in conjunction with Sentinel-1, Sentinel-2 and Landsat-8 data to map CHs in China, reporting a mean bias of -1.46m (Li et al., 2020). A very significant and relevant recent study in 2021 assessed the use of ICESat-2 Ground elevation and CH Estimates in mangroves in Northern Australia using ALS as validation data. The heights of the mangroves that were measured ranged from 25 m in dense areas to 4m in land and marine margins. It found that ICESat-2 Rh98 at about a 20 m resolution overestimated mangrove heights by 0.21 m (mean error) and had a root mean square error of 2.5 m, in addition to night time beams having a better approximation for mangrove heights due to low solar interference (Yu et al., 2022). Yu et al., (2022) found that ICESat-2 had good performance in measuring terrain elevation even in intertidal zones flooded by periodic seawater. However, the mangrove heights in this study were mostly above the 2m range for intact mangroves in a national park- thus with fundamentally different characteristics to shrubs (< 3 m) / restored mangroves (Simard et al., 2019).

The main limitation of the ICESat missions is their limited coverage as their orbits were primarily designed to monitor ice sheets over the poles and are not optimized for forests (Dubayah, Ralph et al., 2021). ICESat-2 GLAS is also a photon counting lidar instrument – and this approach can omit crucial information at times, such as the full vertical structure of the canopy, because it relies on individual photon detection rather than capturing the entire waveform of the reflected signal. Previous attempts at launching lidar satellites specially designed for forest monitoring like Vegetation Canopy LiDAR (VCL) and Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI) failed to launch due to technical or budgetary issues (Tang and Armston, 2019). The latest response in the scientific community to fill the gap in global ecosystems has been to design and launch a new Spaceborne Satellite – Global Ecosystem Dynamics Investigation (GEDI) to address some of ICESat-2's challenges.

2.2.2. GEDI

GEDI is a relatively new addition to the space-borne lidar satellites onboard the International Space Station (ISS) – and was launched on 5 December 2018 on board a Falcon 9. After three months, GEDI started collecting data for scientific use on March 25, 2019 and was poised to end in 2 years. However, the mission was extended until 17th March 2023 and will resume operation in 2024 after

a mission deployed by the Department of Defence payload ends. It was especially designed for applications in forest management and carbon recycling, water resources, weather prediction and topography and surface deformation carrying out high-resolution laser ranging of earth's forests and topography (Potapov et al., 2021). Two of the lasers are full power, and the other is split into two beams, producing a total of four beams. Footprints are separated by 60 m along track and 600 m across track, as shown in Figure 2.1 (NASA, 2023a).

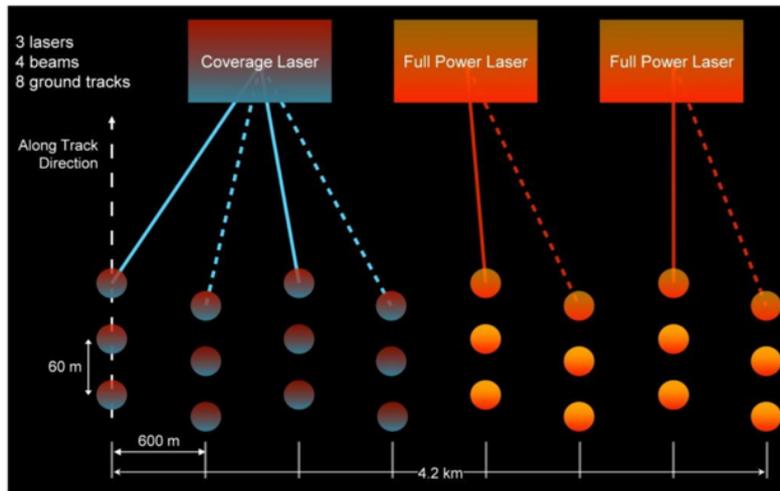


Figure 2.1: GEDI's sampling methodology (NASA, 2020).

GEDI samples only 4% of the entire earth, which could potentially lead to inaccurate estimations of vital parameters such as AGB or CH, due to heterogeneity of forests. Therefore, like ICESat, GEDI too needs data fusion with contiguous data sets. GEDI's data products are summarised in Table 2.2. Level 1B, the Geolocated waveform contains the raw lidar energy returns, which are then analysed to yield subsequent Level datasets. The L2A products (Elevation and Height metrics) is what is commonly used in most studies.

Table 2.1: Summary of available GEDI products.

Data Level	Data Product
Level 1B (L1B)	Geolocated waveform
Level 2A (L2A)	Elevation and height metrics
Level 2B (L2B)	Canopy cover and vertical profile metrics
Level 3 (L3)	Gridded land surface metrics
Level 4A (L4A)	Footprint level aboveground biomass density
Level 4B (L4B)	Gridded aboveground biomass density

Whilst literature on GEDI's use to detect both global and local forest CH is well researched and studied, at the time of writing, literature that uses GEDI use to exclusively study mangrove cover is relatively scarce. From the available scholarly data, two studies stand out; the first of which is an unpublished study that mapped 30 years of mangrove forest height in South Florida by fusion of ICESat-2, GEDI and Landsat data (Xiong et al., 2022).

The second study by Stovall et al., (2021) compared height metrics across 17 different sensor types (including TanDEM-X, GEDI and ICESat-2) over relatively tall (> 2 m) mangrove forests in Gabon, Africa (see Section 2.1, Stovall et al., (2021)). Uncertainty in height metrics increased with forest height across the different sensors and didn't correspond with each other. Additionally, Radar estimates were found to have the lowest calibration error and bias with further improvements using Lidar Fusion. Relative biomass variation was lowest in low-stature mangrove forests. The relative uncertainty across height products was highest in low stature stands (>50% in stands <15m), while in the tallest stands (~45m Hmean) this uncertainty was ~20% or 7-8 m (Stovall et al., 2021). This also translated into AGB estimates; relative biomass variation (a derivative of CH) was highest in low-stature strands (those below. 15m) and locally-calibrated area-wide totals are more representative than generalized global biomass models for high-precision biomass estimates (Stovall et al., 2021). However, a key limitation in this study is that RH100 was used to provide estimates of heights – something that many other studies evaluated against using because of its reported overestimation bias. For example, Potapov et al., (2021) used Rh95 after comparing Rh90 and Rh100 on his data.

2.3. A Summary of Global CHMs

In the subsequent section, now that SRTM, GEDI and ICESat have been introduced, it would be useful to explore the specific methods, data sets and drawbacks that attempts at using these datasets to make Global CHMs bring. The CHMs that will be delineated in this section are summarised in Table 2.2.

Table 2.2: Summary of global attempts at producing Global Canopy Height Maps using Spaceborne lidar.

Data set	Year	Datasets used	Resolution	Spatial Coverage	RMSE	Source	Relative Height Metric(Rh)*
Global <i>Forest</i> CHM	2005	ICESat, MODIS, elevation from SRTM	1 km	Global Coverage	6.1m	Simard et al., (2011)	ICESat Rh100
Global <i>Mangrove</i> CHM	2000	SRTM DEM data & ICESat/ C-Band SAR Interferometry	30 m	-180-180 Latitude 31 - -39.0 Longitude	6.31 m	Simard et al., (2019)	ICESat Rh95*
Global <i>Forest</i> CHM	2019	GEDI & Landsat-ARD	30 m	-180-180 Latitude 50 - -50 Longitude	6.6 m	Potapov et al., (2021)	GEDI Rh95
Global <i>Forest</i> CHM	2020	GEDI & Sentinel-2	10 m	Global coverage	6.6 m	Lang et al., (2022)	GEDI Rh98

*This was formerly Rh100 which was revised to Rh95 in the latest version for better estimates.

*Rh is the height below which the said percentage of energy is returned. i.e. Rh98, Height of 98% of energy return.

Simard et al., (2019)

Simard et al., (2019), as first introduced earlier in Section 2.2 used Primarily C-Band SAR interferometry (SRTM) corrected using ICESat-2 to produce a 30m resolution Global Mangrove CHM. The SRTM mission sampled using radar 60°N and 56°S (about 80% of the earth) to create a Digital Elevation Model (DEM) of the surface of the earth for the year 2000, which was later revised to provide global coverage. However, in forested areas, C-Band (5 cm) has been noted not to fully penetrate the canopy. ICESat, though had limited coverage does fully penetrate the canopy to reach the ground, and the ICESat-1/GLAS derived Rh100, later revised to Rh95 was therefore used to correct the SRTM DEM to represent true mangrove heights (by multiplying by a factor of 1.697). The DEM was spatially constrained to only mangroves using the Global mangrove extent map published by Giri et al., (2011). Regions with SRTM elevation of 0 m but mapped as mangroves in Giri's map were assigned heights of 0.5 m as they were classified as shrubs after

being validated using field data. Validation with field data demonstrated the data had an RMSE of 6.31m, that reduced to 3m if the basal area calculated height (the height value weighted by the basal area of trees) was used instead.

The main limitation of this study is that only 5 countries out of the 115 used to validate the data had mangroves less than 3m – and the mangroves used to calibrate Myanmar's maps had average heights of 5 m. It was explicitly designed to measure taller canopies rather than shorter ones for the purpose of conservation because of an assumption that taller forests contain more AGB.

Potapov et al., (2021)

Potapov et al., (2021) used Landsat imagery at a 30m resolution to train a regression tree to predict GEDI Level 2A, 95th percentile Rh to produce a Global CHM for the year 2019. Landsat's 16-day interval spectral data was employed, and linear regression and data from up to four previous years to fill gaps caused by cloud cover or other inconsistencies were applied. This ensured the spatiotemporal uniformity of the dataset, facilitating the extraction of key metrics like minimum, maximum, and median reflectance values, which are essential for long-term forest monitoring. The GEDI pre-processing steps included filtration by a beam sensitivity of 0.9m, using only full power beam mode, and only collected at night, where the range of the predicted ground elevations was less than 2m. All these measures optimise the Signal to Noise Ratio of the beam. ALS surveys in the United States, DRC Australia and Mexico were used to select the GEDI Rh metric that best corresponded with the field data to be chosen as the predictor variable. The Rh95 was chosen from a selection of Rh 75, 90, 95 and Rh100 to be the best approximate of field heights using an evaluation of the R² (Coefficient of determination), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean difference. Values of GEDI Rh 95 below 3 m were assigned to 0 to be consistent with the standard definition of a tree (Potapov et al., 2021). The authors emphasize that utilizing texture metrics from higher-resolution Sentinel-2 data could potentially enhance the performance in mapping taller canopies, consequently impacting the accuracy of mapping forests with shorter heights (3-5 m).

Lang et al., (2022)

Lang et al., (2022) employed an ensemble of deep Convolutional Neural Networks (CNNs) to transform Sentinel-2 optical images into 10-meter resolution CH predictions. The model was trained using pixels that intersected with GEDI's L2A Rh98th data. The mangroves detected in this study had mean heights of 5m. Unfortunately, like Simard et al., (2019) and Potapov et al., (2021), Lang et al., (2022) too focused on the underestimation problem of tall canopies faced by existing

methods, and implemented methods in the CNN that optimised the predictions to counter the bias from the spatial abundance of short mangroves – like setting parameter weights to inverse of sample frequency, thereby overestimating heights of short canopies.

This is obviously a problem – some mangroves though shrubs (< 3 m tall) have 5 times the same carbon as land-based tropical forests (Donato et al., 2011) and therefore more AGB than these studies that were optimised to detect tall forests give them credit for. Additionally, Potapov et al., (2021) for instance did not process a small portion of land cover within a humid tropical forest (location unspecified). This implies that this map might have either not processed or had a high degree of uncertainty when dealing with humidity, and could indicate a decrease in the reliability of images in the same condition extensively present in coastal mangrove regions.

Overall limitation

It is important to note that many of these studies lose accuracy at local scales and have relatively high RMSEs (over 6 m) which could be significant for regions with vegetation cover that is non-standard and at the tails of allometric parameters such as height. Mangroves can be considered to be such – especially young ones that are scarce in terms of area coverage that might not have provided enough coverage data and even in situ measurements on which many of these models are used to be calibrated. For example, the underlying distribution of mangrove heights studied in Lang et al.'s (2022) study had a mean height of 10m – whereas mean heights in VCS1764 are in the range of 0.8 m – 1.5m. Lang et al., (2022) noted that while his model underestimated heights for tropical grasslands and tropical dry broadleaf forests, it overestimated them for flooded grasslands, and severely overestimated (≈ 2.5 m) for *mangroves, tundra and tropical coniferous forests* (Simard et al., 2019). Mitchard et al., (2013) have separately noted that there are substantial differences among AGB maps with little consistency throughout geographical areas, with errors typically too great to be useful for any practical purpose.

The relatively high RMSEs make it hard to use these findings for local monitoring purposes, and certainly not for REDD+ purposes where carbon credits are estimated per tonne as estimated heights could be 2-3 times their in-situ measurements. The translation between height and biomass is not a linear one, so even if heights are doubled, there is a chance that AGB could be underestimated as mangroves store 5 times more carbon than typical forests (Donato et al., 2011; Stovall et al., 2021). Local CHMs and AGB maps could provide a better estimate for mangrove heights than global ones– however, none were found in Myanmar. However, local studies in

VCS1764 quantifying sample AGB, tree heights and species have been found, which could act as reference points for further study.

2.4. Mangrove monitoring in VCS1764 - Ayeyarwady, Myanmar

Among three main tracts of mangroves in Myanmar—Rakhine coastline, Ayeyarwady delta, and Tanintharyi coastline—the mangroves in the Tanintharyi coastline have now turned into the largest areal extent even though the Ayeyarwady delta had the largest in the past (Aung, 2022). The Ayeyarwady Delta mangroves shrank by 64.2% between 1978 and 2011 to just 92,800 ha and have the least species diversity out of all three coasts. This region faced the deadly impacts of Cyclone Nargis in 2008 when the loss of 75% of the original mangrove cover in the delta turned a natural hazard into a major disaster (UNEP, 2009). 84,537 people were killed in the delta and 1.5 million more were severely affected (Reuters, 2008). In addition, 38,000 ha of natural and replanted mangroves were destroyed and 63% of paddy fields and 43 per cent of freshwater ponds were submerged- severely impairing resiliency and recovery efforts post-disaster (UNEP, 2009).

Efforts in the community to understand and thereby lead to the protection of this essential ecosystem through mapping and study began in 1986, when Aung Myint and Kyaw studied the species of mangrove plants of *Knadone-ka-ni* and *Main-ma-hla* in the Ayeyarwady region. Subsequent studies focused on delineating the species, regeneration yield, photosynthetic biomass and commercial aspects of mangroves, both in Ayeyarwady and in Myanmar in general (Oo, 2020).

The most relevant and recent study by Oo, et al., (2020) assessed the photosynthetic biomass of mangrove species in VCS1764, Ayeyarwadi. This study calculated AGB data including tree height, stem diameter, canopy size and leaf weight of selected mangrove species. The study mapped differences in biomass accumulation with age in species in two distinct areas; restored and natural mangrove areas. From the available data, restored mangrove heights ranged from 0.59 to 1.8 m in measurements made from 2015 to 2019. Height classes of natural mangrove ranged from 1-2ft (0.45 m), 3-4 ft (1.06 m) and 5-6 ft (1.67 m) in the Magyi Tidal Creek (the southern section of VCS1764, Figure 2.2) (Oo, 2020). However, it is important to note that this study sampled only 10 restored mangroves for each species, introducing high uncertainties in the measurements.

VCS1765 Project Location map, Magyi, Thaegone & Thabawkan in the Ayeyarwadi Region

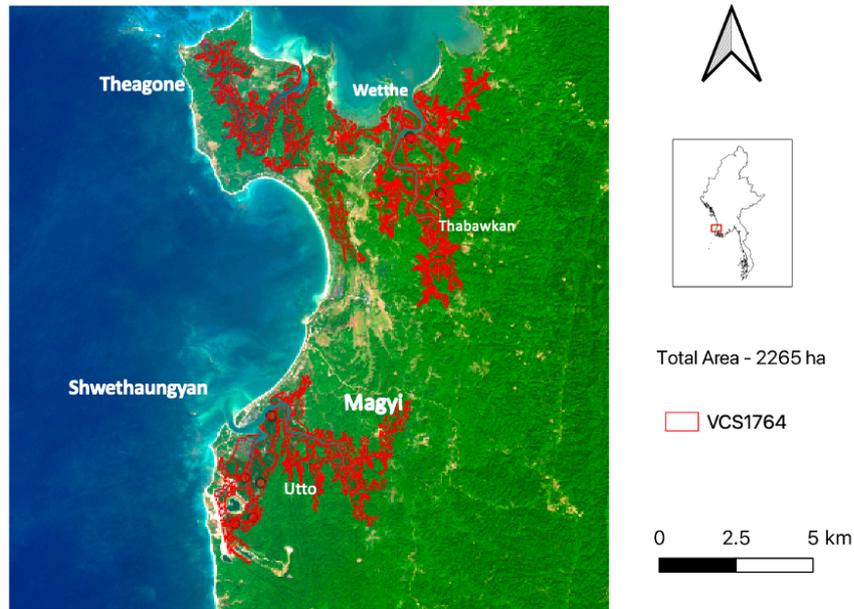


Figure 2.2: VCS1764 Project location in Myanmar.

Another recent study published by Win et al., (2021) investigated the vegetative structure and species distribution of *intact mangroves* in the Southern part of VCS1764 in 2018. The study area in Figure 2.2 shows the location of the project area. Mean tree height in the Utto tidal creek was found to be between 2.2 – 2.5 m, and 1.9 – 2.6 m in the Magyi tidal creek (Win, 2021). A main limitation of this study was that the demarcation between the restored and the natural mangroves in these studies was not clear. It could at best be inferred from the time series analysis with a degree of uncertainty around the demarcation between the intact and restored mangrove using Google Earth Pro in some areas as seen in Figure 2.3.



Figure 2.3: Visible degradation from 1985 imagery in 2010 and restoration in 2020 in the southern part of the site on Google Earth in 2010.

2.5. Research Gap

Lang et al., (2022), seen in Section 2.3 attempted to create a global forest CHM using Sentinel-2 and GEDI. However, its fusion with GEDI for more accurate mangrove heights is relatively unexplored. There have been studies that use GEDI to map Global CH or other forest height (Adrah et al., 2021; Guerra-Hernández and Pascual, 2021; Lang et al., 2022b) and even savannah (X. Li et al., 2023) but there haven't been as many that target mangrove which could open up a new avenue of mangrove structure exploration. Whilst some local studies explored in Section 2.4 capture small-scale species-dependent information and field heights of restored mangroves, they do not cover the whole project area. At the time of writing, though a locally calibrated mangrove extent map is available for Myanmar, Thailand and Cambodia (Baltezar et al., 2023), no locally calibrated Mangrove CHM for Myanmar or VCS1764, normally made using ALS was found.

Since the launch of Sentinel-1 and Sentinel-2, there has been increased spatial resolution of new mangrove maps from 30 to 10 and 20 meters (Campbell et al., 2022). According to Li et al.2020, who compared to Sentinel-1, Sentinel-2 and Landsat data, Landsat-8 showed relatively weaker performance in CH prediction, suggesting that the addition of the backscattering coefficients from

Sentinel-1 and the red-edge related variables from Sentinel-2 could positively contribute to the prediction of forest CH (Li et al., 2020). Additionally, Sentinel offers better resolutions than Landsat (10m vs. 30m), which could be used as the contiguous variable to train the GEDI shots.

Therefore, the objective of this study will be to explore the use of spaceborne lidar - GEDI in monitoring mangrove heights for the localised study site in Myanmar using Sentinel-2 imagery and compare the estimates with Global CHMs.

3. Aims and Objectives

3.1. Research Aim

To fuse GEDI and Sentinel-2 to estimate mangrove canopy heights in a restored area (VCS1764) in Myanmar.

3.2. Research Objectives

- 1) Explore the available GEDI products over mangroves in VCS1764 in Myanmar.
- 2) Fuse GEDI Level 2A Canopy Height (Rh) metrics with Sentinel-2 and use machine learning to create a mangrove Canopy Height Model using field heights as a reference.
- 3) To compare the newly created Mangrove Canopy Height Model with other available global and local datasets.

4. Research Design Methodology

R was the primary Program for handling and training the .tiff files (Sentinel-2) and the shapefiles (GEDI02_A v002) using the Random Forest Model (RF). Python was used to manipulate the GEDI L1B (GEDI01_B v002) waveform and L2B (GEDI02_B_v002) derived plant metrics to visualise them. QGIS was used for most of the maps, while Excel was used to extract the Field data of VCS1764 that was obtained from Vanniarachchy and Jayakody, (2020). All these are open-source software with easily replicable code. All code used in this study will be available on GitHub and fulfil the metadata principles outlined in the Data Management Plan (See Appendix 12.6).

4.1. Project site

The World View International Foundation (Worldview), a Norwegian-based organisation specialising in blue carbon capture set up VCS1764 which is to last from 15th June 2015 to 14th June 2035. The primary goal of the project is to restore and protect 2,065.87 hectares of degraded mangrove forests. Over this 20-year crediting period, the restored mangroves of the project contribute to the GHG emission reductions and removals, which is estimated as 3.6 million tonnes of CO₂e in total amounting to the annual emissions of 0.62 million people in a year in the UK. The mangrove restoration and livelihood improvement activities of VCS1764 are implemented in the three village tracts of Magyi, Thabawkan, and Thaegone in Ayeyarwady Region by 2 tidal creeks, with a total area of 2265 ha. The area is divided into plots with 4 different planting start dates – from 2015, 2016, 2017 and 2018 as shown in Figure 4.1.

Location of Sample Field plots by year of planting in VCS1764, Ayeyarwadi Region

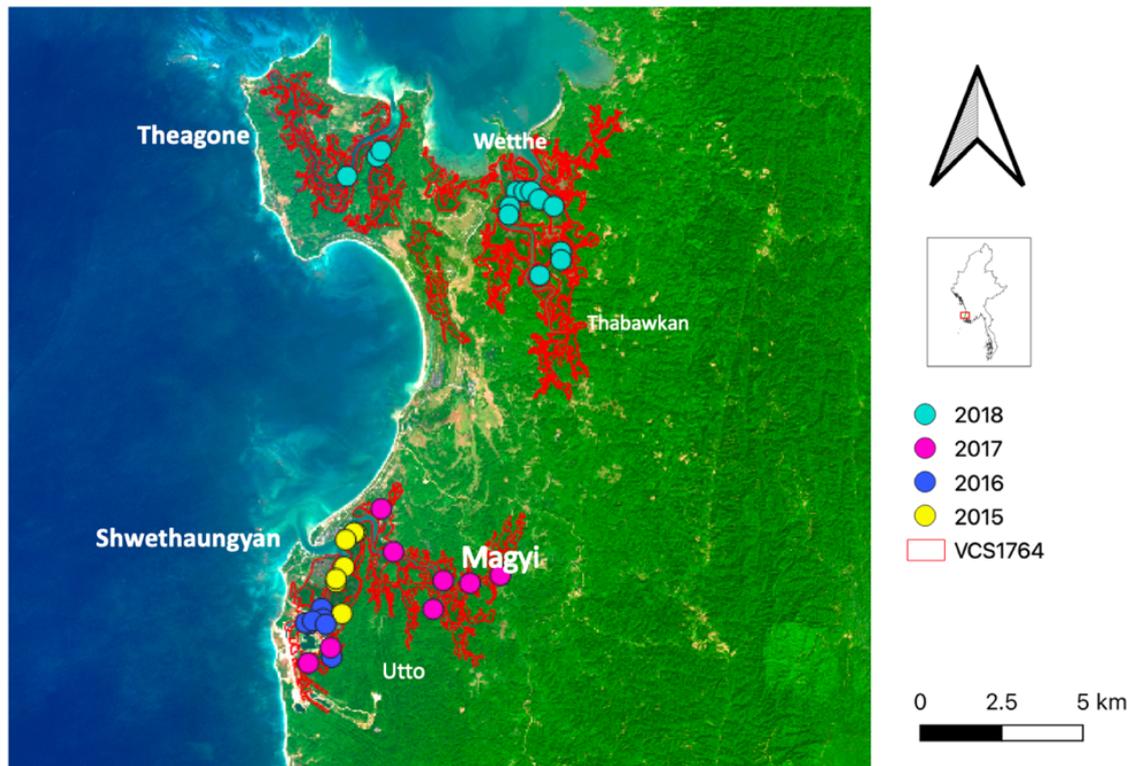


Figure 4.1: Location of sample field plots by year of planting VCS1764 Project location in Myanmar.

Geomorphically, the essential characteristics of a tidal creek are that they are relatively long and narrow, are shallow, and exhibit tidal water level fluctuations and weak tidal currents (Healy, 2005). Like the broader Andaman Sea, the Ayeyarwady region primarily experiences semi-diurnal tides, which means there are two high tides and two low tides each day.

The most significant influence on the Andaman Sea's water levels is the monsoon. Myanmar experiences a tropical monsoon climate with three main seasons (World Bank, 2023).

- 1) Hot Dry Inter monsoonal (February to May)
- 2) Rainy Southwest monsoon (May to November): July and September hold the peak flows from heavy monsoon rains.
- 3) Cool Relatively Dry Northeast monsoon (November to February).

This is important to keep in mind, as Spaceborne lidar techniques involve detection of the ground signal at varied times of the day which could be affected by these seasonal and diurnal tide fluctuations in tide levels.

5. Data and Pre-processing

5.1. Field data

Field data, in terms of Diameter at Breast Height (DBH) and height from 32 permanent sample field plots of 10 m by 10 m is available on the project website. The coordinates of the centre of the plot were recorded and provided as .kml files (Verra, 2023).

Measurements of both are available for the years 2018 and 2019. The heights recorded in May 2019 were used in this study to be comparable to GEDI, as the latter started data collection in 2019. The average height increment measured in a year was about 0.24 m. Site photographs were provided by the Worldview (Vanniarachchy and Jayakody, 2020) at unspecified locations, are shown in Figure 5.1.



Figure 5.1: Field photographs of mangroves at an unspecified location between 2019-2023 in VCS1764 (e) The height of a mangrove looks about $\frac{3}{4}$ of the average height (160cm) of a person— amounting to 120 cm. (Vanniarachchy and Jayakody, 2020).

5.2. GEDI

GEDI has products ranging from L1B-L4B , L1B being the full geolocated full waveform Lidar (Figure 5.2) and subsequent products being derivatives of the first L1B product (see Table 2.1), relating to CH, AGB and even elevation metrics from other satellites such as TanDEM-X (Dubayah, Ralph et al., 2021) as summarised in Table 2.1.

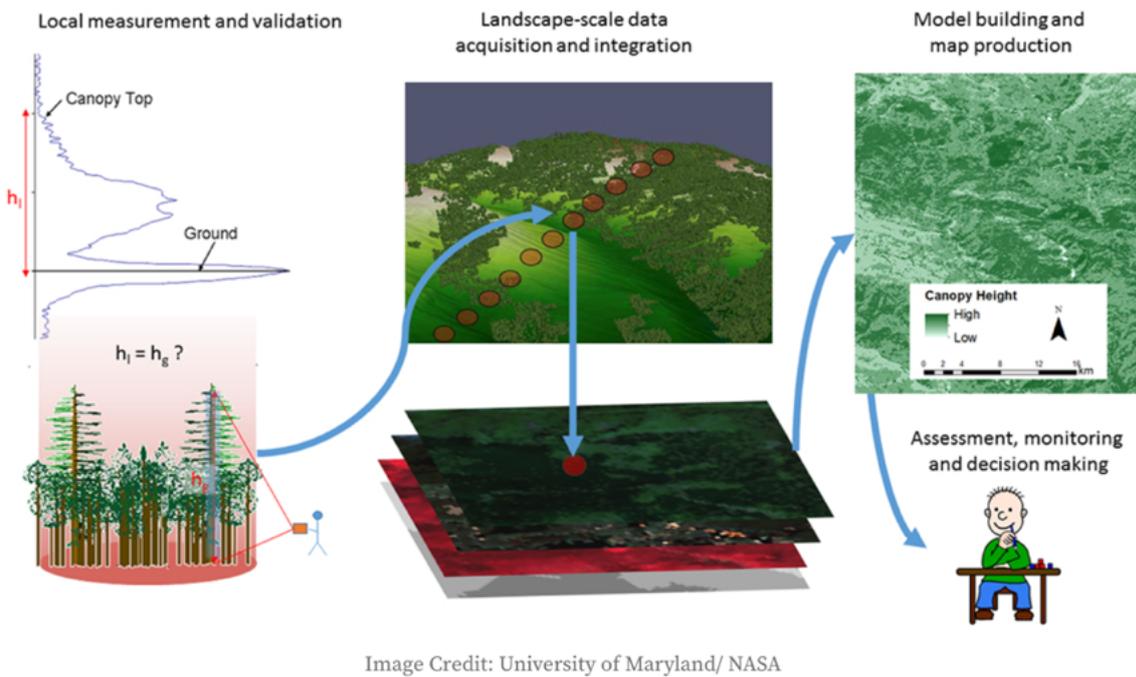


Figure 5.2: The image acquisition process of GEDI (NASA, 2021).

5.2.1. GEDI L1B, L2A & L2B (Objective 1: Explore GEDI products over VS1764)

In this study, the focus started with GEDI's L1B full waveform data which was obtained in .h5 format from NASA's Earthdata search platform (NASA, 2023a). These files targeted a bounding box containing VCS1764, located at WGS84 Lat 17.21311°, Long 94.58789°. The aim was to delve into the foundational assumptions that inform the more detailed GEDI L2A and L2B datasets. The L2A Relative Height metrics were also obtained in the same way in a .h5 format to plot a few visualisations of Rh distributions along the canopy (Section 7.1.2). From the GEDI L2B dataset, the Plant Area Volume Density (PAVD) feature was examined, as this metric provides critical insights into the vertical structure of vegetation and aids in refining the understanding of canopy complexity.

5.2.1. GEDI L2A (Objective 2: Create a Mangrove CHM)

Subsequently, the GEDI L2A (Version 2) Elevation and height metrics, sourced as .shp files through Google Earth Engine (GEE) (note that the .h5 format wasn't used though it is possible to) were utilized. This updated version offered improved horizontal accuracy estimated at 8 m and vertical accuracy of 10 cm, compared to Version 1's 20 m and 50 cm (Dhargay et al., 2022; Guerra-Hernández and Pascual, 2021). For VCS1764, access to a total of 8 transects for the year 2019 and 18 for the year 2020 was available. The GEDI L2A product is the primary focus of this study to make the machine learning model (Objective 2). This will provide the dependent variables to be predicted (Rh1-100), while Sentinel-2 will be the independent variables. A best-fitting Rh metric to the field data will then be chosen to produce the comparisons with the baseline as part of Objective 3.

GEDI L2A product comes with ways in which its data can be checked for accuracy and quality. A way of filtering out low-quality beams is by checking the GEDI-provided TanDEM X elevation, and GEDI's ground elevation (depending on the algorithm setting) and filtering out those that have a significant difference. According to a simulation study, GEDI's 10m horizontal geolocation error could cause up to 50% uncertainty of height estimates (Roy et al., 2021) and lead to spurious results at project boundaries, and therefore studies in the past have filtered for those within 40 m of the study site (Stovall et al., 2021). However, due to the limited number of shots available in 2019, these filtering procedures possible in larger geographical study areas (Guerra-Hernández and Pascual, 2021; Luo et al., 2023; Stovall et al., 2021) could not be carried out. However, the procedures that were are;

The GEDI shots used in model-2019's are shown in Figure 5.3 for demonstrative purposes.

- a) Quality Flag: A quality_flag value, labelled as /{Beam number}/l2a_quality_flag in the L2A product indicates whether a beam meets certain criteria based on energy, sensitivity, amplitude, real-time surface tracking quality, and difference to a DEM (Tang and Armston, 2019). The quality flags of 0 indicate that the shot wasn't used.

Map of GEDI shots with Quality flag over VCS1764 used in Model-2019

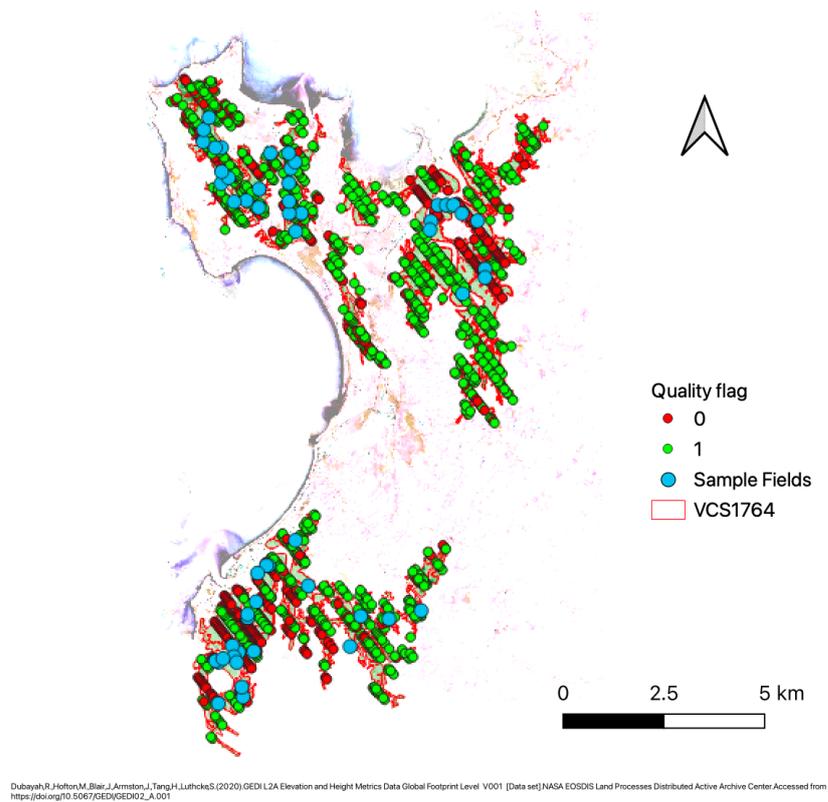


Figure 5.3: Map of GEDI shots with Quality flag 1 and 0 passing over site in 2019.

- b) **Beam Sensitivity:** This is related to the signal-to-noise ratio (SNR) of the laser return and was labelled as $\frac{\text{Beam number}}{\text{sensitivity}}$. A setting of $>95\%$ was used to ensure the signal is not affected by noise or other factors that reduce accuracy.
- c) **Anomaly removal:** These were identified using the Interquartile Range (IQR) method. Specifically, for each variable under consideration, the first (Q1) and third quartiles (Q3) were computed. The IQR, which is the difference between Q3 and Q1, was then used to define the bounds of typical data values. Any GEDI Rh metric lying beyond 1.5 times the IQR from Q1 or Q3 was deemed an outlier and was subsequently excluded from the dataset. This method ensured the removal of extreme values, which could potentially skew the model's predictions. Compared to using a filtration based on the standard deviation, IQR is more robust to skewed data, as it is based on medians and quartiles and doesn't assume the underlying distribution is normal. It also targets the tails of the distribution explicitly, while standard deviation is more sensitive to extreme values.

- d) Daytime beam removal: Night-time beams are more favourable than daytime ones because of the interaction of solar reflectance and Lidar (Tang and Armston, 2019). Model_2020 had 8 available night-time GEDI transects (the others had less than 3 combined), so only those of Model_2020 were filtered for and plotted to see if results could hypothetically be improved.

An issue when interpreting GEDI L2A heights is that it does not discriminate trees from buildings, and therefore can give return height metrics where it is physically not possible for there to be heights; like in water, near sand, etc. Therefore, only shots inside the mangrove region were considered and cropped for the training model. Additionally, GEDI is known to perform poorly on slopes, however, the SRTM DEM in QGIS did not show significant slopes and therefore no methods to account for this were deemed necessary (Potapov et al., 2021).

Tidal effects

The L2A elevations provided used to calculate the L2A products have been corrected for solid earth tides, ocean loading, solid earth pole tide, and ocean pole tide but aren't ocean tides and dynamic atmospheric correction (Tang and Armston, 2019).

Tidal fluctuations in VCS1764 or anywhere near the creeks are not available, however, the nearest gauge site that was found, in Goyangyi Kyun, about 71 km south of the site is shown in Figure 5.4. Since August '23 is a month with traditionally high sea levels, in this location, there is a maximum tidal range in August 2023 of 2.5 m. According to Lagomasino et al., (2016), in microtidal coastal environments, where tidal fluctuations are less than 2 m, overall height estimates may not significantly change the overall distribution of mangrove CHs for the area. Xiong et al., (2022) also found that a 0.3m tidal effect was negligible for their study area in South Florida.

For this study, it was assumed that the Magyi tidal creek on-site, judging by the Goyangyi Kyun's tidal data is a microtidal environment, therefore it is unlikely that tidal effects could largely impact the heights measured. A rough estimate of 0.3 m was chosen as the maximum fluctuations that could exist on-site based on Xiong et al., (2022)'s findings.

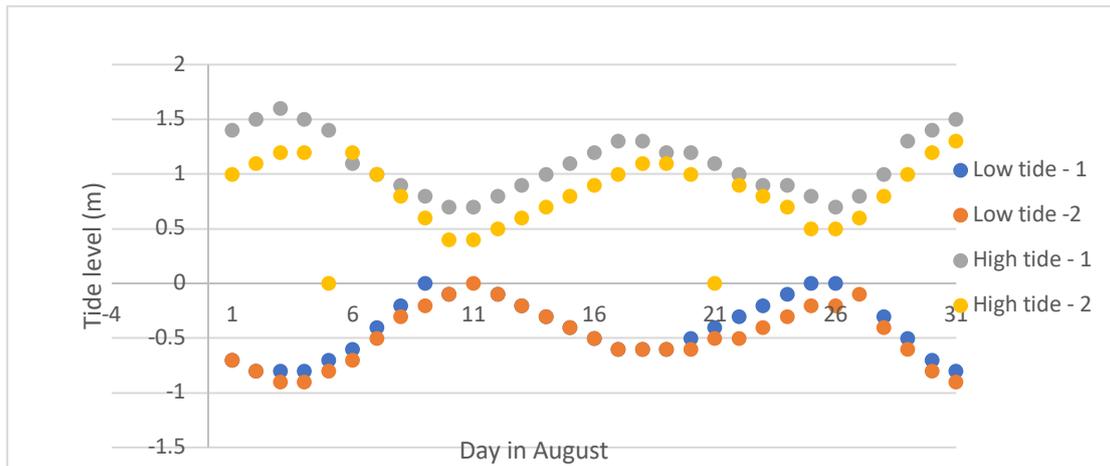


Figure 5.4: Tide conditions in Goyangyi Kyun (Tidechecker, 2023).

Additionally, a preliminary analysis from Google Earth Pro showed that images on site showed maximal tidal inundation in December, and least in February/ March. The RF model that was trained took the median of the pixel values for a period between January and August of 2019, mitigating the impact that tidal inundation would have on the shots that were considered in model-2019.

5.3. Sentinel-2

Sentinel-2 is a pivotal component of the European Union's Copernicus Programme, an endeavour focused on Earth observation to support environmental management, policy-making, and civil security. Launched by the European Space Agency (ESA), the Sentinel-2 mission comprises twin satellites, Sentinel-2A and Sentinel-2B, which were sent into orbit in 2015 and 2017, respectively. These satellites are equipped with state-of-the-art multispectral instruments, a passive remote sensing technology that captures light from multiple specific portions or bands of the electromagnetic spectrum. Sentinel-2 has 12 bands of resolutions ranging from 10m to 60m as described in Table 5.1 with a revisit time of 5 days, suggesting a relatively high temporal coverage. In practice however, cloud cover affected a lot of the imagery in May 2019, when the field height measurements were taken. Therefore, the only Sentinel image that had a cloud cover of less than 15% had to be averaged on Google Earth Engine (GEE) from the period under consideration.

Table 5.1: Table with the description of each Sentinel band variable extracted from GEE.

Band Name	Description/Purpose	Resolution (m)
B1	Coastal aerosol	60
B2	Blue	10
B3	Green	10
B4	Red	10
B5	Vegetation Red Edge	20
B6	Vegetation Red Edge	20
B7	Vegetation Red Edge	20
B8	NIR (Near Infrared)	10
B8A	Narrow NIR	20
B9	Water vapour	60
B11	SWIR (Short Wave Infrared)	20
B12	SWIR	20
AOT	Aerosol Optical Thickness	-
WVP	Water Vapour	-
SCL	Scene Classification Layer	20
TCI_R	True Color Image Red	10
TCI_G	True Color Image Green	10
TCI_B	True Color Image Blue	10
MSK_CLDPRB	Mask for cloud probability	-
MSK_SNWPRB	Mask for snow probability	-
QA10	Quality assessment for 10m resolution	10
QA20	Quality assessment for 20m resolution	20
QA60	Quality assessment for 60m resolution (clouds and cirrus)	60

Covering the Earth's entire surface every five days when working in tandem, Sentinel-2 is instrumental in diverse applications. To obtain the Sentinel-2 imagery, clouds in Sentinel-2 images using the QA60 band, where bits 10 and 11 represent clouds and cirrus at a 60m resolution in GEE were masked. Images from the 'COPERNICUS/S2_SR' collection were fetched for dates in 2019 and 2020, filtered by geographical bounds and a cloudiness percentage of 15% for 2019 and 2020. All images were then resampled within GEE to a 25×25 m grid to match the spatial scale of GEDI.

Another pre-processing that was tried was to exclude the Scene Cover Classification (SCL) which was underwater/ non-vegetated like other authors have approached (Potapov et al., 2021), however this made the model perform worse and therefore was added back again. The bands that were used in the final analysis were: "B1", "B2", "B3", "B4", "B5", "B6", "B7", "B8", "B8A", "B9", "B1", "B12", "WVP", "SCL", "TCI_R", "TCI_G", "TCI_B" and "NDVI". The Normalized Difference Vegetation Index (NDVI) was calculated and added to the list of predictors;

$$\text{NDVI} = (\text{B8} - \text{B4}) / (\text{B8} + \text{B4})$$

NDVI is widely recognized in scientific literature as a reliable measure of vegetation health (Ghosh et al., 2020). This metric capitalizes on the inherent physical properties of healthy vegetation: the tendency to absorb more red light (B4) while reflecting a greater amount of near-infrared light (B8). Therefore, a higher NDVI value usually indicates healthier and more robust vegetation.

5.4. Baseline data (Objective 3: Comparison with Baseline CHMs)

There are 3 Global CHMs whose data has been extracted (See Section 2.3) to compare with field data and the predicted heights from the models that will be generated.

- a) Potapov- 2019- A 30 m spatial resolution global CHM valid for the year 2019 developed by integrating Landsat Analysis ready data and GEDI Rh95 using a RF algorithm (Potapov et al., 2021)
- b) Simard-2000 - A 30 m spatial resolution Mangrove CHM valid for the year 2000, made by integrating SRTM-DEM and ICESat-2/GLAS Rh100 (Simard et al., 2019).
- c) Lang-2000 - A 10 m spatial resolution Global CHM valid for 2020 developed by training a Convolutional Neural Network (CNN) on Sentinel-2 and GEDI Rh98 (Lang et al.,2021).

Table 5.2: Summary of Data sources used.

Data Source	Source	Temporal Coverage	Resolution	Pre - Processing
GEDI L1A, L2B	NASA Earthdata store/ GEE	November '21	25m	NA
Model-2019_s	GEE	22/05/19	25 m	Beam Sensitivity >0.95 Quality Filter ==1
Model_2019	GEE	21/04/19- 18/08/19	25 m	Beam Sensitivity >0.95 Quality Filter ==1
Model-2020	GEE	25/02/20- 11/12/20	25 m	Beam Sensitivity >0.95 Quality Filter ==1
Sentinel – 2 (2019 models)	GEE	01/01/2019 – 30/08/2019	25 m	Resampled to 10m (comparison with ground data plots) Cloud cover < 15%
Sentinel-2 (2020 models)	GEE	01/01/2020- 31/12/2020	25 m	Resampled to 10m (comparison with ground data plots) Cloud cover < 15%
Ground data	Verra.Org	May - June '21	10m	None
Potapov CHM	https://glad.umd.edu/d ataset/gedi/	2019	30m	None
Lang CHM	https://langnico.github. io/globalcanopyheight/	2020	10m	None
Simard CHM	https://daac.ornl.gov/C MS/guides/CMS_Glob al_Map_Mangrove_Ca nopy.html	2000	30m	None

6. Training the model

6.1.1. Random Forest Model

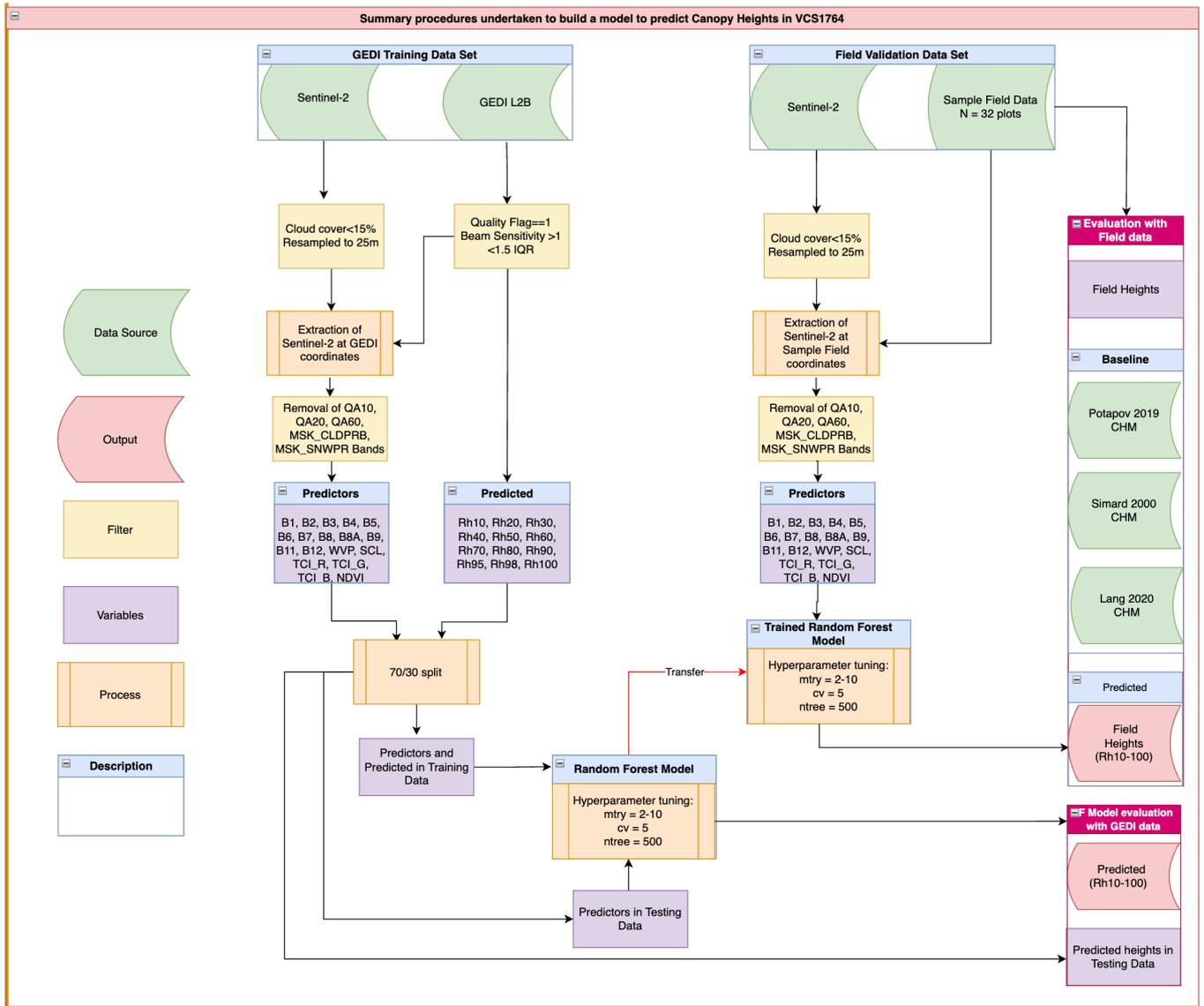


Figure 6.1: Summary of Data Sources, filtration procedures, RF model parameters and outputs of the study.

All coordinates were reprojected EPSG:32646 from WGS84/EPSSG:4326, the CRS commonly used in Myanmar. RF belongs to the class of non-parametric regression methods, i.e. those that do not make specific assumptions about the functional form the variable, unlike linear or logistic regression. This can help capture non-linear relationships between the predictor and the predicted variables in many biophysical processes (Ghosh et al., 2020; Potapov et al., 2019).

A machine learning workflow using RF, an ensemble learning method was used because of the precedent of best-performing model in literature. A study by Luo et al., (2023) in China compared RF, Support Vector Machine (SVM), K- nearest neighbour (KNN), Gradient Boosting Decision Tree (GBT) to link CH in ICESat-2 and multisource remote sensing imagery. RF was shown to have the best predictive performance. Additionally, RF was also used by Potapov et al., (2021) to create a Global CHM, indicating its robustness in prediction and versatility in use.

Initially, a 5-fold cross-validation is established, which means the training data is split into 5 subsets; the model is trained on 4 of them and validated on the remaining one, iteratively. Next, a grid of potential hyperparameters ranging from 2-10 was defined for the RF, specifically the **mtry** parameter, which specifies the number of features (predictors) randomly sampled at each split when building a tree in the forest. This ensures randomness in the model that increases diversity in trees - essential for reliable results. The ones that were provided ranged from 2-10.

The number of trees (ntree) in the forest was set to 500. For a regression problem, each tree in the RF predicts a continuous outcome, and the final prediction is typically the average of the predictions from all the trees – and is thus called an ensemble method.

Typical regression algorithms use variance reduction to calculate the split, i.e. the criterion that maximally reduces the variance of the resulting child nodes. Having multiple trees (like 500 in this case) means 500 random samples from the data with replacement were used to start off each tree and then the results are averaged out for a new piece of data. This helps in capturing complex non-linear relationships in the data and reduces the risk of overfitting by averaging out individual tree predictions. Finally, the trained model is used to make predictions on both the testing set and field data.

There were 8 available GEDI transects in 2019 and 18 in 2020. Because of the greater number of transects in model 2020, it was decided that 3 models would be created (more details in Appendix 12.5 ;

- 2019_s - This comprised of a single GEDI transects data passing over VCS1764 on 22 May 2019 – close to when field height measurements were taken.
- 2019 – Used 6 of the available 8 transects available in the whole year of 2019.

- 2020 - Comprised of 16 of 18 available transects in 2020. 2020 was found to be when GEDI had a lot of passes over VCS1764 and was therefore included to make the findings more robust.

The models, labelled as "2019_s" and "2019", were evaluated on their respective GEDI testing sets based on distinct testing datasets derived from a 70/30 training-testing split in the year 2019. The model designated "2020" was trained utilizing the 2020 Sentinel imagery and subsequently tested against its GEDI testing set shots from 2020. All models were tested on field data extracted from Sentinel-2 imagery captured in 2019.

Table 6.1: Summary of the models and the shots used in them.

Model	Sentinel-2 image specifications	Number of Transects used	Before Filter	After filter	Final no. of Shots after Cropping*
2019_s	The median between Jan-Aug 2019, Cloud cover < 15%	1	2843	631	126
2019	The median between Jan-Aug 2019, Cloud cover<15%	6	9429	1350	225
2020	Median between Jan-Dec 2020, Cloud cover<15%	16	20839	8778	765

*The Cropping was applied to include only VCS1764 mangroves

The number of shots that were filtered before use was more than 50% of the original in all cases. For example, in 2019, 7355 shots had a quality flag of 0, which is almost 80% of the original. 724 had a beam sensitivity of less than 95%. These were subsequently cropped to just the mangrove area to give the final number of shots after cropping (See Table 6.1).

Performance Evaluation: Once trained, the model's predictive accuracy was evaluated on two separate test datasets; Field Heights and test GEDI Rh data. The key metrics are;

- Mean Average Error (MAE) - Provides the average difference between observed and actual means.
- Root Mean Square Error (RMSE)- Providing a measure of the model's prediction error; root of the average summed deviations from the mean.
- Pearson correlation (R) - Calculated by dividing the covariance matrix of the independent variables by their standard deviations. It ranges from -1 to 1 and indicates the strength of a positive or negative relationship.
- Coefficient of Regression (R²) – This indicates the proportion of variance explained by the model.
- P value – This is calculated by looking at the likelihood of a given R² appearing by chance by taking into account the number of data points.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

$$\text{ME} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

$$r_{xy} = \frac{\sum_{i=1}^N (\hat{x}_i - x_i) (\hat{y}_i - y_i)}{\sqrt{\sum_{i=1}^N (\hat{x}_i - x_i)^2} \cdot \sqrt{\sum_{i=1}^N (\hat{y}_i - y_i)^2}}$$

7. Results

7.1. Objective 1: Explore GEDI products over VS1764

7.1.1. GEDI L1B

GEDI L1B Waveform output from a single transect on 22 May 2019 (see Figure 7.1), closest to the Field Height acquisition month (May 2019) will be examined in this section.

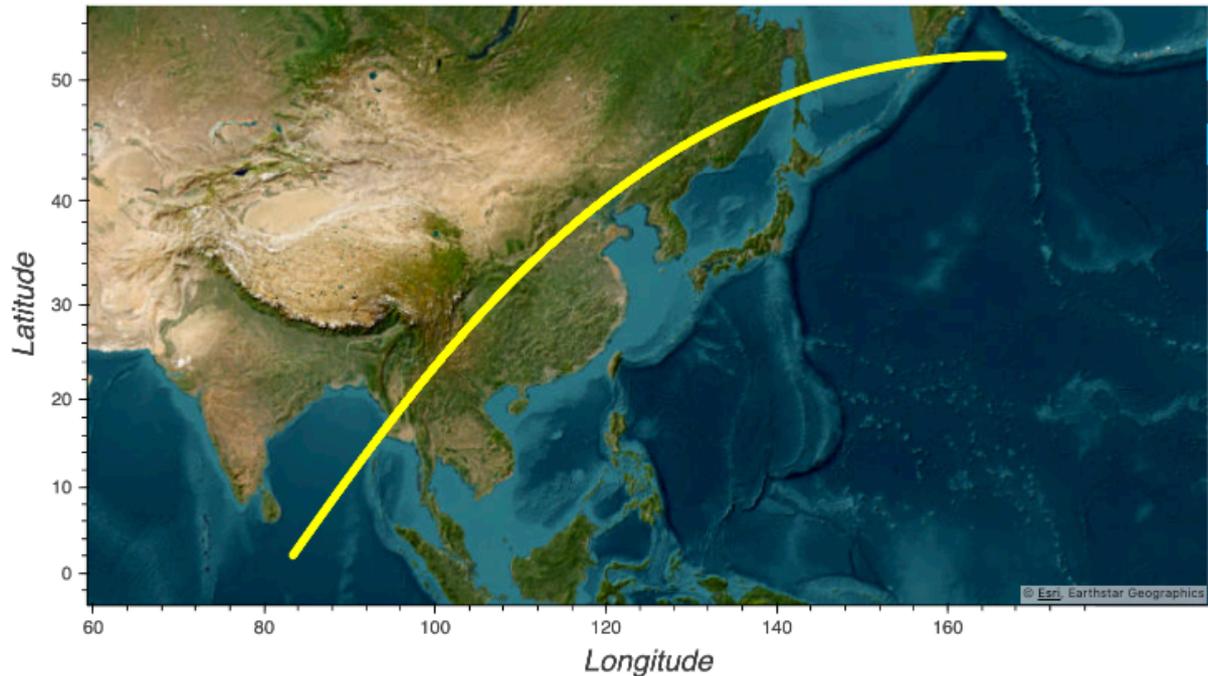


Figure 7.1: GEDI Transect on 22 May 2019 over VCS1764.

Figure 7.2 shows the location of one shot closest to a Sample field plot to view the waveform and relative height metric product from GEDI. Relative Height (Rh) metrics represent the height at which a percentile of the laser's energy is returned relative to the ground. This shot and Sample Field plot was chosen as it is part of BEAM1011, one of the 4 full power beams that operate at a higher power level than the other 4 coverage beams, i.e. have a higher signal-to-noise ratio.

Examining the GEDI Transect closest to a Sample Plot, 22 May 2019

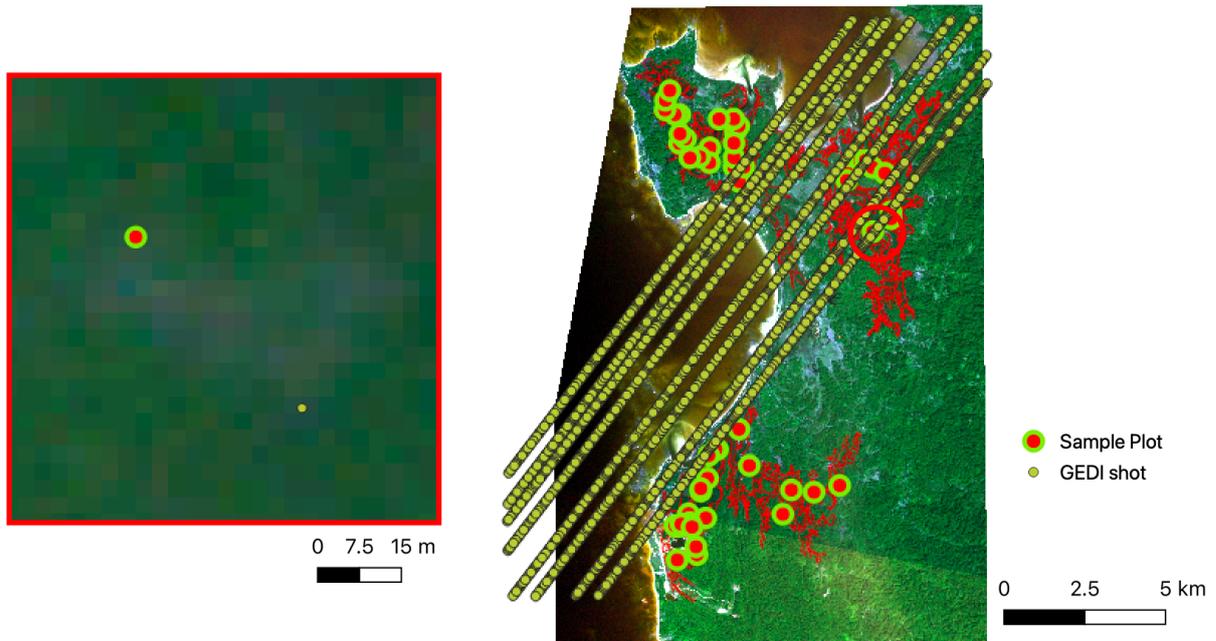


Figure 7.2: Examining a GEDI transect Shot 2494110020003598, closest to a Sample Field plot 12.

The relative height metric of this shot is shown in Figure 7.3 a) which displays a relatively smooth waveform suggesting low variations in mangrove structure. Additionally, it can be observed that Rh 25 is classed lower than the Ground Return, different from the California Redwood forest in Figure 7.3 b) where the first peak gets classed as the ground return. This is calculated from the full waveform information using the 6 different algorithms with different sensitivities/thresholds of ground detection.

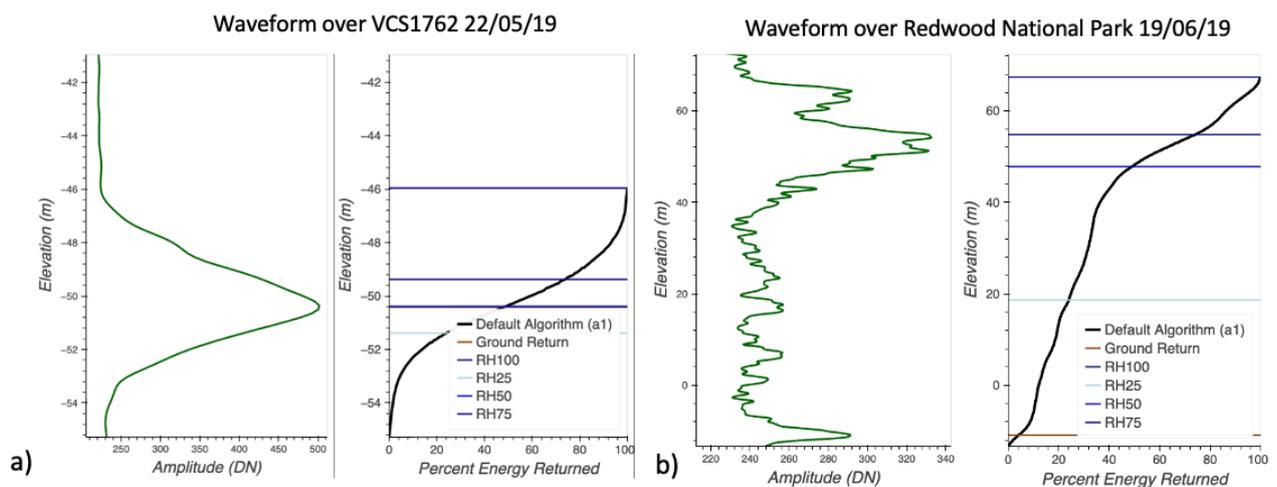
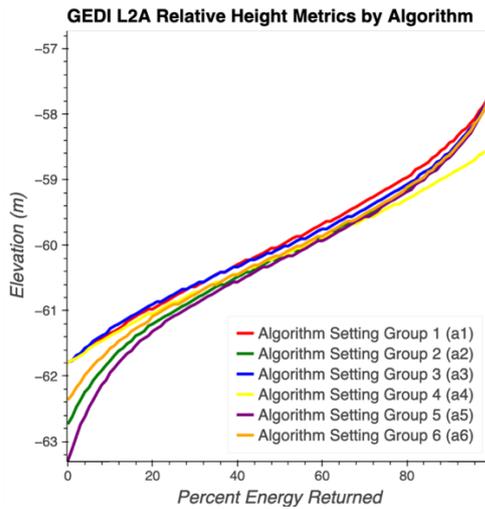


Figure 7.3: a) Comparing the GEDI L1B waveforms of a shot near VCS1764 with that of b) in Redwood National Park.



Setting group	Smoothing width (noise)	Smoothing width (signal)	Waveform signal start threshold	Waveform signal end threshold
1	6.5 σ	6.5 σ	3 σ	6 σ
2	6.5 σ	3.5 σ	3 σ	3 σ
3	6.5 σ	3.5 σ	3 σ	6 σ
4	6.5 σ	6.5 σ	6 σ	6 σ
5	6.5 σ	3.5 σ	3 σ	2 σ
6	6.5 σ	3.5 σ	3 σ	4 σ

Figure 7.4: a) Comparing different algorithms for Shot 2494110020003598. b) σ represents the standard deviation of the background noise level (Tang and Armston, 2019).

In Figure 7.4, some differences in the RH values between the 6 algorithms can be observed, particularly in the lower percentiles that have a lower elevation for the same shot. This is because of different parameter settings that have different ground detection thresholds, as shown in Figure 7.4 b). For example, the more absorbent the ground material, the longer the lower half of the energy return becomes, making it more likely for lower Rh metrics to be classified as negative. To compare the findings of this study with those of globally available datasets (Objective 3 of this study), the default a1 setting was used like them - designed to apply to most places on earth (Tang and Armston, 2019). However, conducting more research into the setting appropriate for mangroves could shift the interpretation of the results. In Figure 7.4, for example, using a5 could provide lower estimates for the elevation by 1.6 m due to its lower waveform signal end threshold.

The GEDI L1B product, containing the raw waveform data and the corresponding L2A metric was plotted for locations representative terrain types, both mangrove and non-vegetated zones within mangrove tracts in Figure 7.5, and the sea and coastal zones near the mangrove Figure 7.6.

In the first set of waveforms, Figure 7.5 for the mangrove and non-vegetated zones, the returns close to the bare ground (shots 91 and 92) have less variation past the 50% waveform energy returns than those with vegetation (Shot 88, 89 and 93), suggesting a unique waveform signature detectable in a mangrove. This is a crucial observation as it conveys essential understory information not conveyed by passive remote sensing sensors. Such spectral signatures are useful in distinguishing plant types (Tommaso et al., 2021) which points to the unique information that spaceborne lidar conveys.

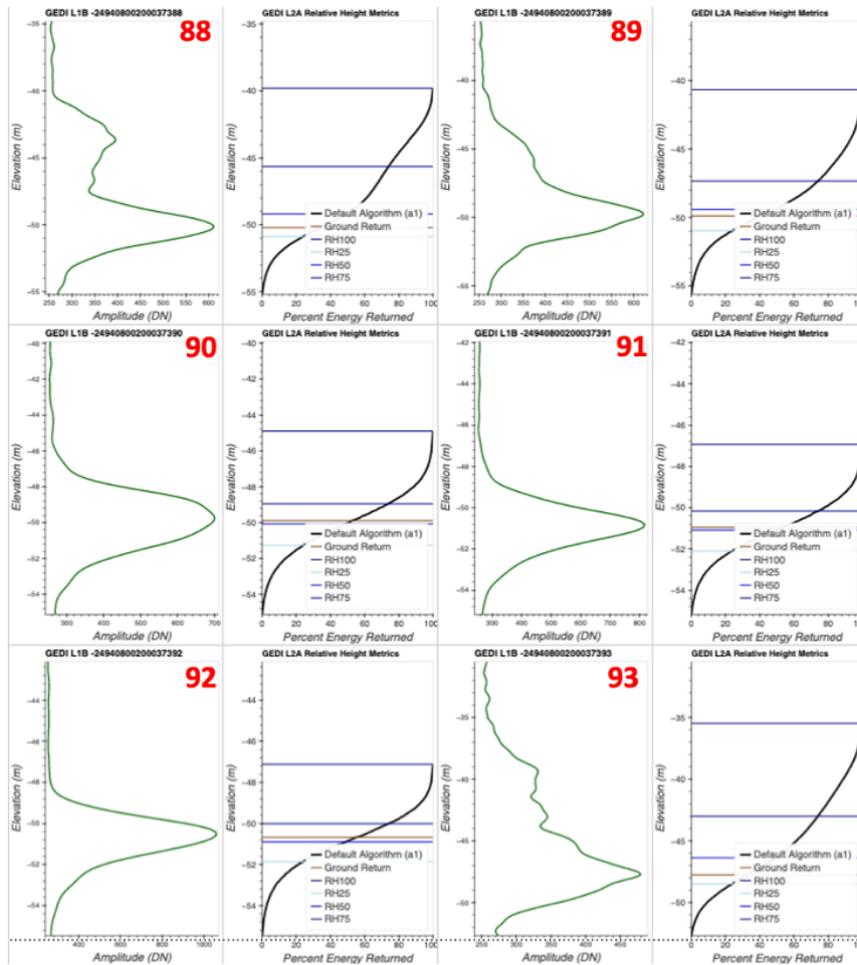
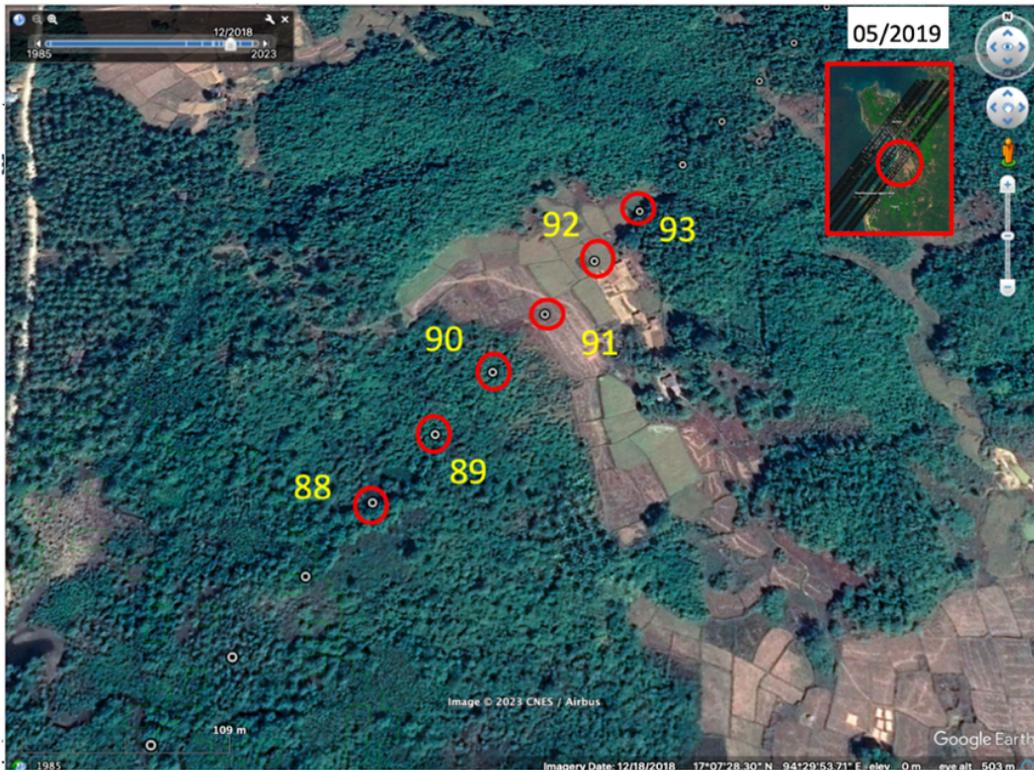


Figure 7.5: The Full waveform and relative height metric over mangrove and bare fields in May 2019.

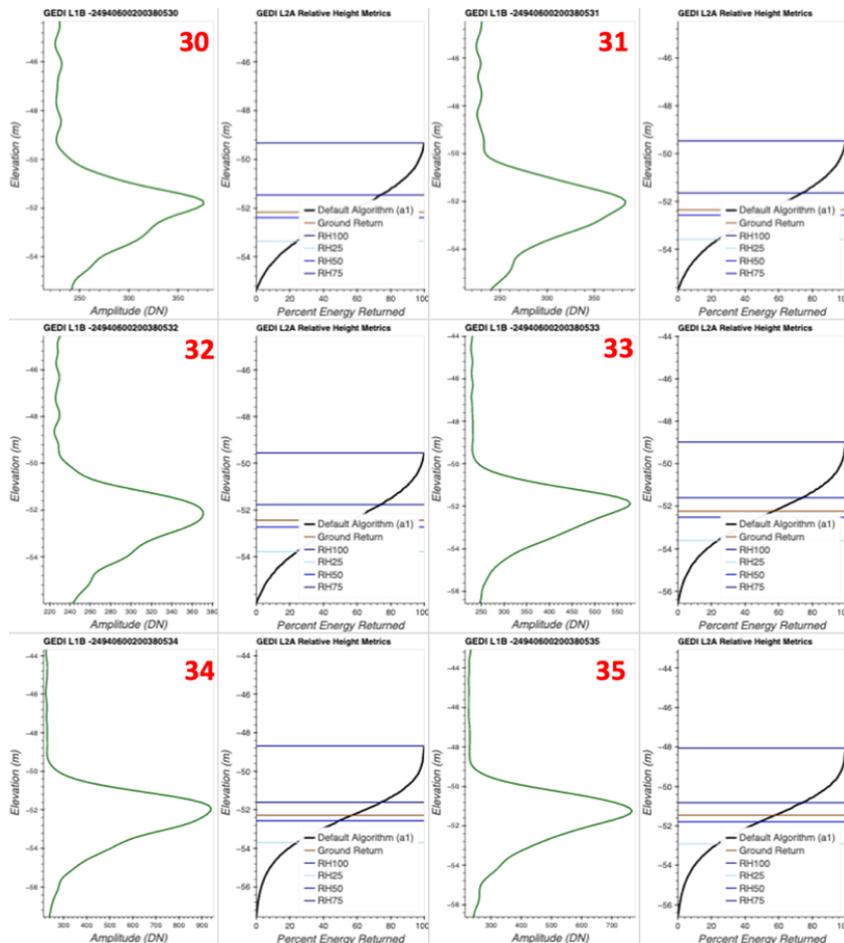
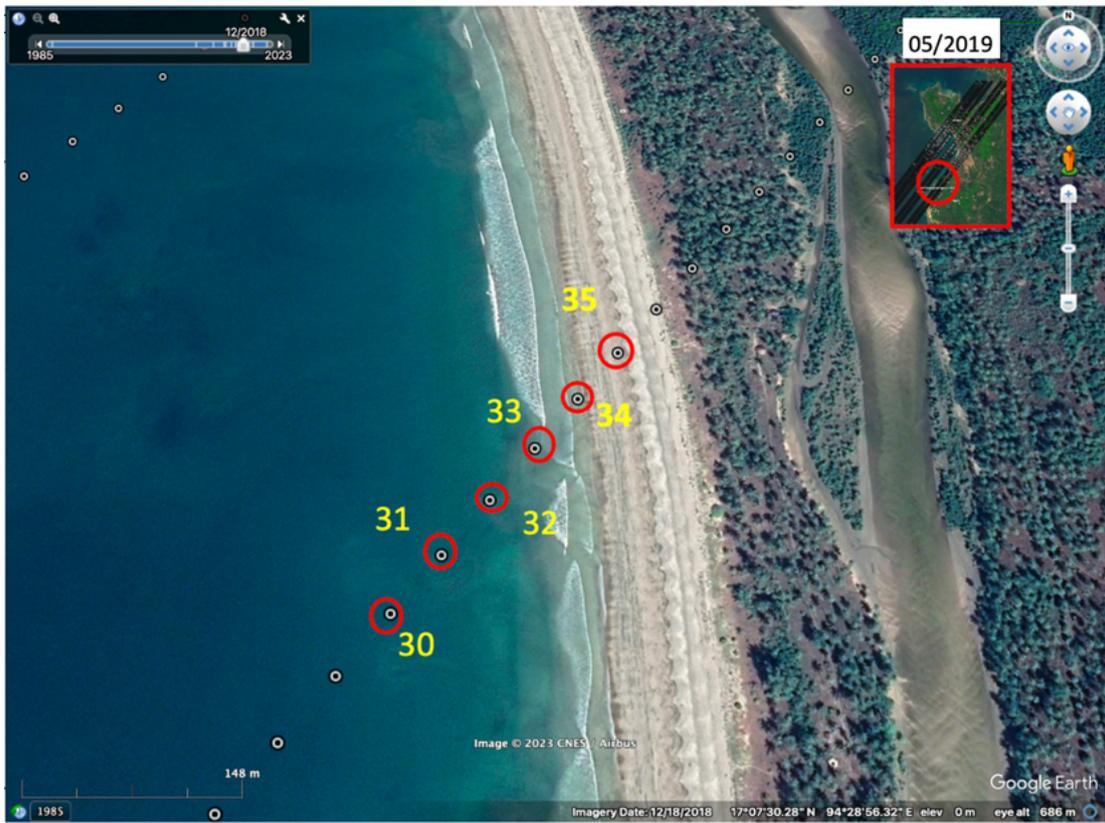


Figure 7.6: Waveform returns over water and coast.

In the second set of waveforms in Figure 7.6, GEDI gives waveform returns even where there is no vegetation/ the area is covered by water. While the common approach in many methodologies is to exclude water and non-vegetated areas through Scene Cover Classification, though this was attempted, did not seem to reduce the model’s accuracy by a great amount (Guerra-Hernández and Pascual, 2021; Potapov et al., 2021). Mangroves are influenced by tidal effects, leading to periodic inundation. Excluding these inundated areas might result in excluding important mangrove regions. Furthermore, the RF algorithm utilized various Sentinel-2 bands, some of which would have captured indications of water presence. This means that the model potentially considers the spectral characteristics of inundated sites, making a strict exclusion based on water presence less critical.

7.1.2. GEDI L2A

Once the waveform of a single shot near the sample field was taken (Figure 7.2), plots of the L2A height metrics were made to view the trends to help formulate the next pre-processing and methodology steps. The Relative height metrics, Rh 0-100 of mangrove in VCS1764, in Figure 7.7 a) is contrasted with that of Tall Redwood trees in California, USA in Figure 7.7 b) to better compare and contrast results on site.

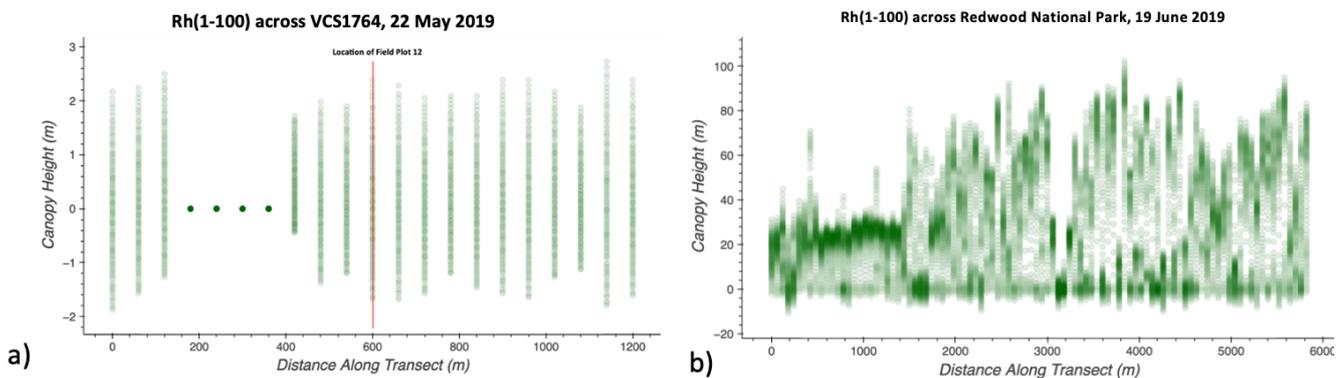


Figure 7.7: a) Rh covering 50 shots to the approximate right and left of the Sample Field Plot 12 in the GEDI in the along-track direction, on 22 May 2019. b) Rh across Redwood National Park

Each vertical green bar shows the vertical position of energy quantiles (0–100%) for a waveform along-track, where dark green indicates higher return energy (ground or more canopy material) and light green indicates lower return energy (less canopy material).

The distribution of the Rh indices over the mangrove region shows less variation in a) than in b)
The canopy over the mangrove in Figure 7.8 a) is not as easily discernible from the rest of the middle growth, when compared to taller Californian redwood trees in Figure 7.8 b).

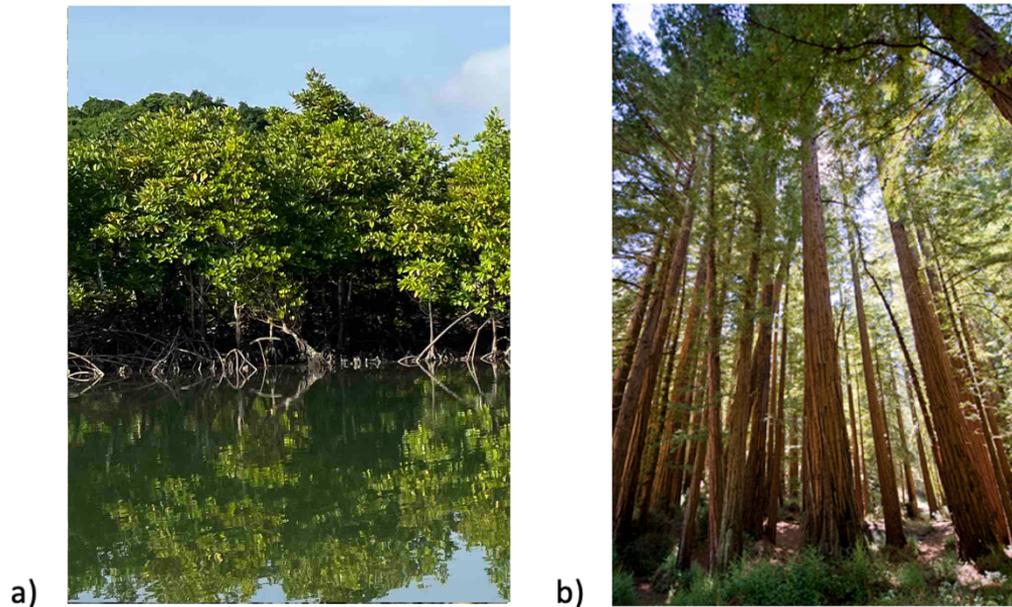


Figure 7.8: Comparison of a) Mangroves on-site in VCS1764 and b) Californian Redwood trees

In Figure 7.9 a) and Figure 7.10, the spatial distribution of the GEDI shots is shown along with the b) Distribution of Rh100 – commonly attributed to CH in literature. The Rh100 metric is considered to be the CH in many studies. Figure 7.10 shows a unimodal distribution indicating areas that had a single layer canopy such as shrub or arbour trees without understory presence, a finding consistent with Q. Li et al., (2023)’s findings when using lidar in China’s mangroves. The full distributions of the rest of the Rh metrics are given in Appendix 12.1, which also show a unimodal distribution like Rh100.

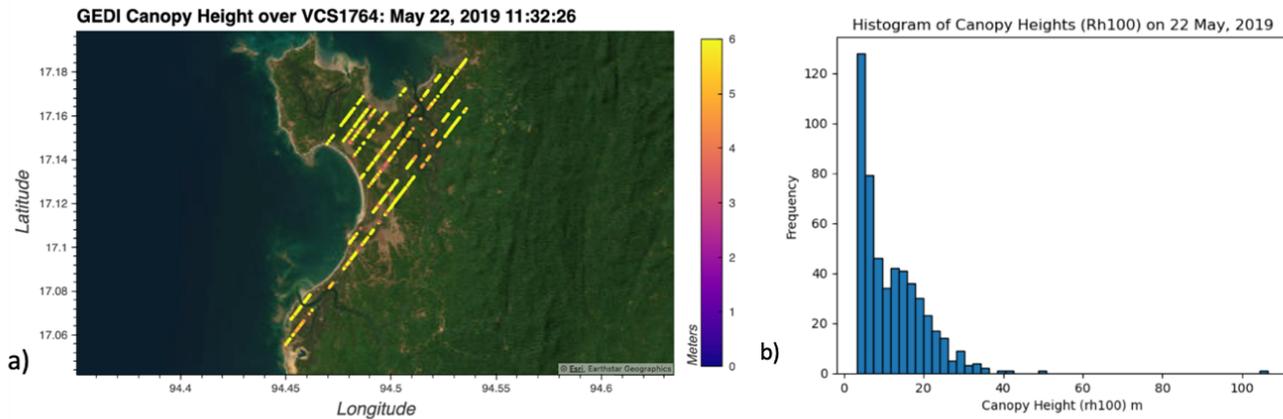


Figure 7.9: a) GEDI Rh100 over VCS1764. b) Distribution of Rh100

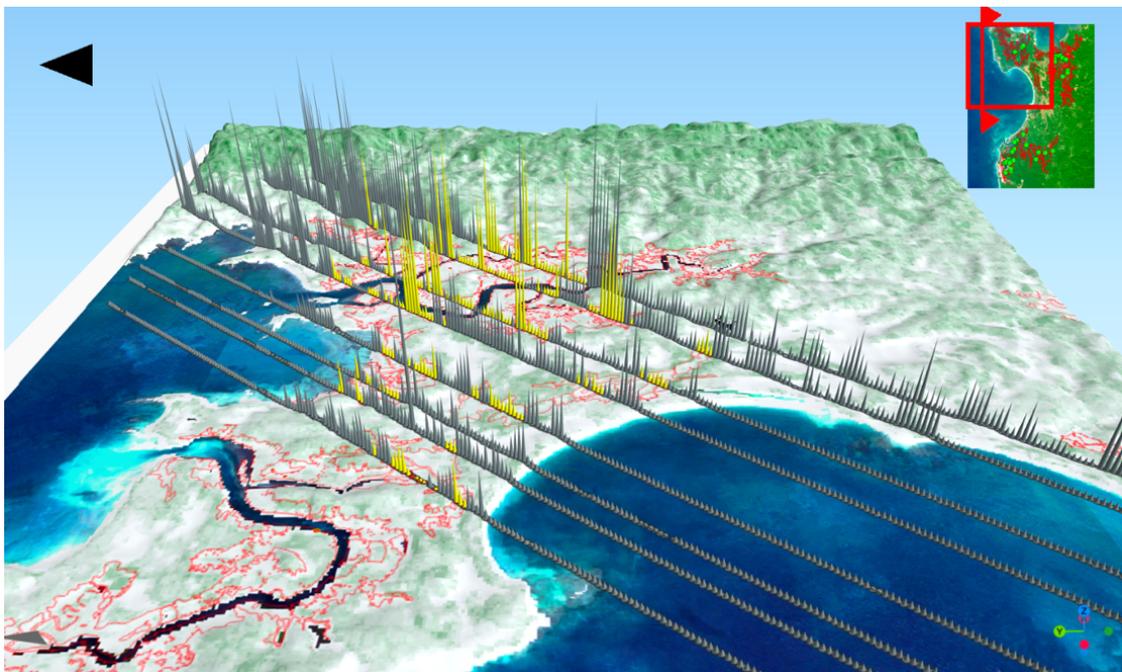


Figure 7.10: 3D Visualisation of Rh100 on 22 May 2019 (Shots in Yellow are those over mangrove in VCS1764) Raster values in Green are Potapov et al., (2021)'s Global CHM.

There is a visible difference in CHs in the mangrove measured by Potapov et al., (2021) as seen in the 3D visualisation. To further investigate some of the other Rh metrics on the same transect from Section Results 7.1.1, QGIS' buffer and intersect tools were used. Out of 1440 filtered GEDI shots obtained for the whole of 2019, only 10 were within a 30 m distance from the 32 sample fields. There were 2 shots with a quality flag of 1 (an indicator of the usability of the shot) which are plotted in Figure 7.11, with their Rh metrics.

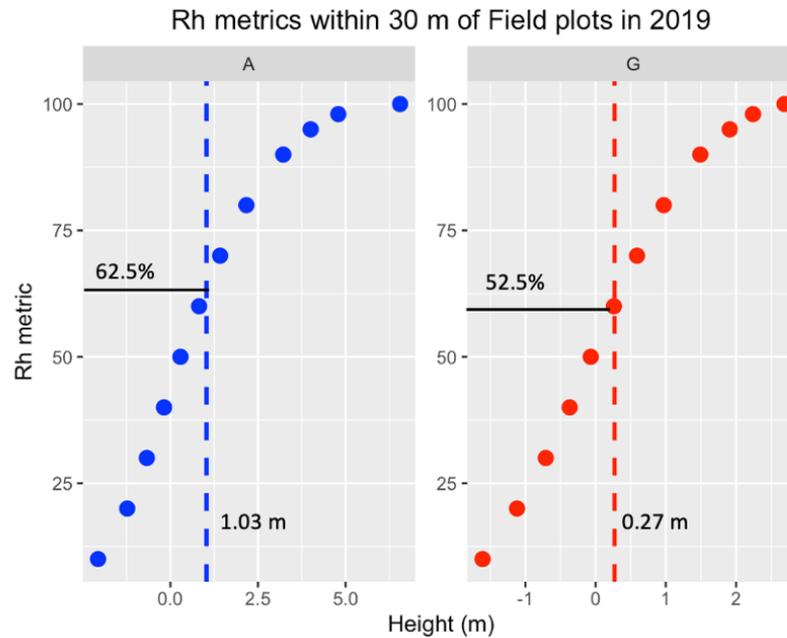


Figure 7.11: GEDI Energy profiles of 2 shots within 30m of their respective Field Plots in 2019.

From the graph, 62.5% and 52.5% respectively seem to be the percentage at which the waveform energy return corresponds best to the two field heights. This is in contrast to commonly used Rh metrics in literature like Rh90, Rh95 and Rh100 that is assumed to represent the top of all types of canopies (Lang et al., 2022b; Potapov et al., 2021).

The primary Rh metrics that were chosen as predictor variables in the model are 12 in number; Rh10, Rh20, Rh30, Rh40, Rh50, Rh60, Rh70, Rh80, Rh90, Rh95, Rh98 and Rh100. The Rh95 and Rh98 were picked too due to their extensive use in literature (Potapov et al., 2021, Duncanson et al., 2022).

7.1.3. GEDI L2B

The purpose of the GEDI Canopy Cover and Vertical Profile metrics product is to extract biophysical metrics from each GEDI waveform. It contains metrics like canopy cover and Planet Area Volume Density (PAVD). The PAVD is defined as the amount of vegetation per unit volume in a given area, providing insights into the structure and density of the forest canopy. To compare the CH metrics, the PAVD is compared with Californian Redwood trees in Figure 7.12 and Figure 7.13 and shows similar patterns to the Rh metrics; a uniform distribution with less stratified vegetation structure.

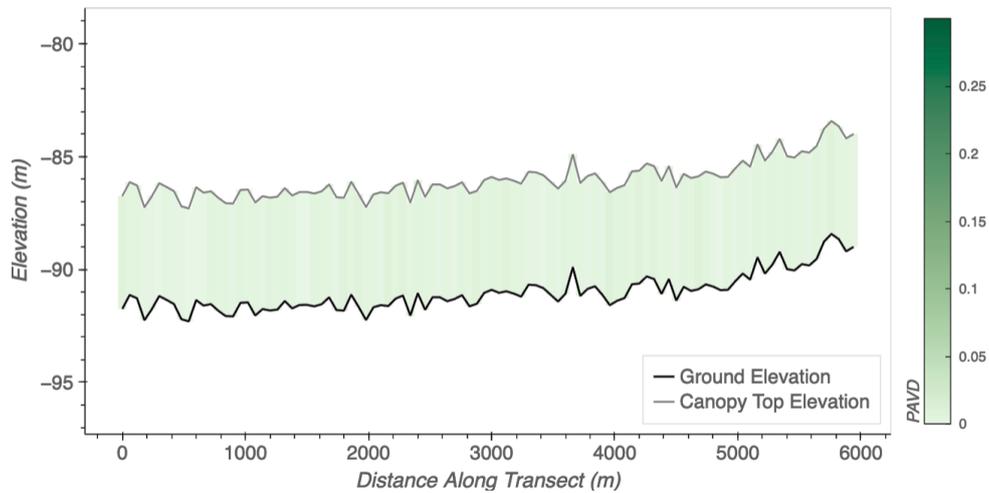


Figure 7.12: GEDI L2B Beam 1011 PAVD over VCS1764 on 22 May 2019.

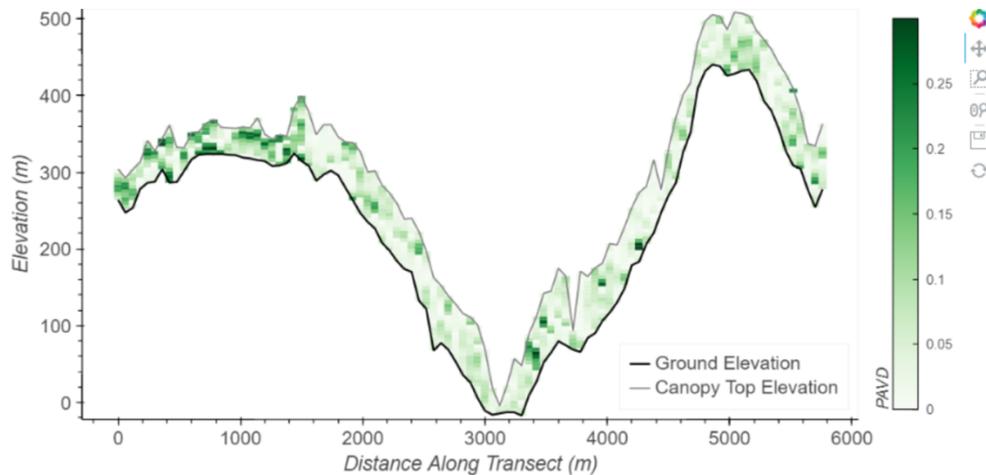


Figure 7.13: GEDI L2B BEAM1011 PAVD over Redwood National Park on June 19, 2019.

7.2. Objective 2: Create a Mangrove CHM

7.2.1. GEDI L2A Rh Exploration with Sentinel-2 bands

A histogram of each of the GEDI L2A Rh metrics, to be used as the dependent variables on 22 May 2019 is shown in Figure 7.14. The distribution suggests that field heights fall within the range of distributions of Rh60 and Rh70. The variance in the calculated Rh metrics from L2A increases with the elevation of return (Rh) – in line with observations from similar studies (Adam et al., 2020; Potapov et al., 2021; Stovall et al., 2021).

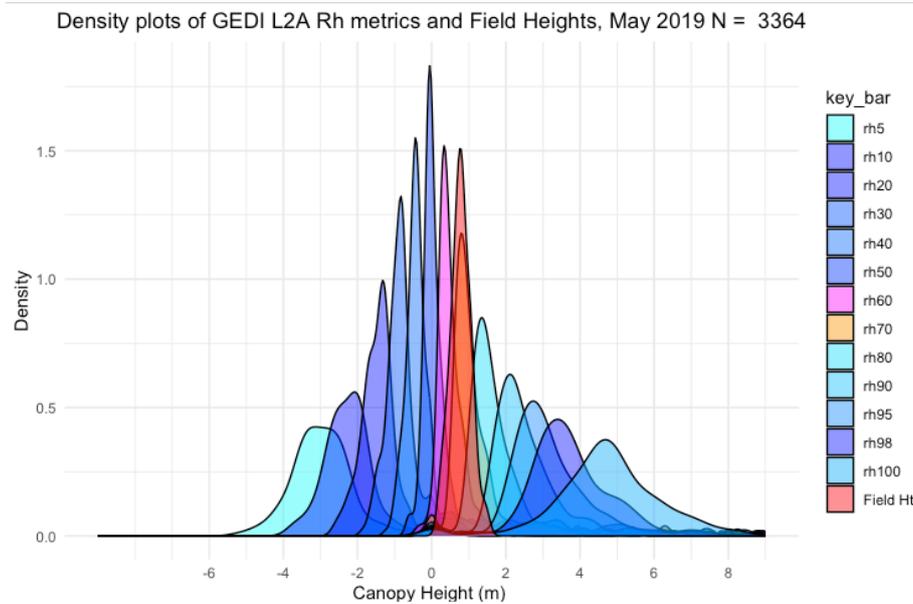


Figure 7.14: Distribution of Rh metrics & Field heights, 22 May 2019.

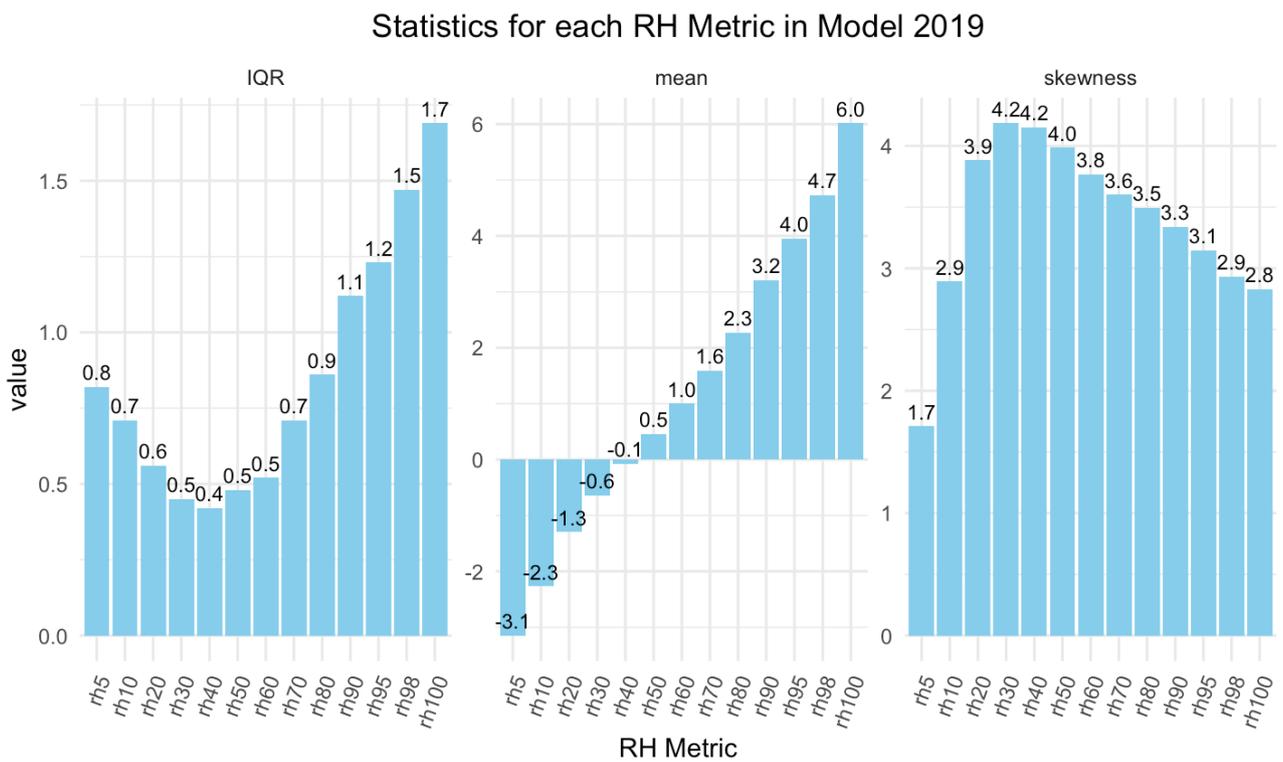


Figure 7.15: GEDI L2A Rh Distribution in the Bounding Region of VCS1764 in model- 2019.

The variability of heights increases with increasing Rh in Figure 7.15. In Figure 7.16, potential correlations that the GEDI L2 metric exhibits with relevant predictor variables are given. There is a noticeable positive inter-correlation among Rh metrics.

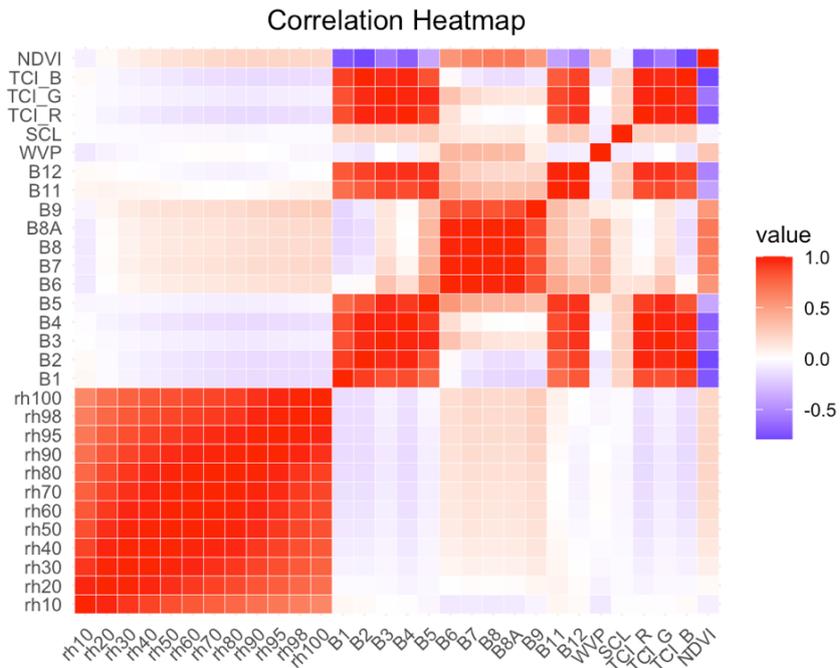


Figure 7.16: Correlation Heatmap for GEDI shots in the VCS1764 region - 2019.

The calculated NDVI and Bands in the Near Infrared Region (B8 – B9) show an increasing positive relationship with all Rh metrics and a negative with solar elevation. For NDVI, this is because when the internal cellular structure of the plant is healthy and intact, it reflects more NIR light than when stressed and dehydrated. A clearer visualisation of how the Sentinel-2 indices and some GEDI L2A products are correlated is in Figure 7.17. As the Rh metric increases, its correlations with Sentinel-2 variables and solar elevation - whether positive or negative get stronger. An elevated NDVI typically signifies dense vegetation cover, implying that the top of canopies are lightly more dense with leaf cover than the lower strata (See the field photographs in Figure 5.1)

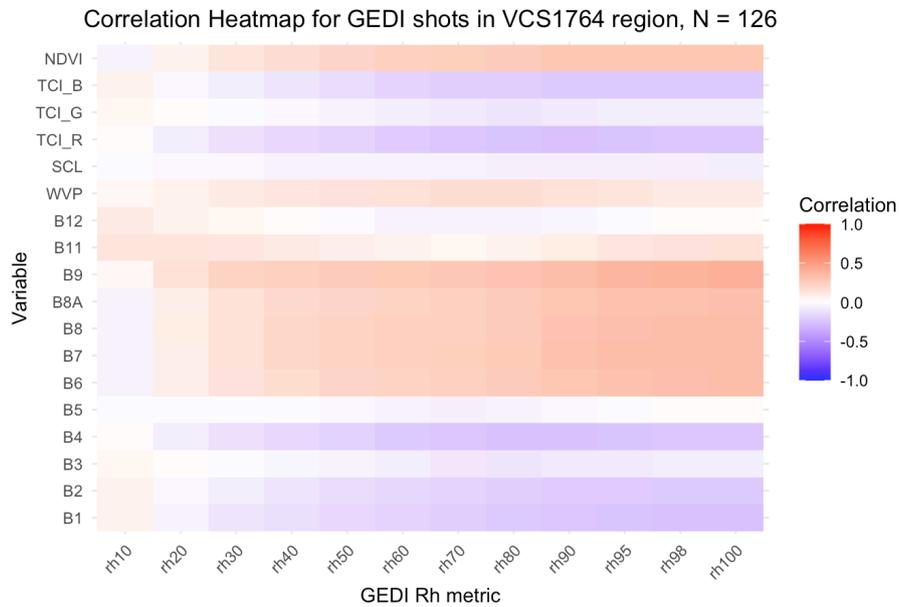


Figure 7.17: Correlation between some Sentinel-2 bands and Rh products.

B9 – corresponding to coastal aerosol has a positive relationship with the Rh that intensifies with the Rh metric. Another notable band is B5, Vegetation Red Edge where the relationship is not as pronounced as for the other Rh variables and is strongest between Rh50 - Rh90.

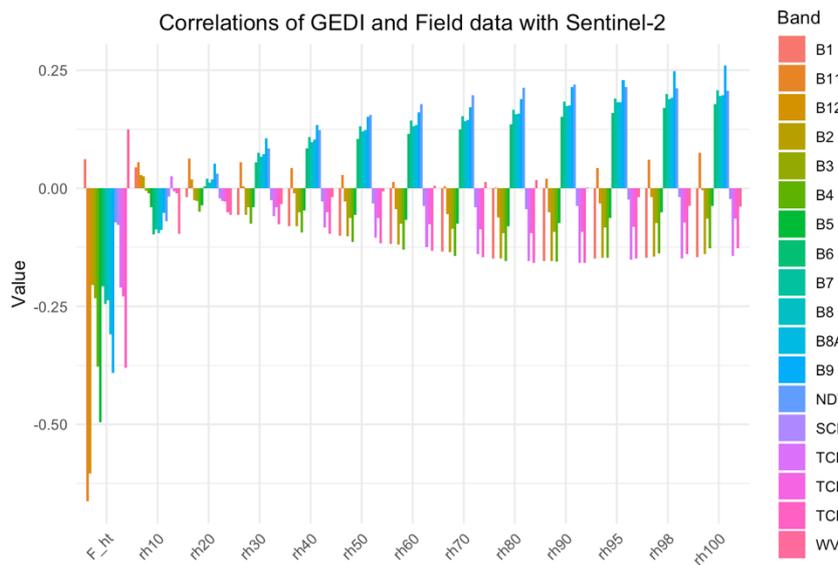


Figure 7.18: Correlations between the Sentinel-2 bands and GEDI-derived Rh values, and field heights.

Figure 7.18 compares the correlations between the Sentinel-2 bands with those that the 32 Field heights have with Sentinel-2 bands. Field_Hts tend to have opposite correlations with Sentinel-2

bands - most noticeably with the NDVI – which is surprising based on theory. This might be because of the small sample size of field plots.

7.2.2. Random Forest Model – Training the model

The `mtry` parameter shows a higher number of parameters sampled for the higher Rh values, as shown in Figure 7.19. Suggesting the higher variability of Rh values caused the RF Model to sample more parameters.

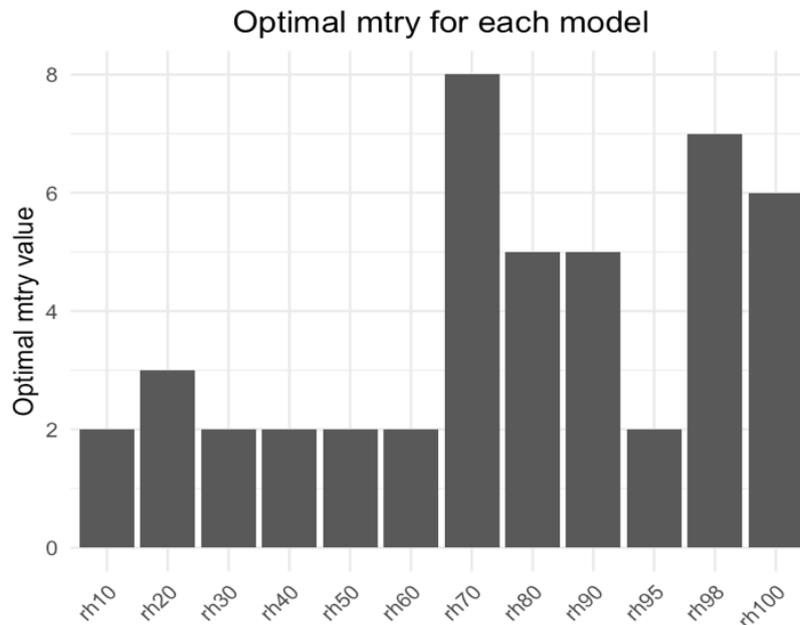


Figure 7.19: Optimal `mtry` for each Rh metric model.

7.2.3. Final results

The three models were compared on their performance on both their field data in Figure 7.20 and their test data in Figure 7.21 (See Appendix 12.3 for more details). The following observations will be classed into two to investigate the differences in;

- Intra-model - performance metrics with Rh.
- Inter-model - performance metrics with the quantity of training data.

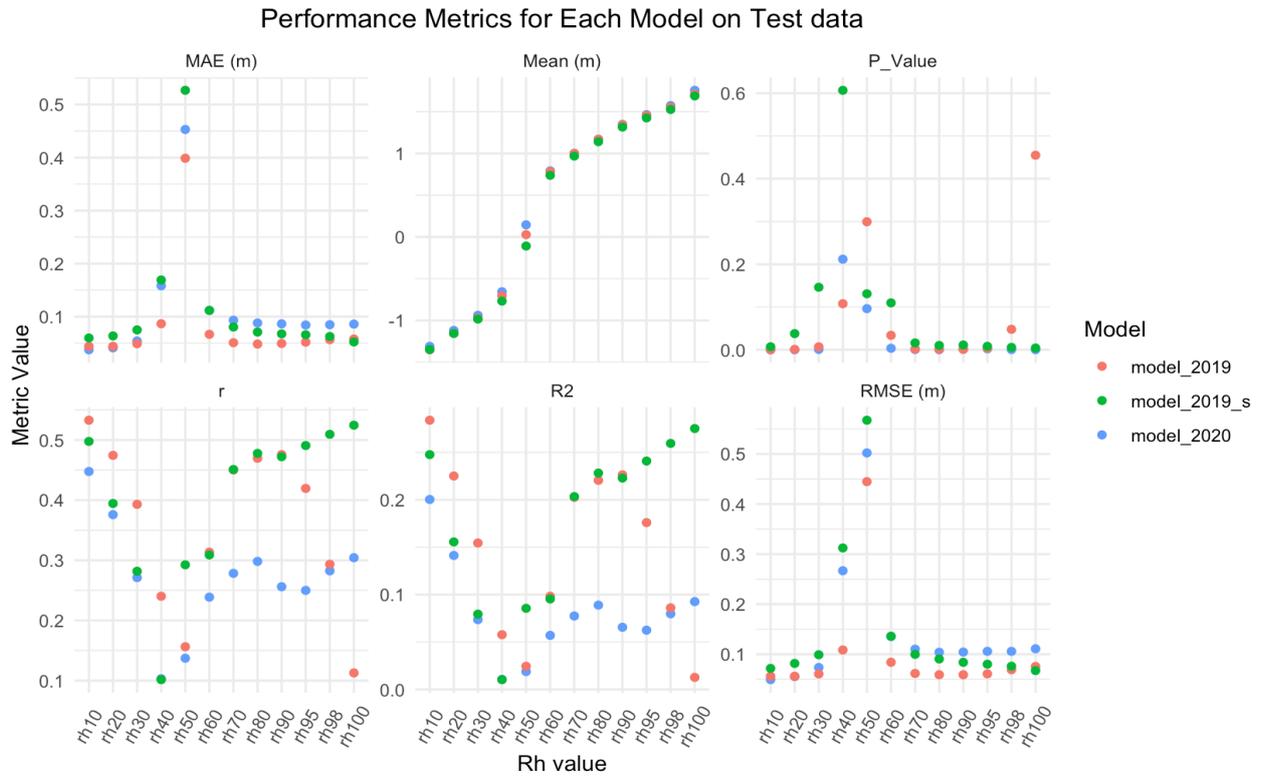


Figure 7.20: Performance metric across the testing set.

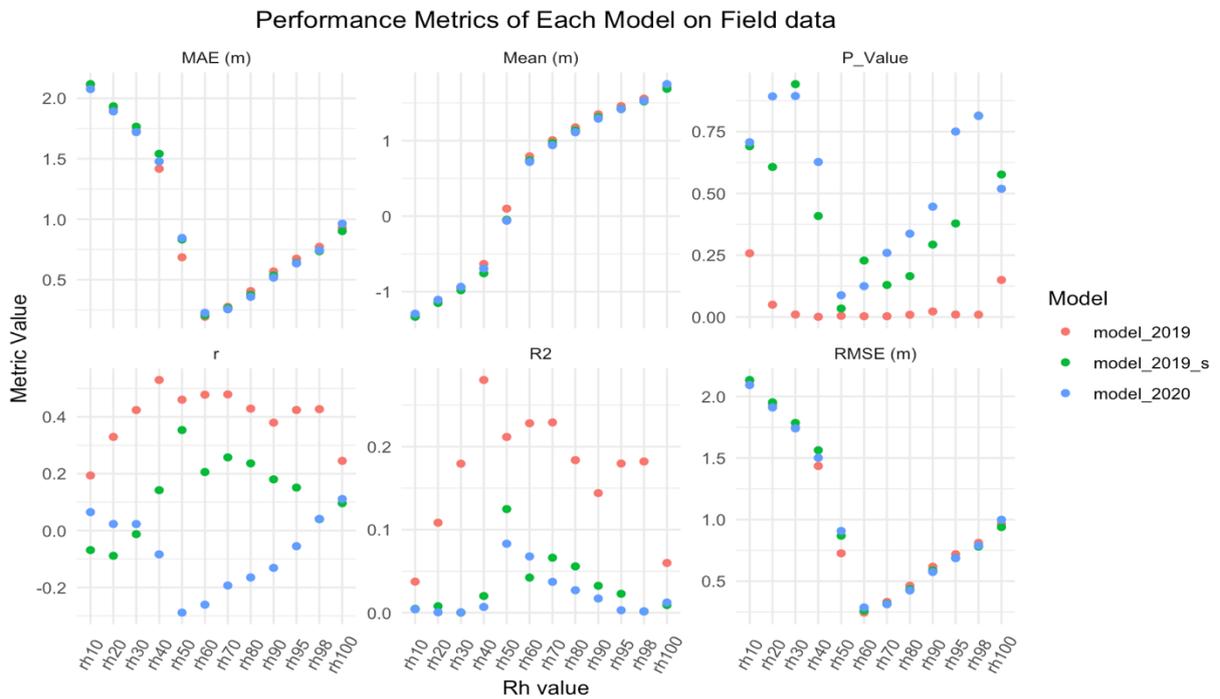


Figure 7.21: Performance metrics of each model on field data.

Intra-model trends

- The mean predicted Rh height increases with the Rh in both the Field and test data.
- The evaluation metrics, namely the R^2 and R exhibit a decreasing trend till Rh50 and then an increasing trend with Rh thereafter. This implies that the predictability of median/average

values is lower in the testing set, and that of the extremes of Rh is high. Conversely in the Field data – this is flipped; predictability in terms of R^2 and R peaks for the median/average values.

- c) The MAE and RMSE – indicative of errors show a steady decline in the testing set with increasing Rh values, suggesting more predictability in the higher Rh values. However, again like the R and R^2 – this trend is flipped; MAE and RMSE reach their minima near Rh50-Rh60. This corresponds with the trend seen in Figure 7.11 where Rh50- Rh60 approximate the true Field heights.

Inter-model trends

- a) Increasing the data (model_2019_s, model_2019, model_2020 in that order) does not decrease the MAE or the RMSEs in the field data unlike what would be expected – in fact, they are about the same.
- b) However, the testing set model_2019 performs to expectations – the MAE, RMSE all decline in the testing set compared to model_2019_s which had just one transect from 2019.
- c) Model 2020 (trained on 16 transects) performs the worst in terms of R^2 and R, followed by model_2019_s (trained on just 1 transect) and lastly model_2019.

With regards to the night-time beam quality filtration, the trends in Figure 7.22 show that whilst the R^2 increased, the model still performed worse than model_2019. The night-time model's results were therefore omitted to make a final composite score however, this could be due to unrelated factors to do with Sentinel-2 data.

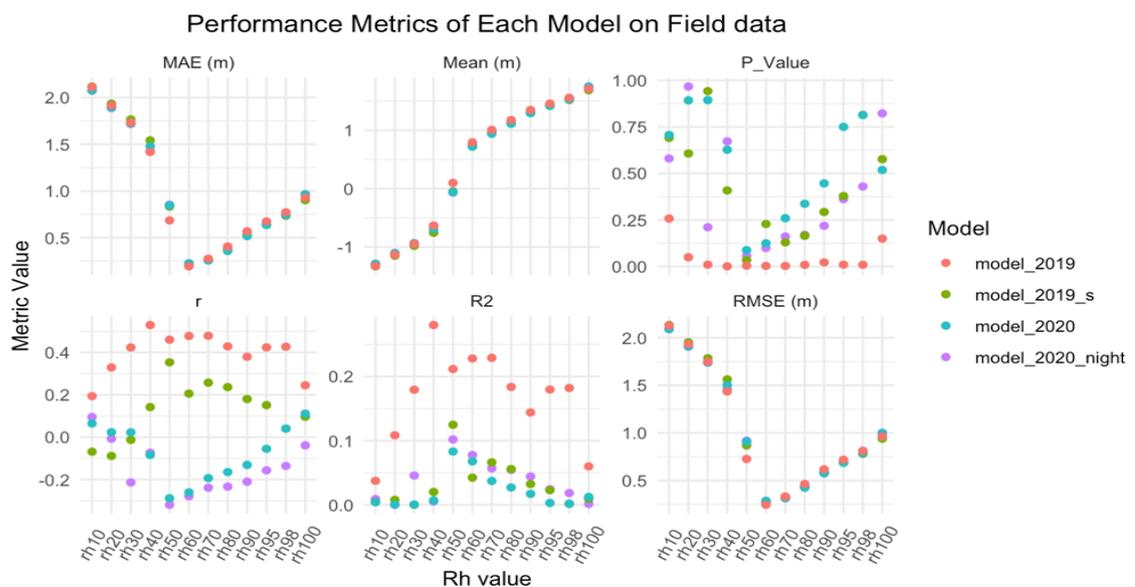


Figure 7.22: Filtering model_2020 for night-time shots.

Composite scores simplify a lot of model evaluation and make trends more salient. Each metric is first adjusted so that lower values are better; for instance, MAE, RMSE, and P-value are inverted by multiplying by -1. These adjusted metrics are standardized through z-score scaling to be on a common scale. Pre-assigned weights are then applied to each metric: R^2 is given a weight of 0.7 as monitoring trends in the field is important, while MAE, RMSE, and P-value each have a weight of 0.4. Finally, a composite score is computed for each model and Rh variable by summing up the product of the z-score and the corresponding weight for each metric. This composite scores in Figure 7.23 and Figure 7.24. provides a holistic measure of model performance, facilitating a more direct comparison between the results in Figure 7.7 and Figure 7.8.

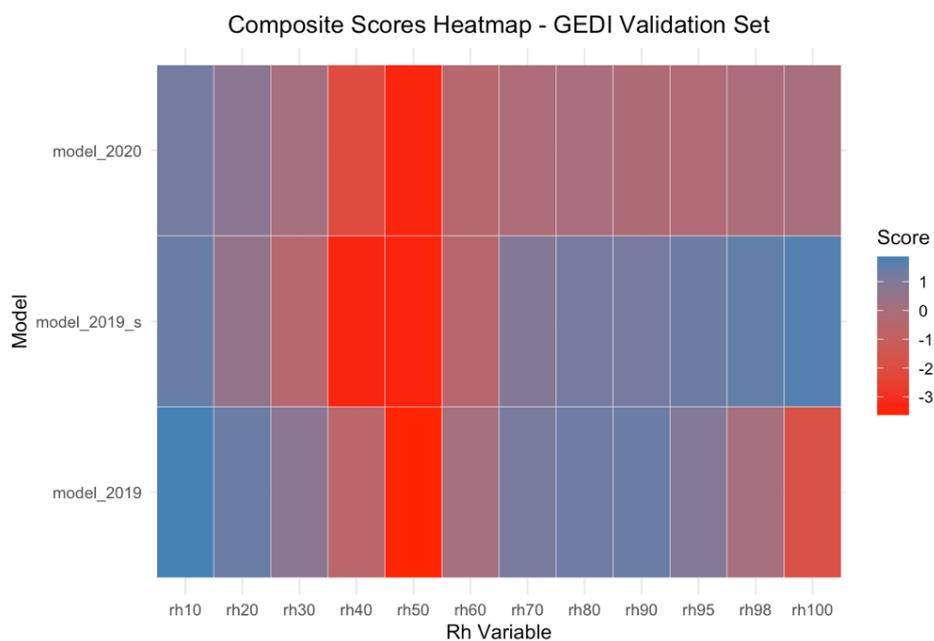


Figure 7.23: GEDI validation set metrics.

The specific scores for each of these models are available in Appendix 12.3. The GEDI validation set in Figure 7.20 and Figure 7.23 shows that the predictions are worst for Rh50, a very different trend from the Field data in Figure 7.24 where median values like Rh50-70 have higher composite scores.

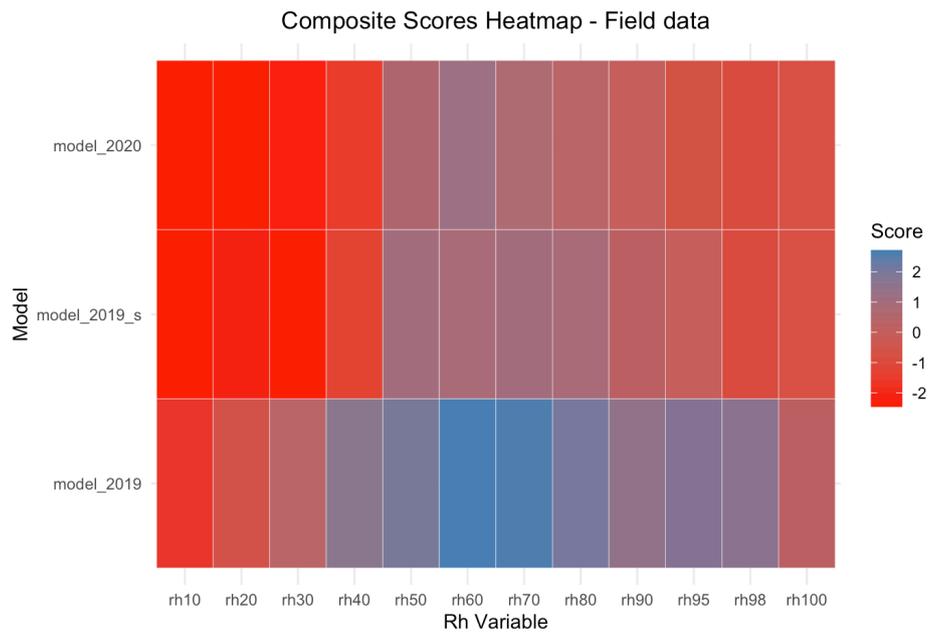


Figure 7.24: Field validation set metrics.

The optimal Rh metric identified that most closely approximates mangrove field heights is Rh60 with the highest composite score. Across all models in Figure 7.24, Rh60 consistently registered the lowest values for both MAE and RMSE in the Field data. Among the three models assessing the Rh60 predictor, the model-2019 exhibited the most promising statistical indicators, with the most significant P-value and the highest R^2 value ($R^2 = 0.2$). The best models across the testing set are for model_2019_s Rh100 and model_2019 Rh 70-80. However, field data will be used to choose the best-performing model. Rh60 from the model_2019, which utilized 8 GEDI transects traversing the site on 22 May 2019, will be the reasonably preferred choice, fulfilling Objective 2 of this research.

7.2.4. Examining Rh60

In this section, Rh60 from model_2019 and its importance scores from the RF model will be explored.

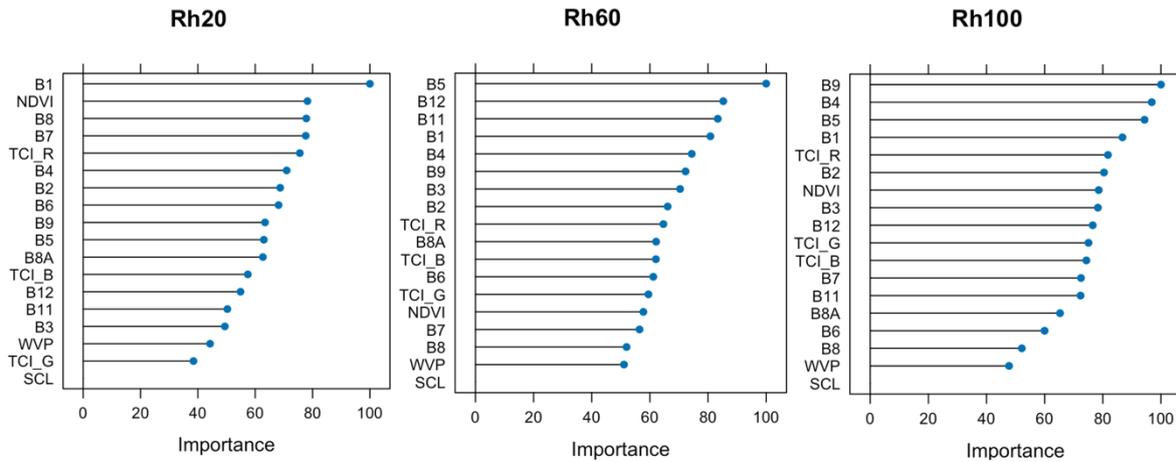


Figure 7.25: The Importance Scores of some Rh variables in the trained RF model.

The importance scores indicate distinct predictor preferences between Rh20, Rh60 and Rh100. B1, Coastal aerosol is useful in distinguishing between the coastal and inland water observations, and features as one of the top 4 predictors in all the Rh metrics shown in Figure 7.25.

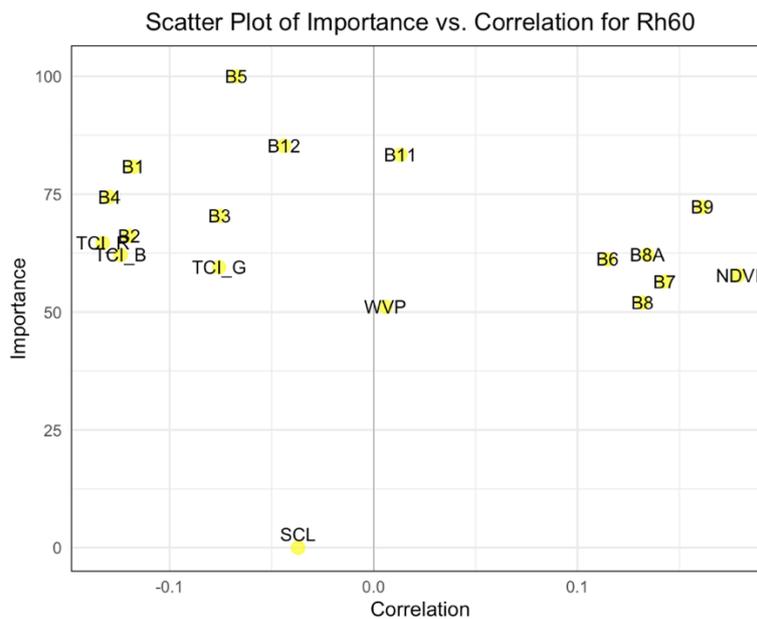


Figure 7.26: The importance score vs. correlation for the Rh60 model.

Of great significance is B5 - The Vegetation Red edge, which has widely been noted in the literature to be sensitive to chlorophyll content and vegetation trees (Li et al., 2020). Additionally, B11 and B12, corresponding to Shortwave infrared bands are useful for soil and mineral mapping and also for vegetation moisture content. This could indicate that the moisture content of vegetation is crucial for prediction Rh60. This will serve as this study's final resultant predicted Field heights in subsequent analyses when it comes to comparing with other available baseline predictions, Objective 3 of this research.

7.3. Objective 3: Comparison with Baseline CHMs

A brief overview of the available baseline CHMs near the sample field plots of mangroves along with their heights in 2019 are shown in Figure 7.27, while their distributions are shown in Figure 7.28.

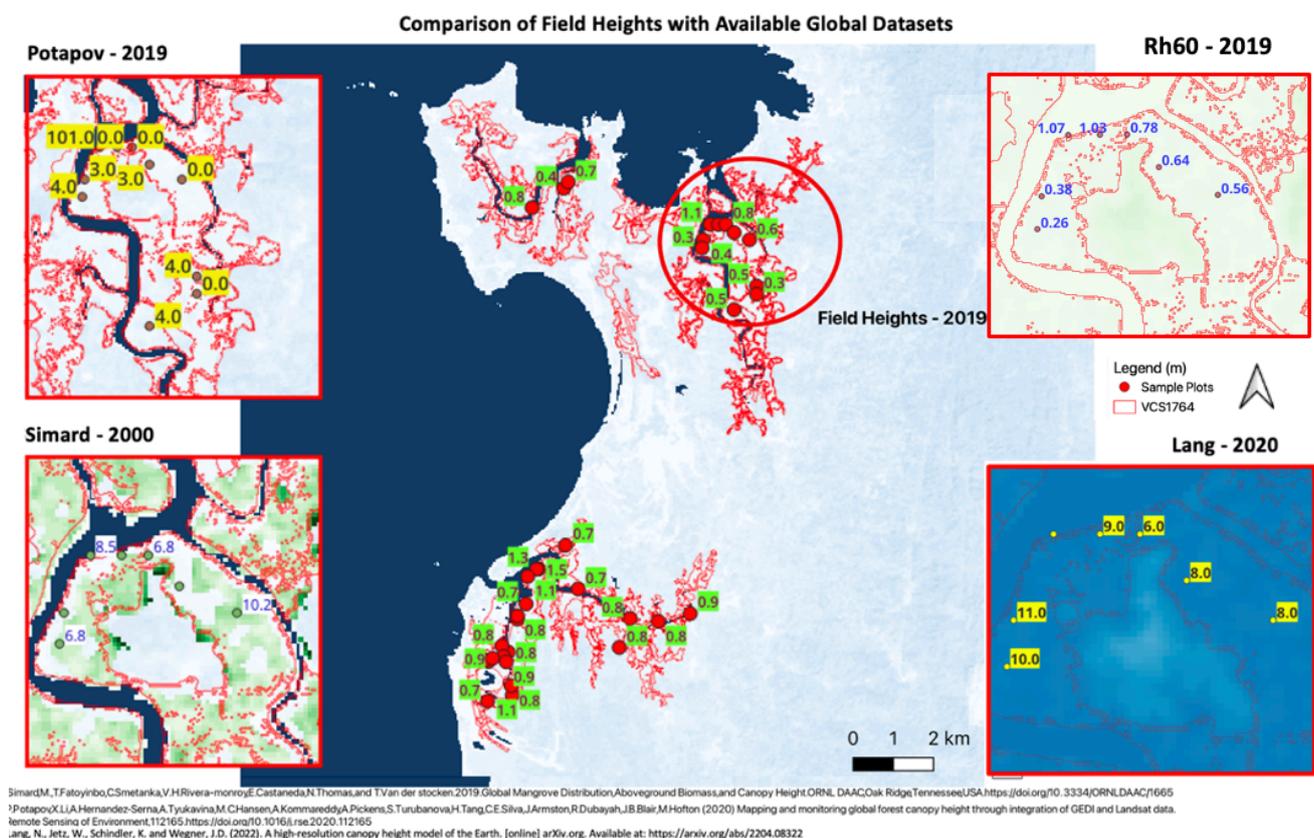


Figure 7.27: Sample values of CH (m) from Rh60-2019, the 3 baseline CHMs & Field Heights.

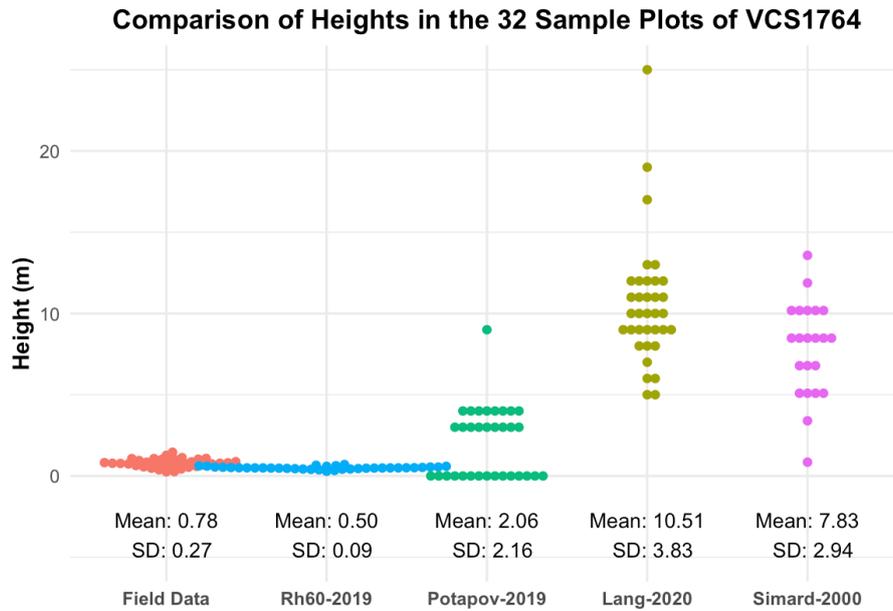


Figure 7.28: Distribution of CH amongst all models; Rh60-2019 & the 3 global CHM baselines.

In Figure 7.29, this information is summarised by comparing the results from Rh60-2019 with those available from the 3 baseline global (Potapov-2019, Lang-2020 & Simard-2000) CHMs by plotting a linear regression line between them and the Field height.

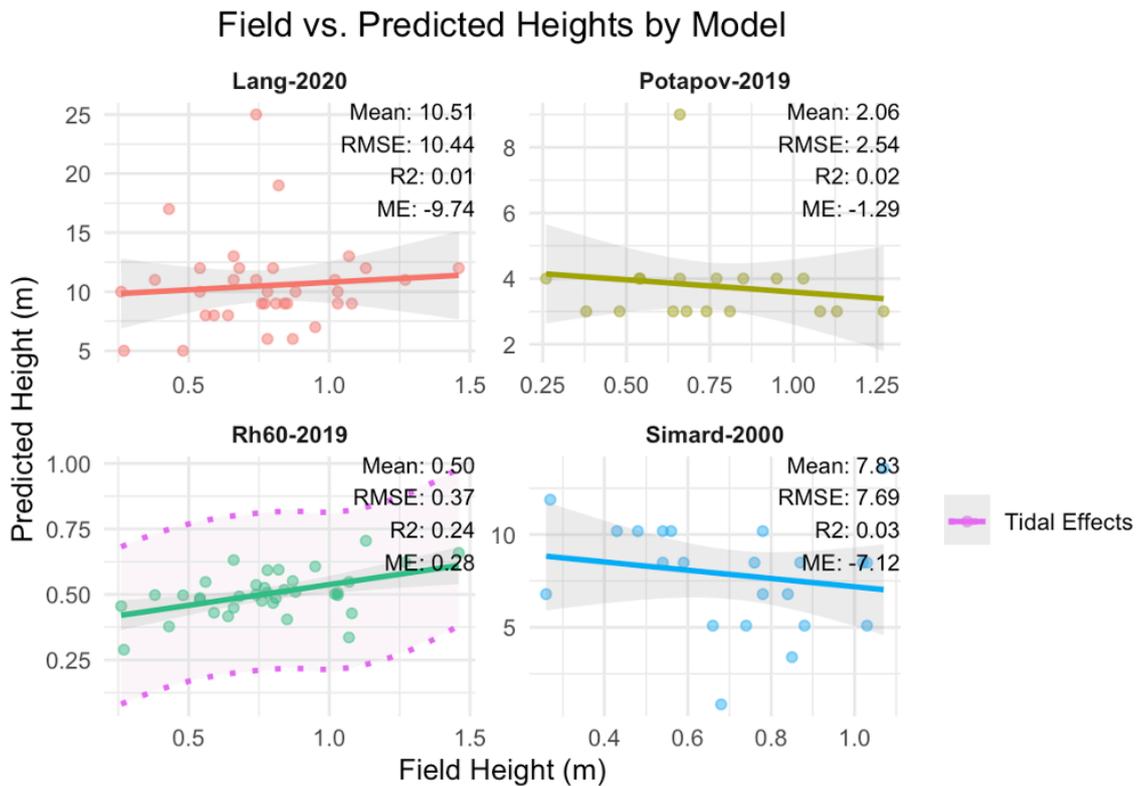


Figure 7.29: Comparison of Field heights with predictions from baseline CHM models and newly created Rh60-2019.

All 3 baseline CHMs when investigated in VCS1764 using QGIS, had many NA pixel values (with the most in Simard-2000), which could mean the mangroves weren't classified/detected as forest. From an initial view of the 2019 shots, after knowing field data, it is surprising to notice that Potapov-2019 - meant to class forests less than 3 m as 0 still had CH values for the < 2 m mangroves. Potapov-2019 is however consistent if the 95th GEDI shot (which the study used) was defined as being the CH, which had a mean of 4m in VCS1764 according to the analysis shown in Figure 7.15.

7.3.1. Comparison with local data

To compare the results from the best model with field data, the distribution of heights for the 32 sample fields in Figure 7.30 is shown. The distribution shows that the heights in the field plots are normally distributed with a mean height of 0.78 m.

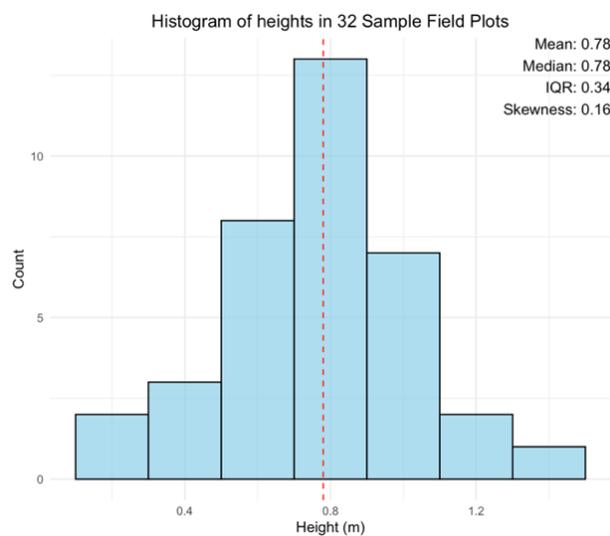


Figure 7.30: Histogram of Field heights measured in 2019 (Vanniarachchy and Jayakody, 2020).

After extracting model Rh60-2019 as the final output of this study, a predicted mangrove CHM in VCS1764 was created as shown in Figure 7.31.

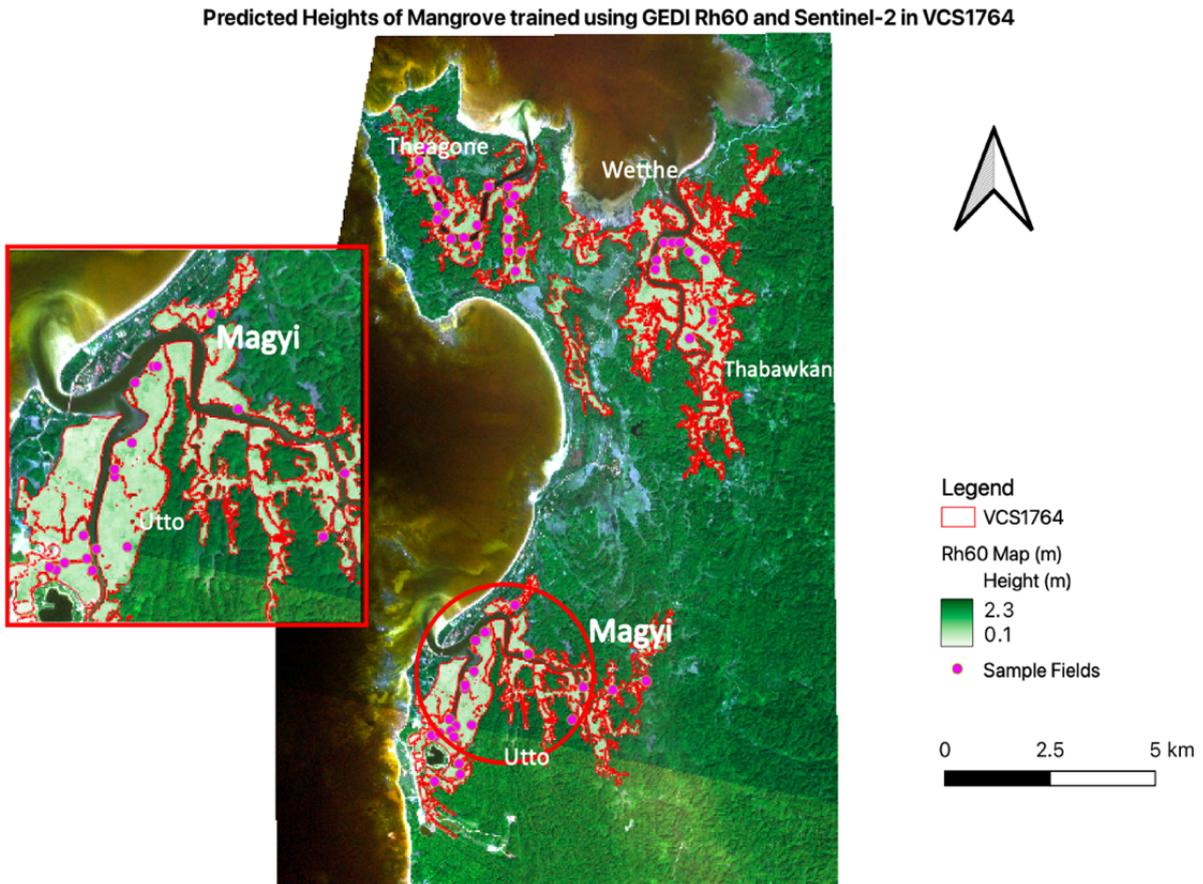


Figure 7.31: Final Model Rh60-2019 Predicted Mangrove Height Map at 25 m resolution.

A sample distribution of predicted heights using Rh60-2019 within VCS1764 is shown in Figure 7.32. The predicted heights for a 25 m resolution map show a mean of 1.01 m and are skewed to the left.

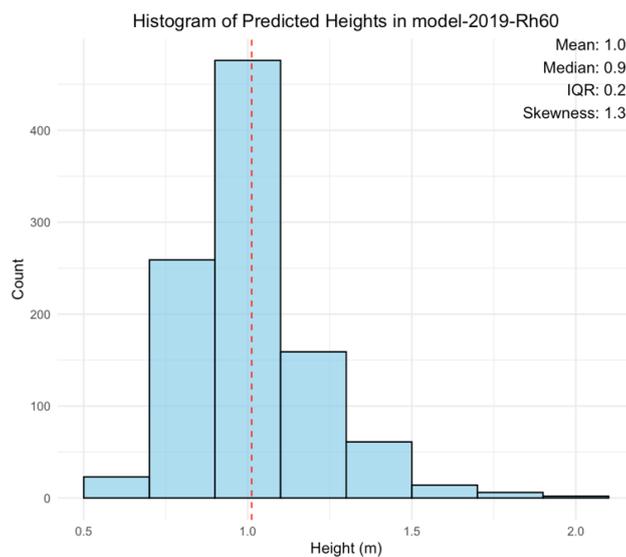


Figure 7.32: Distribution of mangrove heights in VCS1764 using the best-performing model from Rh60-2019.

The main constraint of the available dataset is the lack of direct spatial extent knowledge for the intact mangroves in the region. However, there is an indication that intact mangrove heights in the Magyi Tidal Creek could be up to 1.6 m according to Oo, (2020) and between 1.1. and 2.3 m according to Win et al., (2021) (See Section 2.4). In the predicted data, there are very few instances of places with heights over 2 m in the Magyi Tidal Creek when examined under QGIS– which if the field data of Win et al., (2021) were to go by, would be an underestimate of the true intact mangrove heights. Given that the Rh60 approximately represents the median value, and that only Field height measurements from restored regions were chosen in this model, insights from considering the heights of taller intact mangroves have been omitted entirely.

8. Discussion

8.1. Methods and Model

Lang et al., (2022) noted that the CHs of mangroves have a unimodal distribution with a large spread and heights ranging up to 40 m. This is in agreement with this study’s findings; where restored mangrove was also found to have a unimodal distribution. Although the region was constrained to only those comprising mangroves, further improvements to account for geolocation errors could involve taking a buffer 10 m from the edge of the mangrove to ensure only shots within the mangrove are being used, similar to the method employed by Guerra-Hernández and Pascual, 2021.

It was also observed that there were valid Rh metrics over bare water in the L2A product (unlike the L2B product, where bare water is filtered by NASA). Cropping the GEDI shots to only the mangrove region, while reducing most of the shots that comprised of the Andaman Sea, may not have eliminated all of them. A way to deal with this would be to completely filter shots within a certain meter distance of the shoreline after examining inundation from historical imagery. Additionally, more research into the effects that sparse canopy cover has on flooded young mangroves needs to be conducted. Bare water might be a high fraction of the waveform energy return, and it might be significant to separate the contribution of mangrove return height metrics from that of water. High Rh values greater than 3 m have also been noted to occur in unusual places

such as sand dunes in Saudi Arabia, Alpine grasslands in New Zealand, and an Urban centre in Australia (Potapov et al., 2021, see Fig. 4), which may need closer analysis to be addressed in our study area. On a separate note, one more way of filtering GEDI shots that has been reported in literature is to filter shots with large differences between the elevation of ground return and the TanDEM-X data within the shot. This can help in identifying significant discrepancies/ or help in choosing a better algorithm setting (a1-a5) (Hancock et al., 2021).

All the models used the median Sentinel-2 image that spanned over several months depending on the yearly model. A potential improvement in methodology would be to have smaller time windows in which the Sentinel-2 images are sampled and fit the GEDI time windows more accurately and compare the individual models for each transect in an ensemble. A preliminary look into the available Sentinel-2 images showed that cloud cover was more than 30% in most of them in 2020, which means the process might not be straightforward. While Sentinel-2 provides consistent day-time observations, GEDI's most accurate data points are usually captured at night. This disparity in observation times can introduce variability, especially in dynamic environments where there are diurnal tidal effects and more time and scrutiny might be needed to filter shots to regions further inland. Seasonality has also been noted to change GEDI's accuracy in predicting heights in some types of forests (Rajab Pourrahmati et al., 2023). This has reasons relating to phenology and leaf structure however, whether such seasonality has similar effects in mangroves is still understudied in literature.

On another note, the computed index NDVI, though a reliable indicator of mangrove health, requires caution when interpreting it in wet conditions and coastal conditions with bare soil. When soil gets wet, it gets darker and reflects less red light and less near-infrared light, thus increasing the NDVI without a corresponding increase in vegetation cover that could skew results. Perhaps a better index that could be computed that takes into account this effect and is well suited for sparsely vegetated areas is the SAVI - Soil Adjusted Vegetation Index (SAVI). NDVI also has an effect of saturation – where increasing canopy/ density would not lead to a corresponding increase in NDVI value (Pettorelli et al., 2023). Carrying out more targeted feature engineering, i.e. creating newer and relevant features could help in getting higher R^2 values.

In terms of the model itself, the variables were scaled by 1/3. Though RF algorithms don't theoretically need scaling, the reason they might perform better once scaled is because of noise reduction that increases the regression algorithm's accuracy. Additionally, an R^2 of 0.24, while

better than the baseline Global CHMs, could be improved through other machine-learning techniques. Deep learning (DL) models for instance have been reported to perform better than RF - a study that measured the CHs in China using Sentinel-2 and ICE-Sat found an R^2 of 0.78 for a DL model compared to 0.68 for a RF model (Li et al., 2020).

8.2. Results

8.2.1. Objective 1: Explore GEDI products over VS1764

Across all GEDI products L1B, L2A and L2B, evidence of uniformity of the vegetation structure was found. L1B waveforms and L2B PAVD show less variation than those in typical forests and a vertically uniform structure (See Figure 5.3). In L2A, the Rh metric that best approximates the true field height ranges from 50-65% in Figure 5.12. This could be because in environments where vegetation demonstrates a uniform vertical distribution, the Lidar signals or "bounces" are often quite evenly dispersed throughout the vertical column of the canopy. This uniform distribution means that as the laser pulse penetrates the canopy, it encounters roughly consistent surface areas to interact with as it moves downward. This phenomenon was reflected in the waveform profile captured by full waveform Lidar data (See Figure 7.3). Instead of having pronounced peaks (which would indicate layers of the canopy where the foliage is more dense, like in the Californian Redwood trees), the waveform graph is relatively even and more or less displays a steady decline from the first peak (registered as the ground signal), representing the equal distribution of vegetation at different vertical layers (Harding et al., 1999).

Among the GEDI L2B products, other products that could be explored for further study are canopy cover, Foliage Height Diversity (FHD) and Area Index (PAI) FHD quantifies the vertical heterogeneity of the foliage profile while LAI quantifies the proportion of horizontal leaf coverage in an area. Canopy cover, defined as the fractional area of forest covered by the crown of an individual plant species can be a different indicator for understanding the health of a mangrove ecosystem, particularly young ones that have smaller canopy crowns than developed mangroves. The fractional canopy cover metric could serve as a direct indicator of the area's extent and the health of the young mangrove population, allowing for easy monitoring and comparison to more developed mangrove areas. Further, exploration of GEDI L4 products that measure AGB could capture more information about carbon stocks in mangroves than CH alone.

However, applying GEDI's AGB and derived metrics directly can be laden with uncertainty if products are not locally calibrated. A key study by Duncanson et al., (2022) focused on the development of models for GEDI footprint-level AGB Density (AGBD). This study leveraged both spatially and temporally coincident field and ALS datasets, finding that the inclusion of Rh98, Rh90, and Rh50 significantly improved model accuracy. However, the study did not consider mangroves as a specific plant functional type of their own and grouped them under Evergreen Broad Trees (EBT) in Myanmar's case. The ground data that was used to calibrate GEDI L4 AGB products revealed the closest country near Myanmar that was used to calibrate the products was Indonesia (NASA, 2023b). This leaves a gap in our understanding of how GEDI's own L4 AGB products perform for mangrove ecosystems and presents an opportunity for future AGB comparisons across the available GEDI L4, with other global sources of AGB estimates e.g. that of Simard et al., (2019).

8.2.2. Objective 2: Create a Mangrove CHM

Quite unexpectedly, model-2020, trained on the highest number of shots (765) performed worse than both model-2019 (225) and model-2019_s (126) (see Table 6.1 and Figure 7.24). This could be due to the Sentinel-2 period that was chosen to train the GEDI shots or from varying phenological structure of mangrove detected by the GEDI shots, but the reason cannot be ascertained with confidence for now. The most significant discovery of this study has been that Rh60 might be better in predicting mangrove heights in short-stature forests, in agreement with the initial waveform exploration and analysis of direct GEDI shots near the field highlighted in Section 8.2.1. The strongest predictor was the Red Edge, Band 5, a finding also consistent with that of Luo et al., (2023) who fused ICESat-2 (also a spaceborne lidar) and multisource imagery which included Sentinel-2 to map forest height in China.

This could be due to;

- a) **Distinct Understory Structure:** Unlike typical forests, the mangrove species planted at the study site exhibit "diffused spreading branches," as observed by Oo, (2020) or multiple vertical layers (understory, midstory, overstory). This unique growth pattern could result in a different understory structure from typical forests, which might be more accurately represented by Rh60 rather than Rh90+. Furthermore, it was observed that the upper canopy density, as captured by the PAVD in Figure 4.13, did not vary significantly from the lower

layers. Since Rh95 values are extremely sensitive to small changes and outliers in CH, their use could significantly distort estimates in forests with shorter trees. Rh60, being less sensitive to such noise and outliers, might offer a more reliable central estimate for such ecosystems.

Additionally, Lagomasino et al., (2016) reported that ALS-based methods overestimated CHs by up to 5 m in the Zambezi River delta, exhibiting higher variability compared to SRTM. Interestingly, it was the mean ALS-derived CH—not the Rh100—that closely aligned with field-based measurements (see Table 1, Lagomasino et al.,2016). Lagomasino et al., (2016) also found that the ALS lidar H100 heights were taller by 3 m, whereas this study found a similar overestimate (> 2 m) (See Figure 7.15). This suggests that the use of Rh60 isn't an isolated case in the literature, but a detail that has yet to be incorporated into current baseline CHMs—a gap that can now be filled with more accurate information.

- b) **Methodological Limitations:** The deductions made in the study could be affected by several factors, including the lack of clearly labelled fields to differentiate between intact and non-intact mangroves. Since this model was trained on heights of only restored/ replanted mangroves, those that are intact – with typically greater CH may not have been estimated (See Section 7.3.1). However, there is evidence from the field photographs (See Figure 5.1 a-f) that the intact mangrove is very near/fragmented throughout the restored/replanted mangrove. More field data/ time series analysis to demarcate both is needed to draw further conclusions. A logical next step would involve distinguishing between intact and degraded mangroves using satellite imagery combined with time-series analysis. Preliminary assessments with Google Earth Pro have already revealed discernible changes between the years 2000 and 2020, offering a promising avenue for future research and categorization efforts.

Additionally, the models show slightly different patterns in the testing set compared to training set (see Figure 7.23) this might be because of overfitting but more likely due to differing underlying distributions in the Field data. Even though our testing set implies comparatively lower predictability of Rh60 than Rh70 for model_2019, it does not imply that Rh60 is a wrong choice – just that the confidence in predicting the Rh70 is higher than Rh60 when field heights aren't considered.

8.2.3. Objective 3: Comparison with Baseline CHMs

Apart from the year in which the models provide data, particularly between Simard-2000 with a 19-year gap between the field measurements, differences in the predicted heights between the 3 available baselines can be explained through other factors. The 6 species planted can have maximum heights that range from 15 – 25 m (Allen and Duke, 2006), which may explain why all CHMs overestimated field heights as they could have been measuring intact mangroves instead of degraded ones given that there is a lack of knowledge of the spatial distribution the zones. The MEs were; -1.29 m for Potapov-2019, - 9.74 m for Lang-2020, and -7.12 m for Simard-2000 (See Figure 7.29). However, this conjecture needs to be verified with more field data, as mentioned in Section 8.2.2.

In the final model, dubbed Rh60-2019, an ME of 0.28 m, an RMSE of 0.37 m, and an R^2 value of 0.24 was observed. In comparison, the Lang-2020 model exhibited the highest RMSE, registering at 10.44 m, and also the lowest R^2 value at 0.01. This was followed by the Simard-2000 model, which had an RMSE of 7.69 m and an R^2 value of 0.03, and lastly, the Potapov-2019 model displayed an RMSE of 2.54 m and an R^2 value of 0.02.

- a) Simard-2000 – Simard et al., (2019)'s 30m resolution mangrove CHM precedes the afforestation start date, therefore the heights could be from the former intact regions as the exact period between 1985 and 2000 when the degradation happened isn't known yet. This model was still included as a baseline because it was the only available CHM specific to mangroves. This model had 8 NA values for the 32 field plots, further suggesting its unreliability in being used to predict the mangrove heights, even though this was the only baseline comparator that explicitly targeted mangrove structure. Additionally, it uses Rh95, which might not be applicable for a simple vertical stratified mangrove.

Only five countries out of 115 countries used to get field information in Simard's study have naturally occurring short mangroves. These are Kiribati and the Marshall Islands in Oceania, with mean CHs of 0.8 m; Congo in Africa, with a mean CH of 1.0 m; and the UAE and Bahrain in the Middle East, with mean CHs of 3.7 m and 2 m, and none in the Americas. Additionally, the region-specific model he obtained for Myanmar was calibrated to a mean height of 13.7 m, which may not accurately represent the true distribution in degraded mangroves.

- b) Potapov-2019 – Potapov et al., (2021)'s 30m resolution map used Version 1 GEDI L2A products – known to have 5 times less vertical accuracy than Version 2 used in this study (50 cm as opposed to 10 cm). This study additionally assigned the places where the forest height for the GEDI samples had $Rh_{95} < 3$ m to zero – which could be why there are 3 NA values out of the 32 sites.

- c) Lang-2020 – Lang et al., (2022)'s 10 m resolution map used Rh98, and was optimised around tall trees, as highlighted in Section 2.3, which could have introduced inaccuracies. Like Potapov et al., (2019), this study also used GEDI's Version 1 data as the public release of Version 2 occurred in September 2022 after the study was published in April 2022.

- a) Local data – The underestimation of predicted field heights by Rh60-model-2019 with local data from 2 studies; Win et al., (2021) and Oo, (2020) could be from several possibilities. The field data measured by Win et al., (2021) with the highest predicted mean heights in the creek (2.1m) has no mention of the restored regions or exact physical/spatial characteristics of the trees that were measured to accurately compare results obtained in this study. Due to the highly fragmented nature of mangrove degradation, the proportion and location of intact, degraded and restored mangroves is not known. Mean heights could be very high in a few trees for example, however judging by the fact maximum canopy width of each mangrove is about 0.8 m, they may not cover a large enough spatial extent for GEDI shots with a 25 m resolution to resolve (both measurement and epistemic uncertainty).

The field data with mean heights measured by Oo, (2020) were samples of only 10 trees from each species of the intact mangrove. Whilst this gives a range of heights field heights could have, there is still high uncertainty.

GEDI's 8 m geolocation error captures heights of intact mangroves/vegetation outside the sample field plots (measurement uncertainty), partly enhanced by the differing resolutions of the field plots (10 m by 10 m plots) and the GEDI shots (~25 m).

Uncertainty of the RF model. An R^2 value of 0.24, while higher than the global CHMs (see Figure 7.29) does not capture the full complexity of the system given the fact that the field height correlations are very different from the GEDI height correlations.

8.3. Future work and suggestions

- a) To improve the R^2 of 0.24, more field data, and accurate sampling of Sentinel-2 images specific to the window that the GEDI shot was acquired. Additionally, data from more years (2021 and 2022) could also be compared to the data from 2019 and 2020.
- b) Considering the high costs and time needed to obtain accurate field data in Myanmar, the hypothesis that Rh60 in GEDI is a better predictor of mangrove height could be tested for structurally similar mangroves elsewhere in the world which have openly available ALS data.
- d) An initial examination of ICESat-2 data in VCS1764 suggests its availability and could therefore be compared with GEDI L2A height measurements, offering an alternative avenue for comparative analysis.
- a) A more in-depth examination of GEDI's L2B and L4 gridded AGB, such as FHD and LAI, may provide alternative methods for assessing mangrove growth and health. Indeed, these metrics could serve as more accurate indicators of carbon sequestration, especially considering the varied growth phenologies across mangrove species. A higher CH suggests that current CHMs would overestimate AGB, however, an underestimate is also likely because mangroves have 5 times the carbon sequestration rates than land-based tropical forests (Donato et al., 2011). A report suggested that the Intergovernmental Panel on Climate Change (IPCC) guidelines in 2013 on how to calculate below-ground roots account for just 10% of the biomass of a tree, which was an underestimate for mangrove forests (Evans, 2013). More literature review and field validation of AGB is needed to answer these uncertainties.
- b) A time series analysis using;
 - a. Sentinel-2 data (or other finer resolution Planet data) to map the restored mangrove areas in VCS1764 could be carried out to accurately map the area of mangrove planted. There is evidence that suggests that spectral indices like NDVI and SVI from Sentinel-2 could be used as indicators to map mangrove height directly, which might provide better approximations of true heights, removing the need for a spaceborne lidar product all in all (Ghosh et al., 2020).
 - b. GEDI data fused with Sentinel-2 to map CH changes could be carried out. Though data spans from only 2019-2022, it could theoretically map 0.2 cm changes in

heights (the average yearly increase reported in VCS1764) directly however it is important to be aware that errors could compound from data fusion.

- c) Finally, though GEDI has currently halted its operations, it is set to resume in 2024. Other upcoming satellite launches such as the NASA-ISRO Synthetic Aperture Radar (NISAR), the Surface Water and Ocean Topography (SWOT), and BIOMASS missions could provide more information about the AGB and water changes that can all be helpful in more accurate monitoring of coastal environments. A proposal is also currently underway to expand spaceborne lidar capabilities to offer continuous coverage (The Global Altimeter MISSION: GLAMIS) however its time horizon is uncertain (Hancock et al., 2021).

9. Conclusion

To conclude, some key reflections on the objectives that were set in Section 3.2 are;

- a) Across all GEDI products; L1B, L2A and L2B, evidence of the uniformity of mangrove vegetation structure in VCS1764 was found. GEDI's L1A waveform shows a distinct spectral signature over mangroves, thus it can be further justified to explore derived products like L2A and L2B. The L1B waveforms show less variation than those in typical forests (See Figure 7.3). In L2A, the Rh metric that best approximates the true field height ranges from 50-65% in Figure 7.11. GEDI's L2B product shows that mangroves have fewer understory variation differences than a typical forest in Figure 7.12.
- b) GEDI's L2A metric, Rh60 from model-2019 may be a better approximator for mangrove heights than Rh90+ seen in many global CHM studies. However, this needs to be further validated due to difficulties in mapping the fragmentation of degraded, restored and intact mangroves in VCS1764. Compelling evidence by Lagomasino et al., 2016 suggests mean Lidar heights (close to Rh50-60) might be a better approximation of mangrove heights the Zambezi River delta however, more literature review and field data could help solidify these results.
- c) The Rh60-2019 model created in this study outperformed the others across several statistical measures. Specifically, Rh60-2019 achieved a ME of 0.28 m, a RMSE of 0.37 m, and an R² value of 0.24. Simard-2000 had a ME of -7.12 m and an R² value of just 0.03, suggesting significantly less accuracy and consistency. Likewise, the Potapov-2019 method yielded an ME of -1.29 m and an even lower R² value of 0.02. Lang-2020's approach resulted in the most significant discrepancies, with an ME of -9.74 m and an R² of a mere 0.01. This is

surprising because it had the highest resolution and according to the authors, had higher R^2 and lower RMSE than Potapov-2019.

In summary, this study successfully met several of its key objectives; 1) Assessing available GEDI products in VCS1764, 2) Creating a mangrove CHM for VCS1764, and 3) Comparing the results with baseline CHMs. The vertical profile of short mangroves using GEDI L2B suggested less variation than typical canopies, which prompts a question of whether the use of Rh98 and Rh95 might be not just inaccurate for mangroves, but also for other vegetation types with similar vertical structure e.g., seagrass meadows. It is evident from the literature that the prevalent assumption in leading Global CHM models—which overlooks short forests on the premise of them having low AGB is substantially flawed. This misconception can potentially undermine the effectiveness of a country's estimation of carbon stocks from mangroves, impacting its REDD+ strategy.

The broader implications of this work are manifold: generating precise local estimates of short mangrove forest structure (< 3 m) in VCS1764 not only advances the accuracy of Global CHMs and derived AGB thereafter, but also plays a direct role in enhancing investor trust within the scope of VCS1764, a Verra project. Such trust built through geospatial intelligence is indispensable for fulfilling the goals of the Paris Agreement, improving the implementation of REDD+ strategies and is a timely contribution to the creditability of the Voluntary Carbon Markets.

10. Data availability

A mangrove canopy map using the Rh60th Percentile, along with code will be available for download on GitHub: <https://github.com/excitedmuck/MRes-Mangrove-Height-Detection.git>

11. References

- Adam, M., Urbazaev, M., Dubois, C., Schullius, C., 2020. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sens.* 12, 3948. <https://doi.org/10.3390/rs12233948>
- Adrah, E., Jaafar, W.S.W.M., Bajaj, S., Omar, H., Leite, R.V., Silva, C.A., Cardil, A., Mohan, M., 2021. Analyzing canopy height variations in secondary tropical forests of Malaysia using NASA GEDI. *IOP Conf. Ser. Earth Environ. Sci.* 880, 012031. <https://doi.org/10.1088/1755-1315/880/1/012031>
- Ahmed, S., Sarker, S., Friess, D., Kamruzzaman, M., Jacobs, M., Islam, M., Alam, A., Shuvo, M.J., Nasir, H., Sani, Md.N., Dey, T., Sullibie, C., Naabeh, S., Pretzsch, H., 2022. Salinity reduces site quality and mangrove forest functions. From monitoring to understanding. *Sci. Total Environ.* 853, 158662.
- Allen, J., Duke, N., 2006. CiteSeerX [WWW Document]. CiteSeerX. URL <https://citeseerx.ist.psu.edu/doc/10.1.1.889.7420> (accessed 8.29.23).
- Alongi, D.M., 2014. Carbon Cycling and Storage in Mangrove Forests. *Annu. Rev. Mar. Sci.* 6, 195–219. <https://doi.org/10.1146/annurev-marine-010213-135020>
- Aung, T.T., 2022. Mangroves in Myanmar, in: Das, S.C., Pullaiah, Ashton, E.C. (Eds.), *Mangroves: Biodiversity, Livelihoods and Conservation*. Springer Nature, Singapore, pp. 331–371. https://doi.org/10.1007/978-981-19-0519-3_14
- Aye, W.N., Tong, X., Tun, A.W., 2022. Species Diversity, Biomass and Carbon Stock Assessment of Kanhlyashay Natural Mangrove Forest. *Forests* 13, 1013. <https://doi.org/10.3390/f13071013>
- Ayrey, E., 2022. Elias Ayrey | PhD [WWW Document]. URL <https://www.youtube.com/watch?v=bfj6EkyO77I> (accessed 9.5.23).
- Baltezar, P., Murillo-Sandoval, P.J., Cavanaugh, K.C., Doughty, C., Lagomasino, D., Tieng, T., Simard, M., Fatoyinbo, T., 2023. A regional map of mangrove extent for Myanmar, Thailand, and Cambodia shows losses of 44% by 1996. *Front. Mar. Sci.* 10.
- Bhowmik, A.K., Padmanaban, R., Cabral, P., Romeiras, M.M., 2022. Global Mangrove Deforestation and Its Interacting Social-Ecological Drivers: A Systematic Review and Synthesis. *Sustainability* 14, 4433. <https://doi.org/10.3390/su14084433>
- Bunting, P., Rosenqvist, A., Hilarides, L., Lucas, R., Thomas, N., Tadono, T., Worthington, T., Spalding, M., Murray, N., Rebelo, L.-M., 2022. Global Mangrove Extent Change 1996–2020: Global Mangrove Watch Version 3.0. *Remote Sens.* 14, 3657. <https://doi.org/10.3390/rs14153657>
- Campbell, A.D., Fatoyinbo, T., Charles, S.P., Bourgeau-Chavez, L.L., Goes, J., Gomes, H., Halabisky, M., Holmquist, J., Lohrenz, S., Mitchell, C., Moskal, L.M., Poulter, B., Qiu, H., Sousa, C.H.R.D., Sayers, M., Simard, M., Stewart, A.J., Singh, D., Trettin, C., Wu, J., Zhang, X., Lagomasino, D., 2022. A review of carbon monitoring in wet carbon systems using remote sensing. *Environ. Res. Lett.* 17, 025009. <https://doi.org/10.1088/1748-9326/ac4d4d>
- Dhargay, S., Lyell, C.S., Brown, T.P., Inbar, A., Sheridan, G.J., Lane, P.N.J., 2022. Performance of GEDI Space-Borne LiDAR for Quantifying Structural Variation in the Temperate Forests of South-Eastern Australia. *Remote Sens.* 14, 3615. <https://doi.org/10.3390/rs14153615>
- Donato, D.C., Kauffman, J.B., Murdiyarsa, D., Kurnianto, S., Stidham, M., Kanninen, M., 2011. Mangroves among the most carbon-rich forests in the tropics. *Nat. Geosci.* 4, 293–297. <https://doi.org/10.1038/ngeo1123>

- Dubayah, Ralph, Hofton, Michelle, Blair, James, Armston, John, Tang, Hao, Luthcke, Scott, 2021. GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002. https://doi.org/10.5067/GEDI/GEDI02_A.002
- Duchelle, A.E., Simonet, G., Sunderlin, W.D., Wunder, S., 2018. What is REDD+ achieving on the ground? *Curr. Opin. Environ. Sustain., Environmental change issues* 2018 32, 134–140. <https://doi.org/10.1016/j.cosust.2018.07.001>
- 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., Alonso, A., Andersen, H.-E., Aplin, P., Baker, T.R., Barbier, N., Bastin, J.F., Biber, P., Boeckx, P., Bogaert, J., Boschetti, L., Boucher, P.B., Boyd, D.S., Burslem, D.F.R.P., Calvo-Rodriguez, S., Chave, J., Chazdon, R.L., Clark, D.B., Clark, D.A., Cohen, W.B., Coomes, D.A., Corona, P., Cushman, K.C., Cutler, M.E.J., Dalling, J.W., Dalponte, M., Dash, J., de-Miguel, S., Deng, S., Ellis, P.W., Erasmus, B., Fekety, P.A., Fernandez-Landa, A., Ferraz, A., Fischer, R., Fisher, A.G., García-Abril, A., Gobakken, T., Hacker, J.M., Heurich, M., Hill, R.A., Hopkinson, C., Huang, H., Hubbell, S.P., Hudak, A.T., Huth, A., Imbach, B., Jeffery, K.J., Katoh, M., Kearsley, E., Kenfack, D., Kljun, N., Knapp, N., Král, K., Krůček, M., Labrière, N., Lewis, S.L., Longo, M., Lucas, R.M., Main, R., Manzanera, J.A., Martínez, R.V., Mathieu, R., Memiaghe, H., Meyer, V., Mendoza, A.M., Monerris, A., Montesano, P., Morsdorf, F., Næsset, E., Naidoo, L., Nilus, R., O'Brien, M., Orwig, D.A., Papathanassiou, K., Parker, G., Philipson, C., Phillips, O.L., Pisek, J., Poulsen, J.R., Pretzsch, H., Rüdiger, C., Saatchi, S., Sanchez-Azofeifa, A., Sanchez-Lopez, N., Scholes, R., Silva, C.A., Simard, M., Skidmore, A., Stereńczak, K., Tanase, M., Torresan, C., Valbuena, R., Verbeeck, H., Vrska, T., Wessels, K., White, J.C., White, L.J.T., Zahabu, E., Zraggen, C., 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sens. Environ.* 270, 112845. <https://doi.org/10.1016/j.rse.2021.112845>
- Estoque, R.C., Myint, S.W., Wang, C., Ishtiaque, A., Aung, T.T., Emerton, L., Ooba, M., Hijioka, Y., Mon, M.S., Wang, Z., Fan, C., 2018. Assessing environmental impacts and change in Myanmar's mangrove ecosystem service value due to deforestation (2000–2014). *Glob. Change Biol.* 24, 5391–5410. <https://doi.org/10.1111/gcb.14409>
- Evans, K., 2013. "Dirty Science": Excavating the truth about mangroves and carbon [WWW Document]. CIFOR For. News. URL <https://forestsnews.cifor.org/14197/dirty-science-excavating-the-truth-about-mangroves-and-carbon?fnl=en> (accessed 9.5.23).
- Friess, D., Krauss, K., Taillardat, P., Adame, F., Yando, E., Cameron, C., Sasmito, S., Sillanpää, M., 2020. Mangrove Blue Carbon in the Face of Deforestation, Climate Change, and Restoration 3, 427–456. <https://doi.org/10.1002/9781119312994.apr0752>
- Ganz, S., Käber, Y., Adler, P., 2019. Measuring Tree Height with Remote Sensing—A Comparison of Photogrammetric and LiDAR Data with Different Field Measurements. *Forests* 10, 694. <https://doi.org/10.3390/f10080694>
- Ghosh, S., Behera, M., Paramanik, S., 2020. Canopy Height Estimation Using Sentinel Series Images through Machine Learning Models in a Mangrove Forest. *Remote Sens.* 12, 1519. <https://doi.org/10.3390/rs12091519>
- Giri, C., Ochieng, E., Tieszen, L.L., Zhu, Z., Singh, A., Loveland, T., Masek, J., Duke, N., 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. *Glob. Ecol. Biogeogr.* 20, 154–159. <https://doi.org/10.1111/j.1466-8238.2010.00584.x>
- Greenfield, P., 2023. Revealed: more than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows [WWW Document]. *the Guardian*. URL <https://www.theguardian.com/environment/2023/jan/18/revealed-forest-carbon-offsets-biggest-provider-worthless-verra-aoe> (accessed 1.1.23).

- 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. *For. Ecosyst.* 8, 14. <https://doi.org/10.1186/s40663-021-00291-2>
- Hai, P.M., Tinh, P.H., Son, N.P., Thuy, T.V., Hong Hanh, N.T., Sharma, S., Hoai, D.T., Duy, V.C., 2022. Mangrove health assessment using spatial metrics and multi-temporal remote sensing data. *PLOS ONE* 17, e0275928. <https://doi.org/10.1371/journal.pone.0275928>
- Hancock, S., McGrath, C., Lowe, C., Davenport, I., Woodhouse, I., 2021. Requirements for a global lidar system: spaceborne lidar with wall-to-wall coverage. *R. Soc. Open Sci.* 8, 211166. <https://doi.org/10.1098/rsos.211166>
- Harding, D.J., Lefsky, M.A., Parker, G.G., Blair, J.B., 1999. Lidar Altimeter Measurements of Canopy Structure: Methods and Validation for Closed Canopy, Broadleaf Forests.
- Healy, T.R., 2005. Tidal Creeks, in: Schwartz, M.L. (Ed.), *Encyclopedia of Coastal Science, Encyclopedia of Earth Science Series*. Springer Netherlands, Dordrecht, pp. 949–950. https://doi.org/10.1007/1-4020-3880-1_313
- Hopkins, C., 2014. Lidar archaeology shines a light on hidden sites [WWW Document]. URL <https://www.bbc.com/future/article/20120827-the-laser-archaeologists> (accessed 8.5.23).
- Jennerjahn, T.C., Ittekkot, V., 2002. Relevance of mangroves for the production and deposition of organic matter along tropical continental margins. *Naturwissenschaften* 89, 23–30. <https://doi.org/10.1007/s00114-001-0283-x>
- Lagomasino, D., Fatoyinbo, T., Lee, S., Feliciano, E., Trettin, C., Simard, M., 2016. A Comparison of Mangrove Canopy Height Using Multiple Independent Measurements from Land, Air, and Space. *Remote Sens.* 8, 327. <https://doi.org/10.3390/rs8040327>
- Lang, N., Jetz, W., Schindler, K., Wegner, J.D., 2022a. A high-resolution canopy height model of the Earth.
- Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., Wegner, J.D., 2022b. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sens. Environ.* 268, 112760. <https://doi.org/10.1016/j.rse.2021.112760>
- Li, Q., Bonebrake, T.C., Michalski, J.R., Kit Wong, F.K., Fung, T., 2023. Combining multi-temporal airborne LiDAR and Sentinel-2 multispectral data for assessment of disturbances and recovery of mangrove forests. *Estuar. Coast. Shelf Sci.* 291, 108444. <https://doi.org/10.1016/j.ecss.2023.108444>
- Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., Chen, H., 2020. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *Int. J. Appl. Earth Obs. Geoinformation* 92, 102163. <https://doi.org/10.1016/j.jag.2020.102163>
- Li, X., Wessels, K., Armston, J., Hancock, S., Mathieu, R., Main, R., Naidoo, L., Erasmus, B., Scholes, R., 2023. First validation of GEDI canopy heights in African savannas. *Remote Sens. Environ.* 285, 113402. <https://doi.org/10.1016/j.rse.2022.113402>
- Luo, Y., Qi, S., Liao, K., Zhang, S., Hu, B., Tian, Y., 2023. Mapping the Forest Height by Fusion of ICESat-2 and Multi-Source Remote Sensing Imagery and Topographic Information: A Case Study in Jiangxi Province, China. *Forests* 14, 454. <https://doi.org/10.3390/f14030454>
- Macquarie, R., 2022. Searching for trust in the Voluntary Carbon Markets [WWW Document]. Grantham Res. Inst. *Clim. Change Environ.* URL <https://www.lse.ac.uk/granthaminstitute/news/searching-for-trust-in-the-voluntary-carbon-markets/> (accessed 8.22.23).
- Mitchard, E.T.A., Saatchi, S.S., Baccini, A., Asner, G.P., Goetz, S.J., Harris, N.L., Brown, S., 2013. Uncertainty in the spatial distribution of tropical forest biomass: A comparison of pan-tropical maps. *Carbon Balance Manag.* 8. <https://doi.org/10.1186/1750-0680-8-10>

- NASA, 2023a. Home Page [WWW Document]. GEDI. URL <https://gedi.umd.edu/> (accessed 4.7.23).
- NASA, 2023b. Calibration/Validation [WWW Document]. GEDI. URL <https://gedi.umd.edu/science/calibration-validation/> (accessed 8.29.23).
- NASA, 2020. University of Maryland/ NASA. Mangrove Sci. URL <https://mangrovescience.org/gedi-global-ecosystem-dynamics-investigation/> (accessed 8.8.23).
- Oo, A.Y.Y., 2020. AN ASSESSMENT OF PHOTOSYNTHETIC BIOMASS RELATED TO THE MANGROVE SPECIES FROM THE SHWE THAUNG YAN COASTAL AREA.
- Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C., Stenseth, N.C., 2023. Using the satellite-derived Normalized Difference Vegetation Index (NDVI) to assess ecological responses to environmental change. *Trends Ecol Evol* 20.
- Polidoro, B.A., Carpenter, K.E., Collins, L., Duke, N.C., Ellison, A.M., Ellison, J.C., Farnsworth, E.J., Fernando, E.S., Kathiresan, K., Koedam, N.E., Livingstone, S.R., Miyagi, T., Moore, G.E., Nam, V.N., Ong, J.E., Primavera, J.H., Iii, S.G.S., Sanciangco, J.C., Sukardjo, S., Wang, Y., Yong, J.W.H., 2010. The Loss of Species: Mangrove Extinction Risk and Geographic Areas of Global Concern. *PLOS ONE* 5, e10095. <https://doi.org/10.1371/journal.pone.0010095>
- 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 Sens. Environ.* 253, 112165. <https://doi.org/10.1016/j.rse.2020.112165>
- Rajab Pourrahmati, M., Baghdadi, N., Fayad, I., 2023. Comparison of GEDI LiDAR Data Capability for Forest Canopy Height Estimation over Broadleaf and Needleleaf Forests. *Remote Sens.* 15, 1522. <https://doi.org/10.3390/rs15061522>
- Reuters, 2008. UN says 1.5 million people affected by Myanmar storm | Reuters [WWW Document]. URL https://web.archive.org/web/20140103124710/http://uk.reuters.com/article/2008/05/08/idUKSP48798.CH_.242020080508 (accessed 8.30.23).
- Romañach, S.S., DeAngelis, D.L., Koh, H.L., Li, Y., Teh, S.Y., Raja Barizan, R.S., Zhai, L., 2018. Conservation and restoration of mangroves: Global status, perspectives, and prognosis. *Ocean Coast. Manag.* 154, 72–82. <https://doi.org/10.1016/j.ocecoaman.2018.01.009>
- Roy, D.P., Kashongwe, H.B., Armston, J., 2021. The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. *Sci. Remote Sens.* 4, 100024. <https://doi.org/10.1016/j.srs.2021.100024>
- Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V.H., Castañeda-Moya, E., Thomas, N., Van der Stocken, T., 2019. Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nat. Geosci.* 12, 40–45. <https://doi.org/10.1038/s41561-018-0279-1>
- Simard, M., Zhang, K., Rivera-Monroy, V., Ross, M., Ruiz, P., Castañeda-Moya, E., Twilley, R., Rodriguez, E., 2006. Mapping Height and Biomass of Mangrove Forests in Everglades National Park with SRTM Elevation Data. *Photogramm. Eng. Remote Sens.* 72, 299–311. <https://doi.org/10.14358/PERS.72.3.299>
- Song, C., White, B.L., Heumann, B.W., 2011. Hyperspectral remote sensing of salinity stress on red (*Rhizophora mangle*) and white (*Laguncularia racemosa*) mangroves on Galapagos Islands. *Remote Sens. Lett.* 2, 221–230. <https://doi.org/10.1080/01431161.2010.514305>
- Stovall, A.E.L., Fatoyinbo, T., Thomas, N.M., Armston, J., Ebanega, M.O., Simard, M., Trettin, C., Obiang Zogo, R.V., Aken, I.A., Debina, M., Me Kemoe, A.M., Assoumou, E.O., Kim, J.S., Lagomasino, D., Lee, S.-K., Ndong Obame, J.C., Voubou, G.D., Essono, C.Z., 2021. Comprehensive comparison of airborne and spaceborne SAR and LiDAR estimates of forest

- structure in the tallest mangrove forest on earth. *Sci. Remote Sens.* 4, 100034. <https://doi.org/10.1016/j.srs.2021.100034>
- Su, J., Gastparatos, A., 2023. Perceptions about mangrove restoration and ecosystem services to inform ecosystem-based restoration in Large Xiamen Bay, China - ScienceDirect.
- Tang, H., Armston, J., 2019. Algorithm Theoretical Basis Document (ATBD) for GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics.
- Tidechecker, 2023. Tidechecker [WWW Document]. URL <https://tidechecker.com/myanmar/ayeyarwady/goyangyi-kyun/> (accessed 8.11.23).
- Tommaso, S.D., Wang, S., Lobell, D.B., 2021. Combining GEDI and Sentinel-2 for wall-to-wall mapping of tall and short crops. *Environ. Res. Lett.* 16, 125002. <https://doi.org/10.1088/1748-9326/ac358c>
- UNEP, 2009. Learning from Cyclone Nargis Investing in the environment for livelihoods and disaster risk reduction.
- United Nations, 2023. About the UN Decade [WWW Document]. UN Decade Restor. URL <http://www.decadeonrestoration.org/about-un-decade> (accessed 7.27.23).
- Vanniarachchy, S.A., Jayakody, S., 2020. Reforestation And Restoration Of Degraded Mangrove Lands, Sustainable Livelihoods And Community Development In Myanmar.
- Verra, 2023. Verra [WWW Document]. URL <http://verra.org/about/overview/> (accessed 7.6.23).
- Win, S.-, 2021. Vegetative structure and zonal distribution of true mangroves in Shwe-Thaung-Yan coastal areas, Myanmar. *J. Aquac. Mar. Biol.* 10, 33–39. <https://doi.org/10.15406/jamb.2021.10.00305>
- Woodroffe, C.D., Kumbier, K., Rogers, K., 2020. Use of airborne Lidar to investigate mangrove accommodation space in macrotidal estuaries of northern Australia. *Estuar. Coast. Shelf Sci.* 245, 106988. <https://doi.org/10.1016/j.ecss.2020.106988>
- World Bank, 2023. World Bank Climate Change Knowledge Portal [WWW Document]. *Clim. Change*. URL <https://climateknowledgeportal.worldbank.org/> (accessed 8.21.23).
- Xiong, L., Lagomasino, D., Charles, S.P., Fatoyinbo, T., 2022. Mapping 30 years of mangrove forest height in South Florida by fusion of ICESat-2, GEDI, and Landsat data with machine learning techniques. Presented at the Fall Meeting 2022, AGU.
- Yu, J., Nie, S., Liu, W., Zhu, X., Lu, D., Wu, W., Sun, Y., 2022. Accuracy Assessment of ICESat-2 Ground Elevation and Canopy Height Estimates in Mangroves. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. <https://doi.org/10.1109/LGRS.2021.3107440>
- Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. *Int. J. Image Data Fusion* 1, 5–24. <https://doi.org/10.1080/19479830903561035>

12. Appendix

12.1. Field Heights

a) Distribution of Heights

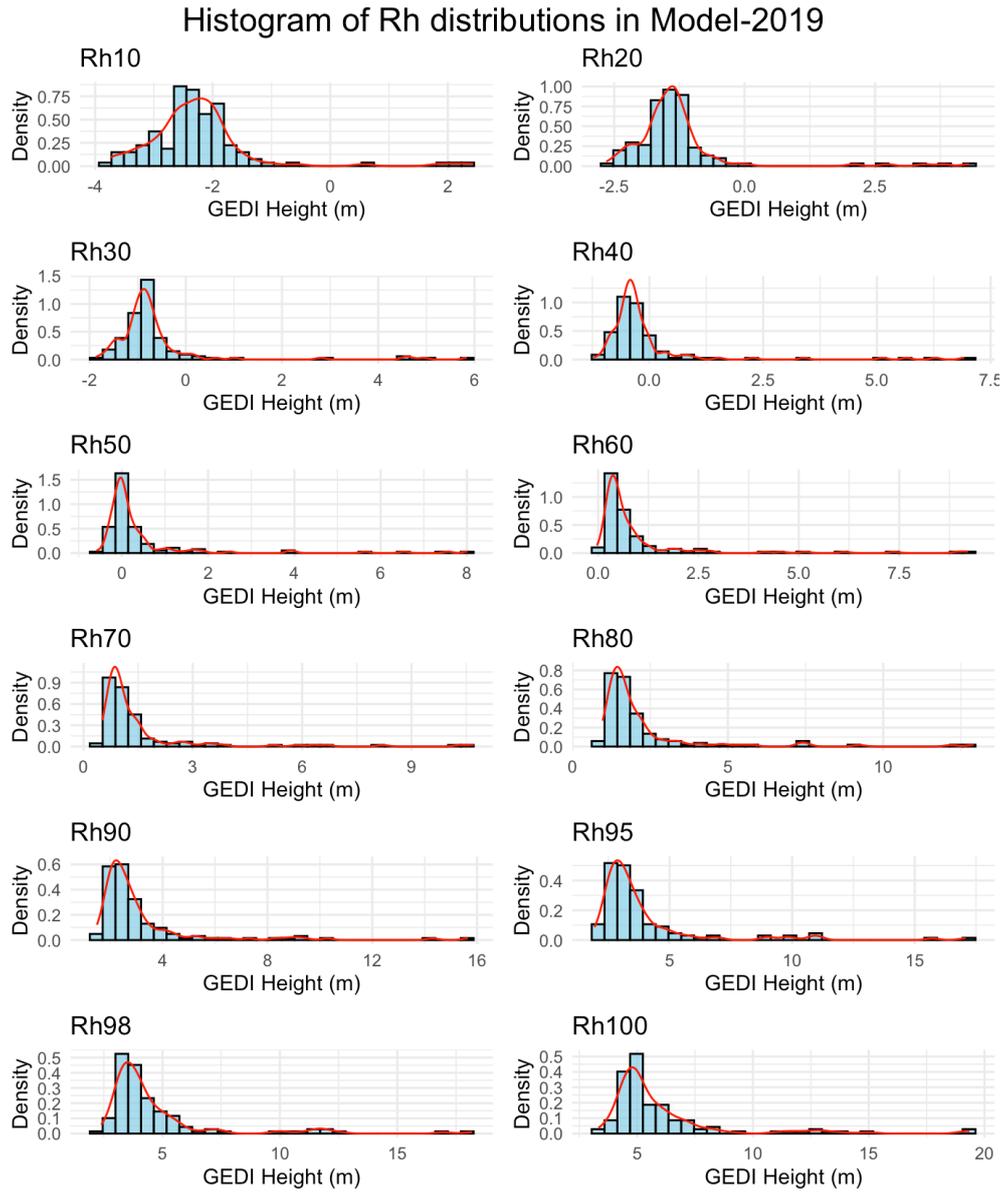


Figure 12.1: Sample distributions of Rh metrics in model-2019.

b) Field heights $^{(0.5)}$

Histogram of Rh distributions in Model-2019

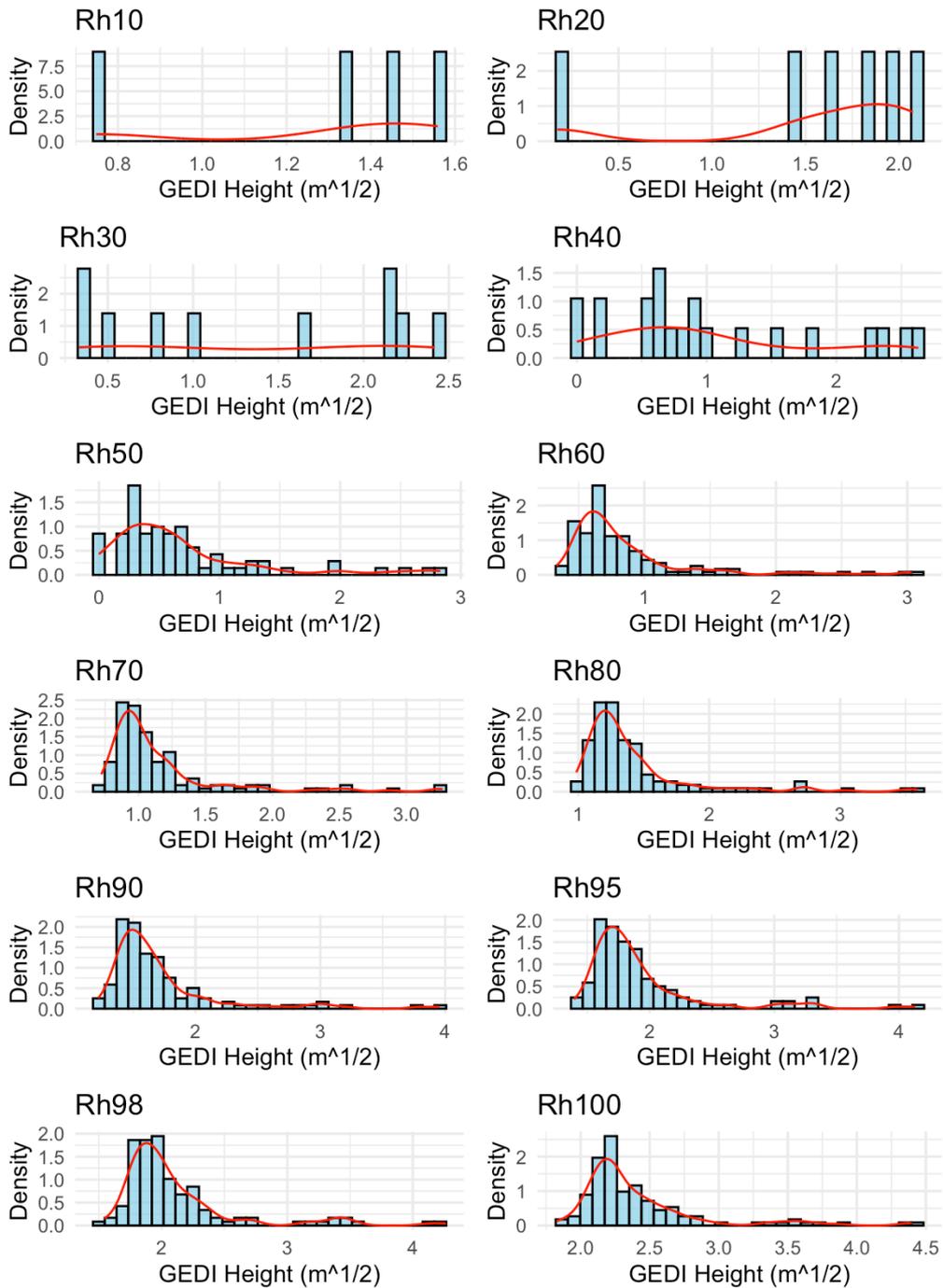


Figure 12.2: Sample distributions of (Rh metrics)^{0.5} in model-2019.

12.2. Model Performance Scores

a) Model 2019_s

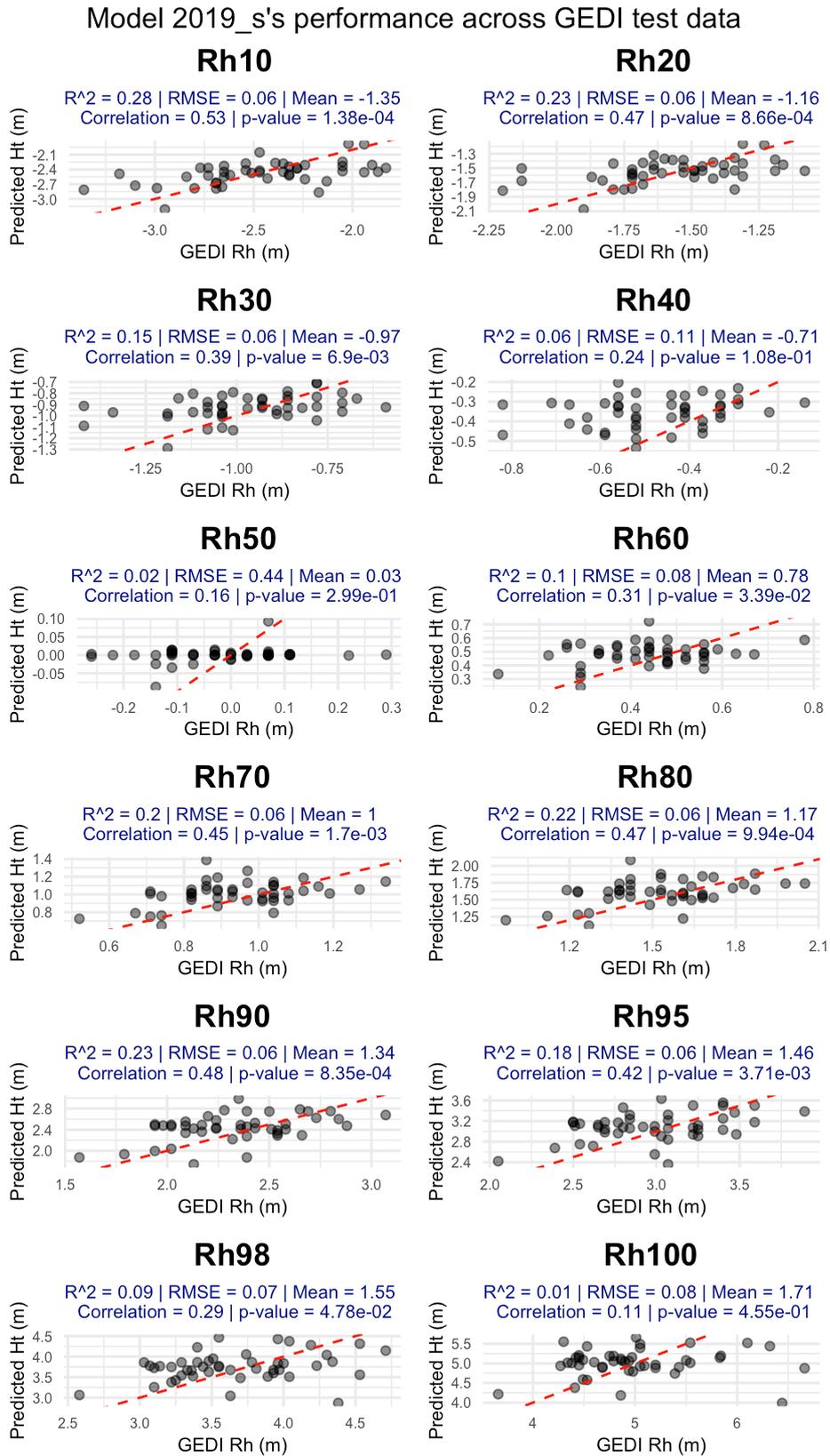


Figure 12.3: Model-2019_s' performance across GEDI testing set.

Model 2019_s's performance across Field data, N = 32

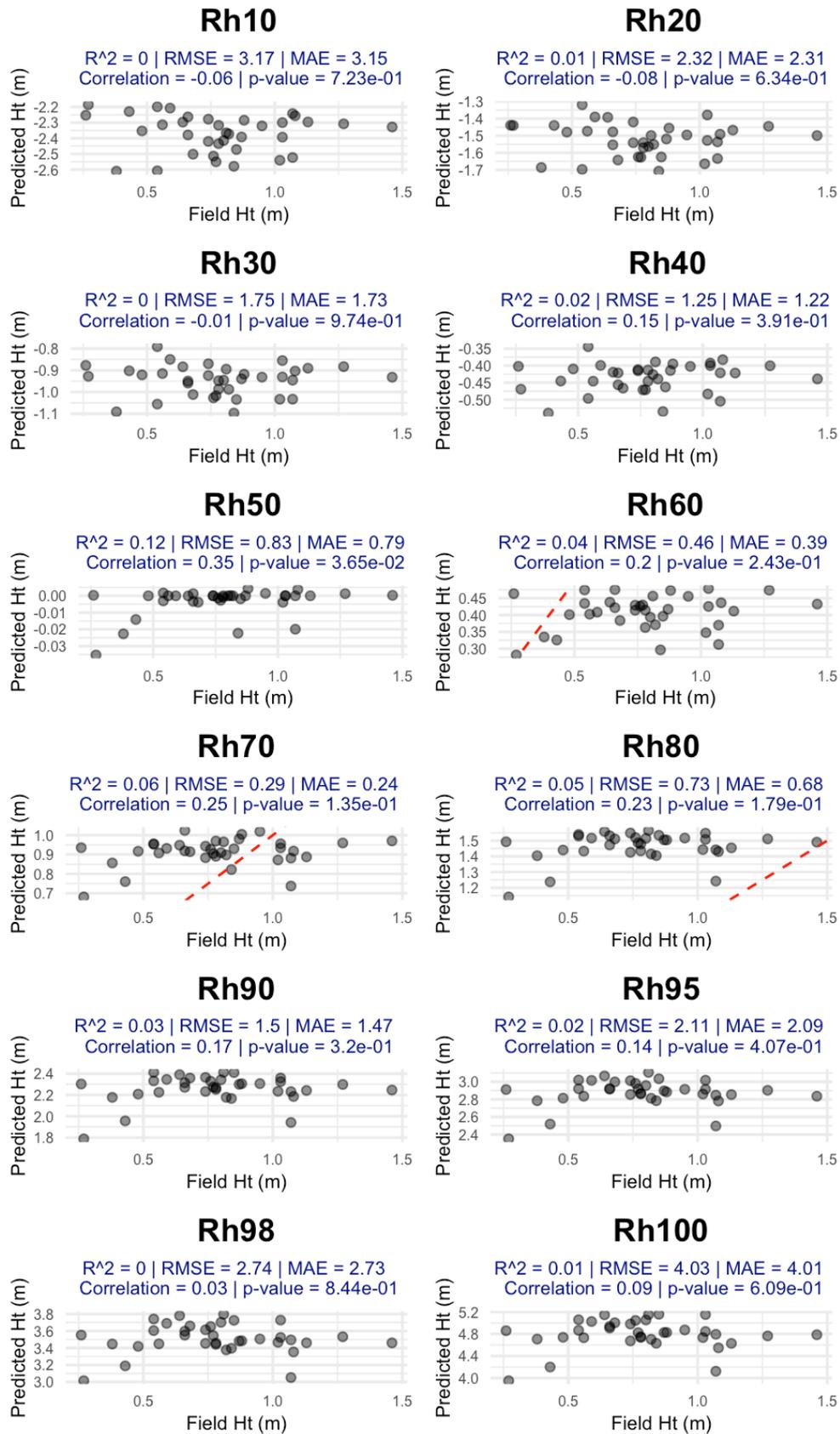


Figure 12.4: Model-2019_s's performance across Field data.

a) Model 2019

Model 2019's performance across GEDI test data

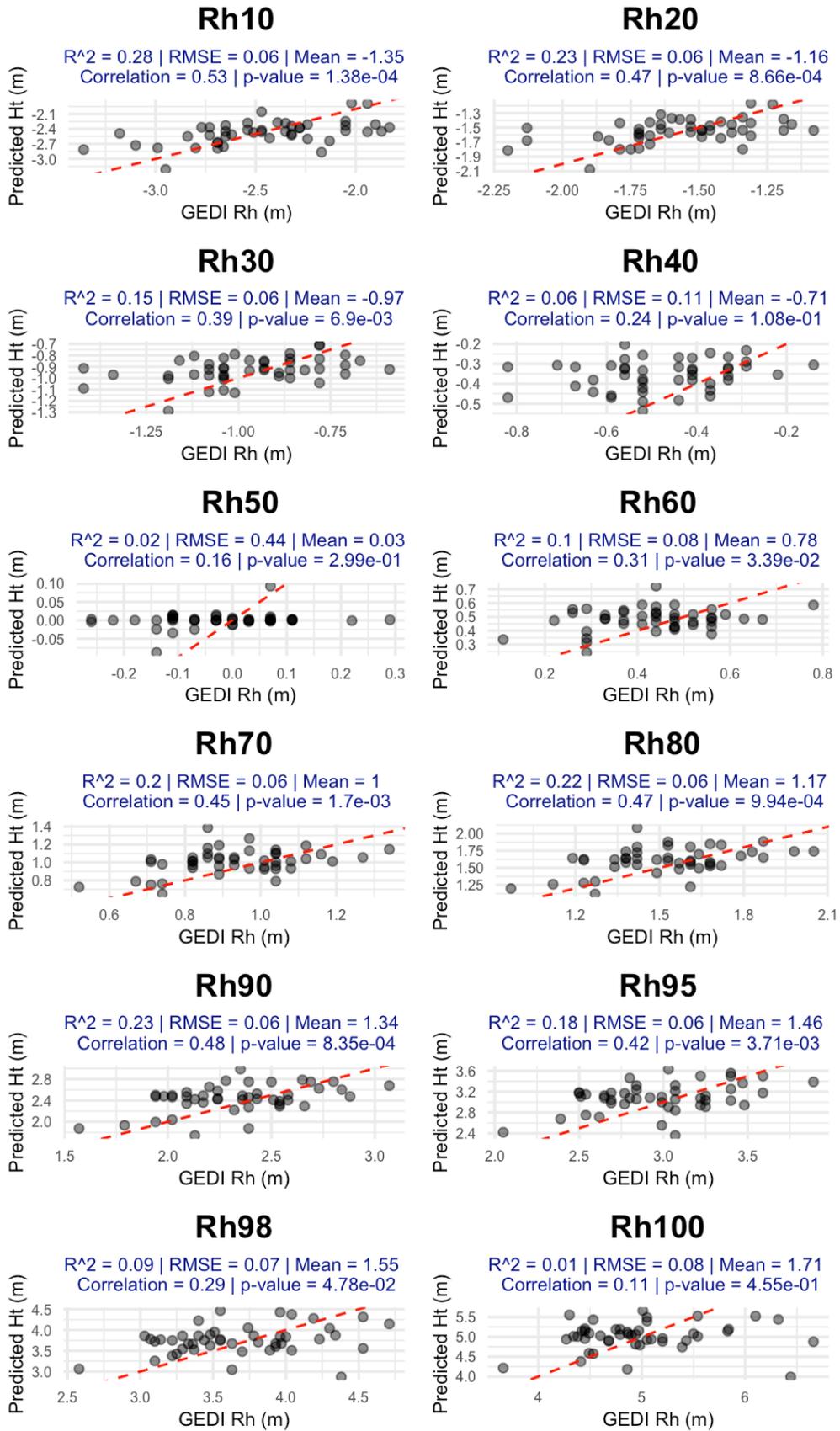


Figure 12.5: Model 2019's performance across GEDI test data.

Model 2019's performance across Field data, N = 32

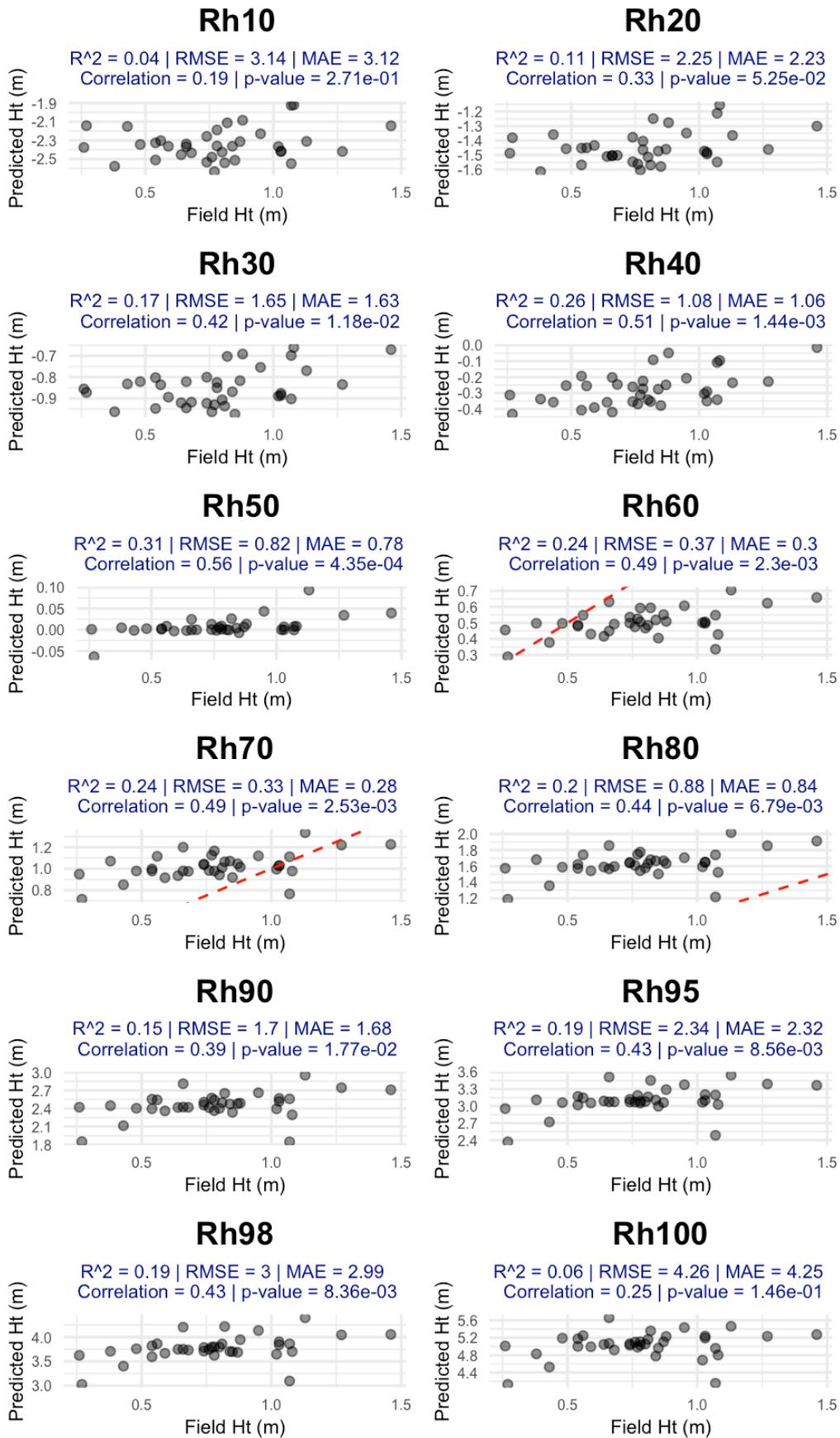


Figure 12.6: Model-2019's performance across Field data.

b) Model 2020

Model 2020's performance across GEDI test data

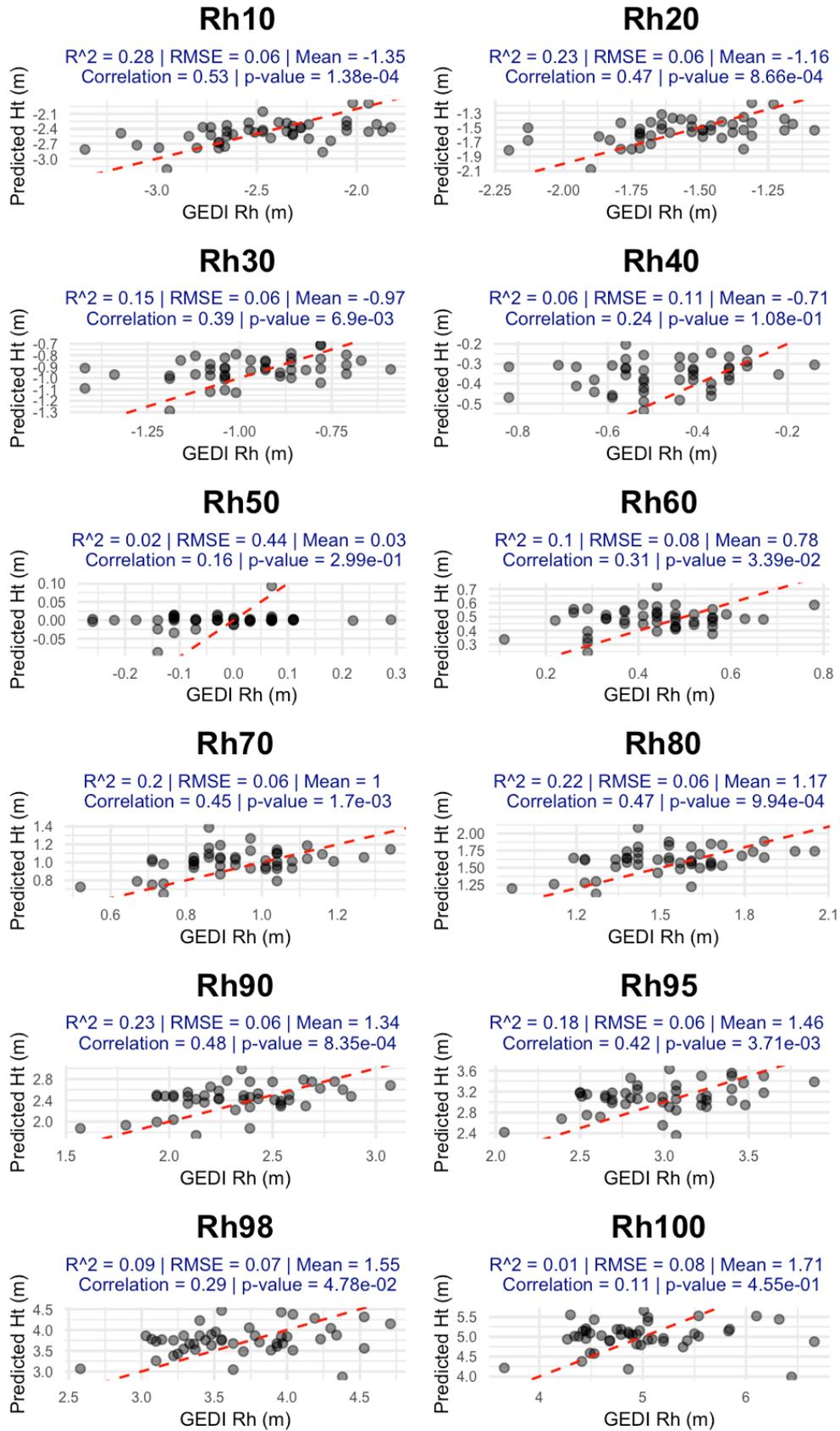


Figure 12.7: Model-2020's performance across GEDI test data.

Model 2020's performance across Field data, N = 32

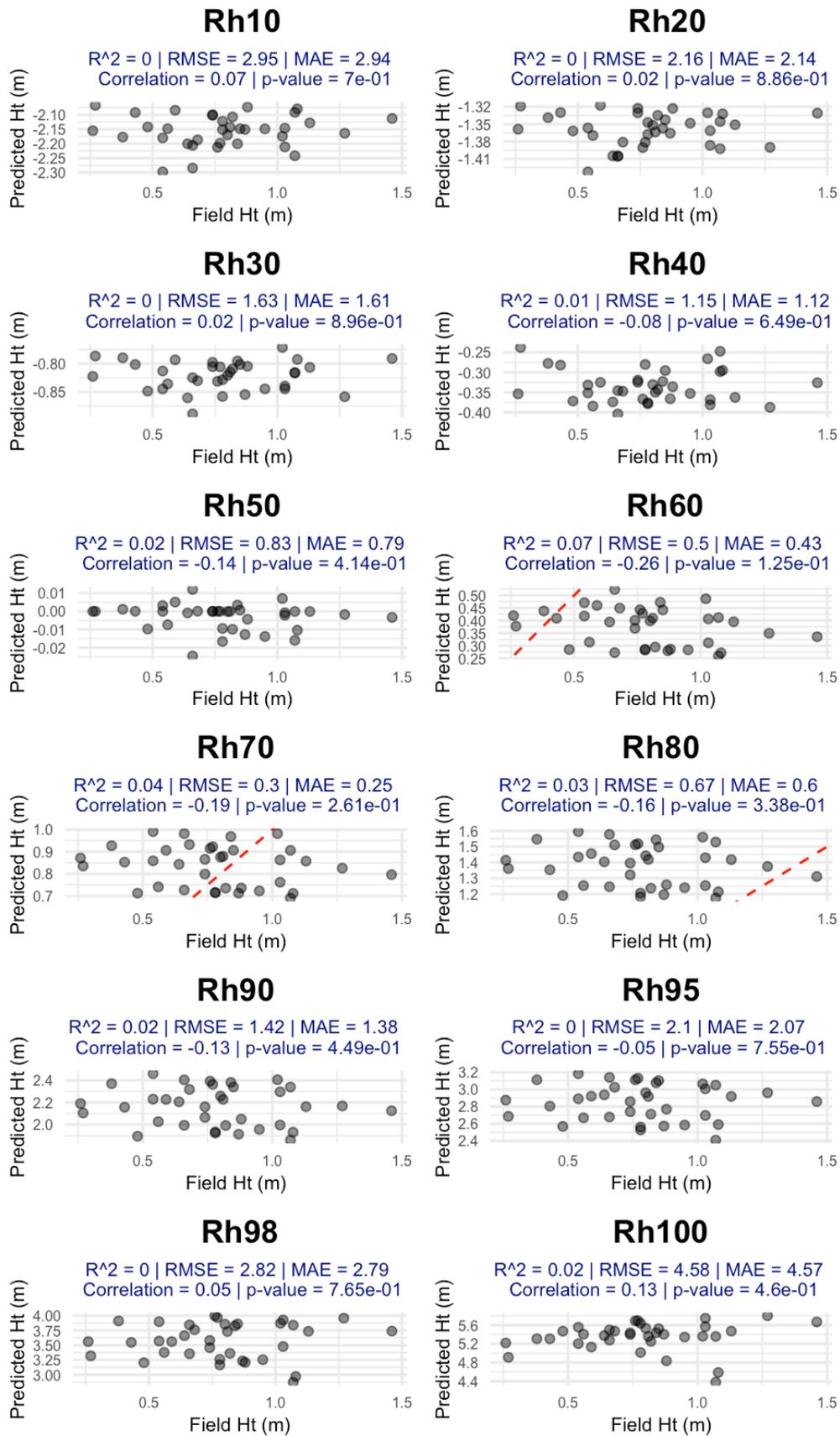


Figure 12.8: Model-2020's performance across Field data.

12.3. Alternative visualization of model results

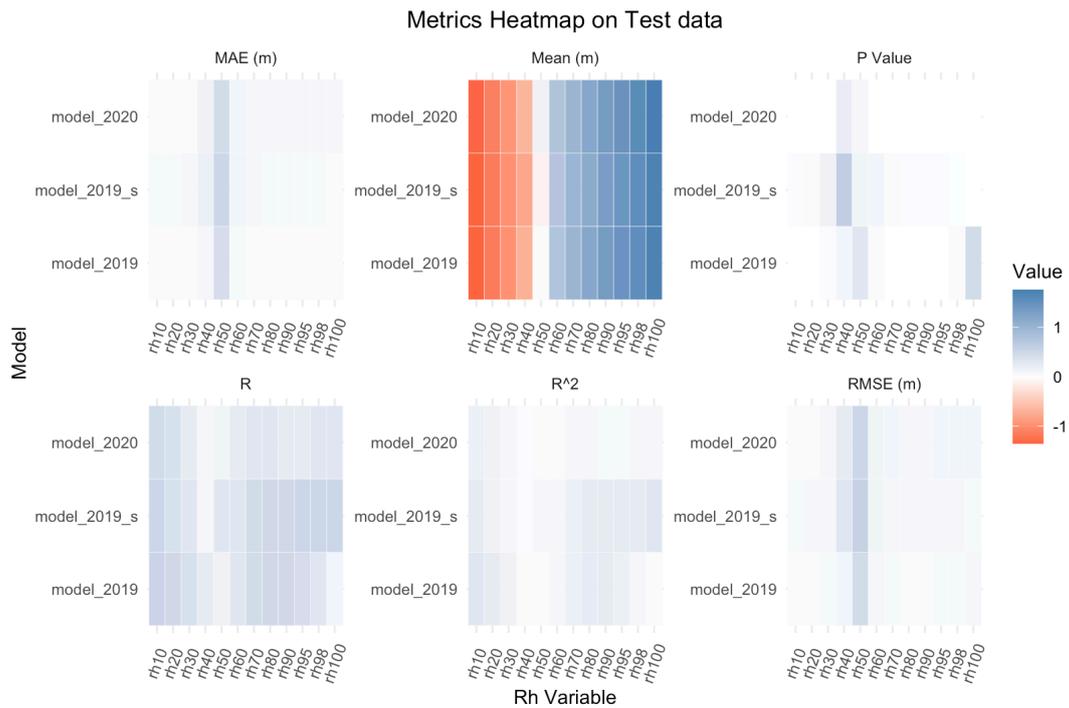


Figure 12.9: Visualisation of metric values across test data.

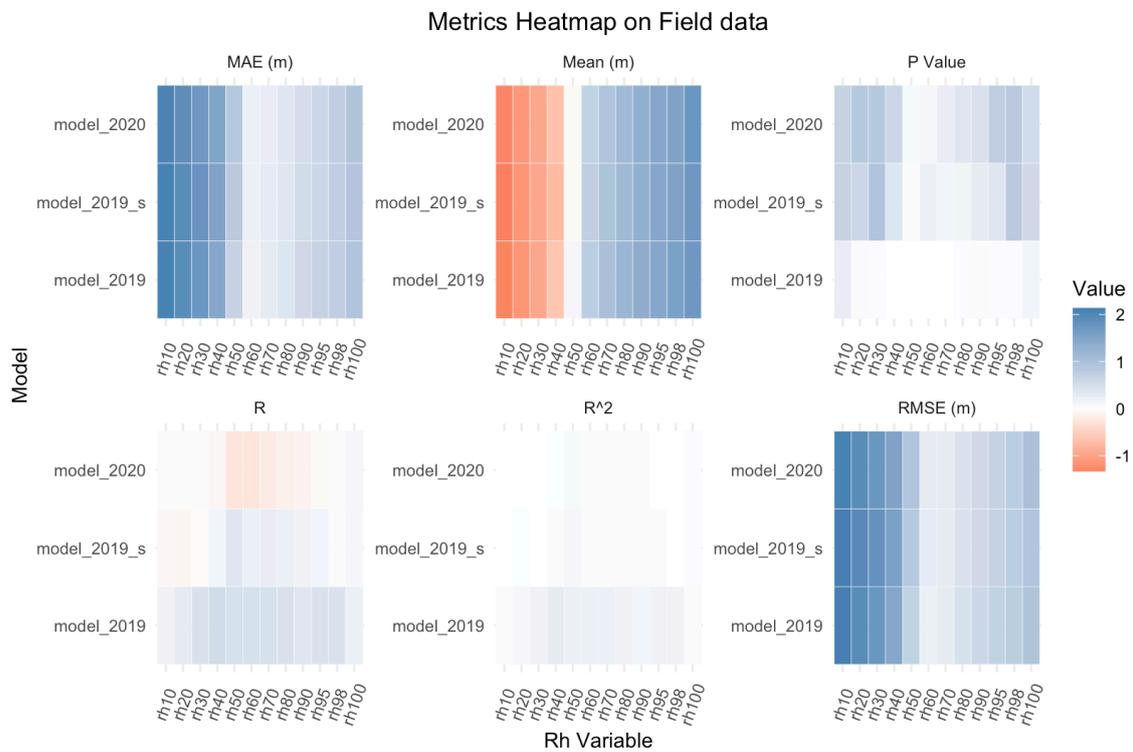


Figure 12.10: Visualisation of metric values across Field data.

12.4. Performance Scores

a) Test data

Model	Rh	composite_score
model_2019	rh10	1.86859849
model_2019_s	rh100	1.71947726
model_2019_s	rh98	1.52316998
model_2019_s	rh10	1.44581448
model_2019	rh20	1.38521436
model_2019	rh90	1.36639456
model_2019_s	rh95	1.33944085
model_2019	rh80	1.32240081
model_2020	rh10	1.22943151
model_2019_s	rh80	1.17888673
model_2019_s	rh90	1.16538472
model_2019	rh70	1.15634347
model_2019	rh95	0.93151461
model_2019_s	rh70	0.89572144
model_2019	rh30	0.75719185
model_2020	rh20	0.71410704
model_2019_s	rh20	0.55615174
model_2019	rh60	0.08108989
model_2020	rh30	0.05433499
model_2019	rh98	0.02198184
model_2020	rh100	-0.01799395
model_2020	rh80	-0.03412630
model_2020	rh98	-0.10382510
model_2020	rh70	-0.16653476
model_2020	rh90	-0.22489944
model_2020	rh95	-0.24856365
model_2019_s	rh30	-0.48507465
model_2020	rh60	-0.48768825
model_2019_s	rh60	-0.49176853
model_2019	rh40	-0.61918290
model_2019	rh100	-1.81675970
model_2020	rh40	-2.06618336
model_2019_s	rh40	-3.42769039
model_2020	rh50	-3.43697194
model_2019_s	rh50	-3.46106917
model_2019	rh50	-3.62431851

Figure 12.11: Composite scores across test data.

b) Field data

Model	Variable	composite_score
model_2019	rh60	2.70933748
model_2019	rh70	2.61132856
model_2019	rh80	2.05453134
model_2019	rh50	1.95451367
model_2019	rh95	1.68949949
model_2019	rh40	1.63283176
model_2019	rh98	1.59067516
model_2019	rh90	1.50223156
model_2020	rh60	1.15099762
model_2019_s	rh70	1.08765004
model_2019_s	rh50	0.99971632
model_2019_s	rh60	0.83349677
model_2019_s	rh80	0.81159156
model_2020	rh70	0.68785757
model_2020	rh50	0.54640419
model_2019	rh30	0.37384683
model_2020	rh80	0.36973865
model_2019_s	rh90	0.25646109
model_2019	rh100	0.18561561
model_2020	rh90	-0.04708273
model_2019_s	rh95	-0.06160930
model_2019	rh20	-0.50956647
model_2020	rh95	-0.69460609
model_2020	rh100	-0.72703717
model_2019_s	rh100	-0.74905703
model_2019_s	rh98	-0.90923634
model_2020	rh98	-0.91803450
model_2019_s	rh40	-1.24108800
model_2020	rh40	-1.54935253
model_2019	rh10	-1.61811465
model_2019_s	rh20	-2.08791834
model_2020	rh30	-2.24348776
model_2019_s	rh30	-2.36384643
model_2020	rh10	-2.42216300
model_2019_s	rh10	-2.44965147
model_2020	rh20	-2.45647343

Figure 12.12: Composite scores across field data.

12.5. Summary of GEDI transects used by model

Table 12.1: Summary of L2A beams with date of acquisition used in each model.

Date	L2A beam
model_2019_s	
2019.05.22	GEDI02_A_2019142104623_O02494_02_T01568_02_003_01_V002
model_2019	
2019.04.21	GEDI02_A_2019111083958_O02011_03_T04619_02_003_01_V002
2019.05.15	GEDI02_A_2019135132930_O02387_02_T01721_02_003_01_V002
2019.05.22	GEDI02_A_2019142104623_O02494_02_T01568_02_003_01_V002
2019.06.27	GEDI02_A_2019178202011_O03059_02_T04261_02_003_01_V002
2019.08.03	GEDI02_A_2019215151748_O03630_03_T03043_02_003_02_V002
2019.08.18	GEDI02_A_2019230092613_O03859_03_T00350_02_003_01_V002
model_2020	
2020.02.25	GEDI02_A_2020056055308_O06819_03_T03196_02_003_01_V002
2020.02.17	GEDI02_A_2020048090026_O06697_03_T01773_02_003_01_V002
2020.02.21	GEDI02_A_2020052072649_O06758_03_T00197_02_003_01_V002
2020.02.29	GEDI02_A_2020060041926_O06880_03_T01773_02_003_01_V002
2020.03.04	GEDI02_A_2020064024540_O06941_03_T04772_02_003_01_V002
2020.03.08	GEDI02_A_2020068011153_O07002_03_T03196_02_003_01_V002
2020.03.27	GEDI02_A_2020087172459_O07307_03_T00350_02_003_01_V002
2020.05.17	GEDI02_A_2020138211220_O08100_03_T04619_02_003_01_V002
2020.05.21	GEDI02_A_2020142193759_O08161_03_T01926_02_003_01_V002
2020.06.14	GEDI02_A_2020166005346_O08521_02_T00145_02_003_01_V002
2020.11.09	GEDI02_A_2020314233134_O10830_03_T10311_02_003_02_V002
2020.11.13	GEDI02_A_2020318215729_O10891_03_T08888_02_003_02_V002
2020.11.17	GEDI02_A_2020322202422_O10952_03_T06042_02_003_02_V002
2020.11.21	GEDI02_A_2020326185117_O11013_03_T07465_02_003_02_V002
2020.12.03	GEDI02_A_2020338141122_O11196_03_T10464_02_003_02_V002
2020.12.11	GEDI02_A_2020346110412_O11318_03_T07465_02_003_02_V002

12.6. Data Management Plan

1) Provide the title and briefly describe the aim and objectives of your MRes project.

Aim: To fuse Global Ecosystems Dynamics Investigation (GEDI) and Sentinel-2 to estimate mangrove canopy heights in Myanmar.

Research Objectives:

- 1) Explore available GEDI's products over mangroves in VCS1764 in Myanmar.
- 2) Fuse GEDI's Rh metrics with Sentinel-2 using machine learning to create a mangrove Canopy Height Model using field heights as a reference.
- 3) To compare the newly created Mangrove Canopy Height estimate with other available global and local datasets.

2) What data will be produced? (Data types, format, standards, scale and method)

Table 12.2: Summary of data products, sources, temporal extents and formats used.

Data	Source	Format	Temporal Extent	Size
GEDI L1B	LPDAAC ¹	HDF5	2019	<500MB
GEDI L2A	GEE	.shp	2019-2020	<1GB
GEDI L2B	LPDAAC	HDF5	2019	<500MB
Sentinel-2	GEE ²	TIFF	2019	<1GB
Field coordinates	Worldview ^{3,a}	.kml	2019	<100MB

^a No field work was undertaken, this data is freely available on the project website.

Quality checks were conducted on the final values using the ground-based measurements by comparing to local studies available (Oo,2021).

¹ [https://search.earthdata.nasa.gov/search/granules?p=C1214470488-ASF&pg\[0\]\[v\]=f&pg\[0\]\[gsk\]=-start_date&tl=1683029451.485!3!!&lat=24.10902370851302&long=-50.484375/](https://search.earthdata.nasa.gov/search/granules?p=C1214470488-ASF&pg[0][v]=f&pg[0][gsk]=-start_date&tl=1683029451.485!3!!&lat=24.10902370851302&long=-50.484375/)

² <https://code.earthengine.google.com/>

³ <https://registry.verra.org/app/projectDetail/VCS/1764>

Methods:

- R, HDFView⁴, QGIS, Visual Studio Code, Excel, Powerpoint were used as primary tools to visualise, extract, manipulate, run machine learning models and share the information that came out from this exercise. Caret, randomForest, gdal, ggplot2, sns, were some of the key packages in R used to generate a lot of the graphics and visuals of the models, while GEDI's Official Tutorial 1B, 2A and 2B were used to generate the full waveform return graphics⁵.

Output files:

- R and Python scripts that reproduce all graphics and models.
- The machine learning models themselves in a (.Rds) format that predict heights from Sentinel-2 images and can be used to create mangrove CHM for any other similar region.
- Final Rh60 .tiff image over VCS1764 showing the Mangrove CHM.

3) What metadata standards will you use? (Metadata content and format)

The metadata was initially documented using the Data Tree template, as was the data dictionary. Each dataset has its own metadata and data dictionary, saved in a separate 'docs' folder. They are named clearly, using the following format: 'datasetname_metadata.txt' and 'datasetname_datadict.txt'. Any other documentation is stored in the 'docs' folder including a summary of any issues.

The generated data will comply with the generic DataTree MetaData format (focused on geographic information and services).

4) How will your data be structured and stored? (Project storage)

- The principal backup used was Onedrive and certain local files on iCloud.
- The folder naming system will be as follows: 'input', 'output', 'processing', '.ipynb scripts', '.py scripts' & README. for each folder. A data dictionary will be used for each file in the Output folder, which will describe the abbreviations, and file conventions of the data labels that are followed. The template provided on DataTree will be used for this, see Appendix A. The MetaData template for all the input and output files that will be followed is attached in Appendix B.

⁴ <https://www.hdfgroup.org/downloads/hdfview/>

⁵ <https://lpdaac.usgs.gov/resources/e-learning/getting-started-gedi-l2a-data-python/>

- Subfolders within 'input' would be labelled into subfolders ('GEDI', 'Sentinel', 'Field') and further subfolders will be created if necessary according to the YYYYMMDDHHSS: format, in which all the input HDF5 and Sentinel in TIFF.
- ISO conventions of alphabetical arrangement will be followed.

5) How will the data be shared during and after the project? (Access, data sharing and reuse)

Data during the project was shared using PowerPoint, Rpubs (for publishing R graphs), emails, and Word documents. The .py scripts and the method to obtain the data (.zip) files will be shared on GitHub during the project, which will also help in version control. GitHub has a 1 GB upper limit on the free tier usage, which is why the original input files will not be directly uploaded but instead a document with a . README will contain the procedure to download the files directly through LPDAAC and The University of Nottingham OneDrive. Within 12 months from the completion of the project, the data will be stored in the Nottingham Data Repository and personal Onedrive.

6) Who has responsibility for implementing the DMP and are resources required?

The sole responsibility for implementing the DMP lie with the CDT student. Ensuring weekly backups, following the 3:2:1 approach will be followed. The Onedrive backups once completed will be the University of Nottingham's responsibility. The responsibility for the archived data will lie with the UKCSS Data and Information Archive, and the versioned code will be Github's responsibility. The resources that might be required will be of the monthly payment plan to Github in case storage requirements surpass 1 GB.

References

Oo, A.Y.Y., 2020. AN ASSESSMENT OF PHOTOSYNTHETIC BIOMASS RELATED TO THE MANGROVE SPECIES FROM THE SHWE THAUNG YAN COASTAL AREA.

Teperek, M. (2015). For EPSRC-funded researchers | Research Data Management. [online] Cam.ac.uk. Available at: <https://www.data.cam.ac.uk/funders/epsrc-funded-researchers> [Accessed 3 May 2023].

GitHub Docs. (2023). About archiving content and data on GitHub - GitHub Docs. [online] Available at: <https://docs.github.com/en/repositories/archiving-a-github-repository/about-archiving-content-and-data-on-github> [Accessed 3 May 2023].

Bgs.ac.uk. (2017). UK Carbon Capture & Storage Research Centre (UKCCSRC) Archive. [online] Available at: <https://www2.bgs.ac.uk/ukccs/home.html> [Accessed 3 May 2023].

