AALBORG UNIVERSITY
DENMARK

Large Scale Mangrove Above-Ground Biomass Estimation using Remote Sensing (RS) & Earth Observation (EO) Data with Machine Learning

Spring 2022 - Masters Thesis

Alexander Boest-Petersen

June 2, 2022

Summary

In recent years, cloud computing and Earth Observation data has significantly grown to allow for large-scale calculation of vital statistics, such as land cover and biomass, to facilitate for the monitoring of the health and development of ecosystems in light of climate change and the United Nations Sustainable Development Goals. With the advancement of cloud computing platforms, such as Google Earth Engine, and the data availability of remotely sensed data on said platforms, can machine learning by utilized in conjunction with high resolution data to estimate mangrove land covers as well as above ground biomass in these threatened areas?

Previous studies estimating mangrove biomass (Simard et al., 2019), utilize data collected at lower resolutions than what is currently available, and/or also collected during a period which is no longer applicable to the conditions of the present. With global datasets, such as Sentinel-2's global imagery at 10m resolution, and active sensors such as NASA's Ice, Cloud, and Elevation Satellite and Global Ecosystem Dynamics Investigation LIDAR platforms sampling the elevation of key features on the Earth's surface at a fine resolution, derivation of high resolution mangrove health statistics, such as area, height, and biomass, should be achievable utilizing GEE.

The workflow developed in this study (Listing 9), within a cloud computing environment, can easily be deployed in any region of the globe (at a country-wide scale) where mangroves may be present with minimal changes in order to generate high resolution mangrove health statistics to facilitate the reporting of these ecosystems and improving stewardship and development of the UN Sustainable Development Goals. Cloud computing and actively sensed data such as GEDI canopy height returns in concert with machine learning techniques have proven to accurately predict key mangrove characteristics, height and above ground biomass, at a large scale (country-wide), allowing for rapid and detailed reporting and management of these ecosystems, proving that they are powerful tools in the fight against future climate change. This study intends to build and develop upon the workflow and techniques derived and utilized in the previous semester's study by Boest-Petersen, 2022b.

Title:
Large Scale Mangrove Above-Ground Biomass Estimation using Remote Sensing (RS) & Earth Observation (EO) Data with Machine Learning.

Project Period:
January 2022 – June 2022

Authors:
Boest-Petersen, Alexander (#20201810)

Supervisor:
Prof. Dr. Jamal Jokar Arsanjani

Number of Pages:
70

Department:
Department of Development and Planning

Address:
A.C. Meyers Vænge 15, 2450 Copenhagen SW, DK

Study Secretary CPH:
Sille Sophia Johannsen

Preface

This report was written during the Spring 2022 Thesis Semester at Aalborg University Copenhagen for the Survey, Planning, and Land Management (Cand. Tech.) Masters program under the supervision of Prof. Dr. Jamal Jokar Arsanjani. The semester took place from January 2022 to June 2022. During the course of the study, large amounts of remotely sensed data (imagery, backscatter, LIDAR, etc.) was processed in custom Python tools or in cloud computing environments (see Appendix) in order to derive mangrove biomass estimations at large scales. The work done in this study was based off of, and an evolution of, the 2021 Fall Semester internship report by Boest-Petersen, 2022b. As such, many similar topics and ideas are presented in this report, but with the aim of improving and furthering the work done previously.

Note: All code listings, figures, and tables are made by Alexander Boest-Petersen specifically for this report unless specified. Cloud computing environments and Python tools can be found in the Appendix of this report.

Acknowledgements

I would like to thank Dr. Kenneth Grogan, head of Remote Sensing Analytics at DHI, for his support and feedback regarding remote sensing data, processing techniques, and feedback for the steps taken in this study.

I would also like to thank Daniel Druce, Remote Sensing Specialist at DHI, for all of his assistance in regards to cloud computing, machine learning, mangrove ecosystems, and processing techniques regarding mangrove analysis.

Finally, I would like to thank Prof. Dr. Jamal Jokar Arsanjani for all of the guidance and supervision provided throughout the extent of this study.

Contents

Acronyms

AGB Above Ground Biomass

ALOS Advanced Land Observing Satellite

API Application Program Interface

ATLAS Advanced Topographic Laser Altimeter System

CHM Canopy Height Model

CNN Convolutional Neural Networks

csv Comma Separated Values

DAAC NASA Distributed Active Archive Center

DBH Diameter at Breast Height

DEM Digital Elevation Model

DSM Digital Surface Model

DTM Digital Terrain Model

EO Earth Observation

ESA European Space Agency

FABDEM Forests and Buildings Removed Copernicus DEM

GEDI Global Ecosystem Dynamics Investigation

GEE Google Earth Engine

GHG Greenhouse Gas Emissions

GIS Geographic Information Systems

GMA Global Mangrove Alliance

GMW Global Mangrove Watch

ICESat-2 Ice, Cloud, and Elevation Satellite

IDE Integrated Development Environment

ISS International Space Station

JAXA Japanese Aerospace Exploration Agency

JERS-1 Japanese Earth Resources Satellite-1

LIDAR Light Detection and Ranging

LULC Land Use Land Cover

MERIT Multi-Error-Removed Improved-Terrain

ML Machine Learning

MMU Minimum Mapping Unit

NASA National Aeronautics and Space Administration

NICFI Norway's International Climate and Forest Initiative

PALSAR Phased Array type L-band Synthetic Aperture Radar

PhoREAL Photon Research and Engineering Library

QGIS Quantum Geographic Information System

RADAR Radio Detection and Ranging

RF Random Forest

RS Remotely Sensed

SAR Synthetic Aperture Radar

SDG Sustainable Development Goals

SDSS Spatial Decision Support Systems

SLR Sea Level Rise

UN United Nations

USGS United States Geological Survey

WCMC World Conservation Monitoring Centre

Glossary

Backscatter "*is the portion of the outgoing radar signal that the target redirects directly back towards the radar antenna. Backscattering is the process by which backscatter is formed*". (European Space Agency, 2021a)

Blue Carbon "*refers to organic carbon that is captured and stored by the oceans and coastal ecosystems*". (Macreadie et al., 2019)

Cloud Computing "*is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet*". (Microsoft, 2021)

Cryosphere "*is the frozen part of the Earth system*". (NOAA, 2022)

Spatial Resolution "*is a measure of the smallest object that can be resolved by the sensor, or the ground area imaged for the instantaneous field of view (IFOV) of the sensor, or the linear dimension on the ground represented by each pixel*". (Liang et al., 2012)

Temporal Resolution "*is defined as the amount of time needed to revisit and acquire data for the exact same location*". (Théau, 2008)

Listings

List of Figures

List of Tables

1 Introduction & Problem Analysis

1.1 Mangroves: Ecological Impact & Climate Change

Mangroves are regions of high productivity, and diverse habitats, for both plants and species, that can be found in tropical and subtropical regions of the world where rivers, intertidal, and coastal waters mix. (Kathiresan, 2021) When compared to other tropical ecosystems, they are relatively species-poor, meaning that they do not support as large of a variety of plant life, and can be homogeneous in structure and make-up. (Duke, 1984) However, mangroves can offer a host of benefits to local ecosystems and communities, such as the storage of carbon via Above Ground Biomass (AGB). With a deforestation rate of 1-2% per year, many areas may disappear within a century, highlighting a need for a detailed inventory to aid with management and restoration. (Alongi, 2002)

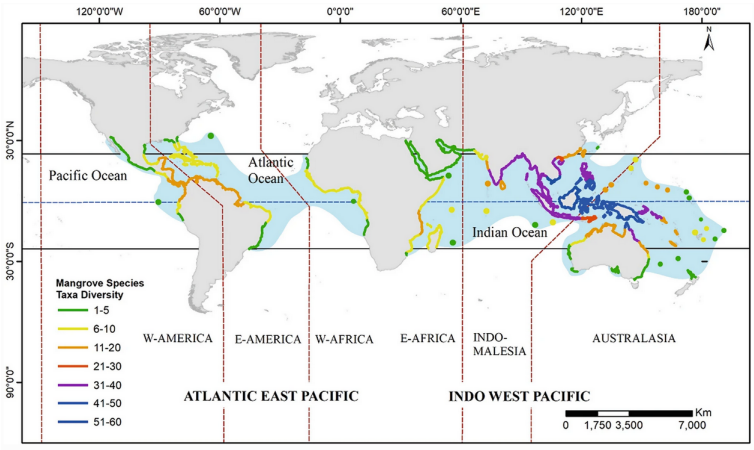


Figure 1: Mangrove locations across the world as per Ragavan et al., 2021.

There are a multitude of factors that can impact the future, as well as current, productivity of mangrove forests on coastlines around the globe. Along with anthropogenic pressures imposed on mangroves, climate change is impacting mangrove ecosystems negatively, compounding the impacts of human-led issues. (Wong et al., 2014)

Anthropogenically induced climate change, that is mainly caused by humans, is a leading cause of rising atmospheric temperatures and thermal Sea Level Rise (SLR). (Zickfeld et al., 2017) SLR is responsible for the loss of land, habitats, and property along the coastlines of the world. Mangroves are a significant ecosystem at the forefront of these recent developments.

SLR and changes in sediment supply to mangroves around the world pose the largest factors of change in regards to mangroves vulnerability, leading to ecosystem regime shifts, often with negative consequences as detailed in Table 1. (Ellison, 2015) Mean sea levels are primarily rising due to thermal expansion of water caused by higher water temperatures, as well as the melting of polar and land ice due to higher atmospheric temperatures. (Stocker et al., 2013) Warmer temperatures and higher sea levels can have varying impacts on mangroves depending on their location. Higher sea levels can inundate mangroves close to the coastline, however, this can also promote the growth of the ecosystems further inland. (López-Medellín et al., 2011) If mangroves are unable to raise surrounding surface elevations via sediment accretion, then they will be submerged and lost. (Krauss et al., 2014) As such, a high resolution (spatially and temporally) inventory of mangroves can prove beneficial to

study the impacts of such events and how mangroves respond to them.

Table 1: Inputs and impacts to mangroves and their potential outcomes as per Ellison, 2015.

Future Impacts	***Processes Impacted***	***Potential Outcomes***
Rising Sea Level	*Forest health* *Forest productivity* *Recruitment* *Inundation period* *Accretion rates*	*Forest mortality, dieback from the seaward edge, migration landward, depending on sediment inputs, topography and lack of barriers.*
Increased waves, wind, and extreme storms	*Forest productivity* *Recruitment* *Accretion rates*	*Forests damaged or spatial area changed, surface elevation change, erosion or excess sedimentation.*
Increased air and sea temperatures	*Respiration* *Photosynthesis* *Forest productivity*	*Reduced productivity at low latitudes and increased winter productivity at high latitudes.*
Enhanced CO_2	*Photosynthesis* *Respiration* *Biomass allocation* *Forest productivity*	*Increased productivity, subject to limiting factors of salinity, humidity, nutrients. Soil elevation increase.*
Increased rainfall and freshwater availability	*Sediment inputs* *Ground water* *Salinity* *Productivity*	*Increased accretion and maintenance of surface elevation, increased groundwater, diversity, productivity and recruitment.*
Reduced freshwater availability	*Sediment inputs* *Ground water* *Salinity* *Photosynthesis* *Forest productivity*	*Reduced ground water, diversity, photosynthesis, productivity and accretion.* *Mangrove migration landward, species change.*

Coastal mangroves can also provide a natural barrier to combat the erosive and destructive forces of worsening storms and their resulting waves. Evidence has highlighted that mangrove forest provide a solution in reducing environmental risks and damages to areas that are particularly vulnerable to coastal erosion via the mechanical transportation of sediment from waves. (Asari et al., 2021) This done by reducing the height of wind and sell waves over short distances, however, this is dependent on the structure, depth, and composition of the mangroves, as such, maintaining their health and size can improve their performance as a natural coastal barrier. (Asari et al., 2021) Mangroves also provide an economic benefits to surrounding areas in the form of provisioning services; food, water, raw materials, etc., as well as regulating services; air quality/climate regulation, disturbance moderation, as well as a host of other factors. (de Groot et al., 2012)

Coastal mangrove areas have also proven to be effective carbon sinks; capturing, storing, and transporting atmospheric Carbon Dioxide (CO_2), a commonly produced Greenhouse Gas Emissions (GHG), into a safer form as mangroves are extremely productive natural ecosystems even though they have a simple forest structure with low diversity as mentioned before. (Lee et al., 2014) This is done by collecting carbon via photosynthesis in the stems, leaves, roots, and branches, transforming it

into the plant's biomass. (Hashim and Suratman, 2021) As such, mangroves are extremely beneficial in the fight against climate change. Mangroves are estimated to contribute approximately 10-15% of carbon storage in coastal regions of the planet. (Simard et al., 2019)

Carbon sequestered and stored in coastal ecosystems is often referred to as Blue Carbon. (Rao et al., 2021) It is estimated that mangrove forest are responsible for around 30% of global blue carbon sequestration, outperforming other ecosystems around the world as shown in Figure 2. (Siikamäki et al., 2012) The falling of foliage,or litterfall, is the principal factor in blue carbon storage and cycling of carbon in mangrove ecosystems, as well as transporting nutrients within and between ecosystems. (Kamruzzaman et al., 2019) With the degradation of mangrove ecosystems, sediment and carbon sources may be transported to neighboring ecosystems, negatively altering the adjacent environments, ultimately impacting a habitats' carbon sinking capabilities in unknown manners. (Asplund et al., 2021)

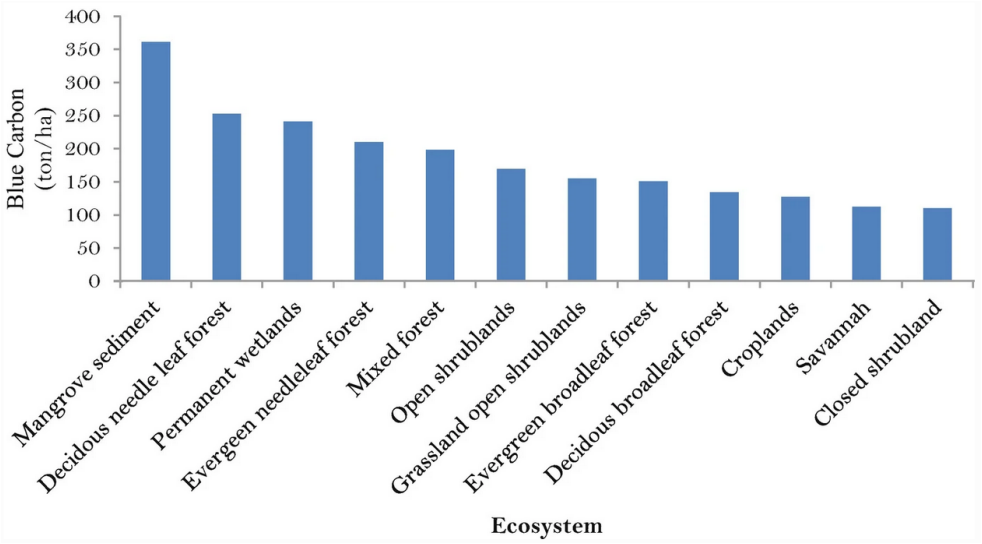


Figure 2: Carbon locating in soils at a depth of 1 meter, highlighting how mangroves are powerful tools in climate change mitigation as per Sanderman et al., 2018.

Irrecoverable carbon is defined as carbon that can be directly influenced by local human action, can be lost due to land use conversion, and is not recoverable within a specified time-frame. (Goldstein et al., 2020 & Bukoski et al., 2022) Mangroves are estimated to contain close to 225 tons of carbon per hectare across the globe with an estimated recovery time of 153 years to recover vulnerable mangrove carbon sources alone. (Goldstein et al., 2020) It is estimated that human development will continue to impact mangroves in the future due to anthropogenic disturbances, making clear a need for consistent and accurate mangrove health measurement and monitoring. (Noon et al., 2021)

Many areas of the world are currently, and historically, experiencing mangrove area loss. Areas of heavy losses can be seen in Malaysia due to rice farming, shrimp aquaculture, and coastal development. (Ong, 1995 & Ong, 2003) Some species of mangroves, such as *Bruguiera hainesii*, are extremely rare with less than 250 matured individuals remaining, and propagate very slowly, headlining the need for study and protection. (Polidoro et al., 2010) More general area loss statistics can be seen in Table 2.

Table 2: Mangrove loss statistics from 1996 to 2016 as per Worthington et al., 2019.

Region	Area (km^2) in 1996	Area (km^2) in 2016	Loss (km^2)	Gain (km^2)
Australia & New Zealand	10,332	10,037	370	74
East & Southern Africa	7,630	7,329	424	122
East Asia	159	159	12	13
Middle East	334	319	19	4
North & Central America & The Caribbean	22,702	21,072	2,196	566
Pacific Islands	6,410	6,327	146	63
South America	19,632	19,063	1,106	537
South Asia	8,701	8,492	435	226
Southeast Asia	46,789	44,060	3,308	579
West & Central Africa	20,107	19,857	422	171
Total	**142,795**	**136,714**	**8,437**	**2,356**

Mangroves require urgent research, management, and attention from the public in order to prevent further losses and allow for rehabilitation of lost and/or damaged areas. (Feller et al., 2017) Even with warming temperatures and higher levels of precipitation allowing for the mangrove expansion, these gains still do not outweigh anthropogenic losses, highlighting a need for detailed and continuous study of mangrove area loss/gain as well as the health of the individual mangrove sites. (Feller et al., 2017)

1.2 United Nations Sustainable Development Goals

The United Nations (UN) Sustainable Development Goals (SDG) are a set of goals adopted by all participating parties that are a part of the 2030 Agenda for Sustainable Development that outlines the targets for peace and prosperity for people and the planet starting now and going into the future. (United Nations, 2021b) Even though mangroves are sensitive to climate change, they can prove to be a valuable tool. Sequestering carbon, enhancing coastline stability, and protecting coastal settlements from storm surges and wave damage are some ways that mangroves can help reduce or mitigate future damages. (Chow, 2018)

Specifically, the conservation and preservation on mangroves falls under goal number 14, to conserve and sustainably use the ocean, seas and marine resources for sustainable development. (United Nations, 2021a) Goals 13, combating climate change, and 15, sustainably manage forests, combat desertification, halt and reverse land degradation, as well as halt biodiversity loss are goals that are directly concerned with the loss and monitoring of mangroves in the future as climate change occurs. (Fakhruddin et al., 2018) Having a detailed, accurate, and up-to-date baseline of local, or global, mangrove biomass stores can allow member nations and concerned stakeholders to have access to

information to enact change to preserve, protect, and restore these critical coastal areas, as well as better understand them and the impacts of anthropogenic pressures. (Rahman et al., 2021)

1.3 Previous Studies & Problem Analysis

In order to track the management, destruction, and changes in coastal mangrove ecosystems, an accurate inventory of biomass is critical in order to properly manage and track the impacts of future climate change and SLR. Previous studies have estimated global (or regional (Fatoyinbo and Simard, 2013)) AGB using sub-optimal elevation data to in order to derive general datasets for worldwide tracking of mangroves. (Simard et al., 2019) Previous studies attempt to use high resolution data to predict mangrove biomass for specific regions of interest ((Pham, Le, et al., 2020), (Suwa et al., 2021), & (Pham, Yokoya, et al., 2020)), excluding impacted areas outside of the study area, or methods of data collection that do not scale to global applications. (Jones et al., 2020) Such studies highlight that an accurate, easily accessible, and regularly updated large-scale (country or countries) dataset of mangrove biomass estimations is clear gap in the ever-important region of mangroves.

A multitude of new satellite and sensor systems such as; the European Space Agency (ESA) Sentinel-1 and Sentinel-2 satellites capable of measuring RADAR Backscatter and high resolution optical imagery respectively, the National Aeronautics and Space Administration (NASA) Ice, Cloud, and Elevation Satellite (ICESat-2) Light Detection and Ranging (LIDAR) system[1], as well as the Global Ecosystem Dynamics Investigation (GEDI) system[2] upon the International Space Station, up-to-date, accurate, and high resolution (Spatial Resolution and Temporal Resolution) data is easily and freely accessible to use from an array of robust active and passive sensors within various cloud computing environments. Such advancements in data availability and data processing platforms have allowed the scientific community, as well as more intermittent users, the ability to perform large scale analysis utilizing petabytes of remotely sensed Earth observation data in the realm of above ground biomass alone in recent years quickly and efficiently. (Sanderman et al., 2018, Yang et al., 2019, Thieme et al., 2020, Li et al., 2019, & Filippelli et al., 2020)

Traditionally, mangrove biomass estimates have been conducted using time-consuming and costly field, in-situ, measurements. (Salum et al., 2020, Jachowski et al., 2013, Aslan et al., 2016, & Vaghela et al., 2021) In-situ measurements include the destructive process of cutting mangrove samples (trees and shrubs) in order to measure various attributes of the plant (such as height, diameter of trunk, and overall weight of sample) to derive the carbon carrying capacity of a specific species or area. (Jones et al., 2020) Previous studies have attempted to streamline this process by using Spatial Decision Support Systems (SDSS) (Tang et al., 2017), data from active (e.g. LIDAR or RADAR) aerial or space-borne sensors, as well as passive sensors to determine mangrove biomass estimations with differing levels of accuracy when compared to manual measurements. Such studies allow for the general estimation of mangrove biomass on large scales, providing stakeholders with a clearer view of global mangrove ecosystems. Or collection of high resolution elevation data is only possible at smaller regional scales, prohibiting the detailed estimation of larger study areas. (Salim et al., 2020) This study hopes to build upon large-scale work/studies and improve upon the derived results to generate a reliable biomass estimation at a large scale using freely available data and data processing platforms.

Using previous studies and research, this report intends to utilize high resolution, actively sensed, Earth Observation (EO) data combined with Machine Learning (ML) techniques for classification and

[1] https://nsidc.org/data/ATL03
[2] https://gedi.umd.edu/data/products/

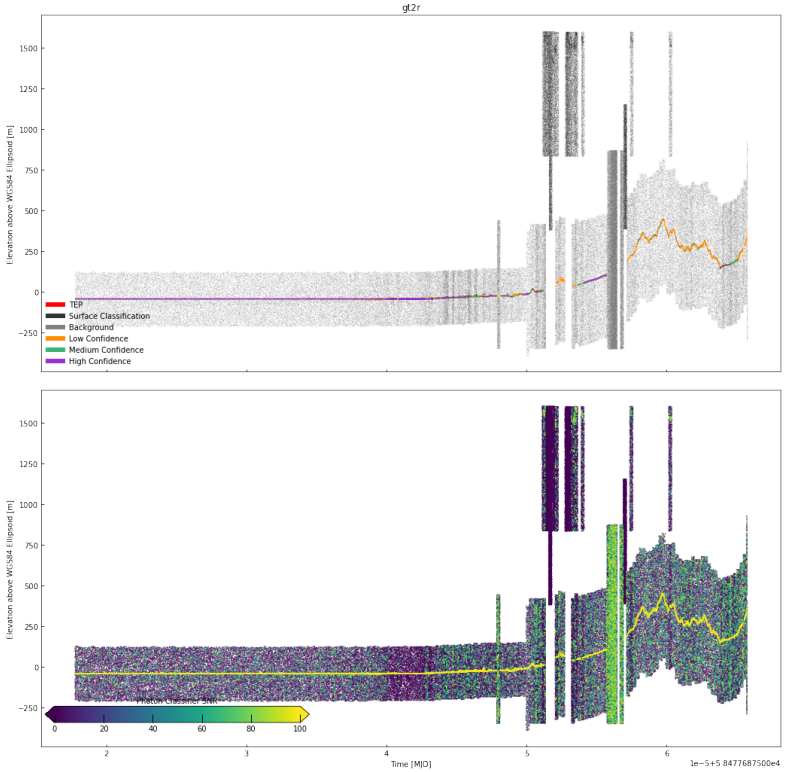


Figure 3: ICESat-2 ATL03 return signal data ('gt2r' track) and surface estimation for a scan of Jobos Bay, Puerto Rico utilizing tool by Siegfried and Sutterley, 2022.

regression of surfaces. Talk about deriving better mangrove masks at smaller than global scales, and better localized heights using multiple AGB formulas. Determining mangrove coverage at smaller-than-global scales and at a resolution of up to 10 meters, mangrove area can be more accurately represented, allowing for more accurately biomass estimations. Combining high resolution actively sensed EO data with regional specific allometric models in a cloud computing setting can allow users to quickly and accurately derive meaningful statistics for their regions of interest. By deriving highly accurate mangrove extent and above ground biomass estimates, at as large of a scale as possible, stakeholders will be able to track the change in mangrove health and mangrove area loss in order to implement strategies of mangrove loss mitigation as well as tracking relevant UN Sustainable Development Goals metrics.

1.4 Research Questions

1. Of the options of actively sensed and derived earth observation data, in regards to canopy heights (ICESat-2 ATL03 & ATL08, and GEDI outlined in Table 3), which performs best when estimating canopy heights for above-ground biomass estimations in a cloud-computing setting?
2. At what scale is it feasible to derive such estimates and at what level (study site, regional, country) is feasible in a cloud-computing environment such as GEE?

2 Data Sources & Platforms

Table 3: Canopy height information sources needed to calculate above-ground biomass. (Boest-Petersen, 2022b)

Data	Description	Resolution	Source	Availability/Temporal Resolution
UAV(Drone) Based	Site/study specific. Costly and not efficient for large scale studies.	Sub-meter per pixel depending on sensor/mission type.	Privately funded research projects from an aerial platform.	Single fly-over. Multiple scans if deemed necessary.
ICESat-2 ATL03	Geolocated photon returns from the Advanced Topographic Laser Altimeter System (ATLAS) sensor aboard ICESat-2. 10,000 pulses per second. (NASA, 2022)	Dependent on terrain type and atmospheric conditions. Average of 70cm resolution.	NASA and OpenAltimetry	September 15, 2018 - Current
ICESat-2 ATL08	Photon counting range estimates separating noise and signal photons to estimate terrain and canopy height from available ATL03 returns. (Neuenschwander and Pitts, 2019)	100 meter along-track resolution.	NASA and UT Austin Research Laboratories.	September 15, 2018 - Current
GEDI	Global Ecosystem Dynamics Investigation laser system (3 lasers) aboard the International Space Station (ISS). (GEDI, 2022)	25 meter pixel resolution.	NASA	December 5, 2018 - 2023 (Estimated)

In order to compare the performance and reliability of raw data (ATL03), as well as derived datasets (ATL08 & GEDI), study areas will be chosen that have local detailed LIDAR scans (see Figures 9 & 13) to allow the generation of small scale Canopy Height Model (CHM)s that represent tree heights to a high degree. These local level datasets will allow for the direct comparison of space-borne platforms and their resulting data to what is actually present on the surface. Once a reliable dataset is determined, a workflow will be developed to allow for the generation of mangrove masks and mangrove canopy estimations using machine learning in a cloud computing environment at as large of a scale as possible.

Existing studies use allometric equations (or regressions) to derive biomass estimates from mangrove or other forested areas. (Fatoyinbo and Simard, 2013, Simard et al., 2019, Jones et al., 2020) Biomass estimates are traditionally derived by incorporating individual tree diameter at breast height (DBH) and canopy height metrics to determine a volumetric biomass statistic. (Kebede and Soromessa, 2018) Existing studies (such as Simard et al., 2019) have incorporated available in-situ measurements to derive region-specific allometric regression equations as seen in Table 7, thus allowing for the estimation of biomass utilizing remotely sensed data with only canopy height estimates. This is possible as relationships between important factors such as tree stem diameter and wood density a linked in a linear fashion to canopy height, facilitating the estimation of AGB with only canopy heights. (Fayad et al., 2014) Allowing for the rapid classification and estimation of carbon stores using machine learning and cloud computing to provide a general estimate as well as track change of carbon stores over time.

2.1 Processing Platforms Utilized

2.1.1 Cloud Computing & Google Earth Engine (GEE)

In previous years, Cloud Computing has enabled Earth Observation scientists to address processing the daily delivery of terabytes of remotely sensed data, allowing for the rapid querying, acquisition, and processing of data in a single environment in some cases. Ultimately, cloud computing provides remote sensing users with services that allow them to leverage data processing and production

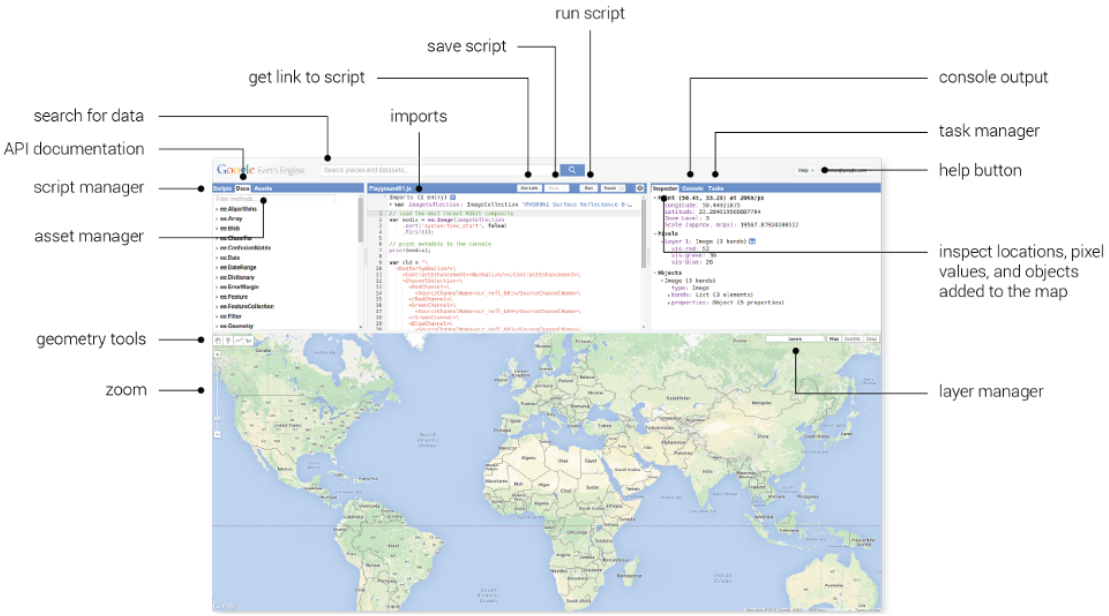


Figure 4: GEE browser-based IDE user interface. (Gorelick et al., 2017)

computing/storage platforms, along with integrated spatial analysis tools, so as to provide solutions in a single interface, such as GEE or the newly introduced Microsoft Planetary Computer platform[3]. (L. Wang et al., 2019) These platforms are made even more accessible by the fact that they are freely available for academic purposes, promoting the widespread use of the platforms, thereby allowing for the monitoring of at-risk ecosystems as well as providing key stakeholders with critical data.

GEE[4] is a web-based Integrated Development Environment (IDE) built upon the Earth Engine JavaScript Application Program Interface (API). (Gorelick et al., 2017) GEE (refer to Figure 4) allows users to query and import petabytes of remotely sensed data quickly and easily, create geospatial data or importing custom shapefiles, tables. These features facilitate the development of algorithms for geospatial analysis, and providing users with information to further conduct analysis outside of the browser environment if needed. (Gorelick et al., 2017)

GEE allowed for the acquisition, pre-processing, and utilization of Sentinel-1 backscatter, Sentinel-2 imagery, FABDEM elevation surfaces, GEDI global canopy height estimates, as well as other datasets such as Copernicus' Global 100m Land Cover to improve mangrove masking performance. Data which was not available in the GEE Data Catalog, such as ICESat-2, was downloaded and pre-processed outside of the GEE IDE and ingested into the editor as an asset in the form of shapefiles for vector data (GMW known mangrove areas), or a tabular (csv) datasets such as ICESat-2 canopy height metric data points.

2.1.2 Python

Python is a programming language that is popular in the realm of geospatial processing and data science. It's array of freely available packages and active community of users enables anyone to query, download, and analyze vast amounts of geospatial information. For this study, Python in conjunction

[3]https://planetarycomputer.microsoft.com/
[4]https://earthengine.google.com/

with various scientific, numerical, and geospatial focused packages, was heavily utilized to access and process ICESat-2 data.

A virtual environment and package manager was used throughout the course of the report in the form of Anaconda[5] to streamline processing environments as many packages require specific dependencies that can easily be cluttered by pre-existing packages. By utilizing virtual environments, workspaces can be streamlined, only having the necessary packages installed, as well as providing the option of easily sharing environments with others for further collaboration. For even leaner processing environments in future studies, Miniconda, a free version of Anaconda, would prove a powerful alternative[6].

2.2 Data Utilized

2.2.1 FABDEM

Traditional elevation datasets, such as the Copernicus GLO 30m elevation dataset, includes all features on the surface of the Earth, such as trees or buildings, which can negatively impact the model performance as it is receiving information that is not representative of what the surface is. Instead, for this study, the Forests and Buildings Removed Copernicus DEM (FABDEM) elevation dataset was utilized to supplement both the mangrove mask classification as well as mangrove canopy height regression in an attempt to increase model performance and overall accuracy of results.

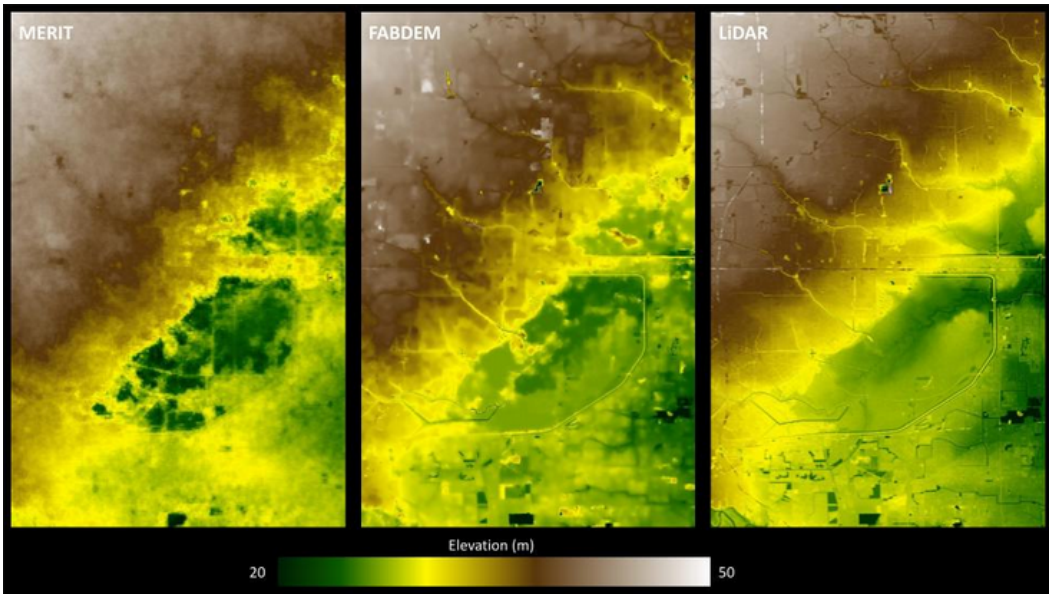


Figure 5: Comparison of elevation FABDEM elevation surface against the Multi-Error-Removed Improved-Terrain (MERIT) DEM (tree removal) and LIDAR elevation surfaces for Houston, Texas. (Fathom, 2022)

Elevation surfaces play a key role in ecology and soil science, allowing users to refine and adjust machine learning or deep learning applications in order to better predict land cover type or other statistics of interest. FABDEM is derived by utilizing machine learning techniques, such as random

[5]https://www.anaconda.com/
[6]https://docs.conda.io/en/latest/miniconda.html

forest regression models, to remove building and tree height biases from the Copernicus COPDEM30 global (mostly) elevation dataset in order to achieve a more accurate Digital Elevation Model (DEM) of the Earth's surface at a global scale. (Hawker et al., 2022) This dataset is novel, with its introduction for educational purposes only being made available in early 2022, with the asset (a 450+GB raster file) being made available on GEE by the community of users. (Roy, 2022)

2.2.2 Copernicus Global Land Cover

The Copernicus Dynamic Global Land Cover is comprised of discrete and fractional land cover layers for 2015 to 2019 at a 100 meter resolution. (Buchhorn et al., 2020) This dataset was used in generating a local mangrove land cover mask. Specifically to provide a more accurate mask, where issues with incorrect classifications of water pixels were experienced in such edge ecosystems by simplifying the classification of water which can be extremely variable in appearance given the atmospheric conditions.

The level 1 dataset utilized in this report is validated to an accuracy of 80.6+/-0.4%, with water classes being mapped as the highest accuracies. (Tsendbazar et al., 2020) This dataset is available in raster format in the GEE Data Catalog for rapid implementation into custom workflows and algorithms.

2.2.3 Historical Mangroves

The Global Mangrove Watch (GMW) is a data portal that was initiated as part of the JAXA Kyoto & Carbon Initiative in 2011 made available by the Global Mangrove Alliance (GMA) to increase access and visibility of possible mangrove habitat extents as seen in Figure 6. (Bunting et al., 2018)

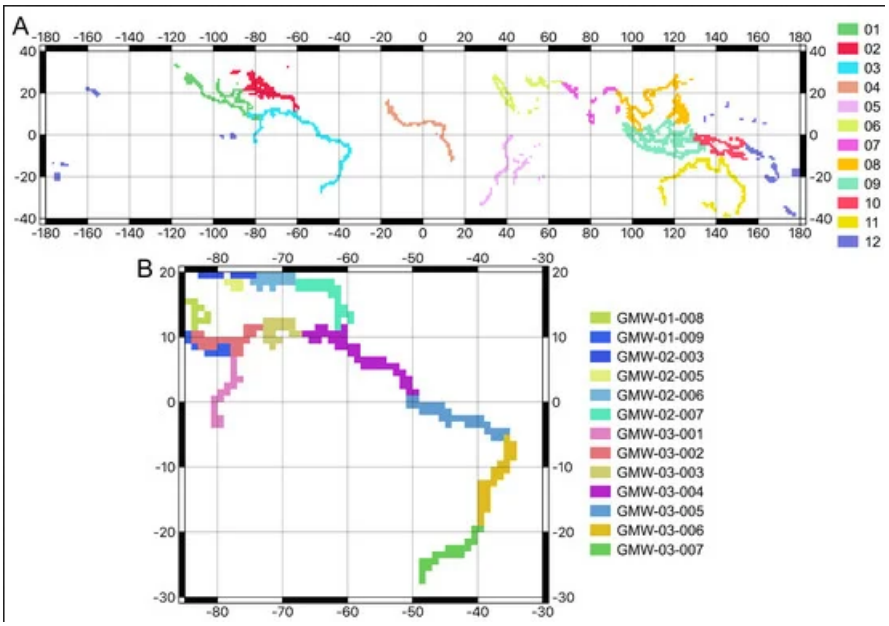


Figure 6: Distribution of GMW project areas across the globe. (Bunting et al., 2018)

This data is made available via the UN's World Conservation Monitoring Centre (WCMC) as a shapefile, with updated annual mapping of mangroves planned from 2018 in to the future derived

from JERS-1 SAR, ALOS PALSAR, and ALOS-2 PALSAR-2 EO platforms. (Bunting et al., 2018) Potential mangrove extents are derived with ALOS PALSAR and Landsat optical data from 2010, with the purpose of providing data to countries that may not have sufficient national mangrove monitoring systems. (Bunting et al., 2018)

A global baseline of areas classified as mangroves in the year 2016 by the GMW was utilized to train ML classification models to generate a local mangrove mask of high-likelihood mangrove land coverage. Classifying land cover on a local level can prove advantageous due to the fact that global datasets are often optimized to meet the complex requirements needed to classify at a global scale even if many study/training sites are taken employed in the overall model. By conducting a mangrove classification, AGB estimates will be more accurate and representative in the study area.

2.2.4 ICESat-2

NASA's Advanced Topographic Laser Altimeter System (ATLAS) instrument onboard the Ice, Cloud, and Elevation Satellite (ICESat-2) satellite system launched in September 2018 and was designed to measure the elevation of the glacial ice-sheets and measure the change in ice coverage/thickness of the northern hemispheres. (NASA, 2021) Even though the sensor was intended to measure frozen/non-vegetated surfaces, the photon returns can be, and have been, used for estimated canopy height in comparison to surface elevation, adding a dataset to the arsenal of AGB estimation.

ICESat-2 can be processing in a Python environment to estimate ground returns and top of canopy returns in order to estimate canopy height at a user-defined resolution (10 meters for this study). Once pre-processed, the estimated canopy height points can be transferred into GEE to facilitate ML regressions of canopy heights using ICESat-2 derived data as training input for the algorithms. With the global coverage of ICESat-2, this sensor system can prove to be a powerful tool in measuring and tracking mangrove ecosystems.

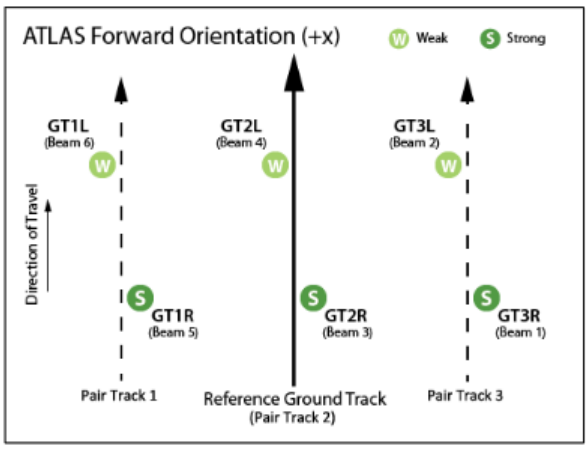


Figure 7: Spot and ground track naming convention utilized by ATLAS in the forward direction. (Neuenschwander et al., 2021)

2.2.5 GEDI

NASA's Global Ecosystem Dynamics Investigation (GEDI) mission consists of a full-waveform LIDAR instrument mounted on the International Space Station (ISS) in late 2018 and is purposely de-

signed to measure forest structure with its primary deliverable being mean aboveground biomass density over forested areas in the range of the ISS (51.6°N & 51.6°S latitudes). (Patterson et al., 2019)

Figure 8: GEDI canopy heights Zambezi River Delta in Mozambique.

One of the crucial datasets from GEDI, is the Level 2A Geolocated Elevation and Height Metrics product (GEDI02_A) which includes ground elevation, canopy top height, and relative height (RH) metrics as visualized in Figure 8. (Dubayah et al., 2021) This dataset is available in GEE in the form of monthly composite rasters with a resolution of 25 meters per pixel and 136 bands of information.

2.2.6 Sentinel-1

Sentinel-1 backscatter data is generated from two polar-orbiting satellites utilizing C-band radar imaging. (European Space Agency, 2021b) The synthetic aperture technique which Sentinel-1 utilizes allows for consecutive time of transmission and reception measurements over a single study area from differing angles incorporating the motion of the platform, essentially creating a large aperture than what is physically aboard the satellite. (Moreira et al., 2013) Synthetic Aperture Radar (SAR), active sensors, are powerful tools in EO as they do not require the Sun to illuminate a scene, or are depending on clear skies which allows for more frequent data collection and shorter return intervals, as well as providing different data than existing sensors, allowing users to have more information for ML applications. (Lemoen, 2022) Sentinel-1 data is made available on GEE at 10 meter resolution for the bands being utilized in this study.

2.2.7 Sentinel-2

Passive imagery, data that is collected from a sensor that relies on the photons emitted by a powerful source (typically the sun) and reflected off of the surface of the Earth, are a popular form of acquiring information about the surface of the Earth. One reliable platform is the Sentinel-2 constellation, a pair of sensors that collect up to 10 meter resolution imagery with the aims of monitoring the variability of surface conditions on our planet. (European Space Agency, 2021c) The bands available from Sentinel-2 also allow for the calculation of a variety of spectral indices as seen in Table 6. These bands, and their derived data, allow for increased information to feed machine learning models for training purposes and prove to be a valuable addition for mangrove masking and height estimations as seen later in this study. This study utilized both the 10 meter surface reflectance product in combination with the cloud probability dataset in order to filter and mask cloudy datasets to provide a clear time series for later calculations. Both of these datasets are freely available within GEE.

2.3 Machine Learning Applications

Generating local-level mangrove masks utilizing global datasets allows users to employ powerful Machine Learning (ML) algorithms to quickly generate a higher accuracy mangrove mask to estimate heights from, as well as also provide the user with the option to monitor the change in mangrove land cover from different time periods of interest. Utilizing GEE's ML library, users have access to many algorithms, such as the Random Forest (Breiman, 2001) and the Gradient Tree Boost (Friedman, 2002) algorithms, both being popular with Land Use Land Cover (LULC) applications. (Abdi, 2020) For this study, both RF and Gradient Tree Boost algorithms were tested in deriving a local-level mangrove masks, with the validity of each model being validated against the previous GMW 2016 global mangrove baseline, as well as visually inspecting the results against Sentinel-2 imagery to compare performance of the two datasets.

Using ML to classify land cover is a straightforward tasks it results in the discrete classification of classes over a study area, resulting in distinct land cover classes. Employing ML algorithms for canopy height estimations requires generating an interpolated, or continuous, surface of estimates of the same type, that is numerical canopy heights, not individual classes (Segal, 2004), being a popular tool for canopy height estimation and biomass prediction (Simard et al., 2011, Bourgine and Baghdadi, 2005, Baccini et al., 2008, Powell et al., 2010, and Baghdadi et al., 2015). Canopy height is a key data point for estimating AGB and thus will be the main form of information used to train a ML regression algorithm to estimate canopy heights where no data exists. (Lang et al., 2022) Canopy heights will be estimated utilizing the RF regression algorithm within GEE to generate a continuous surface of canopy heights. Once heights for a study have been generated, allometric equations will be utilized to estimate AGB based on canopy height for the area in terms of tonnes per hectare (Mg/ha).

2.4 Study Area

In order to compare with performance of canopy height estimations, an area of southern Florida in the United States was chosen as a study area. This area was chosen as it is rich in mangroves and has detailed LIDAR data freely available at high resolutions for much of the state as seen in Figure 9. Such data would allow for the processing and generation of a local CHM that could then be employed to validate the performance of regression canopy height estimations and allow for tuning of parameters to optimize performance to be used on other regions around the world.

A country along the eastern coast of Africa (refer to Figure 10), Mozambique is home to mangroves along a majority of its 2,700km coastline, representing the third largest mangrove area in Africa. (Fatoyinbo et al., 2008) Much of the coastal population of the county is dependent upon the mangroves for food, building material, and fuel, highlighting the need for monitoring the coverage and health of these critical ecosystems. (Fatoyinbo et al., 2008) Due to prevalence of mangroves, data availability of the region, Mozambique was chosen as the study site for this report to refine local mangrove masks, test machine learning algorithms and techniques, as well as estimate above ground biomass for the entire coastline of the country. The sheer length of coastline and volume of data required to be processed in order to generate AGB estimates would necessitate developing, tuning, and refining workflows that can generate accurate estimates in a timely manner over such an extent.

The data utilized and workflows generated for the country of Mozambique, can easily be employed to other countries around the world with minimal changes or alterations to the original code

within GEE, allowing for other users to easily and quickly conduct initial studies of mangrove coverage and health for their countries of interest.

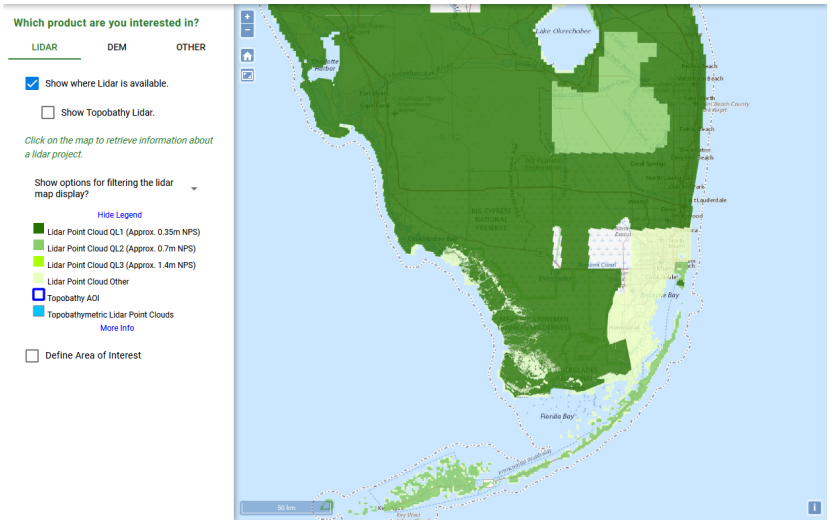


Figure 9: Extensive LIDAR coverage for the state of Florida. (USGS, 2022)

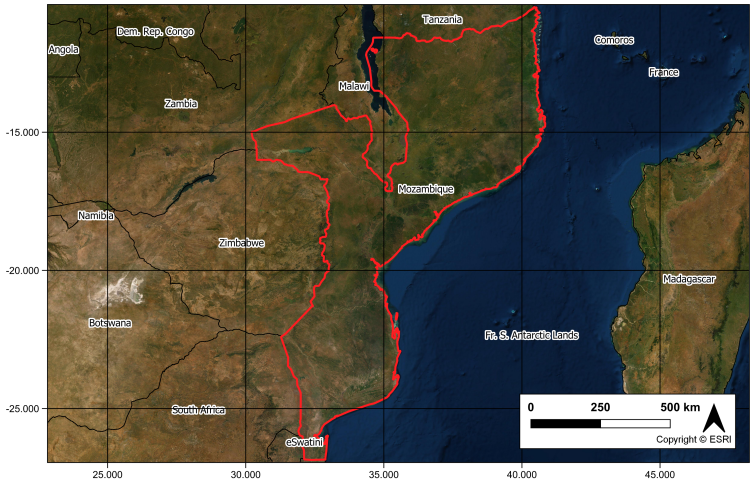


Figure 10: Mozambique study area. Features an extensive coastline with many mangrove-rich areas.

3 Theoretical & Methodological Framework

3.1 Data Processing & Data Fusion

There are many studies and code examples using ICESat-2 that use novel methods to detect surface returns from the myriad of noise returned to the ATLAS sensor, allowing for the detailed mapping of surface elevation along ICESat-2 tracks around the globe outside the intended scope of the mission. However, studies estimating canopy height from ATL03 returns at a higher resolution that the 100 meter official ATL08 data set are few and far between, highlighting a potential gap, or improvement, for canopy height estimations at a fine resolution.

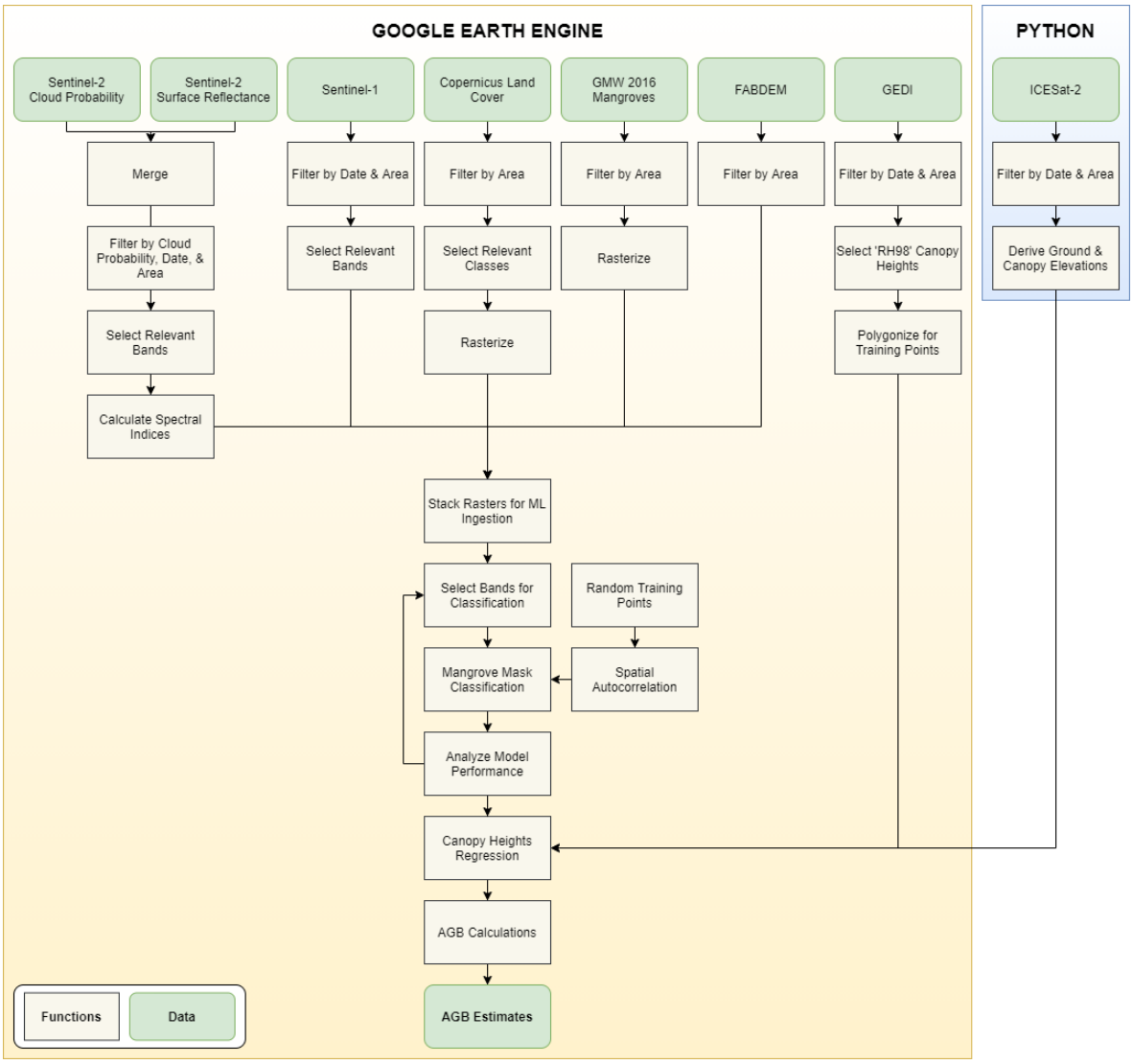


Figure 11: General workflow used. Refer to Figure 12 for canopy height estimation from ICESat-2 returns. (Boest-Petersen, 2022b)

As introduced in the previous section, such a study requires the querying, acquisition, and integration of multiple data types from various EO sensors. Methods will include using tools such as

Python, and its host of geospatial collection and processing libraries as well as machine learning capabilities; Google Earth Engine (GEE) for data collection, data processing, and machine learning capabilities; as well as Geographic Information Systems (GIS) for any further data processing and visualization.

Table 4: Primary data sources used for canopy height estimation. (Boest-Petersen, 2022b)

Data	Source	Purpose	Resolution	Notes
Sentinel-1	ESA Copernicus Mission (through GEE)	Backscatter showing 'texture' of terrain.	10m, 25m, & 40m	C-Band synthetic aperture radar imaging.
Sentinel-2	ESA Copernicus Mission (through GEE)	Visual and NIR bands. Suitable for vegetation indices.	10m to 60m	High resolution imagery.
ICESat-2 ATL03	NASA ATLAS Sensor (through NSIDC)	Deriving 'ground' and 'canopy' elevation to determine canopy height.	Dependent upon ground coverage.	Active LIDAR system with global coverage.
GEDI	NASA (through GEE)	Monthly canopy height composites.	25m	Limited to ISS latitudes.

Data utilized for mangrove masks and canopy height estimates will be restricted to sources that are easily and freely available via GEE or APIs accessed from Python libraries. Doing so will allow for a general data processing framework to be developed and permit future estimates at varying scales as data is made available. Data processing will be done in as few environments as possible, again limited to freely available, and easy-to-use platforms. These factors allow results to be recreated quickly, or for a new area of interest, and easily from any user.

3.2 ICESat-2 Processing

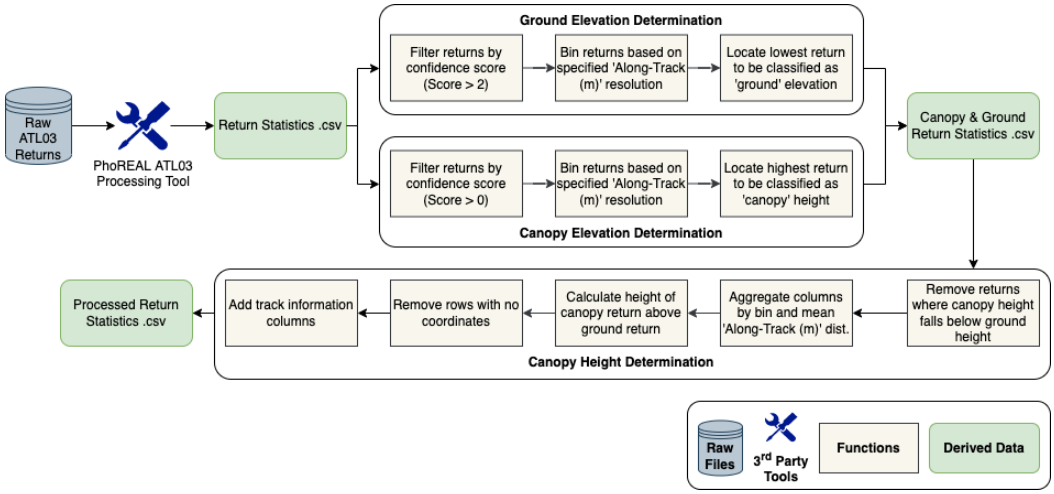


Figure 12: General workflow used to derive canopy heights from raw ATL03 data at user-defined resolution. (Boest-Petersen, 2022b)

In the initial stages of this report, one aim was to determine if ATL03 raw geolocated photon returns could supplement, or replace the existing ATL08, canopy height estimates for a 100 meter seg-

ment, dataset. As mangroves can be located in edge environments or deteriorating, height statistics for a 100 meter segment may prove to be too course for accurately measuring canopy heights.

The ATL03 Raw Geolocated Photons and ATL08 Land and Vegetation Height dataset originate from NASA's Advanced Topographic Laser Altimeter System (ATLAS) sensor on board the ICESat-2 satellite system. It is designed to emit 10,000 laser pulse per site and capture the returns with relevant statistics such as the photons geodetic longitude and latitude, allowing for the calculation of elevations of features across the Earth's surface. (Neuenschwander et al., 2021) This is accomplished by using laser pulses along six tracks as seen in Figure 7, with the three pairs separated by a distance of 14 meters.

In order to facilitate rapid querying, downloading, and processing of ICESat-2 data, a custom Python tool was created for this study. This tool, `icesat2_canopy_heights`[7], allows users to download ATL03, via the NASA Distributed Active Archive Center (DAAC), data for a region and time period of interest and derives estimated canopy heights at a user-defined resolution on the track of the dataset with the parameters of the tool visible in Table 5. Downloading of data is made possible with the `icepyx` Python tool. (Scheick et al., 2019–) In order to make the data more consumable for later processing steps, it was converted from the H5 file format, which can prove to be unwieldy, to a lightweight spreadsheet using the Photon Research and Engineering Library (PhoREAL) developed by the Applied Research Laboratories at the University of Texas at Austin. (Applied Research Laboratories, UT Austin, 2022) `icesat2_canopy_heights` also allows users to query and download raw ATL08 Land and Vegetation Heights for further analysis as well.

Table 5: `icesat2_canopy_heights.py` Python tool parameters for querying, downloading, and processing ICESat-2 ATL03 and ATL08 data. (Boest-Petersen, 2022a)

Property	Description
download	'True/False'. Download raw H5 files from NSIDC. If true, tool will download data. If false, tool will process files in specified working directory. 'False' by default.
data_type	'ATL03' or 'ATL08'. Specify which ICESat-2 dataset to download from NSIDC. 'ATL03' by default.
spatial_extent	Shapefile or coordinates. Extent to download ICESat-2 data for.
date_range	'YYYY-MM-DD, YYYY-MM-DD' Data range for data download. '2018-10-13, today' by default.
username	NASA EarthData registered username.
email	NASA EarthData registered email.
working_directory	Location for file storage/processing.
generate_csv	'True/False' Process raw H5 files with PhoREAL preprocessing tool to generate csv file, True by default.
track_num	If processing ATL03, define which track to derive canopy heights from, all by default.

[7] https://github.com/aboestpetersen/icesat2_canopy_heights

Table 5 continued from previous page

Property	Description
along_track_res	Integer resolution at which to derive canopy heights from. '10' meters by default.

Once the raw H5 ATL03 files have been converted into .csv files, the result spreadsheets were loaded into the script as `numpy` dataframes, allowing for further processing of the datasets. These dataframes are then split into bins whose sizes are dictated by the user-defined resolution they wish to achieve for the end product, for this case, the bins were 10 meters. To remove erroneous photon returns that are likely noise and can negatively skew processing results, only photons classified as medium (3) and high (4) confidence returns were used for ground estimations. Once filtered by confidence and bin size, the lowest elevation photon return was selected from each bin and determined to be the 'ground' elevation for that bin. As such, each 10 meter bin would be assigned a ground elevation of their respective lowest returns as seen in Figure 12.

For canopy height derivations for each bin, the low (2), and background (1), returns were included for the calculations, as it is often the case that canopy returns are mis-classified as noise upon closer inspection. To determine canopy height, the 95th percentile return in regards to maximum height was selected to the be top of canopy elevation for the bin. The 95th percentile was selected to remove returns that were unrealistically high in elevation to be classified as canopy.

Once the lowest, ground, elevation and highest, canopy, elevation was detected for each bin, these two heights could then be subtracted from one-another in order to determine canopy height for each bin along the ATL03 track to be used as training data for canopy height estimates within GEE.

3.3 Canopy Height Model Generation

Processing of local LIDAR datasets for Southern Florida were processed using LAStools[8]. Data was queried and downloaded in bulk for the year 2018 using the freely accessible United States Geological Survey (USGS) National Map Download client v2[9] for the area of interest in .laz format.

In order to generate a CHM, the bare-earth surface must be extracted from the .laz dataset using the `lasground` tool, classifying LIDAR returns into ground points (class=12) and non-ground points (class=1). Once the ground returns were located, the dataset was cleaned of noise using the `lasheight` tool to remove low and high outliers, returns located 3 meters below, and 50 meters above the recently derived surface. This removed any cloud, birds, or any other sources of noisy interference within the LIDAR dataset and increases the accuracy of the resulting CHM.

Next, high vegetation returns (class=5) was identified in the cleaned dataset using the `lasclassify` tool. With vegetation identified, the `lasheight` tool was used to create a normalized point cloud. Normalizing the data allows for more consistent heights across the dataset.Finally, a CHM can be derived by computing the difference between the DSM and DTM for the dataset. First, we thin the datasets to the last (or highest) returns using the `lasthin` tool with creates a DSM, or the surface of the area including canopies. Then, the `las2dem` tool is used to create a DTM of the terrain. These

[8] https://rapidlasso.com/lastools/
[9] https://apps.nationalmap.gov/downloader/

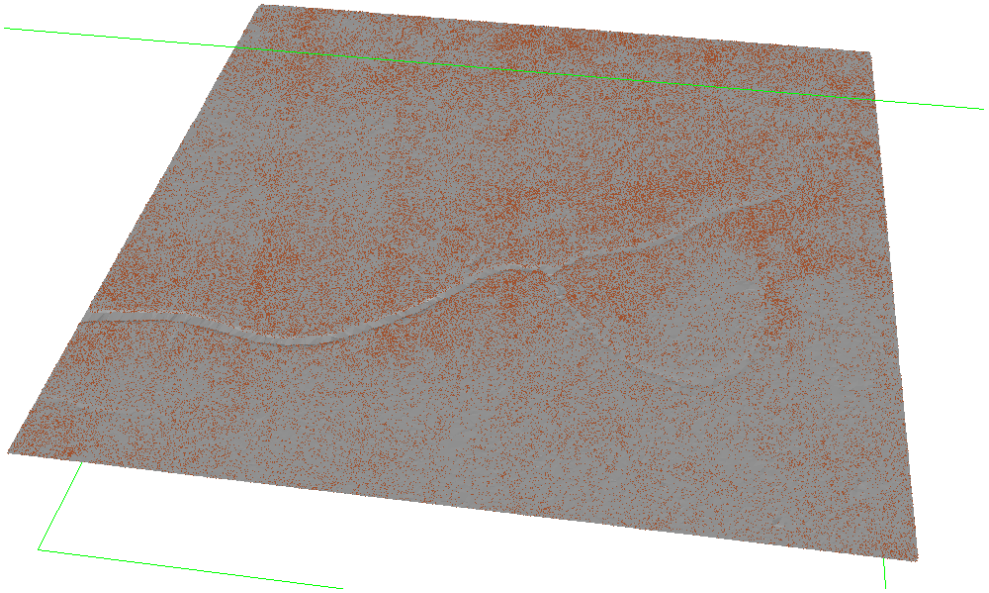


Figure 13: Raw LIDAR data for Florida everglades visualized with LAStools.

two surfaces are subtracted from one-another in order to generate a CHM to be used for data validation. For this study, a CHM was derived at a resolution of 10 meters. Once heights were estimated in GEE, the result was exported for further analysis in QGIS to compare height measurements.

3.4 Grid-Based Processing Areas

Due to large study areas and limitations of processing in GEE, large study areas, such as country-scale, where split into gridded areas to allow the processing of the following workflows on a per-grid basis, instead of the entire study area at once. Doing so allows the user to avoid encountering memory limitation restrictions, as well as avoiding long computation times which can likely time out. This was done using the `coveringGrid` tool in GEE as seen in Listing 1. Using this tool facilitated the study of very large study areas, such as the entire coastline of the country of Mozambique.

```
// Create Grids Over Study Area
var grids = (mozambique_coast.geometry()).coveringGrid('EPSG:4326', 100000);
```

Listing 1: Function used to create a over the study area to allow for the processing of workflows on a per-grid basis instead of the entire study area.

However, even sub-setting the study area into large cells can still offer challenges when processing over such an extent. As with the case of Mozambique, 36 grids proved to be too much for GEE to process in a single workflow. Due to this, only 30% of the created grids were selected for querying imagery and creating training data for the remaining cells as visualized in Figure 14. Doing so, decreases the amount of strenuous tasks needing to be completed on the server-side of GEE and allows for the classification and regression of much larger study areas.

3.5 Sentinel Processing

The approach taken for processing Sentinel 1 and 2 data in this report is outlined in Figure 11. In order for cloud free visual imagery, particularly in equatorial regions, it is best practice to com-

(a) Gridded area of Mozambique coastline.

(b) Randomly selected grids (blue) used for training.

Figure 14: Comparison of original grids versus randomly selected grids used for training.

pile a time-series of data. This involves collecting imagery from multiple dates and merging them together in order to get a more practical dataset of imagery. For this study, imagery was obtained of the course of an entire year as seen in Listing 9.

3.5.1 Sentinel-1

For this study, the vertical transmit/vertical receive (VV) and vertical transmit/horizontal receive (VH) were utilized to allow for cross-band polarization from the interferometric swath product from the Sentinel-1 sensor. After the date range was specified, the resulting imagery was filtered to the relevant area of interest.

```
// S1 Backscatter
var imgVV = ee.ImageCollection('COPERNICUS/S1_GRD')
        .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV'))
        .filter(ee.Filter.eq('instrumentMode', 'IW'))
        .select('VV')
        .map(function(image) {
          var edge = image.lt(-30.0);
          var maskedImage = image.mask().and(edge.not());
          return image.updateMask(maskedImage);
        });
var descChange_VV = (imgVV.filter(ee.Filter.eq('orbitProperties_pass', 'DESCENDING')))
    .filterDate(start, end).mean();
var ascChange_VV = (imgVV.filter(ee.Filter.eq('orbitProperties_pass', 'ASCENDING'))).
    filterDate(start, end).mean();
var imgVH = ee.ImageCollection('COPERNICUS/S1_GRD')
        .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VH'))
        .filter(ee.Filter.eq('instrumentMode', 'IW'))
        .select('VH')
        .map(function(image) {
          var edge = image.lt(-30.0);
          var maskedImage = image.mask().and(edge.not());
          return image.updateMask(maskedImage);
        });
```

```
var descChange_VH = (imgVH.filter(ee.Filter.eq('orbitProperties_pass', 'DESCENDING')))
    .filterDate(start, end).mean();
var ascChange_VH = (imgVH.filter(ee.Filter.eq('orbitProperties_pass', 'ASCENDING'))).
    filterDate(start, end).mean();
```

Listing 2: Function used to create a Sentinel-1 `ImageCollection` within GEE.

Sentinel-1 backscatter was queried and obtained within GEE as seen in Listing 2. This resulted in four bands of data; *'VV', 'VV-1', 'VH'*, and *'VH-1'*, with the ascending and descending pass for each band. Once these bands were selected, they were then added to the stack of bands (see Listing 4) that will be used to train both machine learning models for the mangrove mask as well as canopy height regression as seen in Figure 11. By using backscatter from a Synthetic Aperture Radar (SAR) source such as Sentinel-1, the sensor is able to collect data at a resolution that is finer than tradition radar imagery sources. (van Zyl and Kim, 2011)

3.5.2 Sentinel-2

The majority of raster data used in this study is derived from Sentinel-2 imagery. As stated with Sentinel-1 imagery above, the imagery was collected over the course of a year in order to obtain a cloud free composite image over the study area of interest.

Cloud-free composites were created in GEE by combining the Sentinel-2 Surface Reflectance dataset with the Sentinel-2 Cloud Probability dataset. Merging these datasets allows the user to view the level of cloud coverage in each Sentinel-2 image as well as which types of clouds are present over the study area. Access to this information can allow the user to filter out images that contain cloud cover over a user defined threshold. A threshold of 20% was utilized for this study, meaning that any image that contained over this level of cloud coverage was filtered out and not included in the final year-long composite used for training both ML algorithms. GEE allows users to construct custom functions to apply to datasets, streamlining processing and increasing efficiency of analysis. As seen in Listing 3, any image remaining after the sorting threshold for cloud cover then had this `maskS2clouds` function applied to it, masking any remaining clouds, leaving only clearly visible areas for later training.

```
// Function to mask clouds using the Sentinel-2 QA band.
function maskS2clouds(image) {
  var qa = image.select('QA60');

  // Select bits 10 (clouds) and 11 (cirrus).
  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  // Both flags should be set to zero, indicating clear conditions.
  var mask = qa.bitwiseAnd(cloudBitMask).eq(0).and(
             qa.bitwiseAnd(cirrusBitMask).eq(0));

  // Return the masked and scaled data, without the QA bands.
  return image.updateMask(mask).divide(10000)
      .select("B.*")
      .copyProperties(image, ["system:time_start"]);
}
```

Listing 3: Function used to mask cloud cover present in any Sentinel-2 image.

By utilizing a function, the above formula is applied to every Sentinel-2 image as it is queried, improving memory usage and accelerating processing times within the GEE JavaScript API. Before,

clouds were masked separately, once all of the imagery and been queried and loading into the environment. This often resulting in memory limits being exceed, or computations timing out after taking too long to process. This was due to the fact that the cloud mask was being applied to the entire year-long mosaic of imagery at once, instead of each single image as it was loaded as done with the function.

Once a cloud-free/cloud-masked mosaic was created for the area of interest, spectral indices were calculated using the bands and formulas as seen in Listing 4 and Table 6.

```
// Stack Bands Function
function stackBands(i){
  // S2 Indices
  var selected = i.select('B1', 'B2', 'B3', 'B4', 'B6', 'B8', 'B9', 'B11', 'B12');
  var SAVI = selected.expression('((b(5)-b(3))/(b(5)+b(3))+0.428)*1.428').rename('SAVI');
  var GNDVI = selected.expression('(b(5)-b(2))/(b(5)+b(2))').rename('GNDVI');
  var EVI = selected.expression('2.5*(b(5)-b(3))/(b(5)+6*b(3)-7.5*b(1)+1)').rename('EVI');
  var NDMI = selected.expression('(b(5)-b(7))/(b(5)+b(7))').rename('NDMI');
  var MSI = selected.expression('(b(7))/(b(5))').rename('MSI');
  var GCI = selected.expression('(b(6))/(b(2)-1)').rename('GCI');
  var BSI = selected.expression('(b(7)+b(3)-b(5)+b(1))/(b(7)+b(3)+b(5)+b(1))').rename('BSI');
  var NDWI = selected.expression('(b(2)-b(5))/(b(2)+b(5))').rename('NDWI');
  var ARVI = selected.expression('(b(5)-(2*b(3))+b(1))/(b(5)+(2*b(3)+b(1)))').rename('ARVI');
  var MNDWI = selected.expression('(b(7)-b(2))/(b(7)+b(2))').rename('MNDWI');
  var NDVI = selected.normalizedDifference(['B8', 'B4']).rename('NDVI');
  return i.addBands([SAVI, GNDVI, EVI, NDMI, MSI, GCI, BSI, NDWI, ARVI, MNDWI, NDVI,
    descChange_VV, ascChange_VV, descChange_VH, ascChange_VH, elev])
  .set('system:footprint', i.geometry());
}
```

Listing 4: Function used to calculate spectral indices per Sentinel-2 image and stack resulting calculations into training dataset.

As with the cloud masking/removal function, calculating spectral indices in a function allows the formulas to be applied to each Sentinel-2 image as it is queried, instead of the composited dataset as a whole. This speeds up processing times, and lowers the risk of encountering memory errors or timing out the calculations. The inclusion of these additional bands of information provide data for the machine learning algorithms to estimate land cover from as well as estimating canopy height based on pixel value. After calculating the indices, the information was stacked into the training raster as seen in Figure 11.

3.6 GEDI Pre-Processing

GEDI canopy height data was made available in GEE's Data Catalog in early 2022. This facilitated rapid querying and collection of relevant data for the study area. To filter out any invalid data, the '*quality_flag*' attribute was utilized, values equal to 1 were determined to be valid waveforms returned from the surface of the Earth. Next, the '*degrade_flag*' attribute was set to 0, ensuring that the data selected was not in a degraded state of pointing or positioning information. Finally, to limit the negative impacts of the sun on the quality of data returning to the sensor, returns were limited to those captured during the evening hours utilizing the '*solar_elevation*' attribute as seen in Listing 5.

Table 6: Spectral indices for Sentinel-2 sensor. (Boest-Petersen, 2022b)

Statistic	Band Calculation	Importance
Atmospherically Resistant Vegetation Index (ARVI)	$\frac{(B8–(2\times B2)+B2)}{(B8+(2\times B2)+B4)}$	Vegetation index that is not prone to atmospheric influence. (Tanre et al., 1992)
Bare Soil Index (BSI)	$\frac{(B11+B4)–(B8+B2)}{(B11+B4)+(B8+B2)}$	Quantifies mineral content in soil. (Rikimaru et al., 2002)
Enhanced Vegetation Index (EVI)	$2.5 \times (\frac{B8–B4}{B8+6\times B4–7.5\times B2+1})$	Vegetation index which corrects for some atmospheric influences and canopy background noise for densely vegetated regions. (Somvanshi and Kumari, 2020)
Green Coverage Index (GCI)	$(\frac{B9}{B3}) - 1$	Measure plant health by measuring chlorophyll content in vegetation. (Bian et al., 2020)
Green Normalized Difference Vegetation Index (GNDVI)	$\frac{B8–B3}{B8+B3}$	NDVI that is more sensitive to chlorophyll content. (Gitelson et al., 1996)
Moisture Stress Index (MSI)	$\frac{B11}{B8}$	Canopy stress analysis. (Hunt and Rock, 1989)
Normalized Difference Moisture Index (NDMI)	$\frac{B8-B11}{B8+B11}$	Measures vegetation water content. (Wilson and Sader, 2002)
Normalized Difference Vegetation Index (NDVI)	$\frac{B8-B4}{B8+4}$	Measures plant productivity. (Weier and Herring, 2000)
Normalized Difference Water Index (NDWI)	$\frac{B3-B8}{B3+B8}$	Water body analysis that is sensitive to coastal environments. (McFeeters, 1996)
Soil Adjusted Vegetation Index (SAVI)	$(\frac{B8-B4}{B8+B4+0.428}) \times 1.428$	NDVI corrected for background soil brightness in areas of low vegetation cover. (Huete, 1988)

For this study, the '*rh98*' attribute was utilized as it describes relative canopy height metrics at 98% for each 25m pixel. Using this metric allows for the removal of any potential outlier heights that can negatively skew the dataset and decrease accuracy of later estimations. Then the data was filtered to the area and time period of interest to create a dataset of relevant canopy heights.

```
// GEDI Canopy Information
var qualityMask = function(im) {
  return im.updateMask(im.select('quality_flag').eq(1))
      .updateMask(im.select('degrade_flag').eq(0))
      .updateMask(im.select('solar_elevation').gt(0)); // Select data collected at
    night
};
var gedi_data = ee.ImageCollection('LARSE/GEDI/GEDI02_A_002_MONTHLY')
                  .filterDate(start, end)
                  .filterBounds(grids)
                  .map(qualityMask)
                  .select('rh98')
                  .mean()
                  .rename('canopy_h');
```

Listing 5: Function used to filter GEDI data for canopy height training.

3.7 Copernicus Land Cover

To aid with the classification of mangroves and regression of height estimates by feeding the algorithm more information, the Copernicus Global Land Cover dataset was utilized at 100 meter resolution. This was used in tandem with the GMW global mangrove dataset for 2016 to create pixels that were classified as mangroves, non-mangroves, ocean, or forest as seen in Listing 6.

```
// Copernicus Land Cover
var cop_landclass = ee.Image("COPERNICUS/Landcover/100m/Proba-V-C3/Global/2019").
    select('discrete_classification').clip(grids.geometry());
var cop_ocean = (cop_landclass.updateMask(cop_landclass.eq(200))).rename('
    mangrove_label'); // Select 'ocean' pixels
cop_ocean = cop_ocean.reduceToVectors({geometry: grids.geometry(), maxPixels: 1e12});
var ocean_image = ee.Image(3).clip(cop_ocean); // Assign value of 3
var cop_forest = (cop_landclass.updateMask(cop_landclass.eq(112))).rename('
    mangrove_label'); // Select 'forest' pixels
cop_forest = cop_forest.reduceToVectors({geometry: grids.geometry(), maxPixels: 1e12})
    ;
var forest_image = ee.Image(4).clip(cop_forest); // Assign value of 4

// Mangroves
var mangroves_image = ee.Image(0).clip(grids.geometry());
var mangroves_image = mangroves_image.paint(gmw, 1).rename('mangrove_label').clip(
    grids.geometry()); // Paint Mangroves
var training_mangroves = ee.Image.cat([water_image, ocean_image, forest_image,
    mangroves_image]).reduce(ee.Reducer.firstNonNull()).rename('mangrove_label').toByte
    (); // Combine datasets
```

Listing 6: Function used to filter Copernicus Land Cover pixels and create training data for ML applications.

By incorporating more land cover classes into the training data, the ML model can more accurately predict land cover types. This was done by creating an uniform raster layer label as '0', or areas where mangroves where not present. Then the GMW dataset was converted from a vector to a

raster and 'painted' over the newly created image using the `paint` function in GEE. This essentially burned areas of mangroves into the raster labeled as the value 1. This was also done with subsequent ocean, and forests classes taken from the Copernicus Global Land Cover dataset at increasing values as seen in Figure 15. To reduce data loads and speed up processing times, data was clipped to areas that fell within the grids mentioned in Figure 14.

3.8 Land Cover Classification

A local land cover mask was generated in order to improve upon existing global datasets. Classifying at the local level allows for the classification of areas that may have been missed as well as increasing the resolution of the mask from 30 and 100 meters (GMW & Copernicus Global Land Cover respectively) to 10 meters. Higher resolution allows for increased granularity in monitoring mangrove habitat change. A resolution of 10 meters was selected as much of the training data is derived from Sentinel-2 imagery, which has a resolution of 10 meters. A general overview comparing GMW's mangrove baseline and the derived mangrove mask can be seen in Figure 16.

```
// Stratified Sample Generation
var strat_sample = (compilation.clip(grids.geometry())).stratifiedSample({
  numPoints: 100, // Points to be generated per land cover class
  region: grids,
  classBand: 'mangrove_label',
  scale: 100, // 100m resolution sampling
  geometries: true,
  tileScale: 16 // Overcome memory limitation issues
});
strat_sample = strat_sample.randomColumn(); // Split data for training & validation
var training = strat_sample.filter(ee.Filter.lt('random', 0.7)); // 70% for Training
var validation = strat_sample.filter(ee.Filter.gte('random', 0.7)); // 30% for
    Validation
// Define Classifier
var classifier = ee.Classifier.smileRandomForest(150).train({features: training,
    classProperty: 'mangrove_label', inputProperties: bands_lc});
// Train Classifier
var trained = classifier.train(training, 'mangrove_label', bands_lc);

// Classify mangroves by grid
function classify_image(trained_model, img){
  var classified = img.classify(trained_model);
  var mangroves_masked = classified.updateMask((classified.select('classification')).
    eq(1));
  // Seive Classification
  var mmu = 60;
  mangroves_masked = mangroves_masked.connectedPixelCount(mmu, false);
  mangroves_masked = mangroves_masked.updateMask(mangroves_masked.gte(mmu));
  return mangroves_masked;
}

var classified_mangroves = classify_image(trained, compilation);
```

Listing 7: Function used to randomly sample land cover training classes and predict mangrove cover for coastline of Mozambique.

To train ML models in GEE, training data must be ingested in the form of points. To obtain these points, the `stratifiedSample` tool was utilized, mitigating the need to generate random points

in a separate environment. The tool generates a user-specified number of points, in this case 10, per unique band value. In this case of this study, 10 points were generated per land cover class (4 in total; non-mangroves, mangroves, ocean, and forest) per randomly selected grid in the study area. Due to the large study area in each grid, the number of points able to be generated for training is rather small, at 10 points per class in order to not overcome memory or computation limits of the GEE JavaScript API. Another limitation is the scale of the training data. Instead of sample at 10 meter resolution, that is 1 point per pixel of training data, points were sampled at a resolution of 100 meters. This meant that each point had the values of the pixels in the surround 100 meters averaged and then sampled to the point, saving memory space and shortening computation time. The resulting product were points that contained information of every band in the training raster that lie beneath the point to be used as training data for the mangrove classifier.

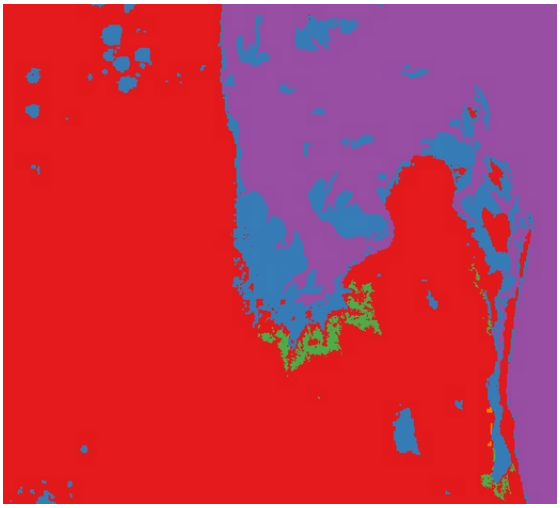


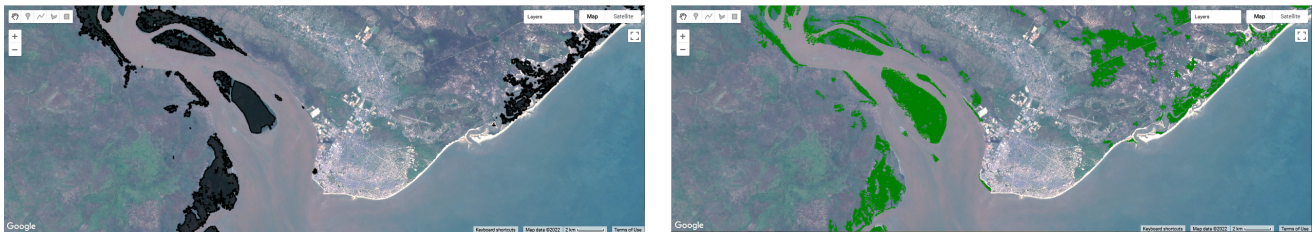
Figure 15: Land cover raster utilizing information from the 2016 GMW mangroves baseline as well as the Copernicus Global Land Cover dataset for the mangrove mask classification for Mozambique coast.

Once training points have been generated for the study area, the points were split into training points, with 70% of the original points being used to train the dataset, and validation points, with 30% of the points used for measuring the performance of the land cover classification model over the study area. In GEE, there a plenty of options in regards to ML algorithms. For this study, the `Random Forest` algorithm was utilized for the land cover classification as it is proven to be a powerful, yet efficient, algorithm in terms of segmented land cover classification. This algorithm was followed closely by, and even exceeded in alternative study areas, by the Gradient Boosted Decision Trees. However, with GEE, these algorithms can easily replace each other, with the user being able to test multiple algorithms quickly and efficiently. For this study, 150 trees were used to train the classifier as increasing the number of trees only results in marginal improvements in classification and significant increases in processing time. By training the classifier from grids randomly selected over the study area, the user can apply the trained classifier over larger, appropriate, study areas. As this model was trained on random grids over the coastline of Mozambique, it was deemed appropriate to use for the remaining areas of the Mozambique coastline that were not used for training due to computation limitations.

As seen in Listing 7, as well as previous listings, the machine learning algorithm has been placed

in a function in order to decrease computation time and reduce processing failures to due memory errors. The function in question, `classify_image` (Line 19 of Listing 7), requires only 2 inputs, the previously trained classifier model, and the images that are to be classified (the raster dataset of stacked bands). By wrapping the classification in a function, the classifier is only working with a single image at a time instead of attempting to classify the entire coastline in a single attempt which will likely result in unnecessarily long processing times or even GEE failing to classify the area as a result of memory errors.

After the image was classified for mangroves, the resulting data was seived, or a Minimum Mapping Unit (MMU) was defined. Doing so removes the classifications of areas that are too small reliable be defined as a mangrove ecosystem, as well as removing small erroneously classified pixels. For this study, the MMU was defined to be 60m^2. By using the `connectedPixelCount` tool on line 24 of Listing 7, GEE sums any neighboring pixels which are connected to each other, locating groups of pixels. Once these groups are located, the `updateMask` tool is used to remove groups that are below the MMU threshold, resulting in a more cohesive, accurate mangrove mask dataset.



(a) GMW classification of mangrove areas.

(b) ML classified mangrove mask.

Figure 16: Comparison of GMW mangrove areas versus ML derived mangrove mask for potion of the Mozambique coast.

Once the classification is complete, the validation points can be employed to measure the performance of the land cover classification. In combination with GEE's band importance metrics, the user can fine-tune the parameters of the ML algorithm for their study area of interest, in order to improve the overall performance of the land cover classification, resulting in more accurate biomass estimates over a region at a resolution of 10 meters per pixel instead of the existing 30 meters per pixel.

3.9 Canopy Height Estimation

GEE's machine learning library also proved to be a powerful tool for estimating mangrove canopy heights over large study areas. These canopy heights can then be employed to estimate AGB using height-dependent allometric equations.

```
// Function to estimate heights and biomass
function mangrove_biomass (gridded_area) {
  // Convert GEDI Raster to Points
  var height_data = gedi_data.sample({
    region: gridded_area.geometry(),
    scale: 25, // Resolution of GEDI data
    geometries: true, // Maintain geometries
    numPixels: 2500, // Limit # of samples
  });
  // Build Training Data to Join Band Info to GEDI Points
  var training_height = compilation.select(bands_height).sampleRegions({
```

```
    collection: height_data.limit(1000), // Limit points sampled
    properties: [training_properties], // Attribute to sample 'canopy height'
    scale: 25, // Resolution of GEDI data
    tileScale: 16, // Avoid memory limit
  });
  // Train a ML Classifier
  var trained_height = ee.Classifier.smileRandomForest(150).setOutputMode('regression'
    ).train(training_height, training_properties, bands_height);
  // Estimate Heights
  var classified_heights = compilation.classify(trained_height);
  // Mask Mangroves
  var masked_heights = classified_heights.updateMask((classified_mangroves.select('
    classification')).eq(60)); // Select sieved mangrove areas
  // Calculate AGB from Simard et al. 2019 Supplemental Table 8
  var agb = (masked_heights.expression('0.44*(height)', {'height': masked_heights.
    select('classification')})).pow(ee.Number(2.1578));
  return agb;
}

// Apply function on gridded coastline
var mangrove_heights = mangrove_biomass(grids);
```

Listing 8: Function used estimate canopy height and subsequent AGB for coastline of Mozambique.

In order to estimate heights, the GEDI L2A Raster Canopy Height (Version 2) was employed as it has more granular resolution than ICESat-2 ATL08 (25 meter versus 100 meter resolution), performed better than derived ATL03 canopy heights, and is part of a mission that's sole purpose is to measure canopy heights and estimate biomass within the latitudes of the ISS with good coverage as seen in Figure 8. In GEE, GEDI canopy heights are made available in raster format and therefore must first be converted to points to allow for ingestion into a ML algorithm. This was accomplished utilizing the `sample` tool, generating a point at the center of each canopy height pixel with the relevant statistics required to train a model to predict canopy heights (can be seen on line 2 of Listing 8). To prevent performance issues, the number of canopy height points generated was limited to 2,500 measurements for the study area.

After obtaining GEDI points, these points were then sampled against the stack raster dataset containing raw Sentinel-1 and 2 bands, spectral indices, elevation, and land cover types. Doing so requires using the `sampleRegions` tool in GEE. As with the `stratifiedSample` tool, this collects the band information from the stacked raster datasets per canopy height point generated from GEDI data. This results in points with the measured canopy heights and band statistics as well that can be employed to train a ML regression model for canopy height estimations. As with previous steps, sampling over such a large area can easily overwhelm the users' memory limits within GEE, as such, the number of points sample was limited to the first 1,000 points from the 2,500 available canopy height points.

Next, the `Random Forest` algorithm was used to train a classifier to estimate canopy heights that could be employed on other images or grids. A continuous surface can be generated by setting the `setOutputMode` parameter to '`regression`', allowing the user to generate a range of estimated heights over the study area instead of a discrete classification. To train the algorithm requires three parameters; the sampled points with band statistics and canopy height data, the attribute to classify (canopy height), and which bands to use for training the classifier. The trained classifier was then used to train each image in the gridded coastline dataset to estimate canopy heights an areas that

were previous classified as mangroves using the `classify` tool in GEE as seen in line 20 in Listing 8. Once canopy heights have been generated, an allometric equation can be applied to the area to generate a biomass estimate.

Canopy heights generated from the ML model are then exported from GEE as a raster to be compared against locally measured CHM to compare the performance of the model estimations to field data. Locally-sampled field measurements from Simard et al., 2019 were also utilized to measure height estimate performance against in-situ data points.

3.10 Mangrove Biomass Estimation

3.10.1 Allometric Equations

Above Ground Biomass can be estimated using allometric equations and valuable tools that provide a cost-effect alternative to biomass estimates when compared to measuring the data in the field. (Ravindranath and Ostwald, 2008).

Table 7: **AGB calibration models.** Where $H_{max} = 1.697 \times H_{SRTM}$ (with ICESat RH100) (Simard et al., 2019).

Allometric Model	**R^2**	**RMSE (Mg ha^{-1})**	**Allometry name**	**Region Covered by Model**
$AGB = 2.572 \times {H_{max}}^{1.5191}$	0.70	180.0	Global Hmax power	Global
$AGB = 0.440 \times {H_{max}}^{2.1578}$	0.85	66	East Africa Hmax power	Somalia to South Africa
$AGB = 0.745 \times {H_{max}}^{1.6228}$	0.70	65	Americas Hmax power	North, Central and South America

As seen in Listing 8, the AGB regressions visible in Table 7 can easily be implemented in a function within GEE once canopy heights have been estimated in order to derive AGB. With region specific, and global, formulas, the user can fine-tune the estimation of biomass depending on the location of the study area in question.

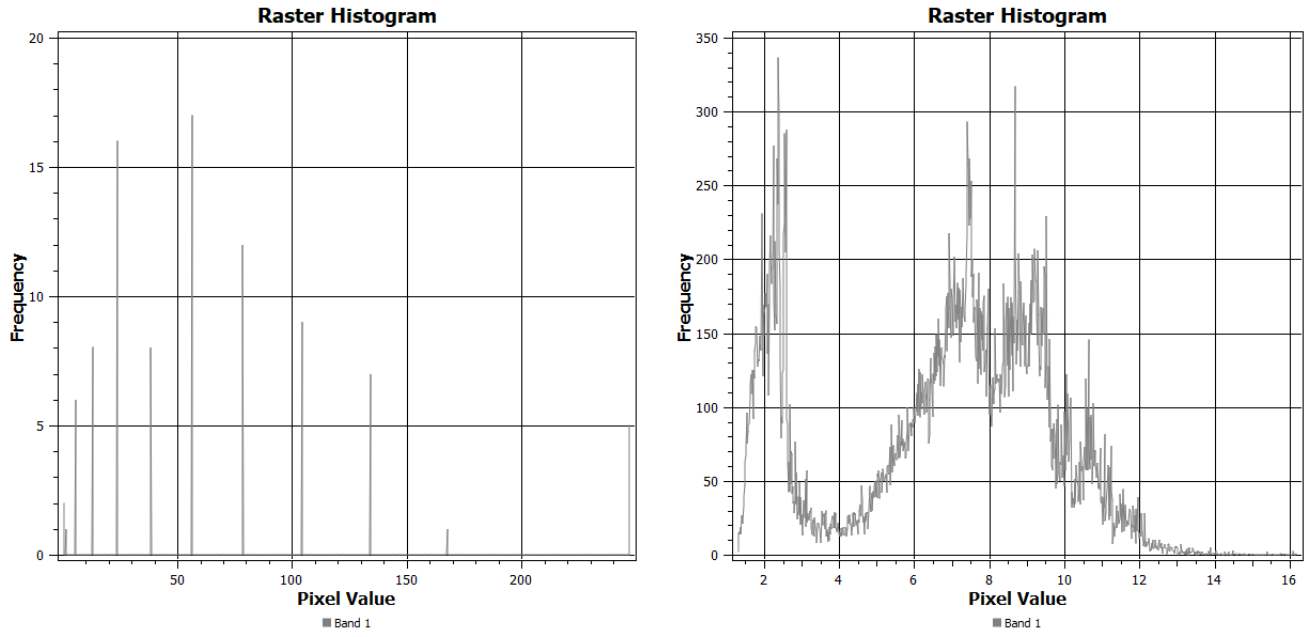
4 Results & Analysis

4.1 Results

Table 8: AGB Change statistics for the Eastern coast of Africa.

	AGB (Mg/Ha)		
	2019	**2020**	**2021**
Somalia	361	361	370
Kenya	162,661	21,362	20,232
Tanzania	24,282	192,300	74,437
Mozambique	90,145,978	N/A	N/A
South Africa	4,168	2,361	7,784

Simard et al., 2019 estimate that there is 261,387,504 Mg/Ha of carbon found in Mozambique with there study in comparison to the results seen in Table 8. However, there are high levels of AGB variability when exporting results from GEE, resulting in inconsistent results when working with derived data outside of the browser API. Over such large study areas, it can prove challenging to obtain AGB statistics for a study area, and thus it is recommended to work at smaller scales when possible for more accurate figures and quicker processing times.



(a) Simard et al., 2019 AGB histogram.

(b) Histogram derived in this study.

Figure 17: Comparison of mangrove histogram distribution for coastline of Mozambique. Pixel value represents AGB in Mg/Ha.

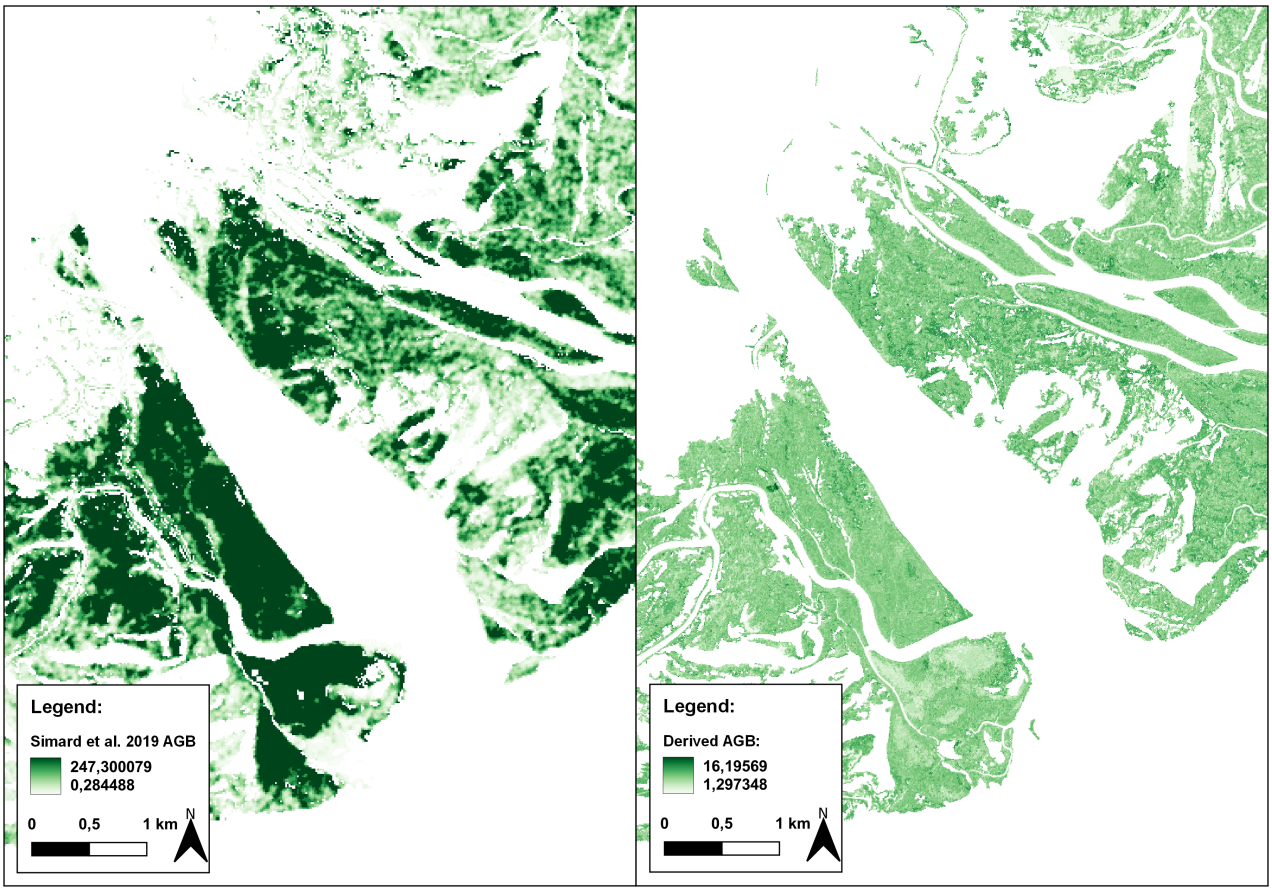


Figure 18: Comparison of Simard et al., 2019 AGB estimates (left), with AGB estimates derived from this study for a portion of the Mozambique coastline.

4.2 Processing Method

As seen in the previous sections, and Listing 9, GEE is a powerful and capable tool for querying, processing, and analyzing large amounts of remotely sensed EO data throughout the entire globe. However, considerations must be made when processing over large study areas, such as the entire coastline of the country of Mozambique. The use of functions, wrapping complex functions and calculations in a format that allows for the implementation of efficient workflows, allows users to effectively implement large-scale, high resolution spatial analysis over vast areas and incorporating many different data sources and formats. Conducting analysis in GEE, also affords users the possibility of generating workflows for a study area whilst have the flexibility of implementing their workflow over different study areas with the change of a few parameters within the browser-based editor. Without the use of functions, an analysis on a country-wide scale would not be possible.

4.3 Mangrove Mask

When using the `stratifiedSample` tool within GEE to generate 100 random points per four land cover classes, a validation accuracy of 90-95% is consistently achieved for the coastline of Mozambique depending on whether the `Random Forest` or `Gradient Tree Boost` (both with 150 trees) al-

gorithm was utilized (results can be seen in Tables 9 & 10). As seen in Figure 19, the ML model is able to catch significantly more detail along the coast of the study area, classifying small mangrove ecosystems which the GMW fails to classify.

(a) Sentinel-2 yearly composite. (b) GMW 2016 baseline. (c) ML mangrove classification.

Figure 19: Comparison of mangrove masking for portion of Mozambique coastline.

As seen in Table 10, the `Gradient Tree Boost` algorithm slightly out-performs the `Random Forest` algorithm for the coastline of Mozambique utilizing the same parameters. However, with the ease of use allowing by operating in the GEE JavaScript API, algorithms can quickly and easily be swapped for testing of performance given different study areas or input parameters. By generating a local-level mangrove mask generated from a global baseline dataset, we can more accurately estimate the amount of AGB within a study area, as the physical area of mangrove ecosystems on the surface is more accurately represented. as this study generates a mask at a resolution of 10 meters, the edges of such environments have more detail and can allow for higher detailed tracking of mangrove loss/gain.

Table 9: Confusion matrix for Random Forest mangrove mask for Mozambique coastline using Gradient Tree Boost algorithm (92% Validation Accuracy). Actual classifications are represented on the x-axis with predicted classifications on the y-axis.

Classification	**Non-Mangroves**	**Mangroves**	**Water**	**Ocean**	**Forest**
Non-Mangroves	29	1	0	0	3
Mangroves	1	29	0	0	0
Water	2	1	26	0	0
Ocean	0	0	4	23	0
Forest	0	0	0	0	35

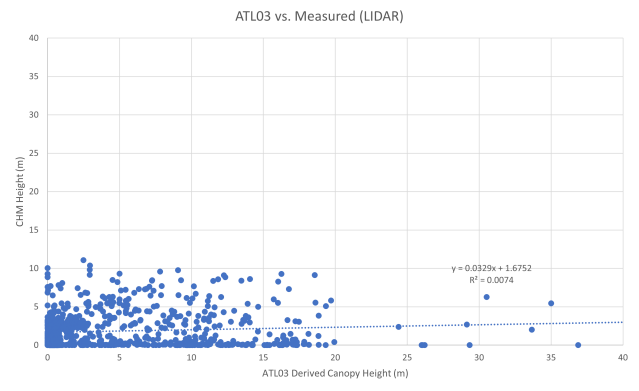
4.4 Canopy Height Measurements

In order to measure the performance of ATL03 canopy height derivations, they were directly compared with CHM generated from locally sourced LIDAR scans. For this study, comparisons were made in the area of Jobos Bay, Puerto Rico and in the coastal Everglades of southern Florida as there is ample ICESat-2 coverage as well as readily available local LIDAR data. The performance for ATL03, ATL03, and GEDI data against local CHMs can be seen below.

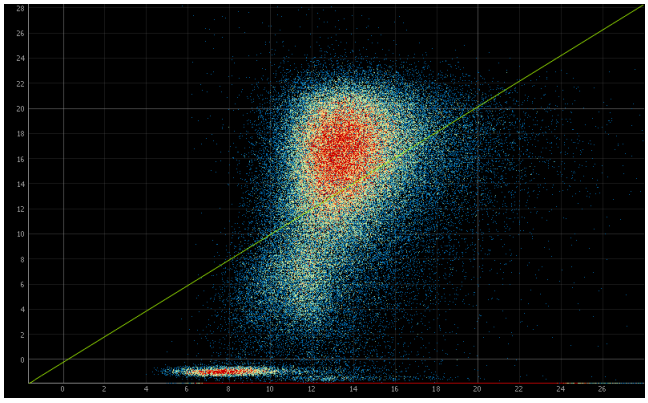
For this report, the CHM was generated at a resolution of 10 meters per pixel for efficient processing times. Canopy heights were derived at a resolution of 10 meters along the track of the dataset

Table 10: Confusion matrix for Gradient Tree Boost mangrove mask for Mozambique coastline using Gradient Tree Boost algorithm (94% Validation Accuracy). Actual classifications are represented on the x-axis with predicted classifications on the y-axis.

Classification	Non-Mangroves	Mangroves	Water	Ocean	Forest
Non-Mangroves	28	0	0	0	2
Mangroves	1	30	1	0	0
Water	1	1	26	2	0
Ocean	0	0	0	22	0
Forest	2	0	0	0	29



(a) Derived canopy heights for Jobos Bay, Puerto Rico.



(b) Scatter/density plot comparison of derived heights for Southern Florida. Y axis: CHM, X axis: ATL03.

Figure 20: Performance of derived canopy heights from ATL03 compared against heights obtained from local LIDAR datasets for Jobos Bay & Southern Florida.

represented by a point with the canopy statistics for the center of that 10 meter bin. This required that zonal statistics be utilized within QGIS to measure performance. Once in QGIS, each point with canopy information had a 10 by 10 meter polygon constructed around it. From these polygons the average canopy height pixel value of the CHM beneath was measured, allowing the user to compare the derived measurements of the ATL03 custom workflow with ground measurements as can be seen in Figure 20.

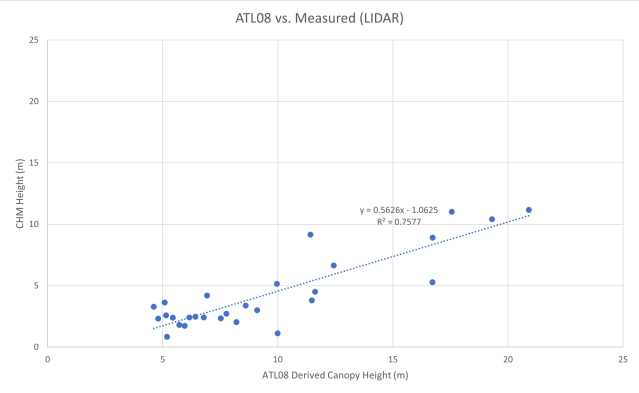
As seen in Figure 20, the derived canopy heights greatly over estimate the actual heights of the canopy on the ground. Even when ICESat-2 returns are taken for the same period in which the local LIDAR data was collect, 2018, and with only the 95th percentile of heights being chosen for canopy heights in the ATL03 data, heights will still much higher than actual measurements with an R^2 value of 0.0074 for Jobos Bay, showing little correlation between the two datasets.

The same workflow was employed to compare ATL08 measurements with local data as was done with ATL03 canopy height derivations. Instead of a 10 by 10 meter polygon being drawn around each data point, ATL08 required different considerations. As canopy height metrics in ATL08 datasets are average heights along a 100 meter segment, a line following the path of the along-track point collection was drawn along the points. From this a 16 meter wide, corresponding with the width of an ICESat-2 track, was drawn around all ATL08 points following the direction of the path. Finally, the

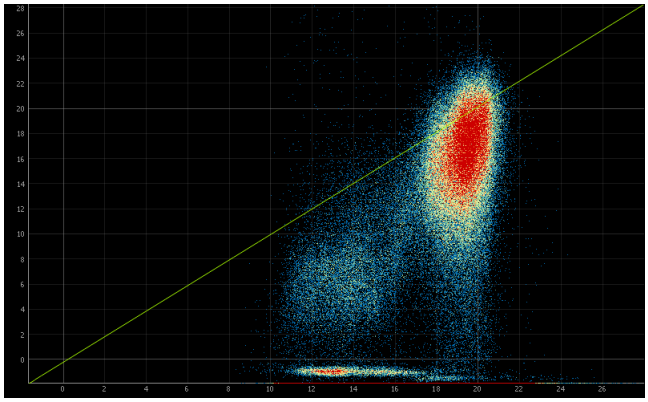
line was split into 100 meter segments, with each ATL08 point being located at the center of the segment as seen in Figure 21.

Figure 21: ATL08 polygons against CHM used to collect zonal statistics to measure data performance.

From polygons above, the average canopy height from the CHM for the area was taken in order to compare the performance of ATL08 versus local LIDAR data for the area. As seen in Figure 22, ATL08 performs relatively well compared to local measurements even when estimating height for a 100 meter segment. With an R^2 value of 0.7577, ATL08 can prove to be a powerful alternative to provide estimated canopy heights for a region with quick turnaround time.



(a) Derived canopy heights for Jobos Bay, Puerto Rico.

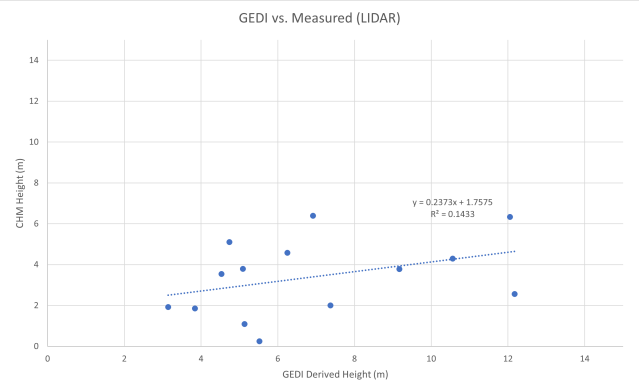


(b) Scatter/density plot comparison of derived heights for Southern Florida. Y axis: CHM, X axis: ATL08.

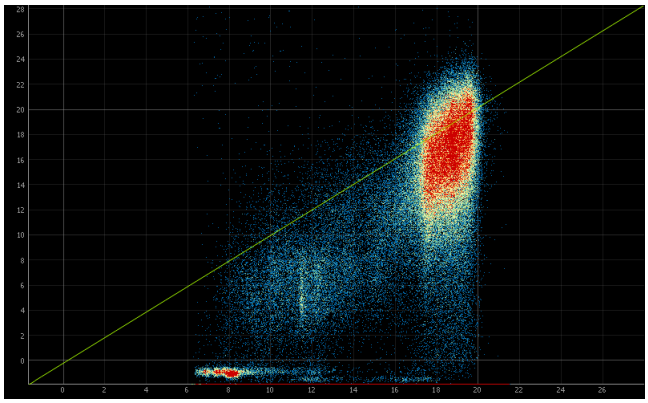
Figure 22: Performance of derived canopy heights from ATL08 compared against heights obtained from local LIDAR datasets for Jobos Bay & Southern Florida.

4.5 Canopy Height Estimates & AGB

Based on the performance of canopy heights that can be seen in the above section, as well as ease of availability in GEE, the GEDI, canopy heights raster dataset was chosen for canopy height model



(a) Derived canopy heights for Jobos Bay, Puerto Rico.

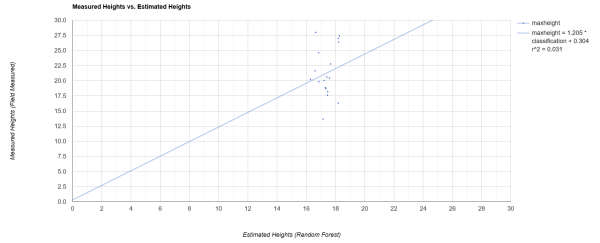


(b) Scatter/density plot comparison of derived heights for Southern Florida. Y axis: CHM, X axis: GEDI.

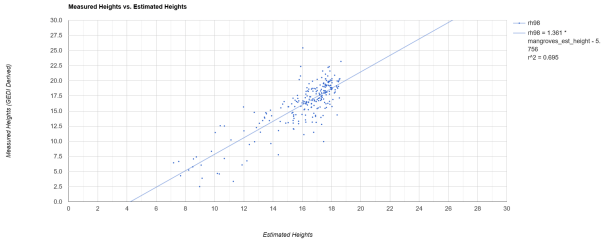
Figure 23: Performance of derived canopy heights from GEDI compared against heights obtained from local LIDAR datasets for Jobos Bay & Southern Florida.

training in future processing and study areas.

The allometric equations presented in Table 7 allowed for the rapid derivation of AGB estimates in GEE (as seen in Line 24 of Listing 8) when other ground-based tree measurements were not available and/or could not be acquired which would allow for more precise estimations.



(a) GEDI heights compared to field measurements.

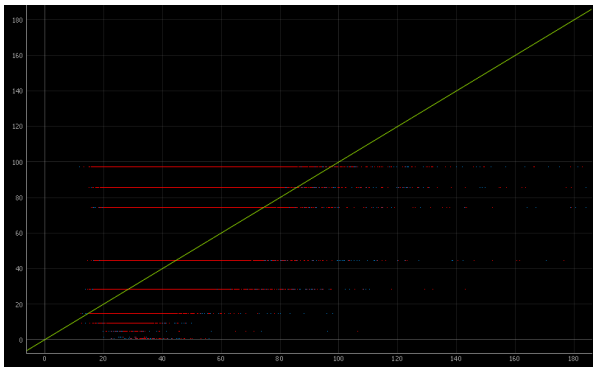


(b) GEDI height performance.

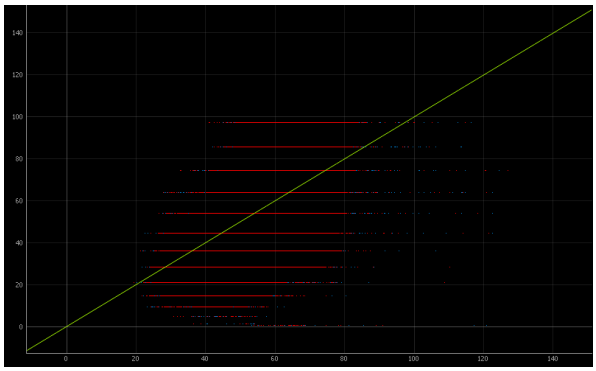
Figure 24: GEDI performance in Southern Florida (m).

With an R^2 value of 0.731 as seen in Figure 24b, GEDI performs well (see Figure 30) in tandem with the `Gradient Tree Boost` ML algorithm to estimate canopy heights for a sample area of southern Florida. This score is followed closely by the `Random Forest` algorithm with an R^2 score 0.724 with both algorithms being trained with 250 trees. Both results were trained with 726 points, and validated against 338 GEDI points with the '*rh98*' canopy height return.

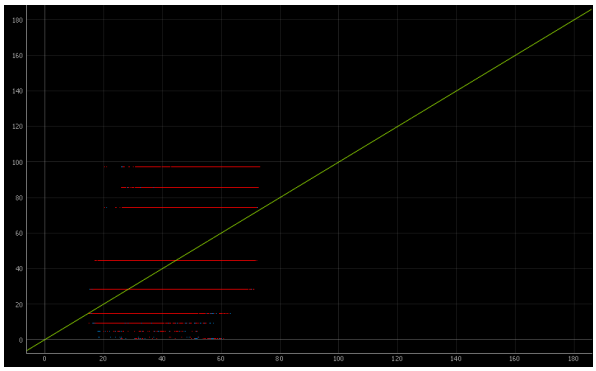
In the southern Florida study area, the derived canopy heights from ICESat-2 ATL03 performed rather poorly, with an R^2 of 0.075 with the `Gradient Tree Boost` algorithm and an R^2 of 0.094 with the `Random Forest` algorithm. In both scenarios the canopy height model was trained with 8,693 ICESat-2 ATL03 returns, and validated against 3,745 returns. Finally, using ATL08 ground and vegetation height returns, an R^2 of 0.178 was achieved with the `Gradient Tree Boost` algorithm, and an R^2 of 0.147 was generated with the `Random Forest` algorithm with 1,674 returns for training and



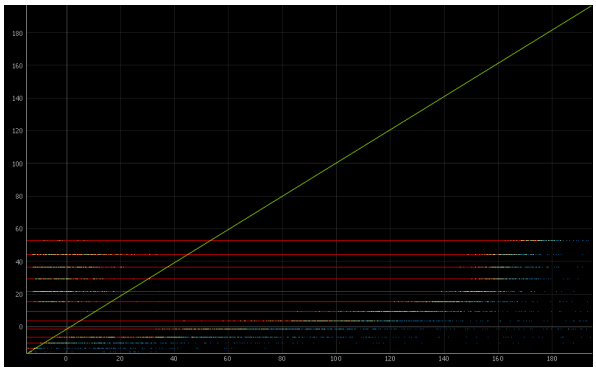
(a) ATL03 AGB estimations.



(b) ATL08 AGB estimations.



(c) GEDI AGB estimations.



(d) CHM AGB estimations.

Figure 25: Scatter plots comparing Simard et al., 2019 AGB estimations (Y axis) for southern Florida against available canopy height datasets. Height datasets used in this study are represented along the X axis.

768 returns for validation the models. Further AGB comparisons for southern Florida can be seen in Figures 26, 27, 28, and 29 located in the Appendix.

From these results, we can see that GEDI derived canopy heights scored the highest when compared to ICESat-2 ATL03 derived heights, and ATL08 out-of-the-box heights. With the ease of use and accessibility of GEDI heights being available in a cloud platform such as GEE, these points were opted for when generating biomass estimates for other study areas. Overall, the `Gradient Tree Boost` algorithm scored slightly higher than the `Random Forest` algorithm, and was thus used in future scenarios. However, users can easily test and swap different algorithms within the GEE environment.

With canopy heights quickly and easily generated with the above-mentioned accuracy, allometric equations to estimate AGB can be applied in one line within GEE. As there are limited resources available to compare large-scale mangrove biomass predictions against (i.e. single-plant biomass estimates vs. ecosystem biomass estimates), Simard et al., 2019 study of mangrove biomass estimates will be utilized to compare with and gauge the performance of this study.

As seen in Figure 25, when compared to Simard et al., 2019's global mangrove biomass estimates, the different canopy height datasets, except when compared to a local LIDAR-derived CHM, all of the datasets result in similar scatter plots. When compared to the Simard et al., 2019 dataset, we can

see that course 30m resolution dataset results in gaps between canopy heights, where using ATL03, ATL08, or GEDI canopy height estimates provides a more linear result range for biomass estimations.

5 Discussion & Conclusion

Utilizing cloud-computing and remotely-sensed Earth observation data can allow interested parties to quickly query, acquire, process, and analyze data to allow for the tracking and management of at-risk ecosystems such as mangroves and their associated biomass. These tools also facilitate the tracking and management of UN SDG progress, and if policy changes are positively impacting these goals and at what scale, and pace, the management and rehabilitation of ecosystems is occurring, allowing for stakeholders to manage SDGs in detail with regular updates. The tools and workflows utilized and developed in this report will allow users and stakeholders to conduct analysis on large scales with quick results, highlighting where issues lie or where improvements are being made.

Overall, Google Earth Engine has proven to be an invaluable tool when computing mangrove area classifications and canopy height regressions for large-scale (up to country-wide) studies. This study proves that, with minimal concessions, such computing and detailed analysis can be conducted quickly and efficiently, allowing others to leverage the raw computing power of cloud services and data to gap fill data poor regions or areas. Such data allows for the close monitoring of ecosystems that are critical local economies as well as mitigating effects and impacts of future climate change across the globe with these powerful carbon storage mangrove environments. Using this platform allowed for the quick calculation of spectral indices over large time-series datasets which proved to be crucial in both land cover masking and mangrove height estimations.

There are multiple canopy height estimates available for users to incorporate into biomass estimate studies, some even at a global scale. Combined with new cloud-computing platforms, large-scale predictions and estimations of parameters regarding the natural world have made large strides of progress in recent years. One of these pivotal platforms is NASA's ATLAS sensor aboard the ICESat-2 satellite platform. Originally designed for measuring the Cryosphere, this LIDAR sensor has proven that it is capable of providing elevation data on a global-scale, however some work is required to make it competitive with other options. One of the goals of this study was to determine if the Raw Geolocated Photon Returns (ATL03) dataset from ICESat-2 could be used to derive canopy heights for training a machine learning regression algorithm. By using ATL03, the aim is to allow users to generate canopy height estimates at a very high resolution (user-defined in the custom Python tool) to provide as an alternative to GEDI 25m resolution estimates, or ATL08 (Ground and Vegetation Heights) ICESat-2 points at a 100m resolution.

Raw ATL03 returns were downloaded and processed in a Python environment before being transferred to GEE, requiring some manual work before being ready for analysis. After preliminary results, ATL03 derived canopy heights performed poorly compared to other datasets available to users (see Boest-Petersen, 2022b). In an attempt to remove over-estimating of canopy heights and be rid of any potential outliers from the derived dataset, the 95th percentile was selected from each 10 meter bin, reducing the height estimates to a potentially more reasonable level. However, as seen in Figures 26, 26b, and 26a, the derived ATL03 dataset still did not outperform, or prove to be a reasonable atlernative to existing datasets with the processing steps taken in Boest-Petersen, 2022a. Further work has been done utilizing deep learning techniques to classify ground points from ATL03 returns in Siegfried and Sutterley, 2022, as well as the processing techniques employed in ATL08 with Neuenschwander and Pitts, 2019. As of writing, there are still significant steps that need to be taken to pre-process ATL03 returns in order derive canopy heights at high resolutions, and as such, other data sources should likely be utilized for AGB estimates.

As an alternative to ATL03, Neuenschwander and Pitts, 2019 have made ATL08, Ground and

Vegetation heights, returns freely available for users via the DAAC. Points can be easily queried and downloaded in a Python environment and then brought into GEE for further processing. As seen in Figures 22b, 21, and 28a, these points perform well and are comparable with GEDI returns as well as in-situ data. These results show that the ICESat-2 platform has a promising future for elevation measurements across the globe even though that is not its intended purpose. However, in edge environments such as mangrove ecosystems, where there is rapid change or degradation, the 100 meter of along-track resolution from ATL08 can prove to be too course for some regions of the globe. To supplement this, GEDI data can fill the gap.

Another popular option for global scale canopy heights is the GEDI system aboard the ISS. As a purpose built sensor for measuring canopy heights, the returns can prove to be a sound alternative for canopy height data. As seen from Figures 23b, 24, and 30, regressions of canopy height and resulting AGB perform well when compared to ATL08 and locally-derived CHM at a small scale. Another development was made earlier in 2022, when the raster GEDI canopy height estimate dataset was made available for all users globally on GEE from 2019 on-wards. With the ease of access of data across large regions of the globe combined with its high performance, GEDI has proven to be a viable and efficient alternative. The datasets availability also negates the need to pre-process EO data in a separate environment before ingesting into GEE. When using the GEDI canopy heights, the only requirement for users is to supply data of known mangroves, such as the GMW 2016 baseline dataset that is freely available and used in this study. Furthermore, utilizing data within GEE results in limited amount of parameters required for tuning to improve performance and minimal coding for any area across the globe at a large spatial scale, such as seen in Listing 9.

In some cases, the `Gradient Tree Boost` algorithm outperformed the `Random Forest` algorithm for mangrove land cover masking. Gradient boosted decision trees can outperform traditional random forests because the algorithm "*builds the weak learner in the direction of the gradient to get the best results in the least amount of time.*" (J. Wang et al., 2018) With the rapid implementation of various classification algorithms in GEE, users can test which works best for their study area. In the case of this study, `Gradient Tree Boost` had higher validation accuracies when validating against the GMW 2016 baseline dataset compared to the `Random Forest` algorithm and as such was utilized for the eastern coast of Africa for this study.

A popular global mangrove area and biomass dataset is detailed by Simard et al., 2019. This study is conducted with reliable, but potentially outdated, data sources (such as SRTM elevation surfaces), or at a resolution which may result in the loss of granularity, such as 30m LANDSAT imagery. This study aims to improve upon this by utilizing higher resolution data and more recent data. Comparisons between this study and the results from Simard et al., 2019 can be seen in Figure 25. When comparing the AGB for southern Florida utilizing newer canopy height sources against previous studies, we can see that the general trend is the same, but we are now able to estimate the biomass of large scale ecosystems with higher detail. Results are also more linear and nature, more accurately representing the health of the ecosystem on the ground, instead of the grouping of results that can be seen in previous studies in Figure 25.

With multiple new canopy height datasets becoming available, whether from specialized sensors such as GEDI, or as a by-product of other missions, ICESat-2, scientists are now able to monitor at-risk ecosystems to provide more information for stewardship and conservation in the age of climate change. When comparing to existing studies, two products stand out in terms of performance. ATL08, ground and vegetation height and GEDI canopy heights are two options for readily available,

global-scale data. With GEDI's purpose-built sensor, higher resolution (25m vs. 100m), and data availability within mangrove ecosystems and accessibility within cloud-computing platforms such as GEE, makes it a powerful tool in the arsenal of studying and monitoring ecosystems at the forefront of climate change mitigation.

Cloud-computing allows users to build a workflow (see Listing 9) and employ it on areas around the globe given data-availability made possible by sensors such as Sentinel-1 & 2, FABDEM elevation, and high computing power afforded by commercial hardware and parallel-ization of processes. By working in Python, users open up possibilities of working with alternative canopy height datasets, such as ATL08 ground and vegetation heights and ATL03 raw geolocated photon returns. This study was able to generate mangrove biomass estimations at a spatial resolution of 10 meters for the entire eastern coastline of Africa utilizing cloud-computing. With the workflow developed from this study, mangrove areas, heights, and AGB can be estimated anywhere in the equatorial mangrove region (refer to Figure 1) with quick turn-around time. Measuring change can efficiently be accomplished by conducting a time series analysis within GEE and utilizing the sensors employed in this study. All of this culminates to provide relevant stakeholders with accurate and detailed information on how these crucial ecosystems are being impacted by human development and climate change. However, care needs to be taken when processing at such a large scale in order to maintain accuracy of derived results, as well as optimizing processing times. As experienced in this study, working at country-wide scales, especially for large coastline extents as seen in Mozambique, Somalia, and South Africa, separation of the study into smaller, more manageable areas, may be beneficial at the cost of longer, most intensive processing times.

5.1 Improvements & Future Studies

With the rapidly developing environment of climate change and available EO platforms and data, tracking mangrove biomass and ecosystem health will constantly evolve at a rapid pace. As such, there will be continual improvements in terms of processing and data in this field that can be incorporated into this study to improve results and the accuracy of the results.

Over the course of this study, multiple developments had been made in terms of EO data availability. One large advancement was the availability of GEDI data on GEE. Another was the sensor issues experienced by Sentinel-1, resulting in decreased data availability, reliance on complex systems can result in periods of decreased data, resulting in lower accuracies, or potentially no predictions due to a lack of critical data.

GEE was the main tool utilized in this study for data querying and processing. However, alternatives are beginning to appear as the world of EO and cloud-computing becomes an attractive alternative to local stored and processed information. Platforms such as Microsoft's Planetary Computer[10] or ESA's OpenEO Platform[11] are some new additions that allow users to leverage petabytes of remotely sensed data in a single line of code, as well as harnessing the power of commercial-grade hardware and parallel-ization of computational processes to derive data for the sake of monitoring climate change and its impacts on the Earth's ecosystems. Such platforms can be crucial in monitoring the progress of the UN Sustainable Development Goals, ensuring that these goals are being met, or quantifying what needs to be done to meet these goals in order to set the stage for climate change mitigation. However, since many of these options are in their infancy compared to GEE, they do not

[10] https://planetarycomputer.microsoft.com/
[11] https://openeo.cloud/

have as extensive of a data catalogue, and thus are limited in what they are able to provide to users. However, this should be quickly resolved with the development speed of such projects, allowing users to leverage the platforms with many dataset options. Implementations of GEE's Python API, such as GEEMap (Wu, 2020), are also becoming available and allowing users to quickly leverage GEE's capabilities within other environments.

Further processing of ATL03 in order to locate and identify ground and top of canopy returns can be done in order to generate more precise training data for further height regressions. Future work could involve using Convolutional Neural Networks (CNN) to apply an iterative process along-track returns to locate such statistics over small areas given enough returns with similar work being done by Siegfried and Sutterley, 2022 for ATL03 ground return locations, and Xie et al., 2021 for locating returns that fall in shallow water for bathymetric mapping applications.

The interest in EO and RS has also led to a large expansion of the private satellite construction and data providers. Active sensors, such as ICEYE's[12] constellation of high resolution SAR satellites, or Norway's International Climate and Forest Initiative (NICFI) high resolution passive monitoring of the globes forests via their Planet program[13], users now have a plethora of free and paid options for high resolution, global scale data to employ in their workflows for monitoring mangroves and tracking SDGs in the face of anthropogenic climate change and land development.

[12]https://www.iceye.com/
[13]https://www.planet.com/products/basemap/

References

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using sentinel-2 data. *GIScience Remote Sensing*, *57*, 1–20. https://doi.org/10.1080/15481603.2019.1650447

Alongi, D. M. (2002). Present state and future of the world's mangrove forests. *Environmental Conservation*, *29*, 331–349. https://doi.org/10.1017/S0376892902000231

Applied Research Laboratories, UT Austin. (2022). *PhoREAL* (Version 3.30). https://github.com/icesat-2UT/PhoREAL

Asari, N., Suratman, M. N., Ayob, N. A. M., & Hamid, N. H. A. (2021). *Mangrove as a natural barrier to environmental risks and coastal protection*. Springer Singapore. https://doi.org/10.1007/978-981-16-2494-0_13

Aslan, A., Rahman, A. F., Warren, M. W., & Robeson, S. M. (2016). Mapping spatial distribution and biomass of coastal wetland vegetation in indonesian papua by combining active and passive remotely sensed data. *Remote Sensing of Environment*, *183*, 65–81. https://doi.org/10.1016/J.RSE.2016.04.026

Asplund, M. E., Dahl, M., Ismail, R. O., Arias-Ortiz, A., Deyanova, D., Franco, J. N., Hammar, L., Hoamby, A. I., Linderholm, H. W., Lyimo, L. D., Perry, D., Rasmusson, L. M., Ridgway, S. N., Gispert, G. S., D'Agata, S., Glass, L., Mahafina, J. A., Ramahery, V., Masque, P., ... Gullström, M. (2021). Dynamics and fate of blue carbon in a mangrove–seagrass seascape: Influence of landscape configuration and land-use change. *Landscape Ecology*, *36*, 1489–1509. https://doi.org/10.1007/s10980-021-01216-8

Baccini, A., Laporte, N., Goetz, S. J., Sun, M., & Dong, H. (2008). A first map of tropical africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, *3*, 045011. https://doi.org/10.1088/1748-9326/3/4/045011

Baghdadi, N., Le Maire, G., Bailly, J.-S., Osé, K., Nouvellon, Y., Zribi, M., Lemos, C., & Hakamada, R. (2015). Evaluation of alos/palsar l-band data for the estimation of ¡italic¿eucalyptus¡/italic¿ plantations aboveground biomass in brazil. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(8), 3802–3811. https://doi.org/10.1109/JSTARS.2014.2353661

Bian, J., Li, A., Lei, G., Zhang, Z., & Nan, X. (2020). Global high-resolution mountain green cover index mapping based on landsat images and google earth engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 63–76. https://doi.org/10.1016/J.ISPRSJPRS.2020.02.011

Boest-Petersen, A. (2022a). *icesat2_canopy_heights* (Version 0.0.3). https://github.com/aboestpetersen/icesat2_canopy_heights

Boest-Petersen, A. (2022b). *Mangrove biomass estimation using rs/eo data and machine learning with dhi group*. Aalborg University.

Bourgine, B., & Baghdadi, N. (2005). Assessment of c-band srtm dem in a dense equatorial forest zone. *Comptes Rendus Geoscience*, *337*, 1225–1234. https://doi.org/10.1016/j.crte.2005.06.006

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus global land cover layers—collection 2. *Remote Sensing*, *12*, 1044. https://doi.org/10.3390/rs12061044

Bukoski, J. J., Dronova, I., & Potts, M. D. (2022). Net loss statistics underestimate carbon emissions from mangrove land use and land cover change. *Ecography*, *2022*. https://doi.org/10.1111/ecog.05982

Bunting, P., Rosenqvist, A., Lucas, R., Rebelo, L.-M., Hilarides, L., Thomas, N., Hardy, A., Itoh, T., Shimada, M., & Finlayson, C. (2018). The global mangrove watch—a new 2010 global baseline of mangrove extent. *Remote Sensing*, *10*, 1669. https://doi.org/10.3390/rs10101669

Chow, J. (2018). Mangrove management for climate change adaptation and sustainable development in coastal zones. *Journal of Sustainable Forestry*, *37*, 139–156. https://doi.org/10.1080/10549811.2017.1339615

de Groot, R., Brander, L., van der Ploeg, S., Costanza, R., Bernard, F., Braat, L., Christie, M., Crossman, N., Ghermandi, A., Hein, L., Hussain, S., Kumar, P., McVittie, A., Portela, R., Rodriguez, L. C., ten Brink, P., & van Beukering, P. (2012). Global estimates of the value of ecosystems and their services in monetary units. *Ecosystem Services*, *1*, 50–61. https://doi.org/10.1016/j.ecoser.2012.07.005

Dubayah, R., Hofton, M., Blair, J., Armstrong, J., Tang, H., & Luthcke, S. (2021). *Gedi l2a elevation and height metrics data global footprint level v002 [gedi02_a]*. https://doi.org/10.5067/GEDI/GEDI02_A.002. (accessed: 06.05.2022)

Duke, N. C. (1984). A mangrove hybrid, sonneratia × gulngai (sonneratiaceae) from north-eastern australia. *Austrobaileya*, *2*, 103–105. http://www.jstor.org.zorac.aub.aau.dk/stable/41739166

Ellison, J. C. (2015). Vulnerability assessment of mangroves to climate change and sea-level rise impacts. *Wetlands Ecology and Management*, *23*, 115–137. https://doi.org/10.1007/s11273-014-9397-8

European Space Agency. (2021a). *Definitions*. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/definitions (accessed: 28.11.2021)

European Space Agency. (2021b). *Sentinel-1*. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1 (accessed: 3.12.2021)

European Space Agency. (2021c). *Sentinel-2*. https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed: 3.12.2021)

Fakhruddin, B., Mahalingam, R., & Padmanaban, R. (2018). Sustainable development goals for reducing the impact of sea level rise on mangrove forests. *Indian Journal of Geo Marine Sciences*, *47*, 1947–1958.

Fathom. (2022). *Fabdem*. https://www.fathom.global/product/fabdem/ (accessed: 06.05.2022)

Fatoyinbo, T. E., & Simard, M. (2013). Height and biomass of mangroves in africa from icesat/glas and srtm. *International Journal of Remote Sensing*, *34*, 668–681. https://doi.org/10.1080/01431161.2012.712224

Fatoyinbo, T. E., Simard, M., Washington-Allen, R. A., & Shugart, H. H. (2008). Landscape-scale extent, height, biomass, and carbon estimation of mozambique's mangrove forests with landsat etm+ and shuttle radar topography mission elevation data. *Journal of Geophysical Research: Biogeosciences*, *113*, n/a–n/a. https://doi.org/10.1029/2007JG000551

Fayad, I., Baghdadi, N., Bailly, J.-S., Barbier, N., Gond, V., Hajj, M., Fabre, F., & Bourgine, B. (2014). Canopy height estimation in french guiana with lidar icesat/glas data using principal component analysis and random forest regressions. *Remote Sensing*, *6*, 11883–11914. https://doi.org/10.3390/rs61211883

Feller, I. C., Friess, D. A., Krauss, K. W., & Lewis, R. R. (2017). The state of the world's mangroves in the 21st century under climate change. *Hydrobiologia*, *803*, 1–12. https://doi.org/10.1007/s10750-017-3331-z

Filippelli, S. K., Falkowski, M. J., Hudak, A. T., Fekety, P. A., Vogeler, J. C., Khalyani, A. H., Rau, B. M., & Strand, E. K. (2020). Monitoring pinyon-juniper cover and aboveground biomass across the great basin. *Environmental Research Letters*, *15*, 025004. https://doi.org/10.1088/1748-9326/ab6785

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics Data Analysis*, *38*, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

GEDI. (2022). *Launch*. https://gedi.umd.edu/instrument/launch/ (accessed: 06.05.2022)

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from eos-modis. *Remote Sensing of Environment*, *58*, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Goldstein, A., Turner, W. R., Spawn, S. A., Anderson-Teixeira, K. J., Cook-Patton, S., Fargione, J., Gibbs, H. K., Griscom, B., Hewson, J. H., Howard, J. F., Ledezma, J. C., Page, S., Koh, L. P., Rockström, J., Sanderman, J., & Hole, D. G. (2020). Protecting irrecoverable carbon in earth's ecosystems. *Nature Climate Change*, *10*, 287–295. https://doi.org/10.1038/s41558-020-0738-8

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

Hashim, T. M. Z. T., & Suratman, M. N. (2021). *Mangroves as a carbon sink/stocks*. Springer Singapore. https://doi.org/10.1007/978-981-16-2494-0_7

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, *17*, 024016. https://doi.org/10.1088/1748-9326/ac4d4f

Huete, A. (1988). A soil-adjusted vegetation index (savi). *Remote Sensing of Environment*, *25*, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Hunt, E. R., & Rock, B. N. (1989). Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment*, *30*, 43–54. https://doi.org/10.1016/0034-4257(89)90046-1

Jachowski, N. R., Quak, M. S., Friess, D. A., Duangnamon, D., Webb, E. L., & Ziegler, A. D. (2013). Mangrove biomass estimation in southwest thailand using machine learning. *Applied Geography*, *45*, 311–321. https://doi.org/10.1016/J.APGEOG.2013.09.024

Jones, A. R., Segaran, R. R., Clarke, K. D., Waycott, M., Goh, W. S. H., & Gillanders, B. M. (2020). Estimating mangrove tree biomass and carbon content: A comparison of forest inventory techniques and drone imagery. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00784

Kamruzzaman, M., Basak, K., Paul, S. K., Ahmed, S., & Osawa, A. (2019). Litterfall production, decomposition and nutrient accumulation in sundarbans mangrove forests, bangladesh. *Forest Science and Technology*, *15*. https://doi.org/10.1080/21580103.2018.1557566

Kathiresan, K. (2021). *Mangroves: Types and importance*. Springer Singapore. https://doi.org/10.1007/978-981-16-2494-0_1

Kebede, B., & Soromessa, T. (2018). Allometric equations for aboveground biomass estimation of *Olea europaea L. subsp. cuspidata* in mana angetu forest. *Ecosystem Health and Sustainability*, *4*, 1–12. https://doi.org/10.1080/20964129.2018.1433951

Krauss, K. W., McKee, K. L., Lovelock, C. E., Cahoon, D. R., Saintilan, N., Reef, R., & Chen, L. (2014). How mangrove forests adjust to rising sea level. *New Phytologist*, *202*, 19–34. https://doi.org/10.1111/nph.12605

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., & Wegner, J. D. (2022). Global canopy height regression and uncertainty estimation from gedi lidar waveforms with deep ensembles. *Remote Sensing of Environment*, *268*, 112760. https://doi.org/10.1016/j.rse.2021.112760

Lee, S. Y., Primavera, J. H., Dahdouh-Guebas, F., McKee, K., Bosire, J. O., Cannicci, S., Diele, K., Fromard, F., Koedam, N., Marchand, C., Mendelssohn, I., Mukherjee, N., & Record, S. (2014). Ecological role and services of tropical mangrove ecosystems: A reassessment. *Global Ecology and Biogeography*, *23*, 726–743. https://doi.org/10.1111/geb.12155

Lemoen, G. (2022). *Synthetic aperture radar (sar) basics*. https://developers.google.com/earth-engine/tutorials/community/sar-basics (accessed: 15.05.2022)

Li, W., El-Askary, H., Qurban, M. A., Li, J., ManiKandan, K., & Piechota, T. (2019). Using multi-indices approach to quantify mangrove changes over the western arabian gulf along saudi arabia coast. *Ecological Indicators*, *102*, 734–745. https://doi.org/10.1016/j.ecolind.2019.03.047

Liang, S., Li, X., & Wang, J. (Eds.). (2012). *Chapter 1 - a systematic view of remote sensing*. https://doi.org/https://doi.org/10.1016/B978-0-12-385954-9.00001-0

López-Medellín, X., Ezcurra, E., González-Abraham, C., Hak, J., Santiago, L. S., & Sickman, J. O. (2011). Oceanographic anomalies and sea-level rise drive mangroves inland in the pacific coast of mexico. *Journal of Vegetation Science*, *22*, 143–151. https://doi.org/10.1111/j.1654-1103.2010.01232.x

Macreadie, P. I., Anton, A., Raven, J. A., Beaumont, N., Connolly, R. M., Friess, D. A., Kelleway, J. J., Kennedy, H., Kuwae, T., Lavery, P. S., Lovelock, C. E., Smale, D. A., Apostolaki, E. T., Atwood, T. B., Baldock, J., Bianchi, T. S., Chmura, G. L., Eyre, B. D., Fourqurean, J. W., ... Duarte, C. M. (2019). The future of blue carbon science. *Nature Communications*, *10*. https://doi.org/10.1038/s41467-019-11693-w

McFeeters, S. K. (1996). The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, *17*, 1425–1432. https://doi.org/10.1080/01431169608948714

Microsoft. (2021). *What is cloud computing?* https://azure.microsoft.com/en-us/overview/what-is-cloud-computing/ (accessed: 25.11.2021)

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, *1*, 6–43. https://doi.org/10.1109/MGRS.2013.2248301

NASA. (2021). *About icesat-2*. https://www.nasa.gov/content/goddard/about-icesat-2 (accessed: 2.12.2021)

NASA. (2022). *Mission*. https://icesat-2.gsfc.nasa.gov/mission (accessed: 06.05.2022)

Neuenschwander, A., Pitts, K., Jelley, B., Robbins, J., Klotz, B., Popescu, S., Nelson, R., Harding, D., & Pederson, D. (2021). *Atlas/icesat-2 l3a land and vegetation height, version 4 user guide*. NSIDC. https://doi.org/10.5067/ATLAS/ATL08.004 (accessed: 21.11.2021)

Neuenschwander, A., & Pitts, K. (2019). The atl08 land and vegetation product for the icesat-2 mission. *Remote Sensing of Environment*, *221*, 247–259. https://doi.org/10.1016/j.rse.2018.11.005

NOAA. (2022). *What is the cryosphere?* https://oceanservice.noaa.gov/facts/cryosphere.html (accessed: 19.05.2022)

Noon, M. L., Goldstein, A., Ledezma, J. C., Roehrdanz, P. R., Cook-Patton, S. C., Spawn-Lee, S. A., Wright, T. M., Gonzalez-Roglich, M., Hole, D. G., Rockström, J., & Turner, W. R. (2021). Mapping the irrecoverable carbon in earth's ecosystems. *Nature Sustainability*. https://doi.org/10.1038/s41893-021-00803-6

Ong, J. (1995). The ecology of mangrove conservation and management. *Hydrobiologia*, *295*, 343–351.

Ong, J. (2003). Plants of the merbok mangrove, kedah, malaysia and the urgent need for their conservation. *Folia Malaysiana*, *4*, 1–18.

Patterson, P. L., Healey, S. P., Ståhl, G., Saarela, S., Holm, S., Andersen, H.-E., Dubayah, R. O., Duncanson, L., Hancock, S., Armston, J., Kellner, J. R., Cohen, W. B., & Yang, Z. (2019). Statistical properties of hybrid estimators proposed for gedi—nasa's global ecosystem dynamics investigation. *Environmental Research Letters*, *14*, 065007. https://doi.org/10.1088/1748-9326/ab18df

Pham, T. D., Le, N. N., Ha, N. T., Nguyen, L. V., Xia, J., Yokoya, N., To, T. T., Trinh, H. X., Kieu, L. Q., & Takeuchi, W. (2020). Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused sentinel-2 and alos-2 palsar-2 data in can gio biosphere reserve, vietnam. *Remote Sensing*, *12*, 777. https://doi.org/10.3390/rs12050777

Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Thu, T., Nguyen, T., Dao, T. H., Thi, T., Vu, P., Pham, T. D., & Takeuchi, W. (2020). Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, vietnam. *Remote Sensing*, *12*, 1334. https://doi.org/https://doi.org/10.3390/rs12081334

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., Salmo, S. G., Sanciangco, J. C., Sukardjo, S., ... 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

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., & Ohmann, J. L. (2010). Quantification of live aboveground forest biomass dynamics with landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, *114*, 1053–1068. https://doi.org/10.1016/j.rse.2009.12.018

Ragavan, P., Kathiresan, K., Kumar, S., Nagarajan, B., Jayaraj, R. S. C., Mohan, P. M., Sachithanandam, V., Mageswaran, T., & Rana, T. S. (2021). *Biogeography of the mangrove ecosystem: Floristics, population structure, and conservation strategies*. Springer Singapore. https://doi.org/10.1007/978-981-16-2494-0_2

Rahman, M. S., Donoghue, D. N. M., Bracken, L. J., & Mahmood, H. (2021). Biomass estimation in mangrove forests: A comparison of allometric models incorporating species and structural information. *Environmental Research Letters*, *16*, 124002. https://doi.org/10.1088/1748-9326/ac31ee

Rao, K., Ranjan, P., & Ramanathan, A. (2021). *Estimation of blue carbon stock of mangrove ecosystem and its dynamics in relation to hydrogeomorphic settings and land use-land cover*. Springer Singapore. https://doi.org/10.1007/978-981-16-2494-0_8

Ravindranath, N. H., & Ostwald, M. (2008). *Carbon inventory methods handbook for greenhouse gas inventory, carbon mitigation and roundwood production projects*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-6547-7

Rikimaru, A., Roy, P. S., & Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology*, *43*, 39–47.

Roy, S. (2022). *Fabdem (forest and buildings removed copernicus 30m dem)*. https://samapriya.github.io/awesome-gee-community-datasets/projects/fabdem/ (accessed: 06.05.2022)

Salim, H. L., Adi, N. S., Kepel, T. L., & Ati, R. N. A. (2020). Estimating mangrove biomass using drone in karimunjawa islands. *IOP Conference Series: Earth and Environmental Science*, *561*, 012054. https://doi.org/10.1088/1755-1315/561/1/012054

Salum, R. B., Souza-Filho, P. W. M., Simard, M., Silva, C. A., Fernandes, M. E., Cougo, M. F., do Nascimento, W., & Rogers, K. (2020). Improving mangrove above-ground biomass estimates using lidar. *Estuarine, Coastal and Shelf Science*, *236*, 106585. https://doi.org/10.1016/j.ecss.2020.106585

Sanderman, J., Hengl, T., Fiske, G., Solvik, K., Adame, M. F., Benson, L., Bukoski, J. J., Carnell, P., Cifuentes-Jara, M., Donato, D., Duncan, C., Eid, E. M., zu Ermgassen, P., Lewis, C. J. E., Macreadie, P. I., Glass, L., Gress, S., Jardine, S. L., Jones, T. G., ... Landis, E. (2018). A global map of mangrove forest soil carbon at 30m spatial resolution. *Environmental Research Letters*, *13*, 055002. https://doi.org/10.1088/1748-9326/aabe1c

Scheick, J. et al. (2019–). *icepyx: Python tools for obtaining and working with ICESat-2 data*. https://github.com/icesat2py/icepyx

Segal, M. R. (2004). *Machine learning benchmarks and random forest regression*. UCSF: Center for Bioinformatics and Molecular Biostatistics. https://escholarship.org/uc/item/35x3v9t4

Siegfried, M., & Sutterley, T. (2022). *read-ICESat-2*. https://github.com/tsutterley/read-ICESat-2

Siikamäki, J., Sanchirico, J. N., & Jardine, S. L. (2012). Global economic potential for reducing carbon dioxide emissions from mangrove loss. *Proceedings of the National Academy of Sciences*, *109*, 14369–14374. https://doi.org/10.1073/pnas.1200519109

Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V. H., Castañeda-Moya, E., Thomas, N., & der Stocken, T. V. (2019). Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nature Geoscience*, *12*, 40–45. https://doi.org/10.1038/s41561-018-0279-1

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*, *116*, G04021. https://doi.org/10.1029/2011JG001708

Somvanshi, S. S., & Kumari, M. (2020). Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data. *Applied Computing and Geosciences*, *7*, 100032. https://doi.org/10.1016/J.ACAGS.2020.100032

Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M. M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., & Midgley, P. M. (2013). Climate change 2013 the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change. *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, *9781107057999*, 1–1535. https://doi.org/10.1017/CBO9781107415324

Suwa, R., Rollon, R., Sharma, S., Yoshikai, M., Albano, G. M. G., Ono, K., Adi, N. S., Ati, R. N., Kusumaningtyas, M. A., Kepel, T. L., Maliao, R. J., Primavera-Tirol, Y. H., Blanco, A. C., & Nadaoka, K. (2021). Mangrove biomass estimation using canopy height and wood density in the south east and east asian regions. *Estuarine, Coastal and Shelf Science*, *248*, 106937. https://doi.org/10.1016/J.ECSS.2020.106937

Tang, W., Feng, W., Jia, M., Shi, J., Zuo, H., Stringer, C. E., & Trettin, C. C. (2017). A cyber-enabled spatial decision support system to inventory mangroves in mozambique: Coupling scientific workflows and cloud computing. *International Journal of Geographical Information Science*, *31*, 907–938. https://doi.org/10.1080/13658816.2016.1245419

Tanre, D., Holben, B., & Kaufman, Y. (1992). Atmospheric correction algorithm for noaa-avhrr products: Theory and application. *IEEE transactions on geoscience and remote sensing*, *30*, 231–248. https://doi.org/10.1109/36.134074

Théau, J. (2008). *Temporal resolution*. https://doi.org/10.1007/978-0-387-35973-1_1376

Thieme, A., Yadav, S., Oddo, P. C., Fitz, J. M., McCartney, S., King, L., Keppler, J., McCarty, G. W., & Hively, W. D. (2020). Using nasa earth observations and google earth engine to map winter cover crop conservation performance in the chesapeake bay watershed. *Remote Sensing of Environment*, *248*, 111943. https://doi.org/10.1016/j.rse.2020.111943

Tsendbazar, N.-E., Tarko, A., Li, L., Herold, M., Lesiv, M., Fritz, S., & Maus, V. (2020). *Copernicus global land service: Land cover 100m: Version 3 globe 2015-2019: Validation report*. Zenodo.

United Nations. (2021a). *Conserve and sustainably use the oceans, seas and marine resources for sustainable development*. https://sdgs.un.org/goals/goal14 (accessed: 28.11.2021)

United Nations. (2021b). *The 17 goals*. https://sdgs.un.org/goals (accessed: 28.11.2021)

USGS. (2022). *3dep lidarexplorer*. https://prd-tnm.s3.amazonaws.com/LidarExplorer/index.html#/ (accessed: 07.05.2022)

Vaghela, B., Chirakkal, S., Putrevu, D., & Solanki, H. (2021). Modelling above ground biomass of indian mangrove forest using dual-pol sar data. *Remote Sensing Applications: Society and Environment*, *21*, 100457. https://doi.org/10.1016/J.RSASE.2020.100457

van Zyl, J., & Kim, Y. (2011). *Synthetic aperture radar polarimetry* (J. H. Yuen, Ed.). John Wiley Sons.

Wang, J., Li, P., Ran, R., Che, Y., & Zhou, Y. (2018). A short-term photovoltaic power prediction model based on the gradient boost decision tree. *Applied Sciences*, *8*, 689. https://doi.org/10.3390/app8050689

Wang, L., Yan, J., & Ma, Y. (2019). *Cloud computing in remote sensing*. https://doi.org/10.1201/9780429488764

Weier, J., & Herring, D. (2000). *Measuring vegetation vegetation (ndvi evi)*. NASA.

Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of landsat tm imagery. *Remote Sensing of Environment*, *80*, 385–396. https://doi.org/10.1016/S0034-4257(01)00318-2

Wong, P. P., Losada, I. J., Gattuso, J.-P., Hinkel, J., Khattabi, A., McInnes, K. L., Saito, Y., Sallenger, A., et al. (2014). Coastal systems and low-lying areas. *Climate change*, *2104*, 361–409.

Worthington, T., Bunting, P., Cormier, N., Donnison, A., Fatoyinbo, L., Friess, D., Hengl, T., Hilarides, L., Krauss, K., Lagomasino, D., Leinenkugel, P., Longley-Wood, K., Lovelock, C., Lucas, R., Murray, N., Narayan, S., Sutherland, W., Ermgassen, P., & Spalding, M. (2019). Mangrove restoration potential: A global map highlighting a critical opportunity. *The 5th international Mangrove Macrobenthos and Management meeting (MMM5)*. https://elib.dlr.de/127024/

Wu, Q. (2020). Geemap: A python package for interactive mapping with google earth engine. *Journal of Open Source Software*, *5*, 2305. https://doi.org/10.21105/joss.02305

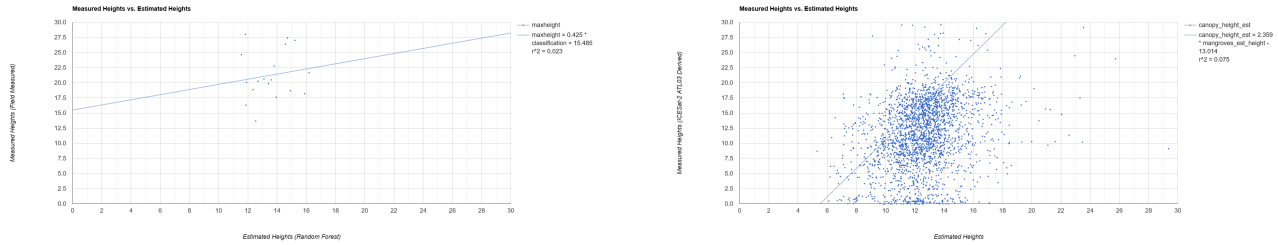
Xie, C., Chen, P., Pan, D., Zhong, C., & Zhang, Z. (2021). Improved filtering of icesat-2 lidar data for nearshore bathymetry estimation using sentinel-2 imagery. *Remote Sensing*, *13*, 4303. https://doi.org/10.3390/rs13214303

Yang, Z., Li, W., Chen, Q., Wu, S., Liu, S., & Gong, J. (2019). A scalable cyberinfrastructure and cloud computing platform for forest aboveground biomass estimation based on the google earth engine. *International Journal of Digital Earth*, *12*, 995–1012. https://doi.org/10.1080/17538947.2018.1494761

Zickfeld, K., Solomon, S., & Gilford, D. M. (2017). Centuries of thermal sea-level rise due to anthropogenic emissions of short-lived greenhouse gases. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 657–662. https://doi.org/10.1073/pnas.1612066114

Appendix

ATL03 Comparisons:



(a) ATL03 heights compared to field measurements.

(b) ATL03 height performance.

Figure 26: ATL03 performance in Southern Florida (m).

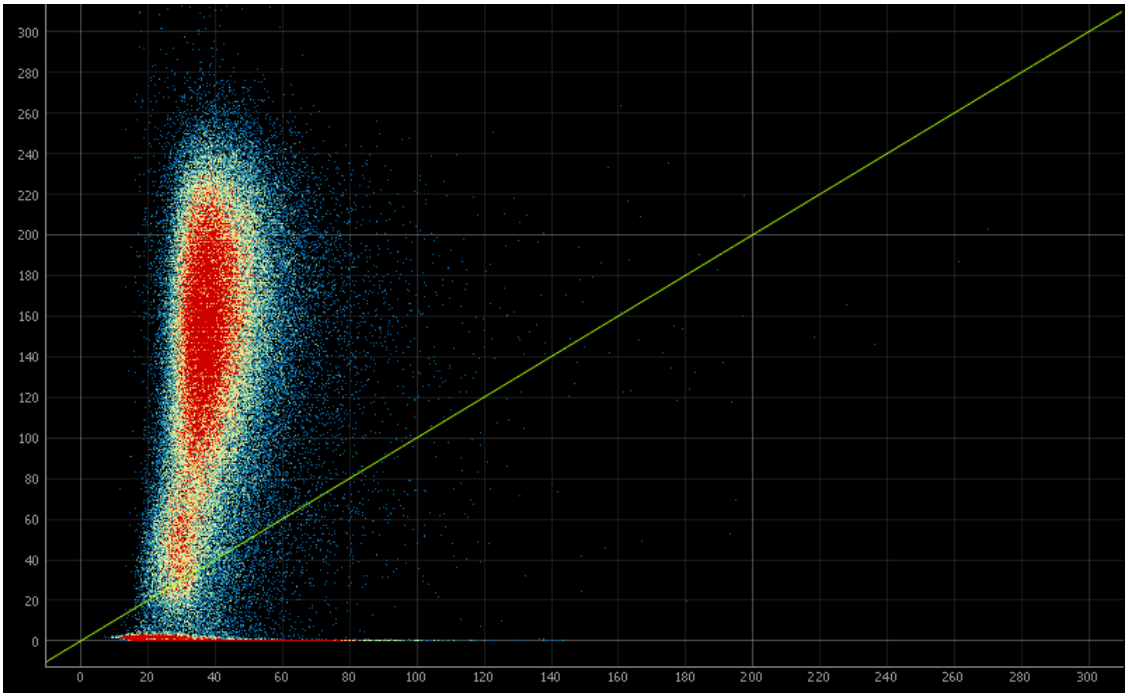
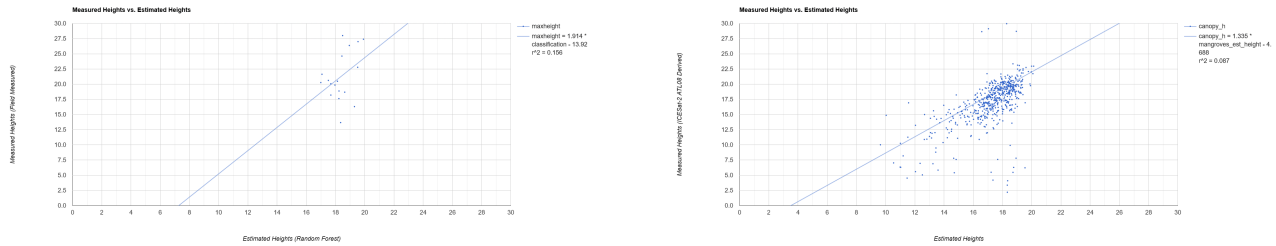


Figure 27: Scatter plot of ATL03 AGB compared with AGB derived from local CHM in Southern Florida in Mg/Ha.

ATL08 Comparisons:



(a) ATL08 heights compared to field measurements.

(b) ATL08 height performance.

Figure 28: ATL08 performance in Southern Florida (m).

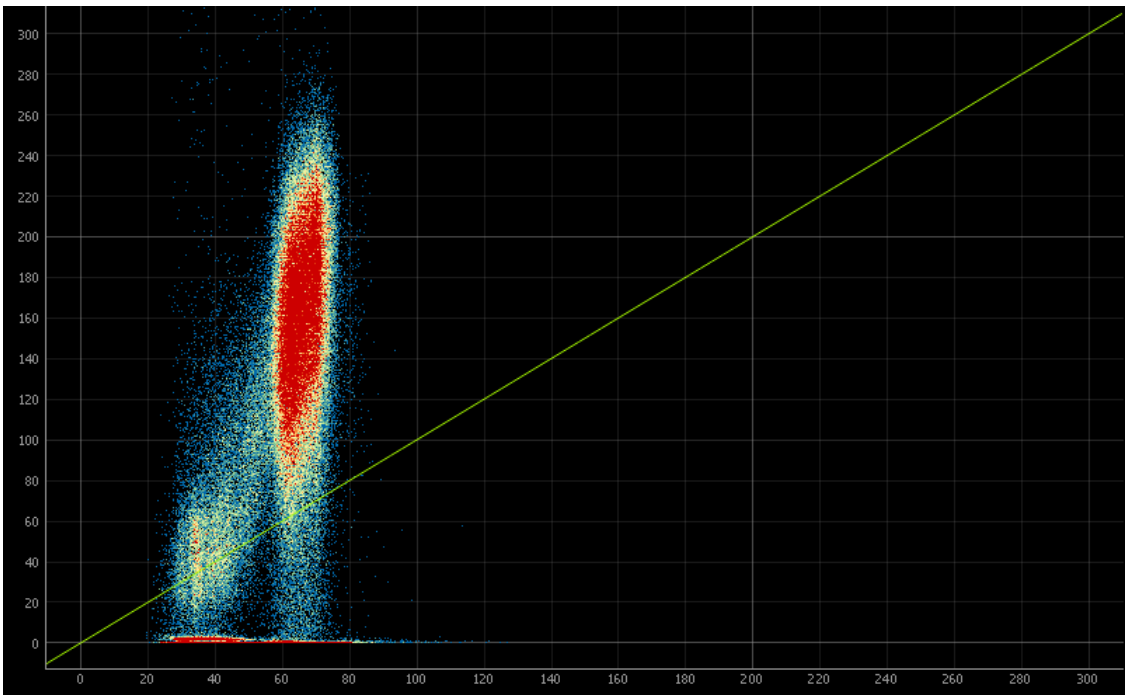


Figure 29: Scatter plot of ATL08 AGB compared with AGB derived from local CHM in Southern Florida in Mg/Ha.

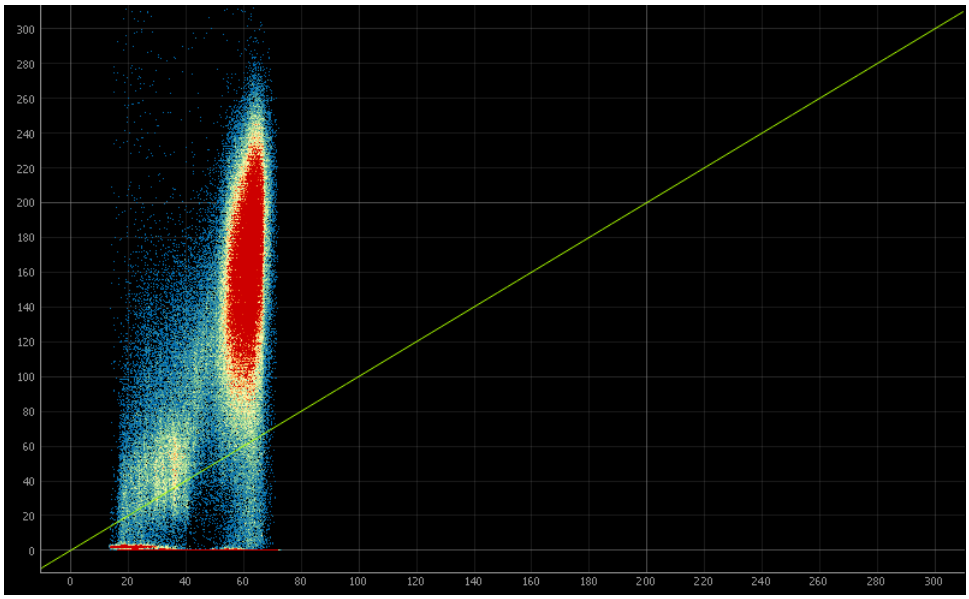


Figure 30: Scatter plot of GEDI AGB compared with AGB derived from local CHM in Southern Florida in Mg/Ha.

Mangrove Biomass Estimate GEE JavaScript Workflow

```
var start = ee.Date('2019-01-01');
var end = ee.Date('2019-12-31');
var bands_lc = ['B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8', 'B8A', 'B9', 'B10', 'B11', 'B12', 'SAVI', 'GNDVI', 'EVI', 'NDMI', 'MSI', 'GCI',
'BSI', 'NDWI', 'ARVI', 'MNDWI', 'NDVI', 'VV', 'VV_1', 'VH_1', 'fabdem']; // Bands to train mangrove mask classifier
var bands_height = ['VV', 'B3', 'B6', 'B9', 'GCI', 'BSI', 'B2']; // Bands to train height regression model

// Function to mask clouds using the Sentinel-2 QA band.
function maskS2clouds(image) {
  var qa = image.select('QA60'); // Select Quality Band

  // Define Cloud Types to be Masked Out
  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  // Set masks to zero to select clear conditions
  var mask = qa.bitwiseAnd(cloudBitMask).eq(0).and(
             qa.bitwiseAnd(cirrusBitMask).eq(0));

  // Return masked image minus 'QA' bands
  return image.updateMask(mask).divide(10000)
      .select("B.*")
      .copyProperties(image, ["system:time_start"]);
}

// Load East Africa Coast Line Dataset (generated by Alexander Boest-Petersen)
var country_select = africa_east_coast.filter('NAME == "Mozambique"'); // Select Mozambique coastline
```

```
var grids = (country_select.geometry()).coveringGrid('EPSG:4326', 100000); // Generate grids over coastline
print('Number of Grids:', grids.size()); // Print grid info
grids = grids.randomColumn(); // Assign random column to grid for random selection
grids = grids.filter(ee.Filter.lt('random', 0.3)); // Select 30% of grids for large Mozambique coastline
print('Grids Test:', grids); // Print total number of grids available
print('Number of Grids to be Processed:', grids.size()); // Print number of grids selected for training models

// GEDI Canopy Information
var qualityMask = function(im) {
  return im.updateMask(im.select('quality_flag').eq(1)) // Select valid wave-forms
      .updateMask(im.select('degrade_flag').eq(0)) // Select data with no comments on quality
      .updateMask(im.select('solar_elevation').gt(0)); // Select data collected at night
};
var gedi_data = ee.ImageCollection('LARSE/GEDI/GEDI02_A_002_MONTHLY')
                  .filterDate(start, end) // Filter by date for the year of 2019
                  .filterBounds(grids) // Filter to randomly selected grids
                  .map(qualityMask) // Map function to filter specific returns
                  .select('rh98') // Select 98th percentile heights
                  .mean() // Calculate mean to generate Image from ImageCollection for later sampling
                  .rename('canopy_h'); // Rename height attribute for later training

// S1 Imagery
var imgVV = ee.ImageCollection('COPERNICUS/S1_GRD')
        .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV')) // Filter out bands that are not of interest
        .filter(ee.Filter.eq('instrumentMode', 'IW')) // Select bands collected in 'IW' mode
        .select('VV') // Select 'VV' Band
        // Define function to calculate edge effects
        .map(function(image) {
          var edge = image.lt(-30.0);
          var maskedImage = image.mask().and(edge.not());
          return image.updateMask(maskedImage);
        });
var descChange_VV = (imgVV.filter(ee.Filter.eq('orbitProperties_pass', 'DESCENDING')))
    .filterDate(start, end).mean(); // Filter by date and Descending pass and average results for later training
var ascChange_VV = (imgVV.filter(ee.Filter.eq('orbitProperties_pass', 'ASCENDING'))).
    filterDate(start, end).mean(); // Filter by date and Ascending pass and average results for later training
var imgVH = ee.ImageCollection('COPERNICUS/S1_GRD')
        .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VH')) // Filter out bands that are not of interest
        .filter(ee.Filter.eq('instrumentMode', 'IW')) // Select bands collected in 'IW' mode
        .select('VH') // Select 'VH' band
        // Define function to calculate edge effects
        .map(function(image) {
          var edge = image.lt(-30.0);
          var maskedImage = image.mask().and(edge.not());
          return image.updateMask(maskedImage);
```

```
        });
var descChange_VH = (imgVH.filter(ee.Filter.eq('orbitProperties_pass', 'DESCENDING')))
    .filterDate(start, end).mean(); // Filter by date and Descending pass and average
    results for later training
var ascChange_VH = (imgVH.filter(ee.Filter.eq('orbitProperties_pass', 'ASCENDING'))).
    filterDate(start, end).mean(); // Filter by date and Ascending pass and average
    results for later training

// FABDEM
var elev = fabdem.mosaic().select('b1').rename('fabdem'); // Select elevation band

// Copernicus Land Cover
var cop_landclass = ee.Image("COPERNICUS/Landcover/100m/Proba-V-C3/Global/2019").
    select('discrete_classification').clip(grids.geometry()); // Select land cover
    classification band and clip to study area
var cop_water = (cop_landclass.updateMask(cop_landclass.eq(80))).rename('
    mangrove_label'); // Select water class
cop_water = cop_water.reduceToVectors({geometry: grids.geometry(), maxPixels: 1e12});
    // Convert raster to vector for burning of data intro training imagery
var water_image = ee.Image(2).clip(cop_water); // Create new image with water cover
    equaling the value of 2
var cop_ocean = (cop_landclass.updateMask(cop_landclass.eq(200))).rename('
    mangrove_label'); // Select ocean class
cop_ocean = cop_ocean.reduceToVectors({geometry: grids.geometry(), maxPixels: 1e12});
    // Convert raster to vector for burning of data intro training imagery
var ocean_image = ee.Image(3).clip(cop_ocean); // Create new image with ocean cover
    equaling the value of 3
var cop_forest = (cop_landclass.updateMask(cop_landclass.eq(112))).rename('
    mangrove_label'); // Select forest class
cop_forest = cop_forest.reduceToVectors({geometry: grids.geometry(), maxPixels: 1e12})
    ; // Convert raster to vector for burning of data intro training imagery
var forest_image = ee.Image(4).clip(cop_forest); // Create new image with forest cover
     equaling the value of 4

// Mangroves
var mangroves_image = ee.Image(0).clip(grids.geometry()); // Create blank image with
    pixels equaling 0 for the study area
var mangroves_image = mangroves_image.paint(gmw, 1).rename('mangrove_label').clip(
    grids.geometry()); // Paint Mangroves into the newly created blank image with the
    value of 1
var training_mangroves = ee.Image.cat([ocean_image, mangroves_image, forest_image]).
    reduce(ee.Reducer.firstNonNull()).rename('mangrove_label').toByte(); // Add
    previous land cover types to training image for model ingestion

// Stack Bands Function
function stackBands(i){
  // S2 Indices
  var selected = i.select('B1', 'B2', 'B3', 'B4', 'B6', 'B8', 'B9', 'B11', 'B12'); //
    Select relevant S2 bands
  var SAVI = selected.expression('((b(5)-b(3))/(b(5)+b(3))+0.428)*1.428').rename('SAVI
    '); // Calculate SAVI
  var GNDVI = selected.expression('(b(5)-b(2))/(b(5)+b(2))').rename('GNDVI'); //
    Calculate GNDVI
  var EVI = selected.expression('2.5*(b(5)-b(3))/(b(5)+6*b(3)-7.5*b(1)+1)').rename('
    EVI'); // Calculate EVI
  var NDMI = selected.expression('(b(5)-b(7))/(b(5)+b(7))').rename('NDMI'); //
    Calculate NDMI
```

```
  var MSI = selected.expression('(b(7))/(b(5))').rename('MSI'); // Calculate MSI
  var GCI = selected.expression('(b(6))/(b(2)-1)').rename('GCI'); // Calculate GCI
  var BSI = selected.expression('(b(7)+b(3)-b(5)+b(1))/(b(7)+b(3)+b(5)+b(1))').rename('BSI'); // Calculate BSI
  var NDWI = selected.expression('(b(2)-b(5))/(b(2)+b(5))').rename('NDWI'); // Calculate NDWI
  var ARVI = selected.expression('(b(5)-(2*b(3))+b(1))/(b(5)+(2*b(3)+b(1)))').rename('ARVI'); // Calculate ARVI
  var MNDWI = selected.expression('(b(7)-b(2))/(b(7)+b(2))').rename('MNDWI'); // Calculate MDNWI
  var NDVI = selected.normalizedDifference(['B8', 'B4']).rename('NDVI'); // Calculate NDVI
  return i.addBands([SAVI, GNDVI, EVI, NDMI, MSI, GCI, BSI, NDWI, ARVI, MNDWI, NDVI, descChange_VV, ascChange_VV, descChange_VH, ascChange_VH, elev])
  .set('system:footprint', i.geometry()); // Add calculated spectral indices to imagery
}

// Training Data
var collection = ee.ImageCollection('COPERNICUS/S2')
    .filterDate(start, end) // Filter by date
    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 20)) // Select images that have less than 20% cloud cover
    .map(maskS2clouds) // Apply cloud masking function
    .map(stackBands); // Calculate and stack spectral indices, and stack elevation data
var compilation = collection.median(); // Calculate median value of each pixel
compilation = compilation.addBands(training_mangroves); // Add mangrove cover after to preserve data type (remain as integer and not converted to float, need to have discrete value for land cover classification)

// Display S2 imagery results.
var viz = {bands: ['B4', 'B3', 'B2'], min: 0, max: 0.2}; // Select visual bands
Map.addLayer(compilation, viz, 'S2 Imagery', false); // Add layer to GEE map

// Stratified Sample Generation for Land Cover Masking
var strat_sample = (compilation.clip(grids.geometry())).stratifiedSample({
  numPoints: 100, // Generate 100 random points
  region: grids, // Region for point generation
  classBand: 'mangrove_label', // Generate 100 points per land cover class for a total of 500 points per grid
  scale: 100, // Average the cells in the surrounding 100m radius for training, required for memory limits
  geometries: true, // Maintain geometries for later training
  tileScale: 16 // Reduce size of calculation tiles on GEE servers to prevent computation timed-out error, 16 is max value
});
print('Stratified Sample:', strat_sample.size()); // Print overall number of randomly generated points
Map.addLayer(strat_sample, {}, 'Stratified Sample'); // Add points to map for visualization of point data
strat_sample = strat_sample.randomColumn(); // Add random column to point data for filtering
var training = strat_sample.filter(ee.Filter.lt('random', 0.7)); // 70% of points reserved for training
var validation = strat_sample.filter(ee.Filter.gte('random', 0.7)); // 30% of points reserved for validation
```

```
// Define Classifier
var classifier = ee.Classifier.smileRandomForest(150).train({features: training, classProperty: 'mangrove_label', inputProperties: bands_lc}); // Can also use smileGradientTreeBoost algorithm
// Train Classifier
var trained = classifier.train(training, 'mangrove_label', bands_lc);

// Classifier coastline by grid
function classify_image(trained_model, img, grids){
  var classified = img.classify(trained_model); // Classify each image in each grid
  var mangroves_masked = classified.updateMask((classified.select('classification')).eq(1)); // Select pixels classified as mangroves
  // Seive
  var mmu = 60; // Define sieve size in square meters
  mangroves_masked = mangroves_masked.connectedPixelCount(mmu, false); // Select groups of pixels smaller than or equal to mmu variable connected using 4-connected rules
  mangroves_masked = mangroves_masked.updateMask(mangroves_masked.gte(mmu)); // Remove areas below mmu threshold, leaving larger areas
  return mangroves_masked;
}

var classified_mangroves = classify_image(trained, compilation, grids); // Apply classification function defined above with provided parameters

var training_properties = 'canopy_h'; // Define height training attribute

// Function to estimate heights and biomass
function mangrove_biomass (gridded_area) {
  // Convert GEDI Raster to Points
  var height_data = gedi_data.sample({
    region: gridded_area.geometry(), // Limit points to study area
    scale: 25, // Resolution of GEDI raster dataset
    geometries: true, // Maintain geometries of points for later training
    numPixels: 2500, // Limit the number of pixels, required to avoid memory limit error
  });
  // Build Training Data
  var training_height = compilation.select(bands_height).sampleRegions({
    collection: height_data.limit(1000), // Limit GEDI points further to 1000 returns, required to avoid memory error
    properties: [training_properties], // Properties to copy from each input
    scale: 25, // Resolution of GEDI point
    tileScale: 16, // Reduce size of calculation tiles on GEE servers to prevent computation timed-out error, 16 is max value
  });
  // Train a ML Classifier
  var trained_height = ee.Classifier.smileGradientTreeBoost(250).setOutputMode('regression').train(training_height, training_properties, bands_height);
  // Estimate Heights
  var classified_heights = compilation.classify(trained_height);
  // Mask Mangroves
  var masked_heights = classified_heights.updateMask((classified_mangroves.select('classification')).eq(60));
  // Calculate AGB with formulas from Simard et al. 2019 Supplemental Table 8
  var agb = (masked_heights.expression('0.44*(height)', {'height': masked_heights.select('classification')})).pow(ee.Number(2.1578));
```

```
  return agb;
}

var mangrove_heights = mangrove_biomass(grids); // Apply above biomass estimation function to study area
print('Mangrove Heights Regression:', mangrove_heights);

// Add relevant layers to map for visualization of results
Map.addLayer(training_mangroves, {min: 0, max: 5, palette: ['#e41a1c', '#4daf4a', '#377eb8', '#984ea3', '#ff7f00']}, 'Training Mangroves', false); // Image for land cover classification training
Map.addLayer(grids, {color: '#5ab4ac'}, 'Grids', false); // Grids used to divide study area
Map.addLayer(gmw.filterBounds(grids.geometry()), {}, "GMW 2016", false); // GMW 2016 Baseline
Map.addLayer(classified_mangroves.select('classification'), {palette: 'green'}, 'Mangroves Classification'); // Classified mangroves
Map.addLayer(mangrove_heights, {min: 1, max: 5, palette: ['ffffcc', 'c2e699', '78c679', '31a354', '006837']}, 'Mangrove AGB', true) // Estimated mangrove biomass
Map.centerObject(grids); // Center map to gridded study area
```

Listing 9: General workflow for mangrove mask and canopy height regression for the coastline of Mozambique. Evolution of workflow from Boest-Petersen, 2022b.

GEE Environment

Link to GEE environment for AGB estimations of Mozambique coastline (refer to Listing 9):
https://code.earthengine.google.com/6dabb48086adcfabd9baf8a34d0b7d8d

Python Environment

Link to Python code for querying, downloading and processing of ICESat-2 returns:
https://github.com/aboestpetersen/icesat2_canopy_heights

Mozambique 2019 AGB GeoTiff File

Link to file via Aalborg University Sharepoint:
https://aaudk-my.sharepoint.com/:i:/g/personal/aboest20_student_aau_dk/EV901afL5txGjNvS3Zjdz1QBqgL
e=pkt1mK