

# Estimation of Aboveground Biomass in Zambia through Integration of GEDI, Sentinel-1 and Sentinel-2 Measurements

Jackson Kalaba<sup>1,2,3</sup>

<sup>1</sup>State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, China

<sup>2</sup>Key Laboratory of Carbon Cycling in Forest Ecosystems and Carbon Sequestration of Zhejiang Province, Zhejiang A&F University, Hangzhou, China

<sup>3</sup>College of Environment and Resources, College of Carbon Neutrality, Zhejiang A&F University, Hangzhou, China  
Email: Kalabajackson20@gmail.com

**How to cite this paper:** Kalaba, J. (2025). Estimation of Aboveground Biomass in Zambia through Integration of GEDI, Sentinel-1 and Sentinel-2 Measurements. *Journal of Geoscience and Environment Protection*, 13, 197-217.

<https://doi.org/10.4236/gep.2025.134011>

**Received:** March 24, 2025

**Accepted:** April 13, 2025

**Published:** April 16, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Accurate estimation of aboveground biomass (AGB) at regional scales remains challenging, particularly in diverse ecosystems like Zambia. This study developed an integrated approach combining spaceborne LiDAR (GEDI), synthetic aperture radar (Sentinel-1), and multispectral (Sentinel-2) measurements to generate footprints of 25 m diameter, 10 and 20 meter resolution maps of AGB across Zambia. Random Forest (RF) regression was employed to establish relationships between GEDI AGB product and Sentinel's variables. Feature importance analysis revealed that Near-Infrared (NIR) and Shortwave Infrared (SWIR) bands were the most significant predictors, particularly in arid regions. The model performance varied significantly across different land cover types, with  $R^2$  values of 0.453, 0.178, and 0.054 for Savanna, Grasslands, and Woody savanna, respectively. However, the overall model validation showed moderate performance with  $R^2$  of 0.508 for AGB estimation. While these results demonstrate the potential of multi-sensor data integration for large-scale vegetation structure monitoring, they also highlight the challenges in achieving consistent accuracy across heterogeneous landscapes. Future research should focus on enhancing model performance through improved calibration strategies, better incorporation of ecosystem-specific characteristics, and collection of reliable ground measurements.

## Keywords

Aboveground Biomass, GEDI, Sentinel-1, Sentinel-2, Random Forest, Multi-Sensor Integration, Zambia

## 1. Introduction

Accurate estimation of aboveground biomass (AGB) is crucial for understanding terrestrial carbon dynamics and supporting climate change mitigation efforts. These structural parameters serve as key indicators of forest ecosystem health, carbon sequestration capacity, and biodiversity (Harris et al., 2021; Lefsky et al., 2005). Traditional field-based methods, while accurate, are resource-intensive and impractical for large-scale assessments, particularly in regions with extensive vegetation coverage like Wulder & Coops (2018).

Zambia presents a critical case for vegetation monitoring, with approximately 59% of its land area covered by forests that are vital for biodiversity conservation, livelihoods, and carbon sequestration (Vincent et al., 2019; Hubau et al., 2020). However, the country faces one of the world's highest deforestation rates, driven by agricultural expansion, charcoal production, and infrastructure development (McNicol et al., 2019, 2020; Chidumayo, 2020). Limited availability of accurate and spatially explicit data on AGB hinders effective forest management and carbon stock monitoring (Ryan et al., 2019). Furthermore, the complex structure of Miombo woodlands, characterized by seasonal variability, diverse canopy arrangements, and high carbon storage potential poses unique challenges for remote sensing-based assessments (Borrelli et al., 2020). Remote sensing has revolutionized forest structure assessments through its diverse sensor technologies and broad coverage capabilities. The integration of active sensors (LiDAR and Synthetic Aperture Radar) and passive sensors (multispectral imagery) provides complementary data streams for comprehensive forest monitoring (Lu et al., 2018; Tang et al., 2019). Among these, spaceborne LiDAR, such as NASA's Global Ecosystem Dynamics Investigation (GEDI), has emerged as a critical tool for canopy height estimation (Dubayah et al., 2020). However, challenges such as spatial discontinuity and saturation in high-biomass regions limit its standalone utility (Simard et al., 2019; Bullock et al., 2020). Sentinel-1's radar capabilities excel in identifying large-scale forest structures with all-weather observation capacity, while Sentinel-2's multispectral data enables detailed characterization of forest types and vegetation conditions (Tamiminia et al., 2018; Bayle et al., 2019). The integration of these complementary datasets, enhanced by advanced machine learning techniques such as Random Forest regression, addresses the limitations of individual sensors and enables accurate, scalable wall-to-wall mapping of AGB (Avitabile et al., 2019; Gupta et al., 2022).

Previous studies have demonstrated the potential of multi-sensor integration for forest parameter estimation. A case study in New York State further validated the integrated approach by combining GEDI, Sentinel-1, and Sentinel-2 data using Random Forest algorithms. Their model, trained on 1.5 million data points, achieved impressive results with  $R^2$  values of 0.74 for 10-meter resolution canopy height estimates and 0.65 for AGB mapping (RMSE = 39.49 Mg/ha) (Bayle et al., 2019). However, most of these studies have focused on temperate or dense tropical forests, with limited applications in seasonally dry woodland ecosystems like those

found in Zambia, highlighting a significant research gap in these important ecological zones.

The integration of these datasets presents several methodological challenges. GEDI's sparse spatial coverage requires sophisticated interpolation techniques, while Sentinel-1's radar backscatter is influenced by soil moisture and surface roughness. Sentinel-2 data, though valuable for vegetation analysis, is frequently affected by cloud cover (Torres et al., 2012; Drusch et al., 2012). Recent machine learning approaches, particularly Random Forest regression, have shown promise in addressing these challenges by effectively combining multiple data sources and handling complex non-linear relationships (Gupta et al., 2022).

The integration of GEDI, Sentinel-1, and Sentinel-2 datasets presents several significant methodological challenges in estimating above-ground biomass density (AGBD). GEDI, while providing precise lidar-derived measurements of vertical forest structure, suffers from sparse spatial coverage due to its sampling design along the International Space Station's orbital paths, leading to gaps in data that require interpolation or fusion with other sources (Dubayah et al., 2020). Sentinel-1's radar backscatter data, though useful for capturing vegetation structure, is sensitive to external factors like soil moisture, surface roughness, and vegetation density, which can introduce noise and complicate its interpretation for biomass estimation (Torres et al., 2012). Similarly, Sentinel-2's optical multispectral data is affected by cloud cover, especially in tropical and subtropical regions, requiring advanced cloud masking and atmospheric correction techniques (Drusch et al., 2012). These challenges are further compounded by the need to align different spatial and temporal resolutions and manage varying spectral and radiometric characteristics across datasets. This study aims to develop an integrated approach for estimating AGB across Zambia by combining GEDI, Sentinel-1, and Sentinel-2 data. Specifically, we seek to: 1) develop and validate a methodology for generating high-resolution (25 m diameter, 10 m and 20 m maps of AGB; 2) assess the relative importance of different remote sensing variables in predicting forest structural parameters across various ecosystem types; and 3) evaluate the model's performance across Zambia's diverse landscapes. Our research addresses critical gaps in current vegetation monitoring capabilities in Southern Africa. By developing and validating methods specifically tailored to Zambia's ecosystems, this study contributes to improved understanding of forest structure dynamics in seasonally dry woodlands. The resulting methodology and products will support sustainable forest management, carbon stock monitoring, and conservation efforts in Zambia and similar ecosystems across Africa.

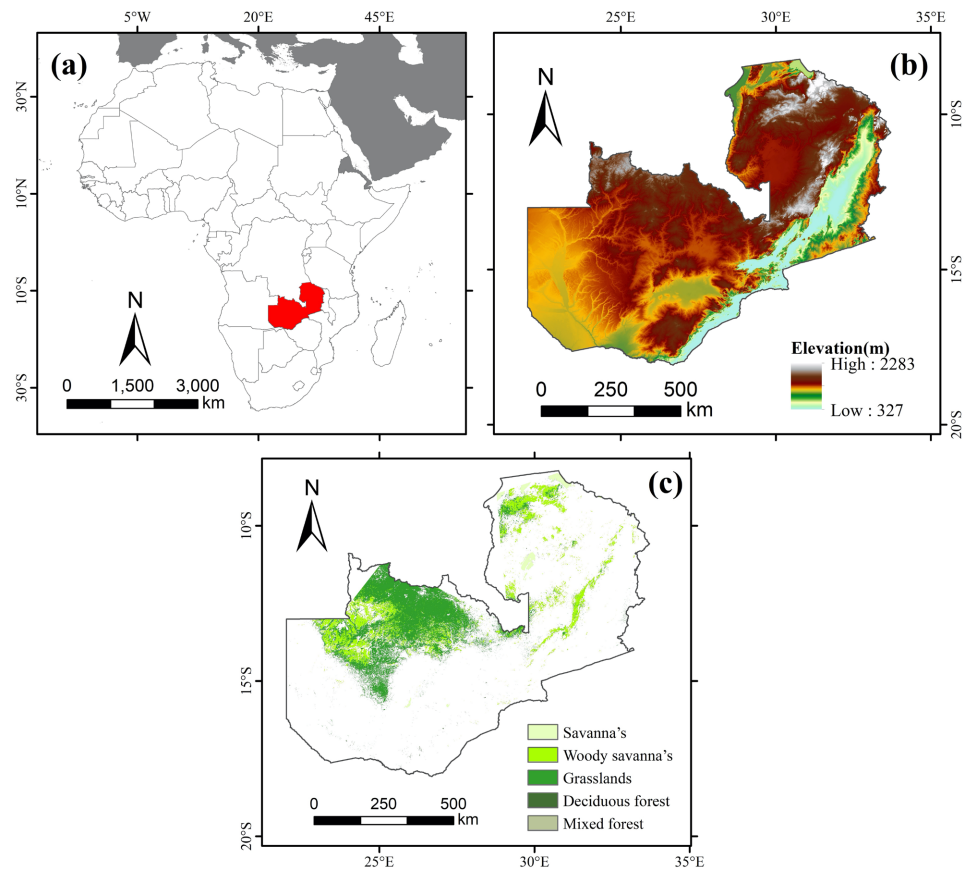
## 2. Data and Methods

### 2.1. Study Area

The study area for this research is Zambia, a landlocked country situated in south-central Africa (Figure 1). Located between the longitudes of 22° E and 33° E and latitudes of 8° S and 18° S, it covers approximately 752,618 km<sup>2</sup>. The country's

diverse topography, with an average altitude of 1200 m, encompasses plateaus, valleys, and highlands, supporting varied vegetation types that cover about 66.5% (450,000 km<sup>2</sup>) of the total land area.

Zambia experiences a tropical savanna climate characterized by three distinct seasons: cool-dry, hot-dry, and rainy (November to April). The average annual rainfall is approximately 1000 mm, varying between 700 mm in lowland areas and up to 1500 mm in the highlands, with the northern regions receiving the highest rainfall, exceeding 1400 mm annually. This climatic variation, combined with topographical diversity, creates distinct vegetation patterns across the country. Northern Zambia is dominated by dense miombo woodlands, primarily composed of *Brachystegia* and *Julbernardia* species, while the southern and eastern regions transition into grasslands and sparsely wooded areas. The vertical distribution of vegetation is evident in elevated regions, where changes in altitude influence species composition and canopy characteristics. The integration of remote sensing technologies such as GEDI, Sentinel-1, and Sentinel-2 in this study offers a unique opportunity to provide accurate and scalable assessments of aboveground biomass (AGB) and forest canopy height across Zambia, which is crucial for sustainable forest management and conservation.



**Figure 1.** (a) Location of Zambia in Africa, (b) elevation of Zambia with the highest elevation of 2283 m and lowest elevation of 327 m, (c) land covers of Zambia mainly covered by savanna lands, grasslands, woody savanna, Deciduous forests and mixed forests.

Zambia's land cover in (Table 1) above is dominated by savannas (340,830 km<sup>2</sup>) and grasslands (205,914 km<sup>2</sup>), with significant coverage of woody savannas (94266.2 km<sup>2</sup>) and deciduous broadleaf forests (35327.20 km<sup>2</sup>). Mixed forests occupy 51824.2 km<sup>2</sup>, while evergreen broadleaf forests are limited to 3553.75 km<sup>2</sup>. Notably, evergreen needleleaf forests are exceptionally rare at just 13.25 km<sup>2</sup>, and open shrublands cover only 607.5 km<sup>2</sup> (Ministry of Lands and Natural Resources, Zambia, 2023; based on MODIS Land Cover data at 500 m resolution). These figures highlight Zambia's prevalence of seasonal tropical ecosystems, with less than 0.1% of total land area comprising needleleaf forests, likely representing plantation outliers or classification artifacts in this predominantly savanna-dominated region.

**Table 1.** Land cover percentage.

Code	Type	Area (km <sup>2</sup> )	Percentage
1	Evergreen Needleleaf Forest	13.25	0.0018%
2	Evergreen Broadleaf Forest	3553.75	0.4703%
3	Savanna	340,830	45.1050%
4	deciduous Broadleaf Forest	35327.2	4.6752%
5	Mixed Forest	51824.2	6.8583%
6	Grassland	205,914	27.2504%
7	Open shrublands	607.5	0.0804%
8	Woody savanna	94266.2	12.4751%

## 2.2. Data

The study integrates three primary datasets: GEDI level 4B, Sentinel-1, and Sentinel-2, accessed through Google Earth Engine for Zambia in 2020. The preprocessed data from all three sensors was integrated to leverage their complementary strengths: GEDI's vertical structure measurements, Sentinel-1's backscatter information, and Sentinel-2's spectral information. This integration forms the foundation for robust estimations of forest structural attributes (Duncanson et al., 2021).

The input dataset was created by merging "GEDI level 4B, S1 and S2 data (Dubayah et al., 2022) which is provided by National Aeronautics Space Administration (NASA) for the year 2020 at 10 m resolution through Google Earth Engine" for the Study area Zambia.

### 2.2.1. GEDI

GEDI, a full-waveform LiDAR system mounted on the International Space Station, provides high-resolution data of forest vertical structure and canopy height (Kellner et al., 2023). The system employs three lasers emitting 242 pulses per second, creating 25 m diameter footprints to generate forest canopy profiles (Qi & Dubayah, 2016). For this study, we utilized the GEDI L4B dataset, specifically focusing on the "Mean Biomass Density", which estimates biomass in Mg/ha. The data was clipped to Zambia's boundaries and quality-filtered using sensitivity

metrics to ensure reliable observations.

### 2.2.2. Sentinel-1 Imagery

Sentinel-1's Synthetic Aperture Radar (SAR) operates in C-band (5.405 GHz) with dual polarization (VV and VH) in interferometric wide swath mode. We utilized Level-1 Ground Range Detected (GRD) products at 10 m spatial resolution, collected from July and August of 2020. The data acquisition employed two polarization combinations: vertical transmit/vertical receive (VV) and vertical transmit/horizontal receive (VH) at 20 m resolution (**Table 2**).

**Table 2.** Dataset.

Satellite	Band name	spatial resolution (m)
Sentinel-1	Co-polarized	20
	Cross-polarized (VH)	20
Sentinel-2	Blue (B)	10
	Green (G)	10
	Red (R)	10
	NIR	10
	Red edge 1	20
	Red edge 2	20
	Red edge 3	20
	Red edge 4	20
	SWIR 1	20
SWIR 2	20	
Elevation	DEM	30

### 2.2.3. Sentinel-2 Imagery

The Sentinel-2A and Sentinel-2B satellites include the multispectral instrument (MSI) and this satellite circles the globe at a height of 786 km, the multispectral sensor possesses a 20.6° field of view equating it to an image swath width of about 290 kilometres (Claverie et al., 2018). The Sentinel-2 consists of 13 spectral bands (**Table 2**) that encompass most areas of the electromagnetic spectrum, it has variations of spatial resolution per spectral band that range from 10 m for visible bands to 60 m for atmospheric bands. This article initially collected S2 pictures from June to August 2020 via the GEE platform. A cloud mask in Google Earth Engine was subsequently used to provide cloud-free photos. Subsequently, the visible, near-infrared (NIR), shortwave infrared-1 (SWIR-1), shortwave infrared-2 (SWIR-2), and red-edge bands were extracted.

### 2.2.4. Data Preprocessing

Google Earth Engine (GEE) was used to preprocess and integrate GEDI, Sentinel-1, and Sentinel-2 data for estimating Aboveground Biomass Density (AGBD) in

Zambia. Preprocessing GEDI, Sentinel-1, and Sentinel-2 data in Google Earth Engine (GEE) is a vital step in deriving accurate above-ground biomass density (AGBD) estimations, leveraging the complementary strengths of these datasets. GEDI data preprocessing focused on ensuring the reliability of vertical canopy metrics. This involved rigorous quality assurance filtering using sensitivity metrics to retain only the most reliable observations of forest structure (Dubayah et al., 2020).

Sentinel-1 synthetic aperture radar (SAR) data preprocessing steps included thermal noise removal, radiometric calibration, and terrain correction to ensure accurate backscatter values in VV and VH polarizations, which are indicative of vegetation structure (Torres et al., 2012). For Sentinel-2 multispectral imagery, cloud masking, atmospheric correction for surface reflectance, and resampling to align with GEDI and Sentinel-1 spatial resolutions were undertaken and are critical. These preprocessing steps ensure consistency, minimize errors, and allow for the seamless integration of datasets, forming the foundation for robust estimations of forest structural attributes in GEE.

Calibration involved training a Random Forest model using GEDI-derived AGBD as reference, with predictor variables including Sentinel-1 backscatter (VV, VH) and Sentinel-2 vegetation indices. In the absence of field data, this study relied entirely on GEE-processed remote sensing data (GEDI, Sentinel-1/2) for AGBD estimation across Zambia's savannas, grasslands, and woody savannas. Site-specific adaptations involved stratifying GEDI samples by ecoregion, the dependence on only 50 pre-selected GEDI samples raises concerns about representativeness, particularly for Zambia's diverse vegetation structures, highlighting the critical need for future field campaigns to validate and refine these satellite-derived estimates, especially in transition zones between biomes.

### 2.3. Random Forest Model

The Random Forest (RF) algorithm is widely recognized for its applicability in diverse domains, including environmental modeling, remote sensing, and forestry, due to its robustness, flexibility, and ability to handle complex datasets. As an ensemble learning method, RF constructs multiple decision trees during training and averages their outputs, reducing the risk of overfitting and enhancing model stability (Breiman, 2001). This characteristic makes it particularly suitable for tasks like estimating above-ground biomass density (AGBD), where data often exhibit high variability and non-linear relationships. One of the strengths of RF is its ability to handle both numerical and categorical variables without requiring extensive preprocessing, such as normalization or transformation. Additionally, it effectively manages missing data and is resistant to noise, making it an ideal choice for real-world datasets like those derived from GEDI, Sentinel-1, and Sentinel-2, which may have inconsistencies or gaps.

The RF algorithm's feature importance metric aids in identifying the most influential variables, allowing researchers to optimize input datasets and interpret

model results more effectively. However, RF is computationally intensive for large datasets and may require substantial processing power when applied to national or global scales, such as modeling forest attributes across Zambia or other countries.

Despite its many advantages, RF models can be prone to overfitting if parameters such as the number of trees, maximum depth, and minimum sample splits are not carefully tuned. Additionally, RF models are non-parametric and inherently opaque, which can limit their interpretability compared to simpler models.

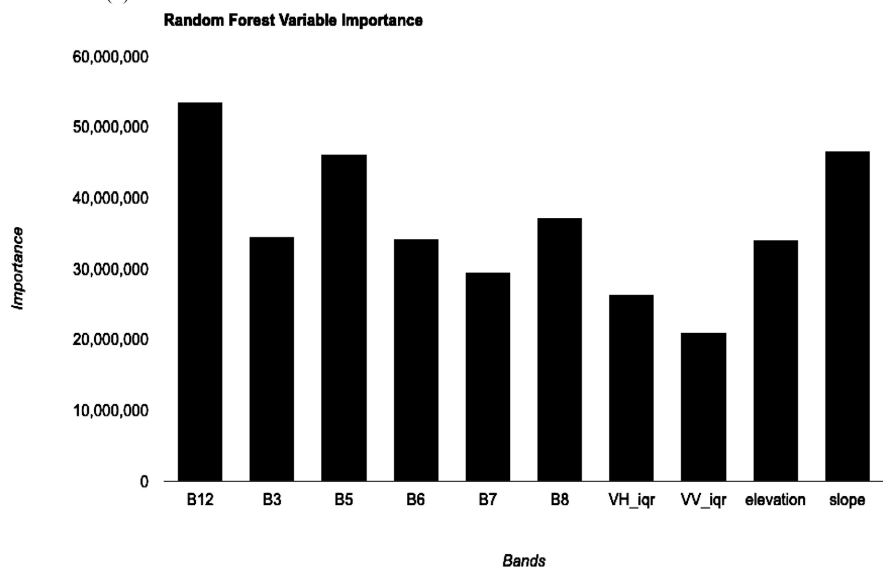
In our Random forest modelling the input parameters included the number of trees set at 50, this determines the ensemble size, with more trees generally improving stability and reducing variance, albeit at the cost of increased computational time (Breiman, 2001), a 70/30 minimum sample split was implemented and it indicates that 70% of the data is used for training and 30% for testing, ensuring sufficient data for model learning and robust evaluation and Maximum depth of 12, a parameter limiting the depth of individual decision trees, prevents overfitting by controlling model complexity, especially in large datasets with high variability.

Random Forest parameter selection in Google Earth Engine (GEE) for AGBD estimation required careful optimization to account for Zambia’s savannas, grasslands, and woody savannas. Key parameters included setting a high number of trees at 50 to stabilize predictions across diverse ecosystems, adjusting variables per split to balance feature selection, to preventing overfitting the limited 50-sample GEDI dataset.

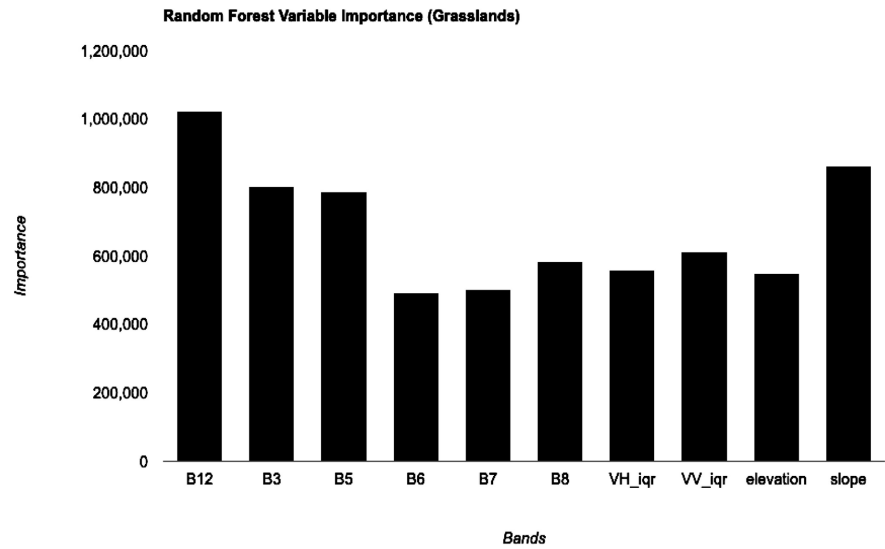
### 3. Results

Figure 2 shows the Random Forest Variable Importance for Savanna’s (a), woody savannas (b), grasslands (c) and overall Zambia (d) and the dominance B12 has on all ecoregions.

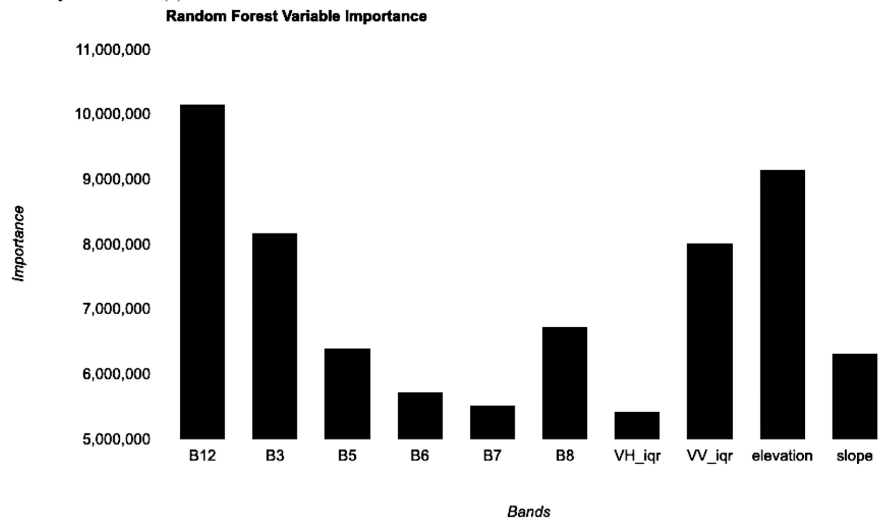
Savanna’s (a)



Grassland's (b)



Woody savanna's (c)



Zambia AGBD (d)

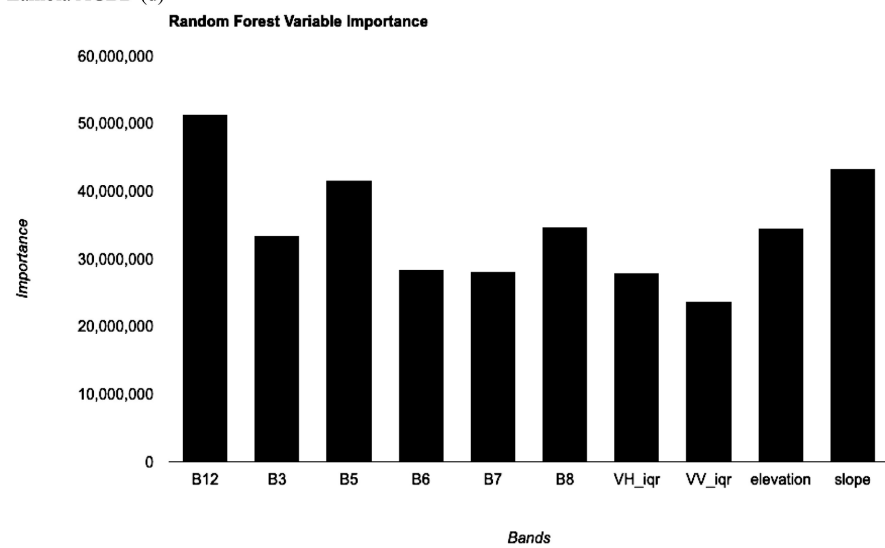


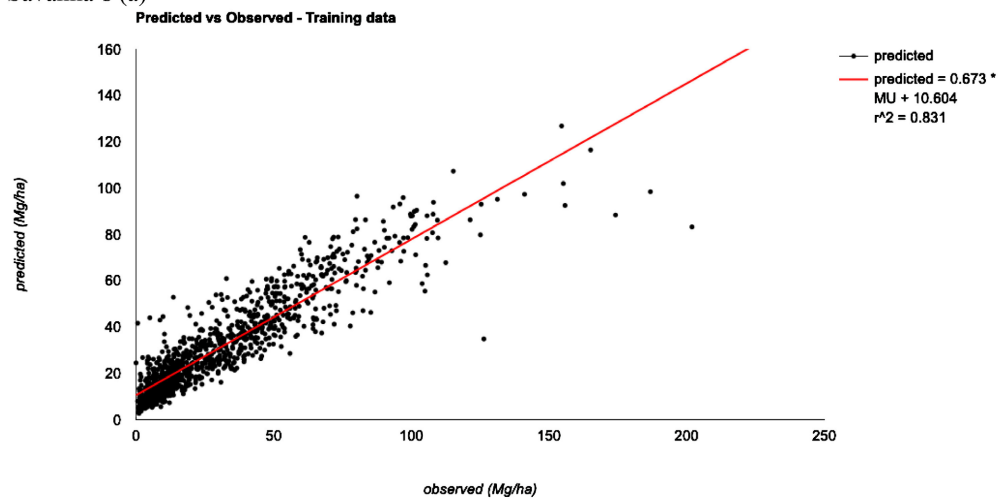
Figure 2. Random forest variable importance.

**Figure 3** observed vs. predicted values for training data 1) savanna  $R^2$  0.743, 2) grassland  $R^2$  0.769, 3) woody savanna  $R^2$  0.791 and Zambia AGBD, 4)  $R^2$  0.836.

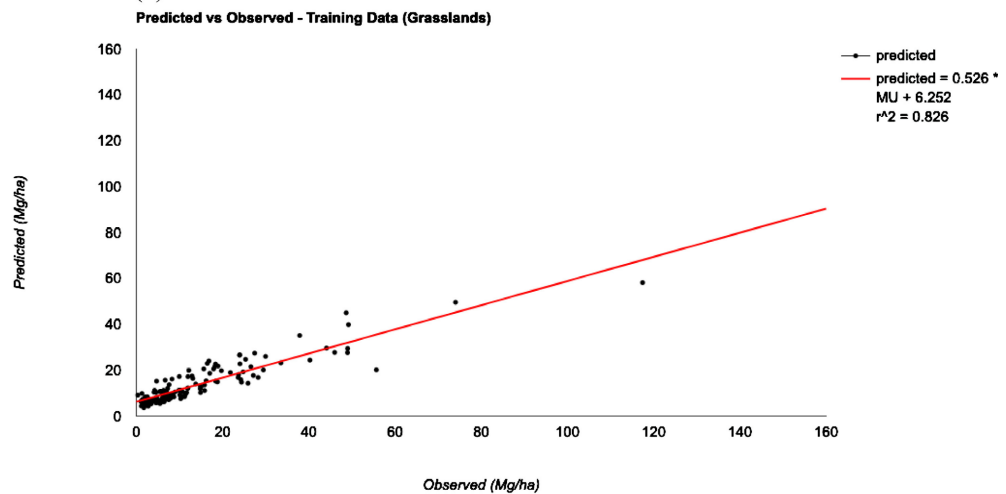
**Figure 4** shows the predicted vs observed values for Validation data in Zambia for (a) Savanna's  $R^2$  0.458 (b) Grasslands  $R^2$  0.178 (c) wood savanna's  $R^2$  0.054 (d) AGBD  $R^2$  0.508.

The mean Aboveground Biomass Density (AGBD) values reveal clear differences in carbon storage across vegetation types of Zambia, with woody savannas (57.26 Mg/ha) having the highest biomass due to their dense woody cover, followed by savannas (26.83 Mg/ha), which mix grasses and trees, and grasslands (13.12 Mg/ha), which store the least due to their lack of woody vegetation. The standard deviations highlight variability within each type: savannas (27.53 Mg/ha) show the highest variability, likely due to uneven tree distribution and disturbances like fire, while grasslands (12.65 Mg/ha) are the most uniform, reflecting their consistent herbaceous structure. Woody savannas (24.92 Mg/ha) exhibit moderate variability, suggesting differences in tree density or species composition. Together, these trends indicate that while woody savannas store the most carbon

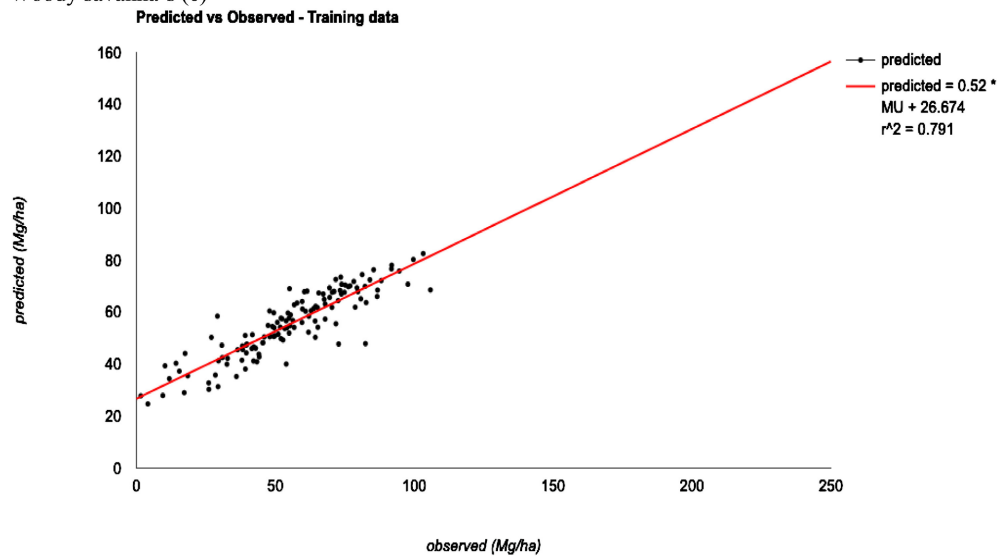
Savanna's (a)



Grassland's (b)



Woody savanna's (c)



Zambia AGBD (d)

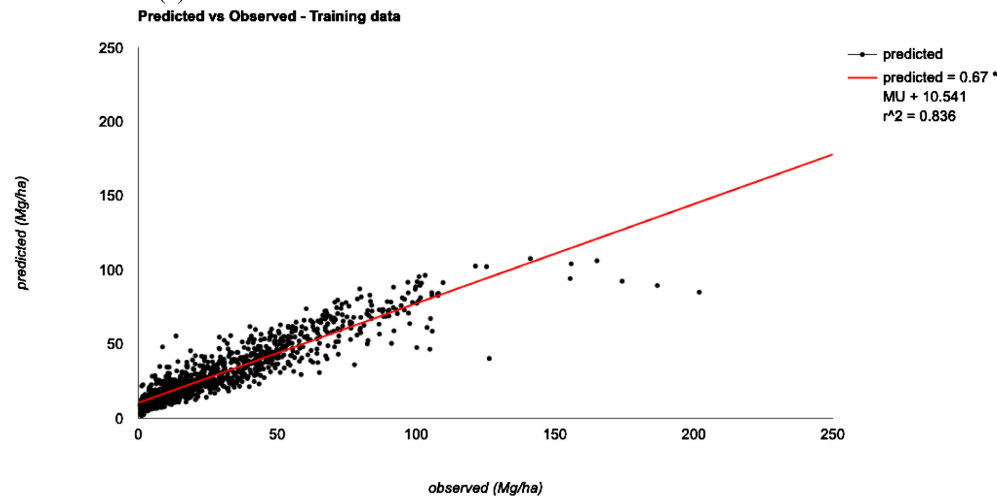
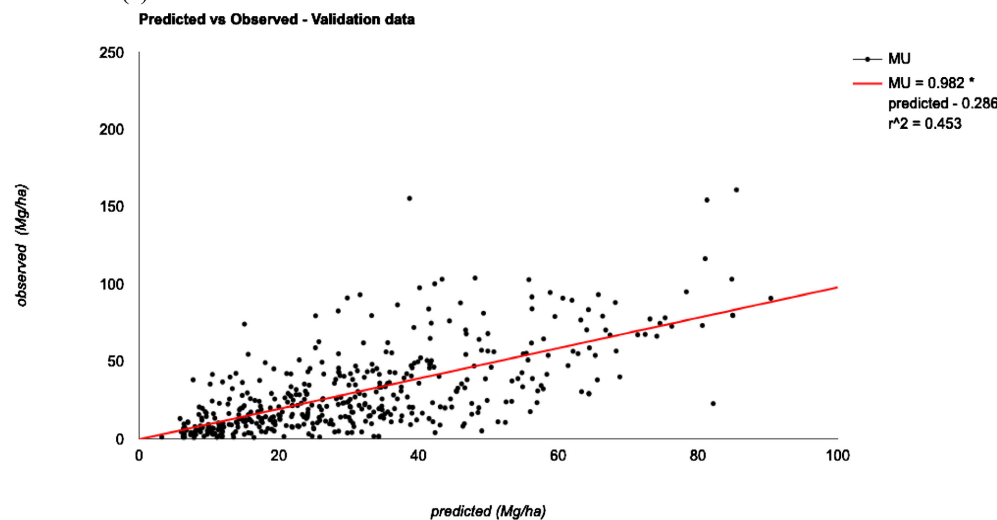
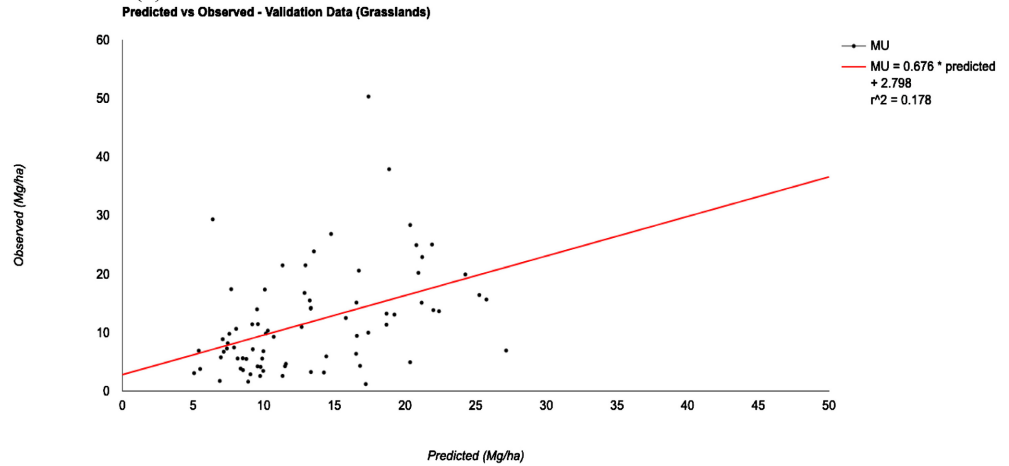


Figure 3. Training data.

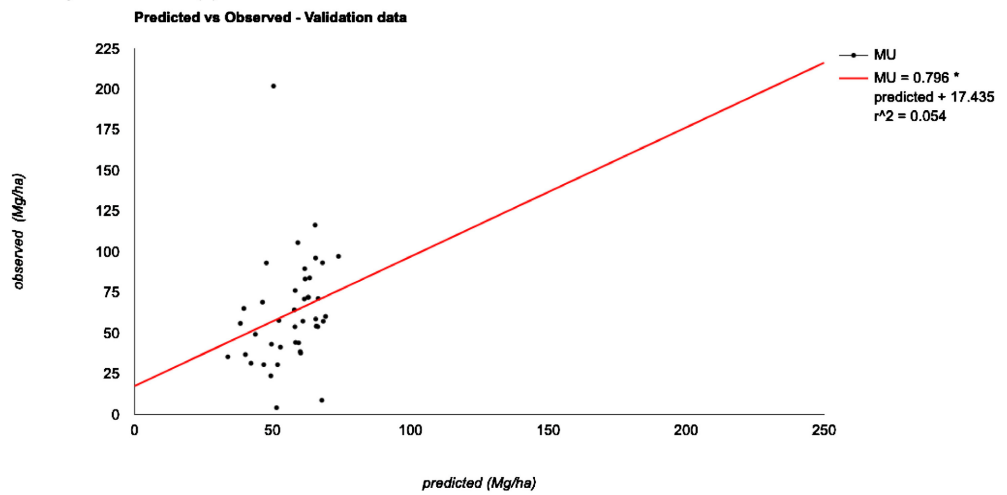
Savanna's (a)



Grassland's (b)



woody savanna's (c)



Zambia AGBD (d)

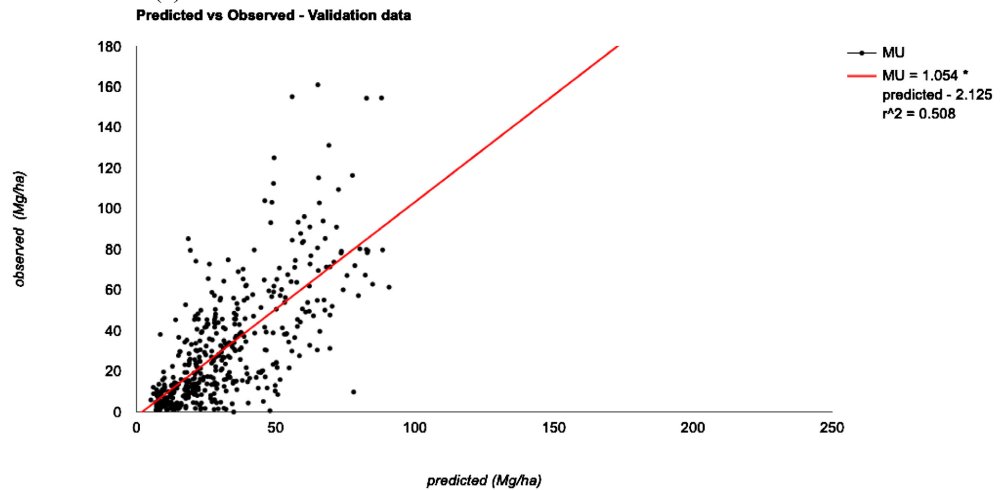


Figure 4. Validation data.

on average, savannas have the most unpredictable biomass, whereas grasslands are the most stable but least carbon-rich. These results in Table 3 show similar results compared to field data results of such ecoregions.

**Table 3.** Descriptive statistics (Zambia).

Vegetation Type	Mean AGBD (Mg/ha)	Standard Deviation (Mg/ha)
Savanna	26.83	27.53
Grassland's	13.12	12.65
Woody savanna's	57.26	24.92

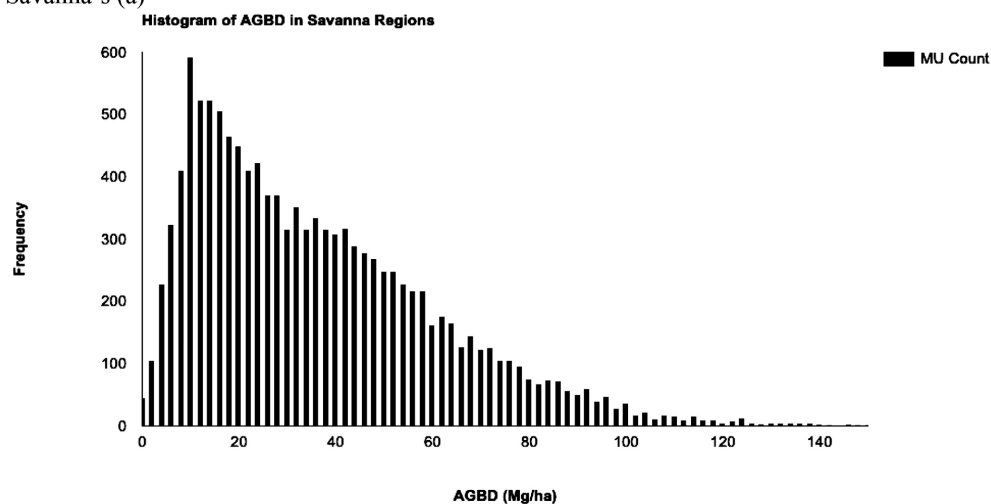
**Figure 5** shows the savanna histograms at a high frequency of 592.06, grasslands frequency at 271.21, woody savanna's at a frequency of 101.55 and AGBD Histogram at a frequency of 924.93.

## 4. Discussion

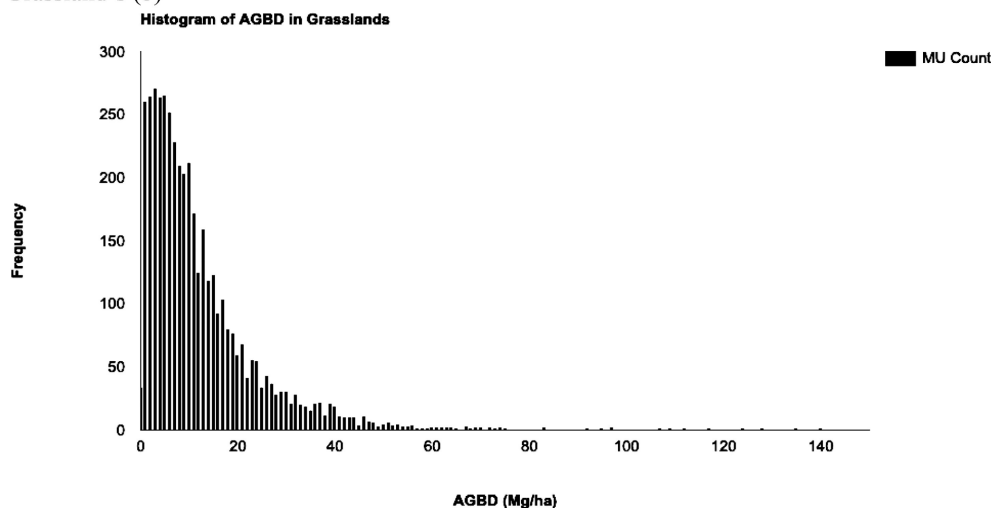
### *Random Forest Variable Importance*

Random forest model feature importance score shows the significance of this feature in the modeling process, according to the results we obtained from our GEE analysis in **Figure 2**, the spectral band played a more important role in the

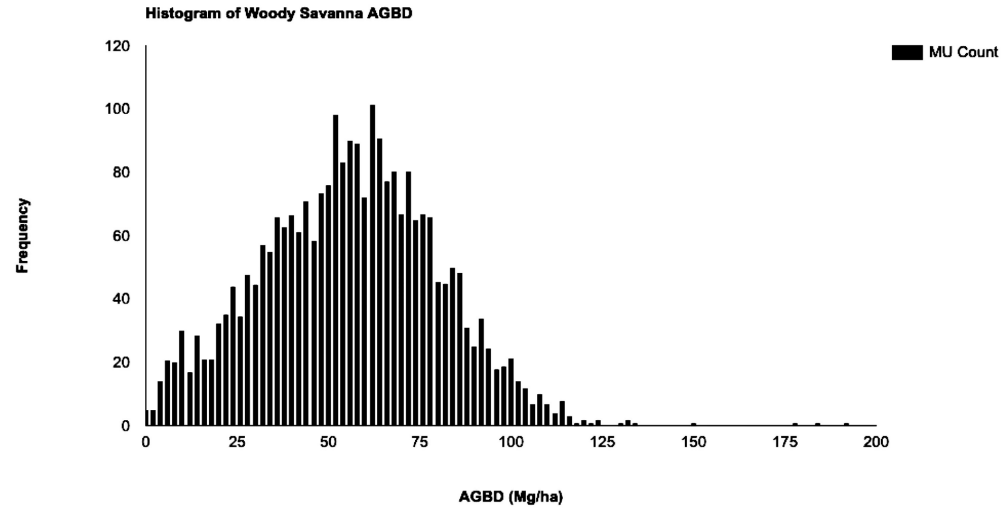
Savanna's (a)



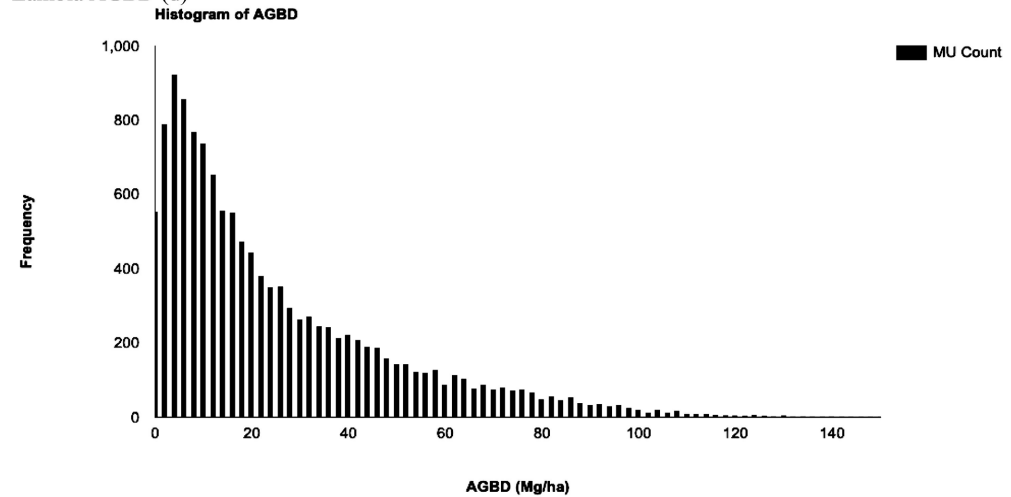
Grassland's (b)



woody savanna's (c)



Zambia AGBD (d)



**Figure 5.** Histogram distribution of predictions and observations.

process of predicting biomass compared to other features. In general, “NIR is especially useful for vegetation analysis because healthy, green vegetation reflects a substantial amount of NIR light”. This reflection allows scientists to measure plant health, biomass, and other vegetation properties by analyzing NIR values alongside other spectral bands (Weier & Herring, 2000). In a research conducted by (Zhu et al., 2015), the NIR had a strong correlation to AGB, but in our study, NIF was the fourth important feature only, and the reason is that SWIR is sensitive to moisture levels in soil, vegetation, and other materials. In dry conditions, bare soil or dry vegetation often reflects more SWIR than NIR. This is particularly noticeable in arid and semi-arid regions where the soil and vegetation have low moisture content.

***Evaluation of RF Model***

The R-squared value was used as a basis for assessing the performance of the model in Figure 3 and Figure 4. In the training data for the predicted versus observed values for Zambia Figure 3 of the results, standard R<sup>2</sup> 0. 836 for AGB in

training set. The result of  $R^2$  is about 83.6%, which demonstrates a strong alignment. This shows that the models' predictions within the training set account for about 83.6% showcasing a high level of predictive accuracy for the subset. The models' values range consistently from 0 - 250 mg/ha, meaning the model performs well within the observed distribution of biomass values in the training data. There were outliers found; these represent instances where the models' predictions deviate significantly from real findings. These anomalies could arise from regional differences in the forest structure or possible inaccuracies in data collection. Although these outliers have little impact on the training data's overall performance, they suggest that there is a possibility that the model may have limitations in capturing some subtleties within certain locations across Zambia.

The results suggest that incorporating more features and adjusting parameters could help in accurately representing the environmental variation responsible for causing biomass disparities. **Figure 4** shows results for validation data. The  $R^2$  in the research showed 45.3% savanna's, 17.8% grasslands and 5.4% woody savanna's after a full land cover evaluation the  $R^2$  percentage gradually improved to 50.8% AGB,  $R^2$  reflects the performance of the models and the closer it is to one the more effective the model is, having an  $R^2$  of 5.08 which is slightly above 50 % show that the model was effective in the estimation of above ground biomass density the unexplained variance from AGBD represents all the factors that are influencing the dependent variables but aren't accounted for by the model. **Figure 4** also shows that the satellites' multi-spectral image meets the needs of remote sensing inversion of the study area. This also shows that the satellites' multi-spectral image meets the needs of remote sensing inversion of the research area. Even though both observed and predictive biomass values still span the 0 - 250 mg/ha range. The validation dataset also contained about 20 outliers in AGBD highlighting regions or instances in which the model's predictions diverge from actual observations. These outliers may indicate varying ecological or environmental conditions not captured during the training phase, such as variations in microclimatic factors, and forest density that current model does not sufficiently account for. The decrease in predictive accuracy from training to validation data emphasizes the need for model refinement which can be accomplished by the inclusion of more environmental predictors or by using different modelling techniques that can better variety of conditions found across Zambia's landscapes. Despite the model's shortcomings in the validation set, the robust training performance provides a solid foundation for estimating aboveground biomass and with further adjustments, the model may prove to be a valuable tool for ecological monitoring and forest management across Zambia. The results we found correspond with the results of the study in Yunnan Province, where a Random Forest model was used to estimate AGB over the whole forest's types and each forest type individually. The GEDI L4A product was used as one of the dataset together with, Synthetic Aperture Radar integrated with Optical Data in this study ( $S1 + S2$ ,  $R^2 = 0.45$ ,  $RMSE = 35.78$  Mg/ha) showed a better performance compared to Palsar data integrated with Optical data ( $PALSAR + S2$ ,  $R^2 = 0.42$ ,  $RMSE = 36.59$  Mg/ha) (Chu et al.,

2023). All investigations concur on the strong association among spaceborne LiDAR and other optical imagery, including MODIS, Landsat, and Sentinel-2 (Baccini, 2008; Chi et al., 2015; Potapov et al., 2021; Simard et al., 2011). Operating other devices and collection of data can be costly, more especially when dealing with large study areas, the challenge of synchronizing and correlating many different dataset ALS etc. data with other datasets arises and when managing many data sources gathered across various time intervals. The alternative of GEDI data offers a uniform dataset since 2019, removing concerns regarding temporal alignment and guaranteeing a defined methodology for AGB estimate.

#### **Distribution of Predictions and Observations**

The distribution of AGB predictions was analyzed through histograms as shown in **Figure 5**. The AGB histogram shows a reasonable match, albeit with some outliers. These outliers suggest ecological conditions or microclimatic factors not captured by the model.

The amalgamation of GEDI, S1, and S2 data provided a robust instrument for generating high-resolution canopy height (CH) and aboveground biomass (AGB) on an extensive scale. The sensors delivered complementing textural, spectral, and vertical data that generated very precise AGB maps. The combination of S1, S2 and GEDI data reduced biasness in the AGB evaluation products, this helped improve the dependability and precision of remote sensing and offer.

The integration of Sentinel-1 SAR and Sentinel-2 Optical data significantly enhanced the Random Forest (RF) model's ability to predict Aboveground Biomass (AGB) and canopy height. This improvement stems from the complementary nature of these datasets. SAR data, particularly through VV and VH polarization backscatter metrics from Sentinel-1 exceeded in capturing vegetation structural attributes, biomass, and moisture content. These attributes are invaluable for estimating biomass in areas with dense vegetation, where optical data can be obstructed by cloud cover or other atmospheric conditions. SAR's penetration ability ensures reliable structural information under such scenarios.

On the other hand, Sentinel-2 optical data offers detailed spectral insights, particularly through the shortwave infrared (SWIR) and red-edge bands. These bands provide physiological and biochemical information about vegetation, such as moisture levels, chlorophyll content, and photosynthetic activity. Vegetation indices derived from these bands, such as the Normalized Difference Vegetation Index (NDVI) and Red Edge NDVI (RENDVI), highlight vegetation health and productivity. The fusion of SAR's structural sensitivity with the physiological information from optical data provides a holistic view of vegetation characteristics, reducing prediction uncertainty in Zambia's heterogeneous landscapes.

The practical application of supplementary research results from estimating above-ground biomass (AGB) using Google Earth Engine (GEE) offers transformative benefits for forest management. One key application is in carbon stock assessment and climate change mitigation efforts. Accurate AGB estimations derived from datasets like GEDI, Sentinel-1, and Sentinel-2 in GEE provide vital data for quantifying forest carbon storage and monitoring changes over time (Dubayah

et al., 2020). These insights are essential for participation in programs like REDD + (Reducing Emissions from Deforestation and Forest Degradation), which require accurate and consistent monitoring of carbon stocks to access funding and validate conservation initiatives (Duncanson et al., 2021). By leveraging GEE's ability to process large-scale datasets efficiently, forest managers can identify regions with significant carbon loss or gain, guiding policy actions and enhancing accountability in climate change mitigation strategies as well as in sustainable forest resource management and restoration planning. AGB data enable forest managers to identify high-biomass areas suitable for selective logging, ensuring sustainable timber harvesting practices while protecting critical habitats (Torres et al., 2012) with the results we found we can easily put it all into practice. The estimation of AGB and canopy height can help prioritize degraded lands for reforestation by identifying low-biomass regions that would benefit most from restoration interventions. Particularly Zambia where forests face threats from deforestation and agricultural expansion, GEE-powered analyses can assist in designing targeted restoration projects, balancing ecological restoration with economic development (Drusch et al., 2012). These applications demonstrate how integrating remote sensing research into forest management practices can support sustainable development goals, ensuring effective conservation and resource utilization.

### Future Research Directions

The research suggests several improvement options to improve prediction performance in the field of biomass estimations. These include investigating complex machine learning algorithms like Gradient Boosting Machines (XGBoost or LightGBM) and deep learning techniques like recurrent neural networks (RNNs) and convolutional neural networks (CNNs). Including additional environmental variables like land-use data, microclimatic variables, and soil characteristics can provide a more comprehensive understanding of the causes of AGB.

Stratifying the study according to Zambia's eco-regions can reduce prediction errors and enhance model resilience. Quantifying uncertainty is crucial to provide more details about prediction accuracy and identify areas needing specific enhancements.

Improving data fusion methods can enhance the vertical resolution and geographical coverage of biomass estimations by combining GEDI data with high-resolution datasets like aerial lidar. Investigating synergies using hyperspectral data can also improve the characterisation of vegetation biochemical characteristics.

Despite the success of the Random Forest (RF) algorithm during training, there may be room for improvement due to decreased accuracy during validation.

To further improve the accuracy and robustness of AGBD estimation in Zambia, future research should also explore the integration of additional remote sensing datasets that complement the existing GEDI, Sentinel-1, and Sentinel-2 framework. Incorporating L-band SAR data from missions like ALOS-2 PALSAR-2 or

the upcoming NISAR mission could significantly reduce saturation effects in dense woodlands, as L-band's longer wavelength (23 cm) enhances canopy penetration and sensitivity to high biomass levels 150 Mg/ha (Mermoz et al., 2022). Hyperspectral data from PRISMA or DESIS could refine species-specific biomass estimates by capturing narrowband spectral features related to lignin and cellulose content (Asner et al., 2021). Leveraging thermal infrared bands from Landsat may improve dry-season AGBD mapping by detecting water-stressed vegetation (Rifai et al., 2022). Finally, assimilating eddy covariance tower data on carbon fluxes could enhance temporal scaling of AGBD, particularly in dynamic savanna ecosystems (Poulter et al., 2023). Such multi-sensor synergies would not only address current limitations but also enable more accurate, spatially continuous, and temporally resolved AGBD monitoring across Zambia's heterogeneous landscapes. Advanced machine learning approaches like deep learning and gradient boosting can improve the ability to capture intricate ecological patterns. Adding time-series data from Sentinel-1 and Sentinel-2 could strengthen the model's resilience.

Future studies should also focus on including climatic and anthropogenic factors as explanatory variables, as these characteristics can help shape policy and better understand variations in AGB in the context of land-use statistics, population density, and climate forecasts.

## 5. Conclusion

The integration of GEDI, Sentinel-1, and Sentinel-2 datasets demonstrated the effectiveness of multisensor approaches in estimating AGB across Zambia's diverse landscapes. Our models achieved high predictive accuracy, with an  $R^2$  of 5.08 for AGBD, reflecting strong relationships between remote sensing data and forest structural parameters. The results reveal significant spatial variability, identifying biomass-rich regions such as intact savanna lands and degraded areas requiring urgent restoration efforts.

While the study achieved promising results, addressing the identified limitations GEDI LiDAR data has limited spatial coverage, particularly in regions with sparse data acquisition due to orbital constraints or dense cloud cover. This may result in gaps in data, which potentially affected the model accuracy in sampled areas and adopting the proposed research directions could significantly enhance the accuracy and applicability of biomass estimation models. These advancements would not only support sustainable forest management and carbon accounting initiatives but also contribute to the broader field of environmental monitoring and conservation, for future studies would suggest Developing region-specific models tailored to Zambia's diverse forest ecosystems and including environmental variables like soil properties and rainfall in the analysis would provide more nuanced insights into biomass distribution.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Asner, G. P., Martin, R. E., Knapp, D. E., Tupayachi, R., Anderson, C. B., Sinca, F., & Martinez, P. (2021). Airborne Laser-Guided Imaging Spectroscopy to Map Forest Trait Diversity and Guide Conservation. *Science*, *371*, 228-231. <https://doi.org/10.1126/science.abd4690>
- Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., & Willcock, S. (2019). An Integrated Pan-Tropical Biomass Map Using Multiple Reference Datasets. *Global Change Biology*, *25*, 1746-1760.
- 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*, Article ID: 045011. <https://doi.org/10.1088/1748-9326/3/4/045011>
- Bayle, A., Tillon, L., & Martin, J. (2019). Effects of Forest Management on Beetle Diversity: Implications for Conservation. *Forest Ecology and Management*, *433*, 729-737.
- Borrelli, P., Robinson, D. A., Panagos, P., Lugato, E., Yang, J. E., Alewell, C. et al. (2020). Land Use and Climate Change Impacts on Global Soil Erosion by Water (2015-2070). *Proceedings of the National Academy of Sciences*, *117*, 21994-22001. <https://doi.org/10.1073/pnas.2001403117>
- Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32. <https://doi.org/10.1023/a:1010933404324>
- Bullock, E. L., Woodcock, C. E., Souza, C., & Olofsson, P. (2020). Satellite-Based Estimates Reveal Widespread Forest Degradation in the Amazon. *Global Change Biology*, *26*, 2956-2969. <https://doi.org/10.1111/gcb.15029>
- Chi, H., Sun, G., Huang, J., Guo, Z., Ni, W., & Fu, A. (2015). National Forest Aboveground Biomass Mapping from ICESat/GLAS Data and MODIS Imagery in China. *Remote Sensing*, *7*, 5534-5564. <https://doi.org/10.3390/rs70505534>
- Chidumayo, E. N. (2020). Deforestation and Biomass Production in Miombo Woodlands: A Carbon Balance Perspective. *Global Ecology and Conservation*, *21*, e00845.
- Chu, T., Guo, X., & Takeda, K. (2023). Global 30 m Impervious Surface Area Dynamics from 1985 to 2020 Using Landsat and Nighttime Light Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *195*, 185-200.
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. et al. (2018). The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sensing of Environment*, *219*, 145-161. <https://doi.org/10.1016/j.rse.2018.09.002>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F. et al. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25-36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Dubayah, R., Armston, J., Healey, S. P., Yang, Z., Patterson, P. L., Saarela, S., & Tang, H. (2022). GEDI L4A Footprint Level Aboveground Biomass Density, Version 2.1. *Earth System Science Data*, *14*, 5727-5754.
- Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S. et al. (2020). The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Science of Remote Sensing*, *1*, Article ID: 100002. <https://doi.org/10.1016/j.srs.2020.100002>
- Duncanson, L. et al. (2021). Aboveground Biomass Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sensing of Environment*, *258*, Article ID: 113147.
- Gupta, A., Singh, K. K., Valbuena, R., & Dubayah, R. (2022). Synergizing GEDI and Sentinel-2 for Wall-to-Wall Mapping of Tall and Short Crops. *Remote Sensing of Environ-*

- ment*, 284, Article ID: 113335.
- Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., de Bruin, S., Farina, M. et al. (2021). Global Maps of Twenty-First Century Forest Carbon Fluxes. *Nature Climate Change*, 11, 234-240. <https://doi.org/10.1038/s41558-020-00976-6>
- Hubau, W., Lewis, S. L., Phillips, O. L., Affum-Baffoe, K., Beekman, H., Cuní-Sanchez, A. et al. (2020). Asynchronous Carbon Sink Saturation in African and Amazonian Tropical Forests. *Nature*, 579, 80-87. <https://doi.org/10.1038/s41586-020-2035-0>
- Kellner, J. R., Armston, J., & Duncanson, L. (2023). Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density. *Earth and Space Science*, 10, e2022EA002516. <https://doi.org/10.1029/2022ea002516>
- Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Del Bom Espirito-Santo, F. et al. (2005). Estimates of Forest Canopy Height and Aboveground Biomass Using ICESat. *Geophysical Research Letters*, 32, L22S02. <https://doi.org/10.1029/2005gl023971>
- Lu, D., Weng, Q., Li, G., Moran, E., & Hetrick, S. (2018). Remote Sensing of Impervious Surfaces in the Urban Areas: Requirements, Methods, and Trends. *Remote Sensing of Environment*, 117, 34-49.
- McNicol, I. M. et al. (2019). Aboveground Carbon Storage and Its Links to Stand Structure in Miombo Woodlands of Tanzania. *Forest Ecology and Management*, 433, 217-225.
- McNicol, I. M., Ryan, C. M., & Mitchard, E. T. A. (2020). Carbon Losses from Deforestation and Widespread Degradation Offset by Extensive Growth in African Woodlands. *Nature Communications*, 9, Article No. 3045.
- Mermoz, S., Bouvet, A., Le Toan, T., Villard, L., Réjou-Méchain, M., & Seifert-Granzin, J. (2022). Biomass Prediction in Tropical Forests: The Canopy Height Effect in L-Band Radar Backscatter. *Remote Sensing of Environment*, 272, 112948.
- Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A. et al. (2021). Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sensing of Environment*, 253, Article ID: 112165. <https://doi.org/10.1016/j.rse.2020.112165>
- Poulter, B. et al. (2023). Global Trends in Carbon Sinks and Their Variability under Climate Change. *Nature Climate Change*, 13, 343-350.
- Qi, W., & Dubayah, R. O. (2016). Combining Tandem-X InSAR and Simulated GEDI Lidar Observations for Forest Structure Mapping. *Remote Sensing of Environment*, 187, 253-266. <https://doi.org/10.1016/j.rse.2016.10.018>
- Rifai, S. W., Li, S., & Malhi, Y. (2022). Savanna Vegetation Structure and Demography under Global Change. *Annual Review of Environment and Resources*, 47, 259-284. <https://doi.org/10.1146/annurev-environ-120920-102939>
- Ryan, C. M., Pritchard, R., McNicol, I., Owen, M., Fisher, J. A., & Lehmann, C. (2019). Ecosystem Services from Southern African Woodlands and Their Future under Global Change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, Article ID: 20150312. <https://doi.org/10.1098/rstb.2015.0312>
- Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V. H., Castañeda-Moya, E., Thomas, N. et al. (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>

- Tamiminia, H., Homayouni, S., & McNairn, H. (2018). A Review of Machine Learning Approaches for Land Use/Land Cover Classification Using Remote Sensing Data. *International Journal of Remote Sensing*, *39*, 5172-5193.
- Tang, X., Zhao, X., Bai, Y., Tang, Z., Wang, W., Zhao, Y., & Liu, H. (2019). Carbon Pools in China's Terrestrial Ecosystems: New Estimates Based on an Intensive Field Survey. *Proceedings of the National Academy of Sciences (PNAS)*, *116*, 8623-8628. <https://doi.org/10.1073/pnas.1812865116>
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E. et al. (2012). GMES Sentinel-1 Mission. *Remote Sensing of Environment*, *120*, 9-24. <https://doi.org/10.1016/j.rse.2011.05.028>
- Vincent, J. B. et al. (2019). Human Disturbance Increases Functional But not Phylogenetic Diversity in Tropical Forests. *Nature Communications*, *10*, 2618.
- Weier, J., & Herring, D. (2000). *Measuring Vegetation (NDVI & EVI)*. NASA Earth Observatory. <https://earthobservatory.nasa.gov/features/MeasuringVegetation>
- Wulder, M. A., & Coops, N. C. (2018). Satellite-Based Time Series Models for Monitoring Forests. *Nature Climate Change*, *8*, 924-928.
- Zhu, Z., Woodcock, C. E., & Olofsson, P. (2015). Continuous Monitoring of Forest Disturbance Using All Available Landsat Data. *Remote Sensing of Environment*, *164*, 152-171.