# Reducing Tropical Forest Biomass Mapping Uncertainty
## – Integrating Field plot data with 3D forest structure from LiDAR

*A thesis submitted*

*in partial fulfilment for the degree of*

**Doctor of Philosophy**

by

**SURAJ REDDY RODDA**


**Department of Earth and Space Sciences**

**Indian Institute of Space Science and Technology**

**Thiruvananthapuram, India**

**March 2024**

# Certificate

This is to certify that the thesis titled **REDUCING TROPICAL FOREST BIOMASS MAPPING UNCERTAINTY – INTEGRATING FIELD PLOT DATA WITH 3D FOREST STRUCTURE FROM LIDAR** submitted by **Mr. Suraj Reddy Rodda** to the Indian Institute of Space Science and Technology, Thiruvananthapuram, for the award of the degree of **Doctor of Philosophy** is a bonafide record of the research work done by him under our supervision. The contents of this thesis, in full, or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Chandra Shekar Jha
Thesis Advisor - ISRO
Former Outstanding Scientist
NRSC/ISRO

Dr. Rama Rao Nidamanuri
Thesis Advisor – IIST
Professor & Head
IIST

Dr. Vinay Kumar Dadhwal
Thesis Advisor – IIST
Former Director
IIST

**Place**: Thiruvananthapuram

**Date**: March 2024

Counter Signature of HOD with seal

# Declaration

I declare that this thesis titled **REDUCING TROPICAL FOREST BIOMASS MAPPING UNCERTAINTY – INTEGRATING FIELD PLOT DATA WITH 3D FOREST STRUCTURE FROM LIDAR** submitted in fulfillment of the degree of **Doctor of Philosophy** is a record of original work carried out by me under the supervision of **Dr. Rama Rao Nidamanuri**, **Dr. Vinay Kumar Dadhwal, Dr. Chandra Shekar Jha** and has not formed the basis for the award of any degree, diploma, associateship, fellowship or other titles in this or any other Institution or University of higher learning. In keeping with the ethical practice in reporting scientific information, due acknowledgments have been made wherever the findings of others have been cited.

Mr. Suraj Reddy Rodda
SC19D041

**Place**: Thiruvananthapuram
**Date**: March 2024

***Dedicated to my little princess "SIARA"***

*My co-conspirator in delaying this thesis…*

*But also the greatest joy of my life...*

# Acknowledgements

I would like to express my sincere gratitude to the following individuals and organizations for their invaluable support throughout my doctoral journey. At the outset, I would like to thank Indian Space Research Organisation (ISRO) to facilitate the opportunity to pursue my PhD studies, and I am forever grateful for their investment in my research. The French Research Institute for Development (IRD) deserves special thanks for their financial support through an International Travel Grant, which facilitated a crucial aspect of my research.

My deepest gratitude goes to my supervisors and mentors. Dr. Rama Rao Nidamanuri, my thesis supervisor, provided invaluable guidance and expertise throughout the research process. Dr. Vinay Kumar Dadhwal, who served as both my thesis supervisor and mentor, offered unwavering support, encouragement, and invaluable advice. I am also grateful to Dr. Chandra Shekar Jha, my thesis advisor, for their instrumental role in helping me register for my PhD program and facilitating my part-time studies at National Remote Sensing Center (NRSC).

I would like to extend my thanks to my colleagues at NRSC – Mr. G Rajashekar, Dr. TR Kiran Chand, Mr. Rakesh Fararoda, and Mr. MSS Praveen. Their constant support, encouragement, and invaluable assistance during fieldwork were essential to the success of this research.

I am also thankful to the members of my doctoral committee, Dr. Samir Mandal, Dr. Deepak Mishra, and Dr. L. Gnanappazham, for their encouragement and insightful comments that significantly improved the quality of my thesis.

I am particularly grateful to my international collaborators, Dr. Maxime Réjou-Méchain, Dr. Pierre Ploton, and Dr. Pierre Couteron. Their support in collaboration, valuable ideas that helped shape my thesis, and specific contributions to Chapter 5 are deeply appreciated.

I extend my deepest gratitude to my parents, whose unwavering love, encouragement, and support throughout my PhD journey were indispensable. Their belief in me, even during moments of doubt, fueled my determination. I am forever grateful for their unwavering presence in my life.

Last but certainly not least, my incredible wife, Ms. Dimple Reddy, deserves a medal (and maybe a lifetime supply of Brownies!!). Her unwavering love, encouragement, and steadfast support were my guiding light through the challenges of doctoral studies. Her selflessness in prioritizing my PhD over her own career is a sacrifice I will always cherish. This accomplishment feels all the sweeter knowing I shared the journey with her by my side.

# Abstract

Tropical forests, with their high carbon storage and productivity, are vital to the Earth's carbon cycle. They act as both carbon sinks and sources, with nearly half their carbon stored in tree aboveground biomass (AGB). Therefore, accurate mapping and monitoring of forest AGB is crucial for designing effective carbon emission reduction strategies and gaining a deeper understanding of the dynamics of Earth's carbon cycle.

However, the large-scale AGB maps over Indian region from 1880 to 2021, generated through integration of Earth Observation (EO) and forest inventory plot data, reports markedly divergent AGB estimates despite accounting for reported uncertainties. These discrepancies stem from the inherent challenges involved in scaling up AGB measurements. While plot-level AGB values act as the primary basis for large-scale EO based AGB mapping, these values are estimated rather than directly measured. Consequently, plot-level AGB estimation introduces several uncertainties that cascade and magnify throughout the upscaling process when linked to EO data.

These uncertainties originate from several sources at both plot- and EO-level including (a) measurement errors inherent to field data collection at plot-level, (b) choice of allometric model to convert tree measurements to AGB, (c) plot size limitations and their representativeness across large area, (d) number of sample plots required for upscaling, (e) geolocation inaccuracies when linking field data to EO proxies, (f) saturation of EO signals with increasing biomass, potentially leading to underestimation. Finally, the generated AGB maps should incorporate uncertainty estimates through nested propagation of errors from various originating sources so as to not only develop defensible estimates at regional/national scales but also to enable robust biomass change detection to assess progress towards climate change goals.

This thesis attempts to address the challenges associated with estimating aboveground biomass (AGB) across large spatial scales in tropical Indian forests. It aims to establish a reference workflow or set of protocols to minimize uncertainties inherent in AGB estimations. Under this overarching aim, the objectives of this thesis are divided into four parts. First, the influence of plot size on the accuracy of plot-level AGB estimates was evaluated. Simulated forests plots were generated based on reference plot network (of 1-ha and 32-ha) across diverse Indian forest sites. The results show that relative error in plot-level AGB decreases with increasing plot size. While a 10x10 m plot yielded a 50% error, it significantly reduced to 5% at 70x70 m (0.49 ha) with a minimal reduction in

further increments to plot-size. Hence, 70x70 m was recommended the optimal plot size for reliable AGB estimation in Indian tropical forests.

Second, the most significant source of uncertainty in spatial AGB prediction – the choice of allometric model used to convert tree measurements to biomass is targeted. By using Terrestrial Laser Scanners (TLS), the study develops a method for non-destructively estimating tree volume using 3D point cloud data. Using the non-destructive tree volume from TLS allometric models were developed for central Indian tropical deciduous forests of India. TLS based models showed greater accuracy in estimating tree-level AGB when compared to the traditional allometric models. However at larger plot sizes (1 ha+) the errors have been minimal, highlighting the need for higher plot size in mitigating AGB uncertainty. However, comprehensive error metrics from the allometric model remain crucial for robust large-scale AGB estimation and uncertainty propagation.

Third, the established high-quality ground reference plots (1-ha each) were utilized to create reference AGB maps using Airborne LiDAR (Light Detection and Ranging) data. LiDAR offers 3D forest structure information and is found to directly link biomass estimated on ground to landscape level with high accuracy. For this, data from 13 sites across global tropics were collected and compiled (5 in Asia and 8 in Africa) and produced reference AGB maps by generating site-specific LiDAR-AGB models. Alongside, spatial uncertainty maps were also generated by propagating random errors from various potential sources through Monte Carlo method with 1000 iterations. This approach addresses the critical need for uncertainty propagation in the hierarchical chain of spatial AGB modeling, leading to robust and defensible regional estimates. The generated reference AGB maps were made available on open-access database and serve as valuable calibration and validation datasets for current and future EO missions (GEDI, BIOMASS, NISAR), ultimately enhancing the accuracy and reliability of large-scale AGB mapping initiatives. Finally, the practical application of these maps was demonstrated by using them to evaluate and refine the accuracy of the ongoing global AGB mapping mission using NASA's GEDI spaceborne LiDAR system, specifically over on Indian forests.

In essence, the current work examines the impact of several uncertainties from various ground and EO sources, highlighting the crucial role of high quality reference data at various stages in hierarchical modelling chain of spatial AGB estimation. The study advocates for adhering to best practices outlined by the Committee on Earth Observation Satellites (CEOS) to generate robust reference LiDAR aided reference AGB maps. This enables calibration of EO data and the creation of reliable and defensible regional and national AGB estimates, fulfilling the Intergovernmental Panel on Climate Change's (IPCC) needs for monitoring carbon stocks and fluxes.

# Abbreviations

| | |
|---|---|
| **3D** | Three dimensional |
| **AGB** | Above-ground biomass; Units - Mg $ha^{-1}$ |
| **AGBD** | Above-ground biomass density ; Units - Mg C $ha^{-1}$ |
| **AIC** | Akaike information criterion |
| **ALS** | Aerial LiDAR System or Aerial Laser Scanning |
| **ATLAS** | Advanced Topographic Laser Altimeter System |
| **BGB** | Below-ground biomass |
| **CCC** | concordance correlation coefficient |
| **CEOS** | Committee on Earth Observation Satellites |
| **CHM** | Canopy Height Model |
| **CO2** | Carbon dioxide |
| **DBH** | Diameter at breast height |
| **DBT** | Deciduous Broadleaf Forest |
| **DR Systems** | Discrete Return LiDAR systems |
| **DSM** | Digital Surface Model |
| **DTM** | Digital Terrain Model |
| **EBT** | Evergreen Broadleaf Forest |
| **EO** | Earth Observation |
| **ETS** | Electronic Total Station |
| **FW Systems** | Full waveform LiDAR Systems |
| **GEDI** | Global Ecosystem Dynamics Investigation |

| | |
|---|---|
| **GHG** | Green House Gases |
| **GLC2000** | Global Land Cover 2000 |
| **GSW** | Grasses Scrubs and Woodlands |
| **ICESAT-2** | Ice, Cloud, and land Elevation Satellite-2 |
| **IPCC** | Intergovernmental Panel on Climate Change |
| **IQR** | Inter Quartile Range |
| **LCM** | LiDAR Canopy Metric |
| **LiDAR** | Light Detection and Ranging |
| **Mg** | Mega gram |
| **Mg C $ha^{-1}$** | Mega gram carbon per hectare |
| **Mg $ha^{-1}$** | Mega gram per hectare |
| **MLS** | Mobile LiDAR System or Mobile Laser Scanning |
| **NISAR** | NASA-ISRO SAR Mission |
| **PFT** | Plant Functional Type |
| **QSM** | Quantitative Structure Model |
| **R** | R programming language |
| **RE** | Relative Error |
| **RMSE** | Root Mean Square Error |
| **TLS** | Terrestrial LiDAR System or Terrestrial Laser Scanning |
| **WD** | Wood Density |

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Forest ecosystems play a crucial role in the global carbon cycle as they constitute the largest terrestrial carbon pool, storing approximately 76-98% of the terrestrial organic carbon (Houghton et al., 2009; Pan et al., 2011). Moreover, these ecosystems absorb greenhouse gases (GHG) and offsets roughly 30% of anthropogenic $CO_2$ emissions through carbon sequestration and thus mitigating climate change (Friedlingstein et al., 2020). Consequently, accurate measurement and mapping of forest carbon (or biomass) represent critical components for quantifying carbon stocks and conducting comprehensive climate change impact assessments.

The Intergovernmental Panel on Climate Change (IPCC) categorized terrestrial ecosystems into five distinct carbon pools. These pools encompass various organic matter components, including above-ground biomass (living trees, plants, etc.), below-ground biomass (root systems), litter (fallen leaves and twigs), woody debris (dead logs and branches), and soil organic matter (humus and other organic compounds within the soil) (Eggleston et al., 2006). Among the five carbon pools, above-ground biomass (AGB) is the most visible, dominant, dynamic and important pool of the terrestrial ecosystem, constituting around ~45% of the total terrestrial ecosystem carbon pool (Bonan, 2008). Although below-ground biomass (BGB) significantly contributes to the overall carbon pool, estimating it presents considerable challenges due to its hidden nature. Estimating BGB often relies on indirect methods, leveraging relationships with AGB and incorporating forest stratification as necessary.

Above-ground biomass (AGB) represents approximately 70 – 90% of the total biomass (Cairns et al., 1997) and remains in a perpetual state of flux due to factors like fire, logging, storms, and land-use changes. This dynamic behavior significantly influences atmospheric carbon fluxes, highlighting the importance of prioritizing research and conservation efforts focused on AGB.

The most accurate way to measure a tree AGB is to fell and weigh all individual components of the tree. However, this method is demonstrably impractical due to its prohibitive costs, time consumption, and logistical challenges. Consequently, tree AGB estimation relies on allometric equations,

relationships developed by destructively measuring a representative sample of trees and linking specific tree dimensions like diameter at breast height (DBH) and height to tree AGB (Basuki et al., 2009; Chave et al., 2014; FSI, 1996). These tree-level AGB estimates are gathered over a fixed area (termed as plot) to provide overall AGB description of that specific region. However, generating a comprehensive spatial description of AGB solely based on data from field plots is not feasible. Achieving a reasonable level of accuracy with minimal uncertainty would necessitate an exceptionally high sampling intensity, making it impractical for large-scale assessments. Therefore, remote sensing is used as a tool for upscaling small scale field measurements to large-scale maps (Lu, 2006).

## 1.1 EO based Above Ground Biomass Mapping

Recent advancement of Earth observation (EO) technologies, encompassing various remote sensing methods, have revolutionized large-scale spatial AGB estimation. By capturing data from satellites and sensors on factors like vegetation height, structure, and spectral reflectance, remote sensing methods allow for cost-effective and efficient mapping of vast areas, including previously inaccessible regions (Franklin, 1986; Weishampel et al., 1996). Large-scale EO based AGB maps are predominantly generated by linking a few field-sampled AGB estimates with appropriate proxy variables from remote sensing data. Numerous studies have demonstrated that various Earth observation (EO) sensors offer distinct advantages for spatial AGB estimation.

Among remote sensing options for large-scale above-ground biomass (AGB) estimation, optical data perhaps offers the most promising alternative due to its unparalleled global coverage, consistent availability, and interpretability. Data from diverse platforms offering spatial resolutions ranging from sub-meter to hundreds of meters (e.g., IKONOS, Worldview, Cartosat, SPOT, Sentinel, Landsat, MODIS) has consistently produced large-scale AGB maps (Baccini et al., 2004; Fararoda et al., 2021; Foody et al., 2003; Hu et al., 2016; Ploton et al., 2012; Reddy et al., 2016). Additionally, radar systems like TerraSAR and ALOS provide valuable cloud-penetrating capabilities and detailed vegetation structure. The advent of cost-effective spaceborne radar acquisition since 2000 has significantly enhanced radar's role in AGB estimation (Liu et al., 2015; Saatchi et al., 2011; Santoro et al., 2021; Santoro and Cartus, 2023; Thumaty et al., 2016).

While traditional remote sensing techniques utilizing optical and microwave data have been effective in estimating aboveground biomass (AGB) through image-based methods and empirical relationships, they often struggle to accurately capture the complexities of forest ecosystems. This limitation results in signal saturation and high uncertainties, particularly in areas with high AGB. The advent of LiDAR (Light Detection and Ranging) systems provided a unique opportunity to capture three-dimensional vegetation canopy structure and

provided direct measurements of canopy height, which is one of the three main descriptors of AGB along with forest closure and forest type (Koch, 2010). Moreover, LiDAR enables the precise extraction of several stand-and tree-level vegetation structure metrics, leading to significantly improved AGB estimates compared to traditional methods (Garcia et al., 2010; Lim and Treitz, 2004; Popescu, 2007; Véga et al., 2015).

## 1.2 Forest Structure from LiDAR

Forest structure transcends a single, readily measured value. It can be defined as the physical and temporal distribution of trees in a forest stand (Oliver et al., 1996). It can be characterized by the position of trees, the vertical layering and the tree species mixture. While not directly measurable as a single quantity, forest structure can be effectively characterized through a variety of quantifiable variables such as species distribution, vertical and horizontal spatial patterns, tree and crown attributes, diameter-at-breast height, stand volume, leaf area density, clumping index and/or combinations of them (Figure 1.1).

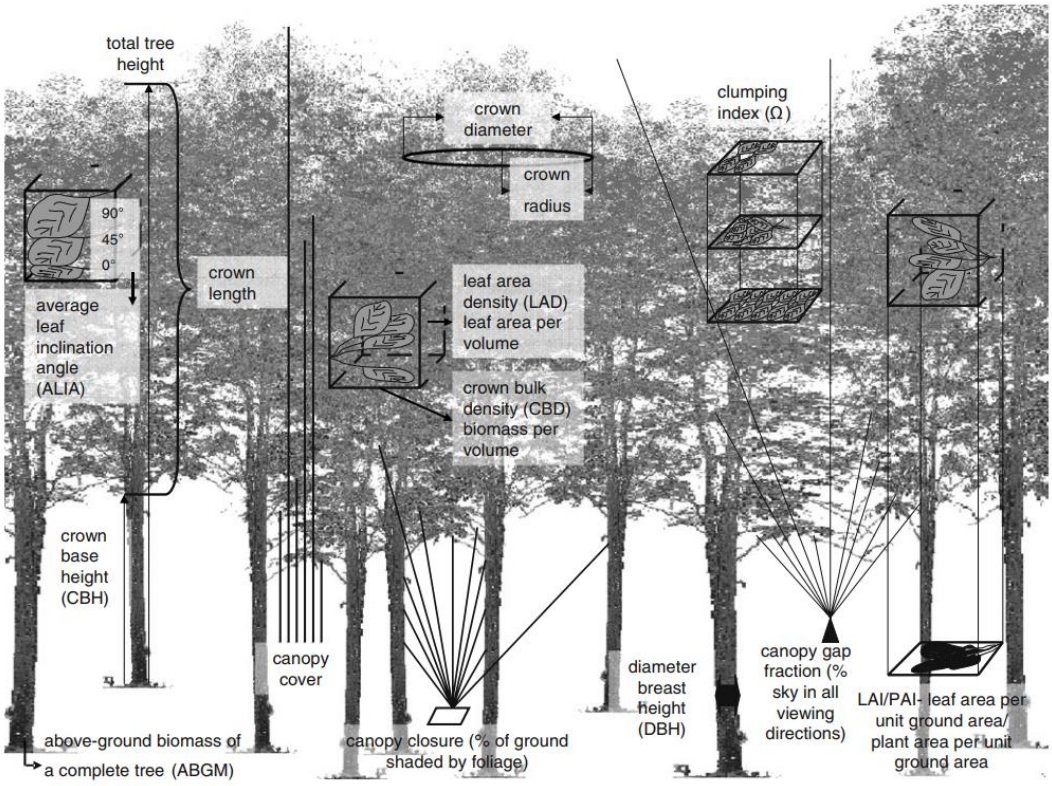


*Figure 1.1. Major structural characteristics of the forest canopy* (Seidel et al., 2011)

The structure of the forest canopy regulates the quantity and quality of light distributions in both space and time with in the stand, which in turn affects the presence or absence of understory vegetation and influences temperature, relative humidity and physiological activity of tree organs and many other organisms within a forest (Jennings et al., 1999; Kobayashi and Iwabuchi, 2008).

*Table 1.1. Potential contributions of forest stand structure elements from LiDARs*

| Forest Stand Element | Structural attributes | LiDAR Derivation |
|---|---|---|
| **Canopy height** | Height of overstorey | Direct retrieval |
| | Vertical distribution of under-storey and other surfaces | Direct retrieval |
| | Sub-canopy topography | Direct retrieval |
| **Foliage** | Vertical foliage profiles | Modeled/In-direct retrieval |
| | Number of strata | Modeled/In-direct retrieval |
| | Foliage density within different strata | Modeled/In-direct retrieval |
| | Canopy Cover / Leaf Area Index | Fusion with other sensors |
| **Tree Measurements** | Diameter at breast height (DBH) | Direct retrieval |
| | Diameter distribution | Direct retrieval |
| | Stem counts per ha | Direct retrieval |
| | Individual tree volumes | Modeled/In-direct retrieval |
| **Stand Biomass** | Stand Basal Area | Modeled/In-direct retrieval |
| | Standing Volume | Modeled/In-direct retrieval |
| | Above-ground Biomass | Modeled/In-direct retrieval |
| **Tree Species** | Species diversity or richness | Fusion with other sensors |
| | Relative abundance of key species | Fusion with other sensors |

Forest structure analysis presents a significant challenge due to the inherent complexity of forest stands. Traditionally, acquiring detailed and accurate information on the vertical and horizontal structure of forest layers relied heavily on datasets obtained through remote sensing techniques like optical, microwave imagery, and aerial photographs (Latifi, 2012). However, these methods primarily provided horizontal information, failing to capture the crucial aspect

of vertical vegetation distribution within forests. Attempts have been made to extract canopy height information using techniques like stereo-photogrammetry and interferometry. However, their practical limitations, including issues with canopy overlap and the accuracy of retrieval, restrict their widespread application.

In contrast to these conventional methods, Light Detection and Ranging (LiDAR) technology has emerged as a transformative tool. LiDAR offers a significant advantage by providing a precise 3-D descriptions of the forest ecosystem and enabling information on forest structure for a variety of purposes. Unlike conventional remote sensing methods that predominantly rely on the passive reflection of electromagnetic waves, LiDAR employs an active approach by illuminating the targeted area with concentrated laser pulses. These pulses propagate outward, and the time it takes for the reflected light to return to the sensor is accurately measured. This time information is subsequently utilized for precise calculation of distances to various objects within the laser beam's path (NOAA, 2012).

In essence, LiDAR operates on a principle analogous to radar technology, with a crucial difference: it employs light pulses instead of microwave waves. The light pulses emitted by a LiDAR sensor when flown over the vegetation system, has the ability to penetrate vegetation gaps and openings within the foliage. This capability enables LiDAR to "see" beyond the top layer of leaves, revealing the intricate vertical structure of the forest. The resulting three-dimensional structural and spatial information, eliminating of the influence of background, under-story, and canopy geometry, can be effectively utilized for characterizing forest structure (Lefsky et al., 1999a; Morsdorf et al., 2006). As highlighted in Table 1.1, LiDAR provides valuable data to in directly measuring and estimating several crucial forest structural attributes at both plot and single tree level (Dubayah and Drake, 2000).

## 1.3 LiDAR Remote Sensing for Forest AGB Characterization

LiDAR technology employs active laser scanners to capture detailed information about the environment, including terrain and physical features like forests. The direct measurements obtained from LiDAR, including canopy height, individual tree parameters (such as diameter, height, branching structures), sub-canopy topography, and stem density, are strongly linked to forest biomass. This inherent link allows LiDAR data to serve as foundational information for generating robust AGB estimates, ultimately enhancing the accuracy of spatial AGB maps. Hence, by providing precise and detailed insights into forest structure, LiDAR technology has emerged as a crucial tool for enhancing the precision and reliability of forest AGB estimations.

However, the attempt to mitigate uncertainty in AGB mapping using LiDARs depends upon various parameters linked to the design of LiDAR systems and the choices made in data acquisition and deployment platforms. Each parameter offers a distinct perspective on describing tree canopies, thereby contributing to the enhancement of forest AGB mapping. In the context of forest AGB mapping, the key classifications of LiDAR systems necessitate understanding, broadly categorized based on (A) the area illuminated by LiDAR systems on the ground, (B) the technology employed to record reflected laser pulses from the target surface, and (C) the platforms on which LiDAR systems are deployed—namely, terrestrial, airborne, or satellite platforms.

### 1.3.1 Small footprint or large footprint LiDAR systems

The size of the footprint, defined as the area on the ground illuminated by a single laser pulse determines the effectiveness of a LiDAR system to characterize canopy structure. The footprint size is influenced by two main factors: sensor distance from the target and laser beam divergence. Satellite-based LiDAR systems typically feature larger footprints, spanning several meters, compared to airborne (centimeters) and terrestrial (millimeters) systems due to varying distances between the sensor and the target area. This distinction between small footprint and large footprint LiDAR systems has direct implications for forest structure analysis, impacting aspects such as canopy penetration, spatial resolution, and data acquisition efficiency.

Small footprint LiDAR systems typically emit laser pulses with narrow beams, resulting in smaller illuminated areas on the ground. These systems offer higher spatial resolution, enabling detailed capture of fine-scale forest features such as individual tree crowns and understory vegetation (Lim et al., 2003). However, their limited footprint coverage may pose challenges in adequately sampling larger forested areas, particularly in dense canopy environments. On the other hand, large footprint LiDAR systems emit laser pulses with wider beams, covering larger areas on the ground per pulse. While these systems provide broader coverage and are more efficient for mapping extensive forest landscapes, they sacrifice spatial resolution compared to their small footprint counterparts. This compromise in resolution may result in reduced detail in forest structure characterization, particularly regarding smaller vegetation elements. Dubayah et al. (2000) highlight that small footprints may not be ideal due to potential oversampling of crown shoulders and missing treetops. Conversely, large footprints offer advantages like reduced cost for mapping extensive areas and capturing vertical forest structure details.

The selection between small and large footprint LiDAR systems depends on the specific objectives of forest analysis and the trade-offs between resolution and coverage requirements. Understanding the impact of footprint size is

essential for optimizing data acquisition strategies tailored to specific research goals and ensuring accurate interpretation of forest structure information by considering the limitations associated with each footprint size.

### 1.3.2 Full waveform or Discrete return LiDAR Systems

The other critical factor influencing the information acquired by LiDAR is the technology employed to record the energy of the reflected pulses from the target surface. They can be categorized as either discrete return (DR) or full waveform (FW) systems (Lim et al., 2003) on the basis whether they record the range to the first return and/or last return (DR systems) or fully digitize the return signal (FW systems) (Figure 1.2) . They differ from one another mainly with respect to the range measurement method, which implies significant differences in sensor hardware design. On the application side, it results in distinctively different number of range measurements recorded for each emitted laser pulse and subsequent substantial differences in data processing and analysis tools. As a result, the data collected by DR and FW sensors representing the same 3D target may look dramatically different (Parrish and Scarpace, 2007).

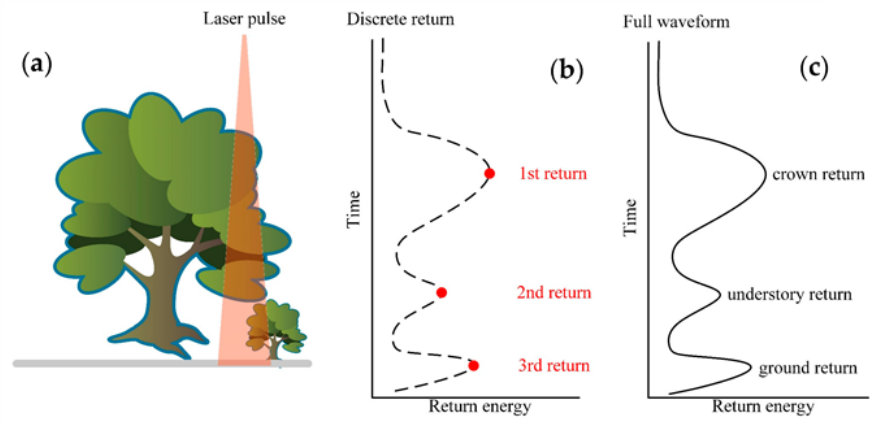


*Figure 1.2. Conceptualisation of different LiDAR systems over a vegetation canopy (a) Intersection of the laser illumination area with the tree crown and signals received with the (b) discrete return, (c) full waveform LiDAR systems*

Full waveform LiDAR captures the entire temporal profile of the returning laser pulse for each transmitted beam. A full waveform (FW) LiDAR measures the full profile of a return signal by sampling it in fixed time intervals, typically 1 ns (equivalent to 15 cm sampling distance), with theoretically quasi-continuous recording of the reflected energy for each emitted laser pulse. This allows for the measurement of the height distribution of illuminated surfaces (Wagner et al., 2008). While offering a complete digitized vertical profile at a sub-meter scale, the spatial resolution is typically coarser, ranging from 10 to 100 meters (Wagner et al., 2008). Due to its ability to provide detailed information about the reflected signal, full waveform LiDAR finds primary application in research (Lefsky et al., 2002).

Discrete return LiDAR, in contrast, focuses on capturing specific points within the returning signal where significant peaks indicate features encountered by the laser beam. These systems can capture information from one to five return signals per pulse (Hofton et al., 2006). The first return typically corresponds to the closest object reflecting the laser beam. This is often the topmost layer of vegetation or the ground surface in open areas. However, in densely forested environments where the tree canopy is close together, the situation becomes more complex. The laser pulse may be reflected by elements like foliage before reaching the ground, resulting in second or third-return signals (Lefsky et al., 2002). The final-return signal usually originates from the ground surface itself (Reutebuch et al., 2005; Wagner et al., 2008). This phenomenon of multiple returns is crucial for understanding the structure of forested areas using discrete return LiDAR.

### 1.3.3 Airborne, Terrestrial and Spaceborne LiDAR systems

The platform chosen for LiDAR deployment significantly influences the type of data acquired. Factors like the intended application, required measurement range, and the size of the area being covered all play a crucial role in platform selection (Lefsky et al., 2002). The major platforms are listed below (Figure 1.3).

- Aircraft (Airborne LiDAR system)
- Tripod or Mobile (Terrestrial LiDAR system)
- Satellite (Space-borne LiDAR system).

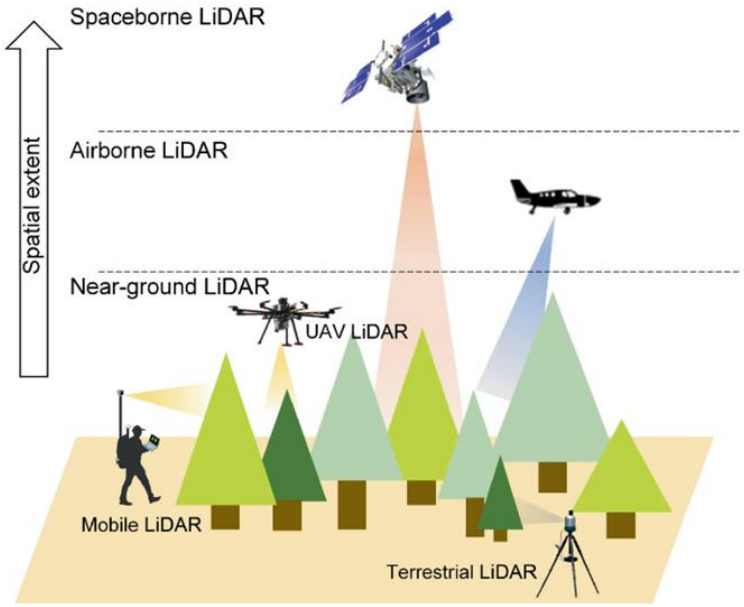


*Figure 1.3. LiDAR platforms used in forestry applications*

Airborne (or Aerial) LiDAR systems (ALS), typically mounted on airplanes or drones, offers exceptional detail and covers large areas with high efficiency.

Airborne LiDAR boasts sub-meter level accuracy and utilizes a small footprint (less than 1 meter) for detailed measurements (Næsset, 2004). Data acquisition can be in either discrete or waveform return format depending on the operating wavelength (typically 1064 nm) and flight altitude.

Due to its ability to cover large regions effectively, ALS is widely used for applications like biomass studies, forest inventory, and habitat modeling. ALS provides direct measurement of canopy height, and has proven their ability to directly measure forest structure, including canopy height and crown dimensions and estimate forest attributes, including tree, plot and stand level estimates for tree height (Falkowski et al., 2006; Maltamo et al., 2004), stem volume (Hyyppa et al., 2001; Naesset, 1997), basal area (Lefsky et al., 1999b; Means et al., 2000) and tree species (Brandtberg, 2007; Van Aardt et al., 2008). It also plays a vital role in large-scale regional assessments for carbon quantification efforts (Asner et al., 2012).

Terrestrial LiDAR (TLS) systems offer a unique perspective for capturing detailed 3D data of specific locations. They come in two main configurations: tripod-mounted scanners, providing high-resolution views from a single position, and mobile scanners mounted on vehicles or carried by personnel for broader area coverage. A key advantage of TLS is its ability to capture discrete objects from multiple angles. TLS is more useful for capturing small (relative to those captured from an aircraft) irregular objects such as buildings, earthworks and landforms such as cliff faces which can be profiled and monitored during mining. For complex objects with occlusions (hidden areas), TLS data acquired from multiple viewpoints can be merged to create a comprehensive 3D representation. However, this process can be time-consuming in dense environments like forests (Popescu et al., 2007).

One limitation of TLS is its operational range. Compared to airborne LiDAR, the effectiveness of TLS is hampered by occlusions caused by foliage and dense vegetation. Additionally, detailed TLS data collection can be expensive compared to other LiDAR technologies. Mobile Laser Scanning (MLS) addresses some of these limitations. Mounted on vehicles or carried on foot, MLS offers broader area coverage while still capturing valuable 3D data. Despite limitations, both TLS and MLS find valuable applications in forestry. They excel at capturing high-resolution 3D data of individual trees and forest structure, enabling the extraction of specific tree attributes like diameter, height, and crown width (Liang et al., 2018; Morsdorf et al., 2009; R Suraj Reddy et al., 2018). Furthermore, TLS/MLS is well-suited for in-depth analyses requiring intricate 3D information, such as tropical forest mapping and monitoring landform changes.

Satellite-mounted LiDAR systems, also known as spaceborne LiDAR, offer a unique vantage point for studying Earth's surface. Compared to airborne LiDAR, spaceborne systems hold a distinct advantage: circumventing limitations

imposed by airspace restrictions. Furthermore, spaceborne LiDAR excels in capturing data from geographically remote locations like the ice caps and polar regions. These areas often pose significant challenges for traditional airborne operations due to harsh weather conditions and logistical limitations. Satellite-based LiDAR can play a crucial role in these areas, enabling vital 3D profiling that contributes significantly to scientific research endeavors.

Despite its reach, spaceborne LiDAR employs a larger footprint, typically ranging from 25 to 80 meters in radius (Lefsky et al., 2002). This translates to the efficient sampling of various forest structures as the satellite traverses its orbital path. Additionally, the number of sampling points progressively increases as the satellite moves across its track, providing a broader data density. Recent advancements in space technology have opened exciting possibilities for forest structure measurement. Earth observation missions like NASA's ICESat-2 and GEDI serve as prime examples. These missions provide invaluable LiDAR data, crucial for analyzing three-dimensional aspects of vegetation biomass and structure (Biradar et al., 2020; Liang et al., 2018).

In brief, Different LiDAR systems mounted on various platforms cater to biomass information at varying scales: at individual tree-level (using TLS), stand-level (using ALS or MLS), at regional scales (using Space LiDARs). A recent review by Borsah et al., (2023) highlighted that airborne LiDAR is the most frequently used platform due to its ability to efficiently cover large areas while maintaining high accuracy. Spaceborne LiDAR usage is less common, and terrestrial LiDAR plays a limited role in large-scale AGB assessments. Notably, most studies focus on local-scale analyses, with fewer investigations conducted at regional or global scales.

## 1.4 Uncertainty in Spatial AGB estimation

The uncertainty in spatial above-ground biomass (AGB) maps stems from several contributing factors. The primary limitation is the inherent uncertainty associated with biomass estimation methods at various stages throughout the data acquisition and analysis process. Ground data or Field data, often collected through plot-based inventory measurements serves as the ground truth for calibrating and validating remote sensing techniques used for spatial AGB estimation. However, limitations exist. Primarily ground-level biomass measurements are seldom conducted directly and instead rely on estimations derived from volume equations, introducing errors due to extrapolation from merchantable volume and overlooking variations in wood density. Moreover, inconsistencies in forest area definitions and spatial layers compound these uncertainties.

Next, field plots represent only a small fraction of the total forest area, potentially leading to biased estimations, particularly in areas with high spatial

heterogeneity (McRoberts et al., 2010). Further, scaling these ground biomass measurements, typically conducted on small plots, to remote sensing pixel resolutions introduces inherent errors due to spatial aggregation. Additionally, geo-location errors in ground measurements contribute to inaccuracies. The sensitivity of different remote sensing data (optical and microwave) also plays a role, with limitations in capturing high-biomass regions due to saturation. Furthermore, the representativeness of ground samples across vast landscapes raises doubts about their ability to accurately reflect regional biomass distributions.

Accurately mapping Above Ground Biomass (AGB) across vast areas requires acknowledging and addressing inherent uncertainties throughout the data acquisition and analysis process. These limitations arise from factors like field measurements, remote sensing data quality, and chosen modelling approaches. Addressing them is crucial for two main reasons. Firstly, mitigating these uncertainties allows for the generation of more reliable spatial AGB maps, reflecting the true distribution of biomass across the landscape. Secondly, precise estimations of uncertainty are essential when using Earth Observation (EO) data to track changes in biomass over time. This ensures that observed changes are demonstrably larger than the inherent margin of error in the data. Addressing these multifaceted challenges is crucial for generating reliable and representative maps of AGB, ultimately enabling effective monitoring and management of this vital carbon resource.

### 1.4.1 Role of LiDAR in reducing Uncertainty in Spatial AGB estimation

The spatial AGB estimation across vast areas necessitates a hierarchical approach (Figure 1.4). This process involves progressively building upon ground measurements to generate regional AGB maps. However, inherent uncertainties arise at each stage, propagating to the final estimations.

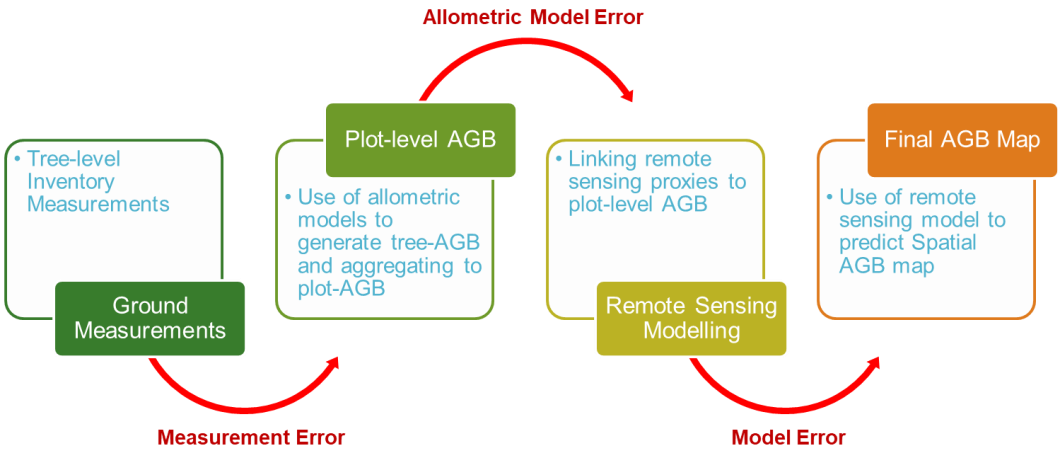


*Figure 1.4. Hierarchical process of Spatial AGB estimation and associated uncertainties*

A. **Ground Measurements**: Forest inventories gather data like tree diameter-at-breast height (DBH), height, and species information. Measurement errors in this initial stage introduce the first layer of uncertainty.
B. **Plot-Level AGB Estimation**: Pre-established allometric models convert these measurements into individual tree AGB and subsequently sum them to obtain plot-level AGB. Errors associated with these allometric models contribute to uncertainties in both tree and plot-level AGB estimates.
C. **Remote Sensing Integration**: Plot-level AGB is then linked to remotely sensed data proxies by establishing models. This stage introduces further complexities:
    - Upscaling plot data (smaller than remote sensing pixels) inherently introduces errors due to the mismatch in spatial resolution.
    - The developed models may have inherent errors that propagate to the final AGB map.

Recognizing and mitigating uncertainties throughout the multi-stage process of spatial Above-Ground Biomass (AGB) estimation is crucial for generating reliable forest biomass assessments. These uncertainties accumulate from ground measurements to the final regional maps, impacting their accuracy. Techniques like LiDAR, which capture the intricate three-dimensional structure of forests through precise laser measurements, offer valuable insights into key characteristics directly influencing biomass, such as tree height, crown size, and density. Utilizing distinct LiDAR systems in different stages has the potential to improve the accuracy of AGB estimation at various stages. However, the high cost and resource limitations associated with LiDAR restrict its widespread use for comprehensive regional coverage. Exploring alternative techniques, potentially in conjunction with LiDAR, remains necessary to achieve broader coverage for reliable large-scale biomass assessments.

Terrestrial LiDAR systems offer dense three dimensional (3D) virtualization of forest systems, which could provide the basis for building 3D tree models using quantitative structure models (QSMs). A QSM is a geometric model that describes a complete tree using a hierarchical collection of cylinder objects estimating topological, geometrical and volumetric details of the tree's woody structure (Raumonen et al., 2013). QSMs are reported to be adequate to estimate stem volumes (or biomass) at individual tree levels with very low bias in a non-destructive manner (Hackenberg et al., 2015; Mayamanikandan et al., 2019). The non-destructive tree volume estimation overcomes nearly every complicating factor associated with destructive sampling, including the issue of insufficient sample size in the development of allometric equations (Chave et al., 2004; Stovall et al., 2018). On the other hand, though QSMs could provide tree-level estimations of tree volume and biomass (when combined with wood density information), estimating tree-level biomass for all trees over a given area is still a daunting task. Nevertheless, the non-destructive QSM derived tree

volumes would aid in the development and improvement of site-specific (and species-specific) allometric equations for tree volume estimation, which are still a major source of uncertainty in field biomass estimation (Réjou-Méchain et al., 2019).

Aerial LiDAR systems provides direct measurement of canopy height, which is one of the three main descriptors of AGB along with forest closure and forest type (Koch, 2010). Its high-resolution laser scanning capabilities provide a detailed three-dimensional (3D) representation of the forest canopy, allowing researchers to extract crucial structural features directly linked to biomass, such as tree height, crown size, and density (Hyyppa et al., 2001; Lefsky et al., 1999b; Means et al., 2000; Naesset, 1997).

Despite its advantages, the high cost and limited coverage of LiDAR necessitate exploring alternative techniques for large-scale forest assessments. Satellite remote sensing emerges as a promising solution due to its broader scalability (Avitabile et al., 2016). Current research focuses on leveraging LiDAR-derived AGB maps from smaller study areas to calibrate regional AGB estimations using textural and spectral properties obtained from multi-spectral and microwave satellite imagery (Duncanson et al., 2021; Labrière et al., 2022). However, further development and testing across diverse forest types are crucial before widespread regionalization.

The recent launch of spaceborne LiDAR systems like GEDI (Global Ecosystem Dynamics Investigation) and ATLAS (Advanced Topographic Laser Altimeter System) offers new possibilities for gathering forest structure information across vast areas. However, it's important to note that these systems primarily collect profile measurements, not comprehensive scans. Therefore, directly scaling up these measurements necessitates the integration of suitable satellite imagery (both optical and microwave data).

By combining LiDAR data with readily available satellite imagery, researchers strive towards not only achieving precise one-time spatial estimates of AGB but also monitor AGB changes over time. This integrated approach holds significant promise for comprehensive and dynamic forest biomass assessment crucial for sustainable forest management and carbon monitoring efforts.

## 1.5 Aim, Motivation and Objectives

Accurate regional Above-Ground Biomass (AGB) estimation necessitates a multifaceted approach that minimizes uncertainties across various stages of the process. This begins with collecting high-quality ground measurements and minimizing their inherent errors by considering all factors influencing the upscaling process. At ground-level data, these factors include plot size, volume equations, geolocation accuracy, and sampling uncertainty.

Systematic collection and analysis of these measurements alongside existing volume equations are crucial. However, further refinement can be achieved by developing volume equations specific to the local forest ecosystem and expanding wood density measurements across diverse tree species. By incorporating spatial variability within trees – the way wood density can differ from trunk base to branches – these expanded measurements provide a more nuanced characterization of biomass distribution. Ultimately, this comprehensive approach, encompassing meticulous ground data collection, local volume equations, and detailed wood density measurements, paves the way for more accurate and reliable regional AGB estimations.

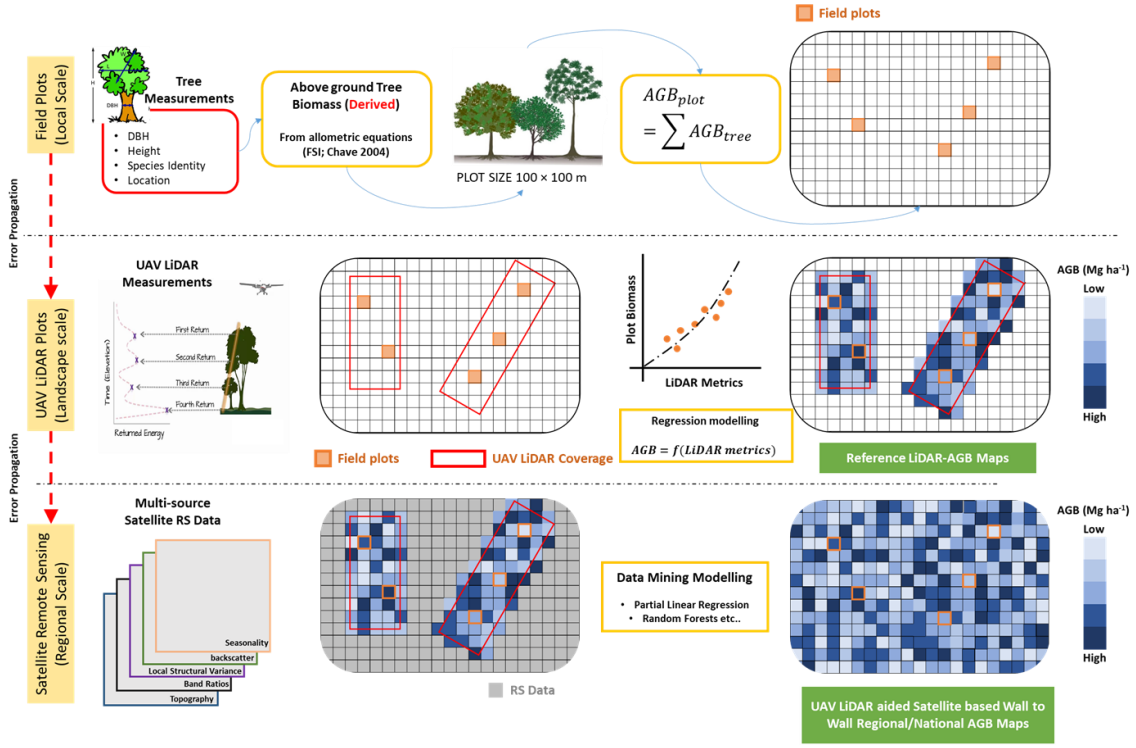


*Figure 1.5. Schematic flow of development of reference sites by collecting high-quality ground measurements and systematically generated reference LiDAR maps to serve as calibration/validation data for EO based spatial AGB estimation*

While high-quality ground measurements are essential for accurate regional Above-Ground Biomass (AGB) estimates, incorporating forest structure data from Aerial LiDAR offers significant potential for further improvement. Unlike traditional methods, LiDAR-derived forest structure metrics exhibit a non-saturating relationship with AGB, allowing for more accurate biomass descriptions, especially when combined with high-quality ground data. This non-saturating relationship enables the generation of high-quality, high-resolution reference AGB measurements at a landscape level, covering the entire area encompassed by the LiDAR data. These LiDAR-based reference maps literally multiplies the ground measurements with a factor of 10 to 100 (based on the LiDAR area coverage) and serve as calibration and validation data for

integrating information from various Earth Observation (EO) sources (Figure 1.5).

However, to fully leverage these LiDAR-derived reference maps, both ground-level and LiDAR-based AGB measurements require thorough characterization, including uncertainty quantification. This enhances the sensitivity of biomass estimation models that utilize multi-source remote sensing data. While implementing these steps comprehensively across vast areas might not be feasible, establishing a network of well-characterized reference sites presents a viable alternative. These reference sites, ideally covering a few hundred hectares each, would require high-quality ground measurements and complete Aerial LiDAR coverage.

Building upon this concept, this study proposes a two-pronged approach for Indian forests: (A) Enhancing Ground-Level Field Data to improve the quality of ground-level field data and understand how various factors influence biomass estimation at the plot level. (B) Generating Reference AGB Maps using high-quality field data and detailed 3D forest structure information derived from Aerial LiDAR. These reference maps will serve as valuable calibration and validation data for existing and future EO missions. Overall, this study would also establish reference protocols for generating such datasets at various stages, ultimately leading to the creation of more accurate regional AGB maps in the future.

In this context, this study aims to address the following research questions.

- **Ground Sampling Size**: What is the minimum ground sampling size required to ensure accurate biomass representation at individual sampling points across diverse forest types?
- **Plot Design and Uncertainty**: Can optimizing plot size or shape (e.g., circular vs. square) reduce ground measurement uncertainty, including geolocation errors, for improved data quality?
- **Optimal Plot Size**: Is there an optimal ground-level plot size for maximizing the accuracy and efficiency of ground-level biomass measurements across different forest types?
- **Non-Destructive Volume Equations**: Can we develop reliable non-destructive volume equations for biomass estimation using detailed 3D point clouds captured by Terrestrial LiDAR data?
- **Allometric Model Selection**: How does the choice of allometric models (traditional vs. locally derived models) affect the accuracy of plot-level AGB measurements, particularly when considering variations in species composition?

- **Sample Representativeness**: How can we improve the representativeness of ground samples collected in the field? Can we leverage metrics derived from Aerial LiDAR data to guide efficient and representative sample generation across a landscape?
- **Uncertainty Propagation**: How can we effectively propagate uncertainties associated with ground measurements throughout the upscaling process to generate reliable regional-scale AGB estimates?

### 1.5.1 Specific Objectives

This thesis investigates the inherent uncertainties associated with Earth Observation (EO)-based aboveground biomass (AGB) estimation in the Indian region. By analyzing and quantifying these uncertainties, the research aims to establish a high-quality reference AGB dataset. This dataset will serve as a crucial tool for improving the accuracy and reliability of current and future EO-based biomass mapping across India.

**I. Optimizing Plot Design for Accurate Plot-level AGB Estimates:**

- Establish tree-marked large plots across diverse Indian forest types.
- Analyze the impact of plot size and shape on plot-level AGB estimates to determine the optimal design for accurate biomass assessments.

**II. Developing non-destructive volume equations using Terrestrial LiDAR**

- Develop methodologies to estimate individual tree volume using 3D point clouds from terrestrial LiDAR.
- Construct new, locally-derived allometric models for improved biomass estimation.
- Quantify the impact of these improved allometric equations on overall plot-level AGB estimates compared to traditional approaches.

**III. Generating High-Quality Reference LiDAR AGB Maps with Uncertainty Estimates:**

- Utilize high-quality ground data and Aerial LiDAR measurements to generate reference AGB maps at the landscape level.
- Quantify and account for spatial uncertainty within the resulting LiDAR-derived AGB maps.

**IV. Calibration and Validation of Existing EO-Based AGB Products:**

- Calibrate and validate existing EO-based AGB products against the high-quality LiDAR AGB maps developed in objective 3.

## 1.6 Thesis Structure

This thesis is organized to address the previously outlined research objectives in dedicated chapters. Each chapter is self-contained and follows a consistent structure: introduction, research problem, data used, methodology, results and analysis, discussion, and conclusions.

**Chapter 1: Introduction**

- The current chapter offers a comprehensive overview of spatial aboveground biomass (AGB) mapping techniques, highlighting the limitations of traditional methods. It underscores the potential of Light Detection and Ranging (LiDAR) technology in overcoming these challenges by providing detailed 3D stand structure information. Additionally, it introduces a novel hierarchical approach for generating high-quality spatial AGB maps, which combines field data and Aerial LiDAR coverage to establish reference sites, crucial for calibration and validation of Earth Observation (EO) data, leading to more accurate AGB estimations.

**Chapter 2: Review of Spatial AGB Estimates over India**

- This chapter delves into a detailed literature review on historical AGB estimates over Indian forests from 1880 to date. It consolidates estimates from various studies, enabling analysis of variability and discrepancies arising from different methodologies, underlying assumptions and EO data used. Moreover, it explores the inherent challenges associated with upscaling AGB estimates along with associated uncertainties and discusses the importance of reference sites in streamlining this process to generate defensible AGB estimates at regional/national scales.

**Chapter 3: Optimizing Plot Design for Accurate Plot-level AGB estimates**

- This chapter focuses on the influence of plot size and design on reducing uncertainty in ground-based field plot AGB measurements. Utilizing a network of established large plots (1-ha) across tropical Indian forests, the chapter investigates the relationship between smaller plot sizes and the large plot reference. This is achieved by generating synthetic plots to determine the optimal field plot size for minimizing field-level errors associated with ground sample variability.

**Chapter 4: Developing non-destructive volume equations using Terrestrial LiDAR**

- This chapter presents the application of 3D point clouds obtained from Terrestrial LiDAR (TLS) data for creating virtual 3D tree models. It illustrates how these models can accurately estimate tree volumes, comparable to

traditional destructive sampling methods. Additionally, the chapter investigates the utilization of TLS data for developing non-destructive tree volume equations tailored for tropical forest ecosystems in India. It conducts a comparative analysis of these equations against conventional methods, emphasizing the potential enhancements in accuracy and efficiency.

**Chapter 5: Generating High-quality reference LiDAR AGB maps with uncertainty estimates**

- This chapter builds upon the insights gained from Chapters 3 and 4, which focused on improving ground-level data collection. It establishes a standardized protocol for compiling field data across diverse tropical Indian forest sites, ensuring consistency and facilitating robust data analysis. Furthermore, the chapter outlines the process of integrating the standardized field data with detailed 3D structural metrics derived from Aerial LiDAR data. This integrated dataset is utilized to generate high-quality reference AGB maps at the landscape level for four key tropical Indian forest sites. Additionally, the geographical scope is expanded, detailing the application of a similar methodology to produce nine additional reference maps for sites in Africa and Thailand.
- Moreover, the chapter discusses on the implementation of a Monte Carlo approach to integrate errors from all potential sources in a hierarchical manner. This method propagates random errors through the estimation process from ground measurements to map level, thereby establishing reference AGB maps along with associated uncertainty measures.

**Chapter 6: Calibration and Validation of Existing EO-Based AGB products**

- This chapter leverages the reference maps generated over Indian forests to demonstrate the application of LiDAR-derived AGB maps for calibrating and validating EO data. Here, the focus is on the Global Ecosystem Dynamics Investigation (GEDI) LiDAR mission data, but the approach can be extended to other EO products.

**Chapter 7: Conclusions and Future Scope**

- This chapter provides a concise summary of the key findings and inferences drawn from the entire research endeavor. It reiterates the significance of the research in improving the accuracy and reliability of EO-based spatial AGB estimates for tropical forest ecosystems, emphasizing the value of Aerial LiDAR integration for this purpose.

# Chapter 2

# Review of Spatial AGB Estimates over India

The critical role of forest aboveground biomass (AGB) in regulating global carbon cycling and mitigating climate change necessitates its accurate assessment. In the current time, several researchers (both in India and across the globe) have utilized various data sources, like satellite imagery and field measurements, to estimate forest AGB at National scales. These efforts have yielded a plethora of regional and global biomass maps encompassing India. However, a comprehensive review and evaluation, particularly of India-centric maps, remains conspicuously absent from current scientific literature. This chapter aims to bridge this knowledge gap by rigorously examining a wide range of existing forest AGB maps for India.

The endeavor to conduct a meaningful comparative analysis across these maps is not without its inherent challenges. Discrepancies exist due to the utilization of diverse methodologies within different studies. These methodological variations manifest in maps with varying units (e.g., tonnes per hectare, megagrams), spatial resolutions, and the application of distinct conversion factors. Additionally, some maps exclusively present AGB estimates, while others offer total biomass assessments. Furthermore, discrepancies exist in the forest masks used by different studies.

## 2.1 Current Forest Biomass Maps over India

Over Indian forests, numerous studies have generated above-ground biomass (AGB) estimates as part of the global or regional assessments. Presently, there exist approximately 18 global and 15 regional (India-specific) AGB assessments (Figure 2.1). It is noteworthy that some assessments attempt to estimate AGB for multiple years, providing valuable temporal data. The initial AGB assessment for India dates back to 1880, employing conventional bookkeeping methods. Over time, methodological advancements have facilitated the incorporation of Earth Observation (EO) data. This integration has significantly increased the number of national AGB maps, particularly following the 2000s with the enhanced accessibility of EO data.

However, ensuring data consistency and comprehensiveness remains a significant challenge. Not all studies generate spatially explicit AGB maps. Some studies restrict their output to regional estimates, hindering in-depth analysis at the grid level. Additionally, methodological discrepancies exist, with individual studies employing unique approaches, utilizing different grid sizes, and incorporating diverse input parameters.

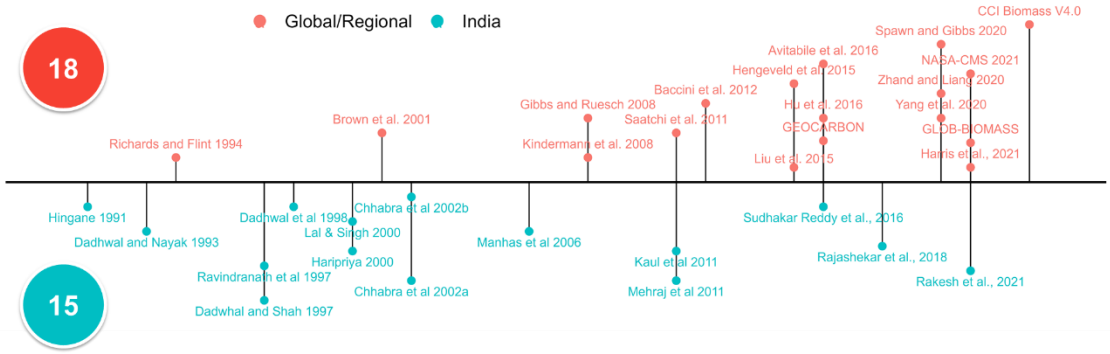


*Figure 2.1. List of Studies providing Biomass Estimates over India (1991 - 2023)*

To facilitate a structured analysis, the available map categories can be broadly classified into three groups based on the methodologies employed for map generation:

(A) Conventional Methods (Stratification and Multiplication)

Conventional methods rely on field sampling or bookkeeping approaches to generate initial AGB or biomass assessments (Brown et al., 2001; Rajashekar et al., 2018; Richards et al., 1994). Field sampling involves the establishment of sample plots where detailed inventory measurements are subsequently conducted. Conversely, bookkeeping methods utilize pre-established conditions and attempt to estimate AGB based on factors influencing biomass, such as human activity and population density. Once a collection of AGB estimates is obtained, it is multiplied by the total forest area (derived from remote sensing imagery, topographical sheets, or ground surveys) to generate regional or national-scale AGB estimates. This method essentially functions through a two-step process of stratification and multiplication. First, the forest area is stratified using factors such as forest area, density, or other relevant descriptors. Sample estimates of AGB are then collected within each strata. Finally, based on the average and statistical distribution of AGB within each strata, the total regional/national AGB is estimated through multiplication.

(B) Remote Sensing Approach

The remote sensing approach leverages forest characteristics detectable from satellite data, such as structure, reflectance, backscatter, and other properties (Baccini et al., 2012; Fararoda et al., 2021; Saatchi et al., 2011; Santoro and Cartus, 2023). Here, field samples collected from traditional forest

inventories are linked to corresponding remote sensing variables. This calibration process allows the available remote sensing data to be extrapolated to regional/national scales. With the increasing availability of data and advancements in methodologies, the linkages are established using a wide range of sophisticated techniques, including machine learning models, support vector machines, and deep learning algorithms. Overall, various remote sensing parameters are linked to the field-estimated AGB to generate spatially explicit AGB maps.

(C) Map Fusion Approach

As the name suggests, this method aims to combine individual AGB maps generated from different methodologies to create a new, improved map with potentially higher accuracy (Avitabile et al., 2016; Spawn and Gibbs, 2020; Zhang and Liang, 2020). Here, multiple AGB maps are fused into a single map by addressing potential biases between them using an independent dataset. This combined map is expected to be more accurate and reliable compared to the individual input maps.

### 2.1.1 Spatial Biomass Estimations over Indian Forests: 1880 – 2021

A comprehensive review identified a total of 85 biomass (AGB or total biomass) assessments generated by 33 studies encompassing India within their scope (both global and national). Notably, some studies included multi-temporal assessments, providing valuable data across different time periods.

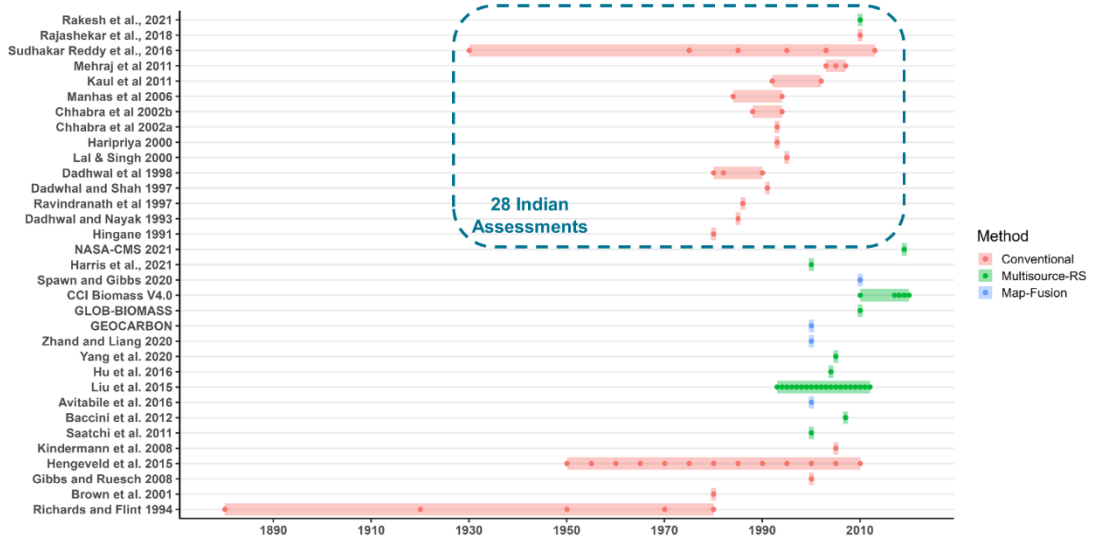


*Figure 2.2. Historical Timeline of Forest Aboveground Biomass (AGB) Estimations in India by various studies.*

Figure 2.2 depicts a historical timeline of aboveground biomass (AGB) estimations conducted in India by various studies. Each data point (dot) represents the year of an assessment, with multi-temporal assessments

(assessments conducted for multiple years) connected by a band. The color coding of both dots and bands corresponds to the specific method employed by the study for AGB estimation.

Of the total 85 assessments, 28 originated from regional Indian studies. These studies focused solely on India and did not produce spatial biomass maps. Their outputs were primarily tabulated data with national-scale biomass estimates. In contrast, most global studies generated spatial biomass maps and made them available through open-access platforms. However, a significant limitation across both global and regional studies was the lack of uncertainty estimates associated with the AGB assessments.

However, a significant challenge to data comparability lies in the heterogeneity of these assessments. Each map exhibits variations in several key aspects:

- **Spatial Resolution**: Spatial resolution varies considerably, ranging from coarse (1°) to fine (30 meters).
- **Units**: The units employed for biomass quantification differ across studies.
- **Methodological Discrepancies**: The methodologies used for AGB estimation vary between studies.
- **Forest Mask Inconsistencies**: The forest masks employed to define forest areas differ across studies.
- **Conversion Factor Variations**: The conversion factors used to convert different biomass components (e.g., leaves, stems) to total biomass or AGB also vary.

Further complicating comparisons, some studies solely report AGB estimates, while others provide total biomass assessments. This heterogeneity necessitates standardization efforts to facilitate meaningful comparisons between these AGB assessments and enable a more comprehensive understanding of forest biomass in India.

## 2.2 Standardization of Biomass maps

To facilitate a meaningful comparative analysis of the AGB maps across India, we implemented a two-pronged standardization approach. First, all maps were standardized to a common grid resolution of 5 kilometers. This ensures consistent spatial representation and facilitates direct comparisons between maps with varying original resolutions. Second, we addressed the inconsistency in reported biomass values due to issues related to units and also forest masks. Since directly converting all maps to a single unit (e.g., Mega gram) might introduce additional errors, we opted to generate consistent total biomass density maps. These maps express biomass per unit area (e.g., Mega gram per hectare) by dividing the original biomass estimates by the corresponding forest

area within each grid cell. This normalization process accounts for discrepancies in the forest masks used by different studies and allows for a more accurate comparison of the actual biomass content within defined forest areas. While the issue of using diverse forest masks cannot be fully resolved, generating biomass density maps offers a robust approach for comparative analysis across these datasets.

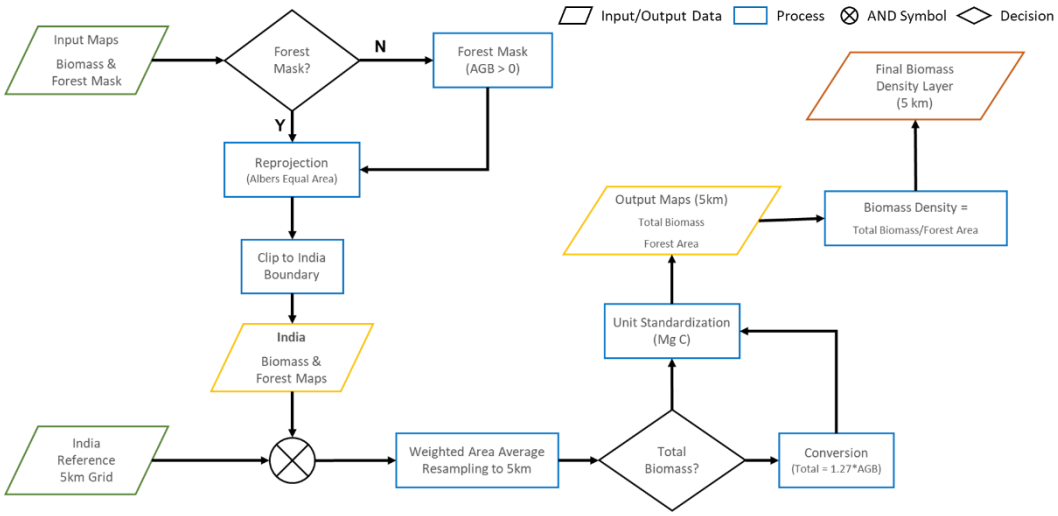


*Figure 2.3. Flowchart depicting the methodology used for standardizing the biomass maps over India*

The key steps involved in the standardization process are as below.

A. **Spatial Reference System**: All maps were first reprojected to the Albers Equal Area projection. This ensures consistent spatial representation and facilitates accurate comparisons, regardless of the original projection used by each study.

B. **Geographic Extent**: Each map was then clipped to the boundary of India, ensuring all analyses focus solely on forest areas within the country.

C. **Forest Mask Consistency**: Ideally, all studies should provide the forest mask to define forest areas. However, if a forest mask was not provided with a particular map, a new mask was generated by assuming pixels with non-zero biomass values represent forest cover.

D. **Reference Grid**: A reference 5-kilometer grid for India was generated to serve as a common spatial framework for all standardized maps and facilitate further analysis.

E. **Grid Resolution and Resampling**: Both the forest mask and biomass maps for all studies were then aggregated to a common reference 5-kilometer grid scale using a weighted area average resampling technique.

This resampling process ensures consistent spatial resolution across all maps and minimizes information loss during the aggregation process.

F. **Biomass Unit Standardization**: Following resampling, the units of biomass were standardized. Studies reporting AGB estimates were converted to total biomass by multiplying the AGB value by a factor of 1.27, which represents a typical conversion factor for forests (IPCC). Also, the biomass from Mg ha-1 is converted into Mg C ha-1 by using a common multiplier of 0.47 across all studies.

G. **Biomass Density Maps**: Finally, at the standardized 5-kilometer grid scale, the total biomass for each grid cell was divided by the corresponding forest area. This generates biomass density maps, which express biomass per unit area (e.g., Mega gram per hectare). This normalization process accounts for discrepancies in the original forest masks and allows for a more accurate comparison of the actual biomass content within defined forest areas.

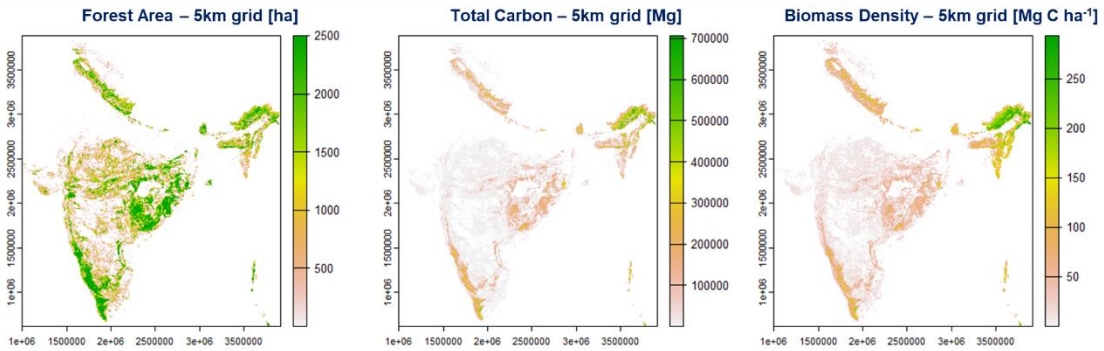


*Figure 2.4. Standardized biomass density estimate for India based on the original available AGB map from Avitabile et al., 2016.*

Figure 9 presents a standardized biomass density map for India. This map originated from a global AGB map created by Avitabile et al., (2016). The original map boasted a finer spatial resolution of 1 kilometer and employed the GLC 2000 forest mask. However, it solely offered estimates of aboveground biomass (AGB). To facilitate its inclusion in the comparative analysis, a two-step standardization process was implemented. First, the map was resampled to a common grid resolution of 5 kilometers using a weighted area average resampling technique, ensuring consistency with other standardized maps. Second, since the original map lacked total biomass estimates, a total biomass density map was generated by multiplying the AGB values by a factor of 1.27, a typical conversion factor for Indian forests. The resulting standardized map depicts total biomass density (e.g., Mega gram Carbon per hectare) across India at a 5-kilometer grid resolution to enable direct comparison with other standardized biomass density maps.

## 2.3 Variability of Biomass Estimates

Following the standardization of biomass maps, the total biomass carbon density for India was estimated. This involved dividing the total biomass carbon derived from the standardized 5-kilometer maps by the total forest area obtained from the same standardized maps. This calculation yielded a national-scale biomass carbon density value for each respective assessment year included in the analysis.

For studies that lacked biomass maps entirely, a two-step approach was employed. First, their total biomass estimates were standardized based on whether they represented AGB or total biomass (using the conversion factor of 1.27 for AGB). Second, this standardized total biomass estimate was then divided by the total forest area reported by the specific study. This comprehensive approach facilitates a direct comparison of biomass carbon density across time, as estimated by various studies, despite potential inconsistencies in their original methodologies and map formats.

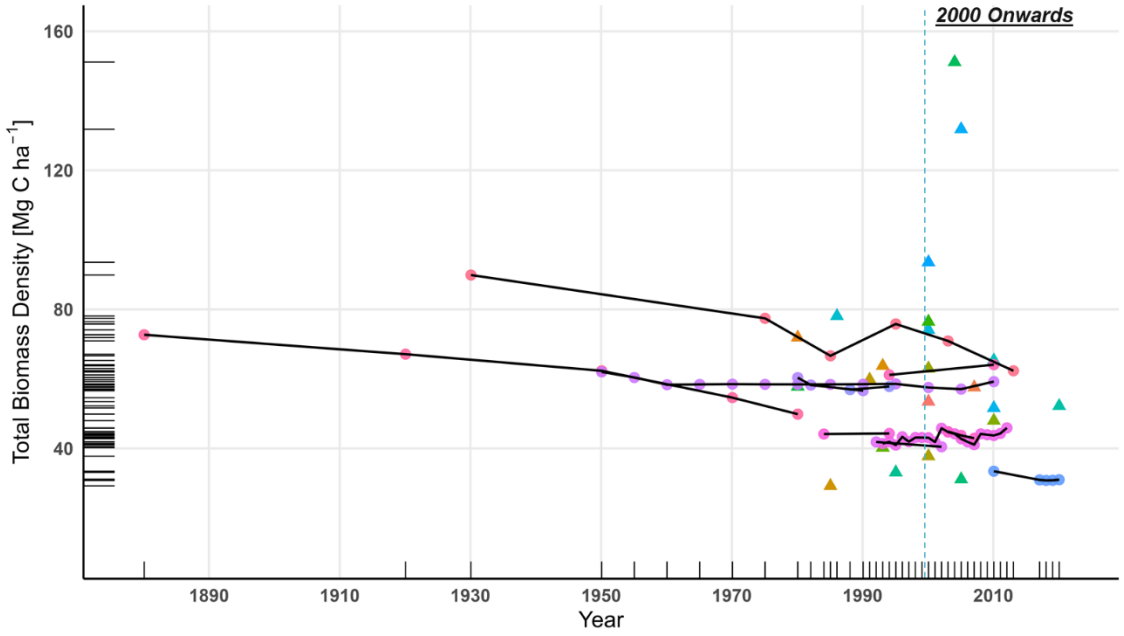


*Figure 2.5. Temporal Variability of Biomass Carbon Density across Indian Forests*

Figure 2.5 depicts the temporal variability of biomass carbon density across Indian forests, as estimated by various studies. Each data point (dot or triangle) represents a single biomass assessment for India conducted by a specific study. The color coding differentiates the data points based on the study that generated the estimate (The color is linked to the methods used as in Figure 2.2). Lines connect data points from the same study when they represent multiple assessments conducted over time. For example, the line connecting the dots/triangles for 1880, 1920, 1950, 1970, and 1980 indicates that these biomass assessments were all derived from a single study by Richards et al., (1994).

The Figure 2.5 visually highlights the substantial variability observed in biomass carbon density estimates across different studies and time periods. This variability underscores the need for standardized methodologies and data collection approaches to facilitate more robust comparisons and a clearer understanding of forest biomass dynamics in India.

### 2.3.1 Inconsistencies of Biomass Estimates over India: 2000 – 2023

Separate analysis was conducted for biomass assessments conducted after the year 2000. This period coincides with the increased use of methodologies that link field biomass or AGB estimates with remote sensing parameters. Studies within this timeframe employed a wide range of Earth Observation (EO) data, including optical, passive and active microwave sensors, and LiDAR measurements. This segregation facilitates visualization of the dispersion patterns observed in AGB estimates across India.

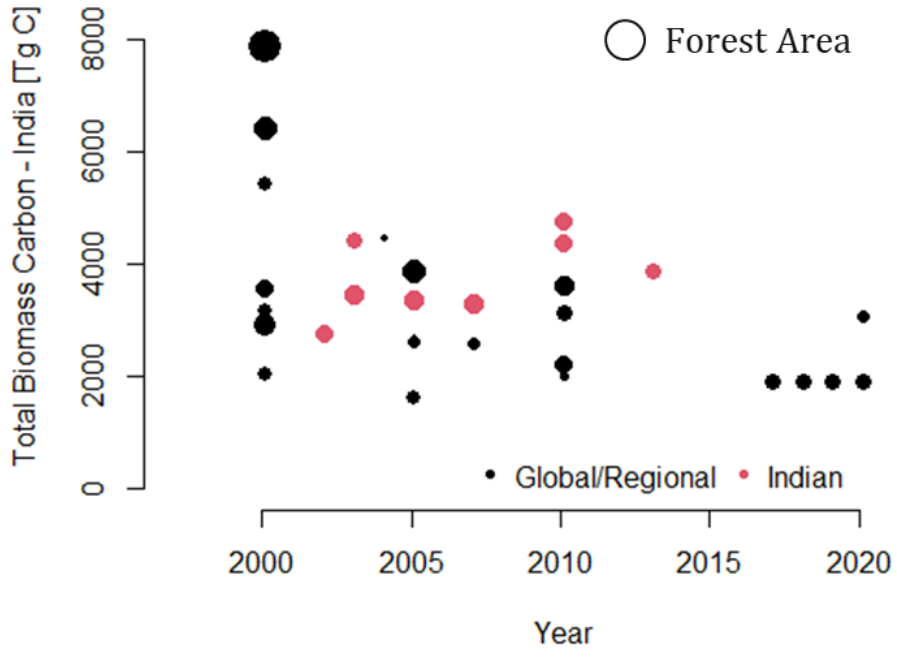


*Figure 2.6. Total Biomass Carbon Estimates over India from 2000 - 2023 as estimated by various studies.*

A total of 30 biomass assessments for India were identified for the period 2000-2023. Significant variations were observed across these assessments in terms of both total biomass carbon and associated forest area (Figure 2.6). Reported values for total biomass carbon ranged from a minimum of 1625 Tg C to a maximum of 7885 Tg C, with a coefficient of variation (CV) of 43.7%. This substantial variability can be largely attributed to disparities in total forest area estimates reported by individual studies. Forest area values ranged from a

minimum of 20.68 million hectares (Mha) to a maximum of 125.18 Mha, with a CV of 31%.

Furthermore, analysis of standardized total biomass density revealed similar patterns of variation (Figure 2.7). Values ranged from a minimum of 30 Mg C $ha^{-1}$ to a maximum of 151 Mg $ha^{-1}$, with a CV of 48%. These findings highlight the ongoing challenges associated with achieving consistent and comparable estimates of forest biomass across different studies. The observed variations emphasize the importance of promoting standardized methodologies and data collection approaches for future assessments.

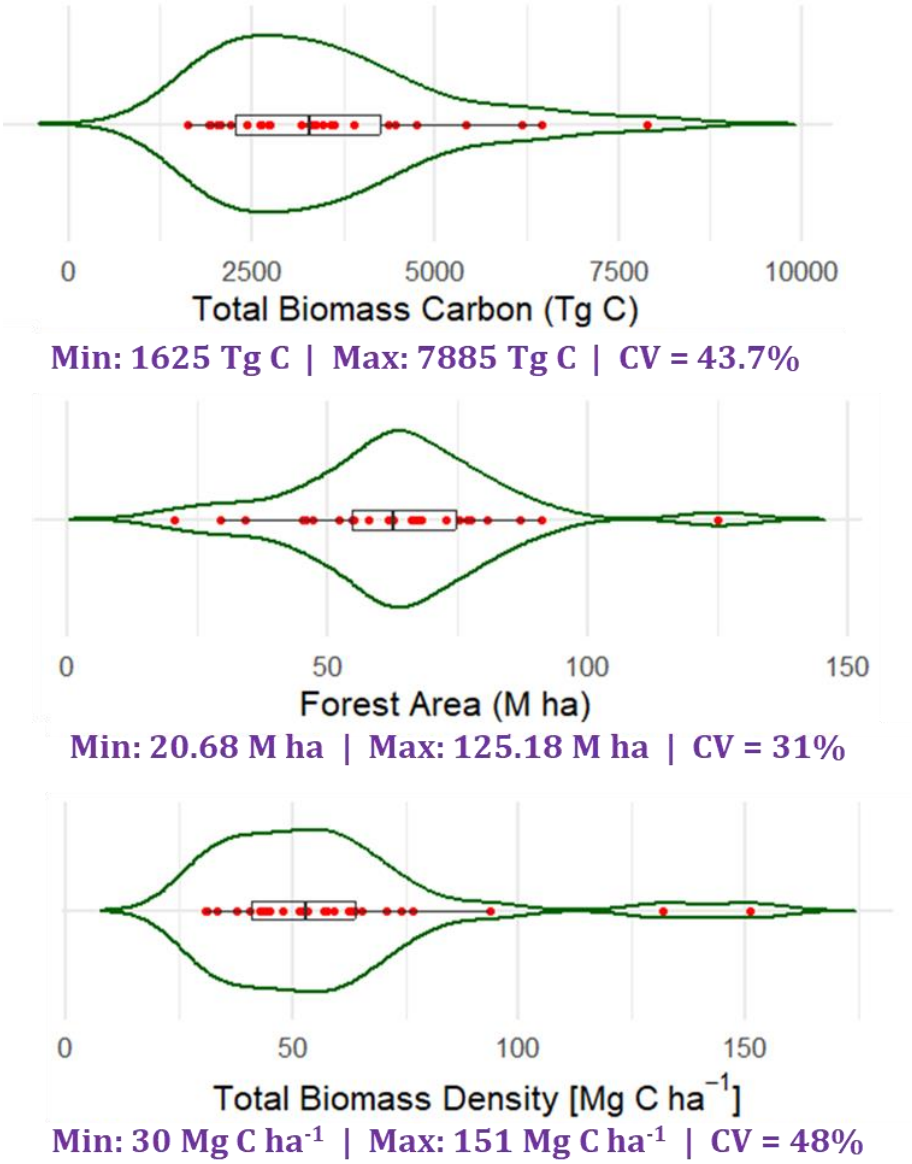


*Figure 2.7. Violin box-plots showing the estimates of Total Biomass Carbon, Forest Area and Total Biomass Density over Indian forests from 2000 to 2023*

### 2.3.2 Snapshot of Spatial Variability from Indian Biomass Maps: 2000 – 2023

Among the 30 biomass assessments available for the 2000-2023 period, we were able to acquire standardized maps for 19 (18 global and 1 Indian). These standardized AGB maps were then utilized to calculate the coefficient of variation (CV) for biomass carbon density at a 5-kilometer grid scale.

The analysis of spatial variability within these standardized biomass estimates reveals even more pronounced disparities. The CV across the 5-kilometer grid cells ranged from 0% to a concerning 250%, with an average CV of 94% and a median CV of 91% (Figure 13). These wide variations are particularly significant in regions with low biomass, where accurate estimation is especially critical.

These findings underscore the critical need for standardized AGB estimates, not only at the regional level but also across spatial scales. Establishing and adhering to common protocols and methodologies are essential for building a robust and consistent understanding of India's forest biomass and its temporal changes. This standardized approach is crucial for effective monitoring of forest health and dynamics.

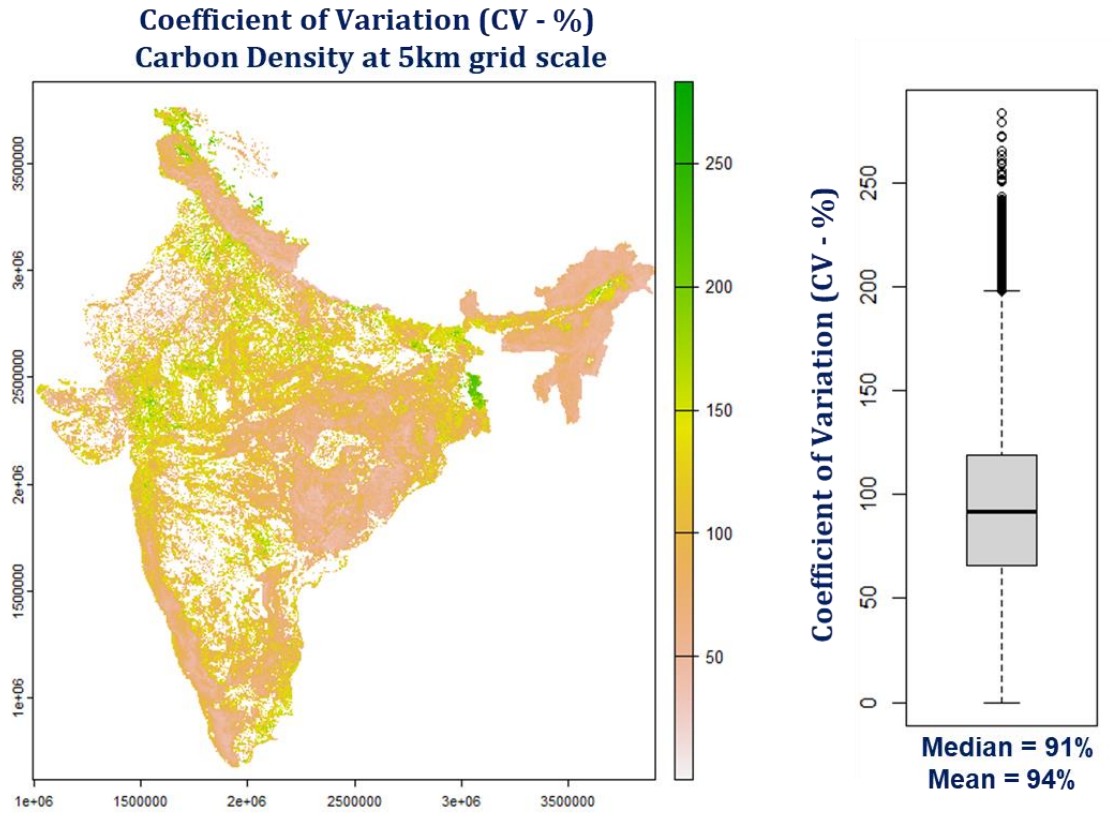


*Figure 2.8. The coefficient of variation (CV) in biomass carbon density at 5km grid scale, calculated across 19 Maps generated by different studies between 2000 and 2023. Higher CV values indicate greater variability (Units: %)*

## 2.4 Challenges in Regional Forest Biomass Mapping

Accurately mapping forest biomass or above-ground biomass (AGB) across diverse regions in India remains a complex undertaking due to inherent limitations and uncertainties associated with current methodologies. The uncertainties in EO-based biomass mapping stems from several contributing factors (Figure 2.9) and differ for different methods and assumptions and hence causing a substantial variability in the overall AGB estimates over India.


*Figure 2.9. Steps involved in Hierarchical AGB mapping from ground to regional scale.*

The Phase-1 typically involves estimating AGB using traditional volume equations derived from tree height and diameter measurements. While seemingly straightforward, this method introduces errors. These equations are often tailored to estimate merchantable volume, potentially neglecting variations in wood density across different tree species and parts. Furthermore, the absence of a comprehensive wood density database for various tree species and regions introduces additional uncertainties. Scaling ground-based measurements conducted on small plots to the larger resolutions of remote sensing pixels inherently introduces spatial aggregation errors. Compounding these issues are potential geolocation errors associated with the ground measurements themselves.

Remote sensing data, while a valuable tool, also presents its own limitations. Optical and microwave sensors, have varying sensitivities that can hinder accurate biomass estimation. For instance, optical data can saturate in regions with high biomass, rendering it less effective. The representativeness of ground samples used for calibration purposes becomes questionable when extrapolating them across vast and ecologically diverse landscapes. These samples might not fully capture the true regional variability in biomass distribution.

Data availability and consistency pose further challenges. Inconsistencies in forest area definitions and spatial layers across different studies lead to discrepancies in the spatial extent of the resulting biomass estimates. The absence of uncertainty estimations in many existing AGB assessments makes it difficult to assess the reliability and confidence intervals of the generated biomass maps. Furthermore, while numerous regional biomass assessments exist, many lack spatially explicit maps, hindering in-depth analysis at the regional or grid level, and often provide only national-level estimates.

Finally, characterizing changes in biomass over time using Earth Observation (EO) data presents a unique challenge. This approach requires highly accurate individual biomass estimates for each time period being analyzed. However, for meaningful change detection, the actual change in biomass needs to be significantly larger than the inherent uncertainty associated with the individual estimates.

Addressing these multifaceted challenges is paramount for generating reliable and representative maps of regional AGB in India. Overcoming these hurdles will ultimately enable effective monitoring and management of this vital carbon resource, informing critical decision-making related to forest conservation and climate change mitigation strategies.

## 2.5 Towards Improved Forest Biomass Estimates

The ever-increasing importance of accurate forest biomass estimates necessitates the development of more robust methodologies. However, uncertainties introduced at various stages of the hierarchical biomass mapping process can propagate errors to the final results (Figure 2.10). Minimizing these errors is essential for generating reliable estimates.

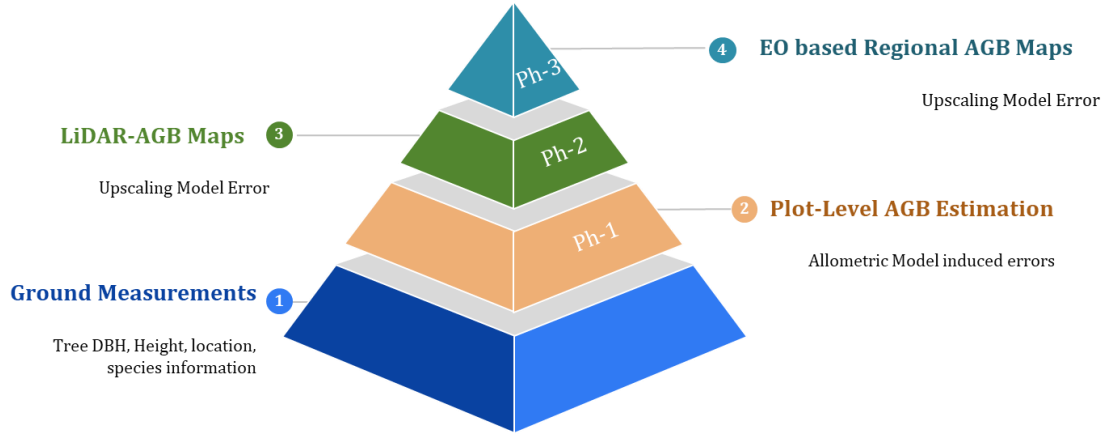


*Figure 2.10. Phased approach transitioning from ground measurements to EO based regional biomass maps*

The initial phase focuses on plot-level above-ground biomass (AGB) estimation (Figure 2.10). Here, efforts should prioritize improving the accuracy of allometric models and potentially increasing plot sizes. Larger plots can address issues like sample representativeness, geolocation errors, and difficulties in linking ground data to EO data. High-quality ground data from larger plots can significantly reduce uncertainties at the outset.

However, challenges persist when linking ground measurements to EO data across vast regions. These include limitations in total sample size, sampling intensity, and potential signal saturation in EO data itself. Airborne LiDAR (Light

Detection and Ranging) emerges as a potential solution. LiDAR-derived AGB maps can act as a powerful intermediary, providing high-quality training data for calibrating EO data. This approach can facilitate the generation of reliable and defensible AGB estimates at national or even global scales.

A critical element is establishing a network of "super-sites" that can generate extensive forest inventory data over large areas. Ideally, these sites would undergo multiple censuses over extended periods and be combined with long-term airborne LiDAR data collection (Figure 1.5). This comprehensive approach would provide a robust foundation for reliable AGB mapping and enable long-term monitoring of global change impacts on forest ecosystems.

By implementing these solutions at each stage of the hierarchical biomass mapping process, we can significantly reduce uncertainties and generate more accurate and reliable regional biomass maps. This will ultimately enhance our understanding of forest biomass distribution and dynamics, supporting informed decision-making related to forest conservation and climate change mitigation strategies.

This thesis explores the possibility of overcoming these limitations and developing a more robust methodology for generating reliable regional forest biomass maps in India. By focusing on improvements at various stages of the hierarchical biomass mapping process, the study aims to minimize uncertainties and generate more accurate estimates. The proposed solutions encompass refinements in allometric models used for plot-level biomass estimation, potentially increasing plot sizes, and leveraging airborne LiDAR data as a bridge between ground measurements and Earth Observation (EO) data. Through these advancements, this thesis strives to contribute to a more comprehensive understanding of forest biomass distribution and dynamics in India, ultimately supporting informed decision-making for forest conservation and climate change mitigation strategies.

# Chapter 3

# Optimizing Plot Design for Accurate Plot-level AGB estimates

Accurate estimation of aboveground biomass (AGB) at the plot level is critical for reducing uncertainties in large-scale AGB quantification. This, in turn, is essential for achieving the goals of international initiatives like REDD+ (Reducing Emissions from Deforestation and Forest Degradation) and INDCs (Intended Nationally Determined Contributions) that address climate change mitigation. While remote sensing plays a vital role in spatial AGB mapping, the accuracy of these estimates can be significantly influenced by several factors related to the design of field plots used for data collection. This chapter specifically investigates the impact of field plot size and shape on the accuracy of AGB estimates. We analyze data collected from a network of spatially explicit field plots established across three distinct tropical forest sites in India. Here, we use data from 1-hectare plot network to evaluate the influence of these plot design characteristics.

## 3.1 Introduction

Forests represent the largest aboveground carbon stock in the terrestrial biosphere, and strongly influence global $CO_2$ exchange through forest degradation and re-growth (Pan et al., 2011). Consequently, there has been an increasing interest in accurate assessment of this carbon in recent times (Harris et al., 2012; Saatchi et al., 2011).

Advancements in remote sensing and GIS-based AGB mapping hold promise for reducing uncertainties in forest AGB estimates, crucial for effective emissions mitigation strategies. However, the accuracy of these broad-scale estimates heavily relies on the quality of reference field inventory data (Mitchard et al., 2011; Réjou-Méchain et al., 2019). Since remote sensing variables are calibrated against field data, any uncertainties in field biomass measurements can propagate to large-scale estimates (Pelletier et al., 2011; Réjou-Méchain et al., 2014). While previous studies have explored uncertainties at national and global scales (Baccini et al., 2008; Pan et al., 2011; Rajashekar et al., 2018; Saatchi et al., 2011; Santoro et al., 2021), uncertainties related to field inventory data in tropical forests remain relatively understudied.

Advancements in remote sensing and GIS-based AGB mapping hold promise for reducing uncertainties in forest AGB estimates, crucial for effective emissions mitigation strategies. However, the accuracy of these broad-scale estimates heavily relies on the quality of reference field inventory data (Mitchard et al., 2011; Réjou-Méchain et al., 2019). Since remote sensing variables are calibrated against field data, any uncertainties in field biomass measurements can propagate to large-scale estimates (Pelletier et al., 2011; Réjou-Méchain et al., 2014). While previous studies have explored uncertainties at national and global scales (Baccini et al., 2008; Pan et al., 2011; Rajashekar et al., 2018; Saatchi et al., 2011; Santoro et al., 2021), uncertainties related to field inventory data in tropical forests remain relatively unexplored.

As the estimation of tree-level AGB, and subsequently plot-level AGB (sum of AGB of all trees within the plot), is derived from tree inventories rather than direct measurement, achieving accurate AGB estimates hinges on various factors. These factors encompass determining the optimal reference plot size, obtaining precise measurements of tree variables like diameter at breast height (DBH) and height, assigning appropriate wood density values for each tree species, and selecting suitable biomass allometric and height–diameter (H–D) models, dependent upon the availability of tree height measurements. Notably, one of the most significant factors influencing accuracy is plot size (Chave et al., 2004; Gobakken and Næsset, 2009).

The field plot size significantly impacts the accuracy of AGB estimates due to boundary effects, where smaller plots expose a large portion of trees to the plot boundary (Mayamanikandan et al., 2020). Conversely, larger plot sizes mitigate these boundary effects by reducing the perimeter-to-area ratio, resulting in more precise AGB estimates (Asner et al., 2012; Chave et al., 2004). Furthermore, larger plots help mitigate errors arising from field plots to satellite image co-registration problems (Frazer et al., 2011). Despite these advantages, large plots present certain limitations because they are cost and time intensive and more complicated to measure. The cost of forest field inventories are predominantly influenced by sample size and plot size. Hence, it becomes crucial to optimize the sample size, and plot size to achieve a balance that minimizes the inventory cost while maintaining an acceptable level of estimation accuracy for forest inventory attributes.

Réjou-Méchain et al., (2014) highlight significant local spatial variability in AGB density for standard plot sizes, with average deviations of 46.3 % for replicate 0.1 ha subplots within a single large plot, and 16.6% for 1 ha subplots. They recommend using field data from large plots for calibration of remote sensing products. Similarly, a study conducted in productive forests in Norway found that increasing plot size from 200-250 $m^2$ to 1,000-4,000 $m^2$ resulted in a notable reduction in standard deviation from 20-25% to 10-15% (Næsset, 2002).

Similarly, Chan et al. (2021) and Navarro-Cerrillo et al. (2017) uses circular plot of varying radius to predict the AGB using airborne LiDAR based canopy height measurement and reported higher accuracies with larger plot radius, suggesting better suitability of the large plots for AGB estimation. Zolkos et al., (2013) conducted a comprehensive review of studies on AGB estimation by various remote sensing platforms and sensors and observed a strong correlation between model errors and plot size. Enlarging plot size resulted in rapid decrease in model estimation error.

Currently over Indian forests, Forest Survey of India carries out field Inventories using nested sampling design with sample area of 0.08 ha (< 0.1ha) for national AGB estimates (FSI, 2021). Further, remote sensing based regional and national AGB estimates using national scale field inventories are mainly based on the conventional 0.1 ha size plot (Fararoda et al., 2021; Rajashekar et al., 2018) which produces larger estimation error due to smaller plots sizes (Mayamanikandan et al., 2020). Hence there is a justified need to optimize the field inventory plot size for improved AGB estimates which are consistent with global efforts (Duncanson et al., 2021).

In this context, this study focuses on investigating how plot size and shape influence plot-level AGB estimation across various forest types in India. Additionally, we aim to identify the optimal plot size for accurate AGB estimates within Indian forests.

## 3.2 Materials and Methods

### 3.2.1 Study Area

This study utilizes field inventory measurements collected across three geographically distinct sites within the Tropical Indian Region. These sites encompass a diverse range of forest types, as shown in Figure 3.1. Field campaigns conducted between 2015 and 2020 captured data from various forest ecosystems.

- **Betul (BTL), Madhya Pradesh**: represents a dry deciduous forest system dominated by *Tectona grandis*, with additional co-dominant species like *Diospyros melanoxylon*, *Lagerstroemia parviflora*, and *Madhuca longifolia*
- **Achanakmar (ACK), Chhattisgarh**: features a moist deciduous forest is characterized by the dominance of *Shorea Robusta* alongside *Terminalia alata*, *Anogeissus latifolia*, and *Ougeinia dalbergioides*.
- **Yellapur (YLP), Karnataka**: encompasses a transition zone from dry deciduous to moist deciduous and wet evergreen forests. . The dominant tree species in YLP's deciduous forests include *Tectona grandis*, *Terminalia paniculata*, and *Xylia xylocarpa*, while the evergreen forests are

characterized by *Diospyros buxifolia, Syzygium gardneri,* and *Knema attenuata*.

A comprehensive description of these study sites, including details on their geographical location, climatic conditions, and specific tree measurements, can be found in Table 3.1.

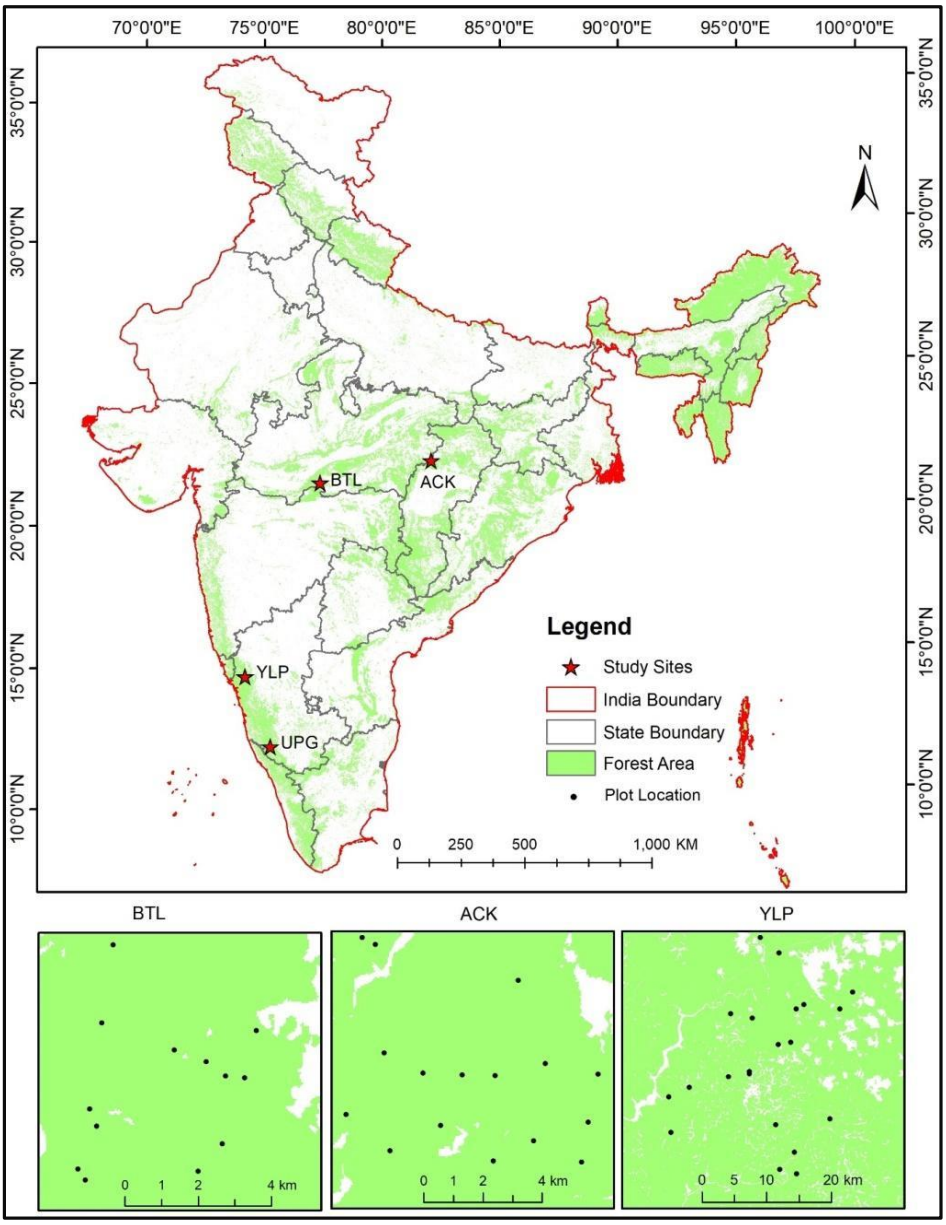


*Figure 3.1. Study Area Map showing the distinct study sites and spatial distribution of the 1-ha size field plots in the Betul (BTL), Achanakmar (ACK) and Yellapur (YLP) site is also shown.*

A total of 51, 1-ha (100m x 100m) plots were established (BTL – 13, ACK – 16 and YLP – 22) across these diverse sites, ensuring accurate geolocation and north orientation using Differential Global Positioning System (DGPS) and Electronic Total Station (ETS). Each plot was further subdivided into 25 quadrats (0.04 ha each) using ETS for conducting inventory measurements. Within each

quadrat, detailed information was collected for trees exceeding a diameter at breast height (DBH) of 9.55cm (circumference > 30cm). This data included species identification, DBH, and precise relative positioning within the quadrat (accuracy ±0.5m achieved using ropes and measuring tapes). Additionally, tree heights were measured for a subset of trees within each plot using a NIKON Forestry Pro Laser range finder.

*Table 3.1. Sample plots statistic and climate parameters across study sites. $N_{xxx}$ indicates total number of plots, DBH measurements, and Height measurements. Mean Annual Temperature and Mean Annual Precipitation are computed using WorldClim Version 2.1 data.*

| **Site** | **$N_{Plots}$ (1ha)** | **$N_{DBH}$** | **$N_{Ht}$** | **$H_{range}$** | **$Density_{range}$ (Stem/ha)** | **MAT ($^0$C)** | **MAP (mm)** |
|---|---|---|---|---|---|---|---|
| BTL | 13 | 3972 | 642 | 7-29 | 150 - 437 | 25.6 ± 0.3 | 1266 ± 27 |
| ACK | 16 | 5750 | 962 | 4.4-34.5 | 186 - 509 | 23.3 ± 0.7 | 1328 ± 42 |
| YLP | 22 | 8519 | 4847 | 1.7-45 | 140 - 749 | 24.4 ± 0.5 | 2383 ± 421 |

### 3.2.2 Impact of plot-size on plot-level AGB estimation

To comprehensively assess how plot size and shape influence aboveground biomass (AGB) estimates, we generated various synthetic plots with different dimensions based on actual tree locations (Figure 3.2). These plots ranged from small squares (10m x 10m, 0.01 ha) to larger squares (90m x 90m, 0.81 ha). The additional plots upto 6.76 ha were also generated for BTL based on the large 32-ha continuous plot data. Additionally, we generated circular plots with seven different sizes (0.08 ha to 0.5 ha) to isolate the impact of shape while maintaining constant area (Figure 3.2).

To solely evaluate the uncertainty caused by plot size, we controlled for errors arising from the allometric model by using the equation by Chave et al., (2014) for tree-level AGB calculations. For each synthetic plot, we calculated the total AGB by summing the AGB of all trees within its boundaries. We then scaled the AGB values to represent a 1-hectare area and calculated the relative error (RE) as an indicator of estimation accuracy (Figure 3.3). To ensure reliable estimates, we randomly generated 100 synthetic plots for each size class.

$$Relative\ Error\ (\%) = \frac{Scaled\ AGB - RefAGB}{Ref\ AGB} * 100$$

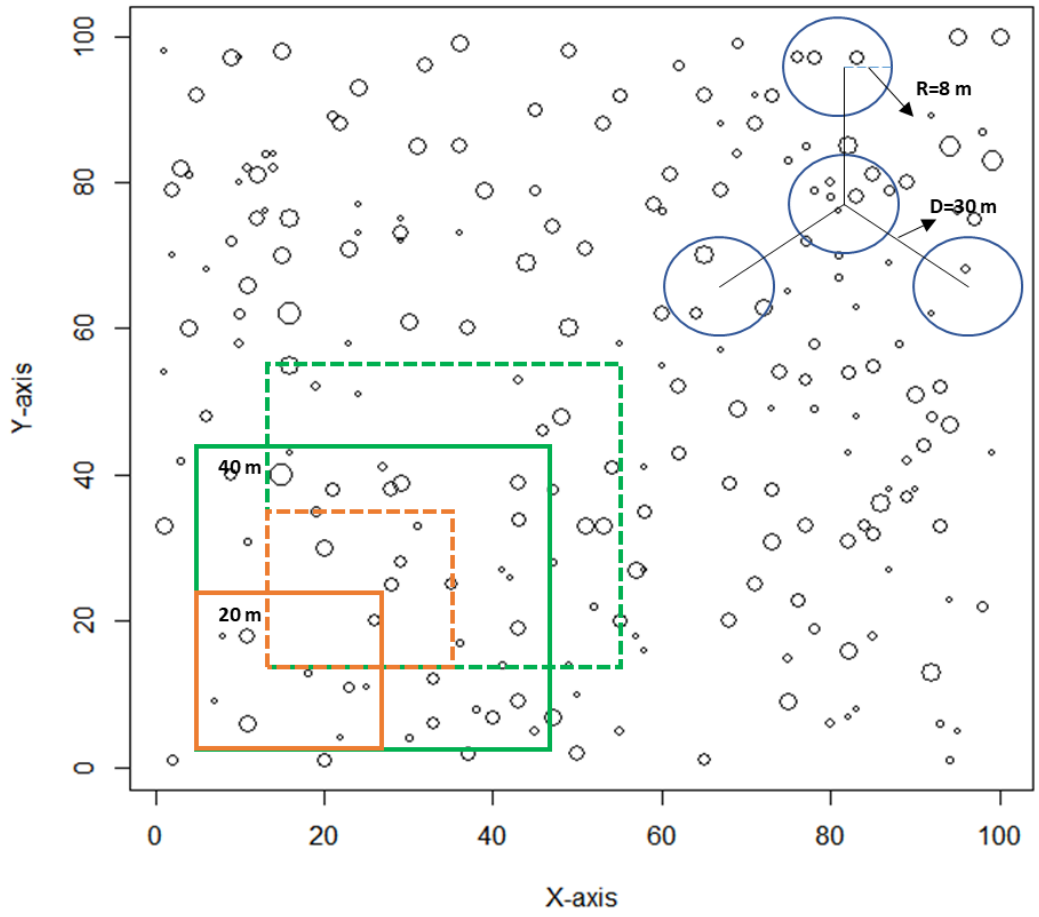


*Figure 3.2. Schematic illustration of synthetic plot generation using mapped tree locations within a 1-ha reference plot. Circle centers represent individual tree positions, and their radii are proportional to the corresponding tree diameters at breast height (DBH).*

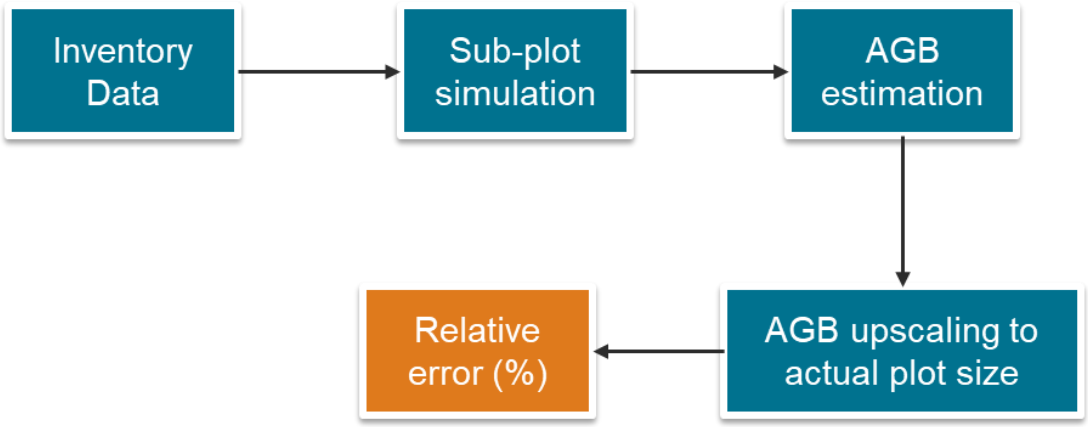


*Figure 3.3. Flow chart of methodology used for estimating impact of plot-size on plot-level AGB estimation.*

## 3.3 Results

### 3.3.1 Betul, Madhya Pradesh

We began the analysis by focusing on the Betul site in Madhya Pradesh because it has a unique advantage - a very large, continuous field plot spanning 32 hectares. This extensive plot allowed us to generate synthetic plots of various sizes, ranging from a small 0.04 hectares to a larger 6.76 hectares (Figure 3.4). Notably, plots exceeding 0.64 hectares were derived from this large 32-hectare plot for practical reasons. The analysis revealed significant variations in relative error (RE) across different plot sizes, particularly for smaller plots, based on the 100 simulations we conducted. Figure 3.4 illustrates this variability in RE for each simulated plot size.

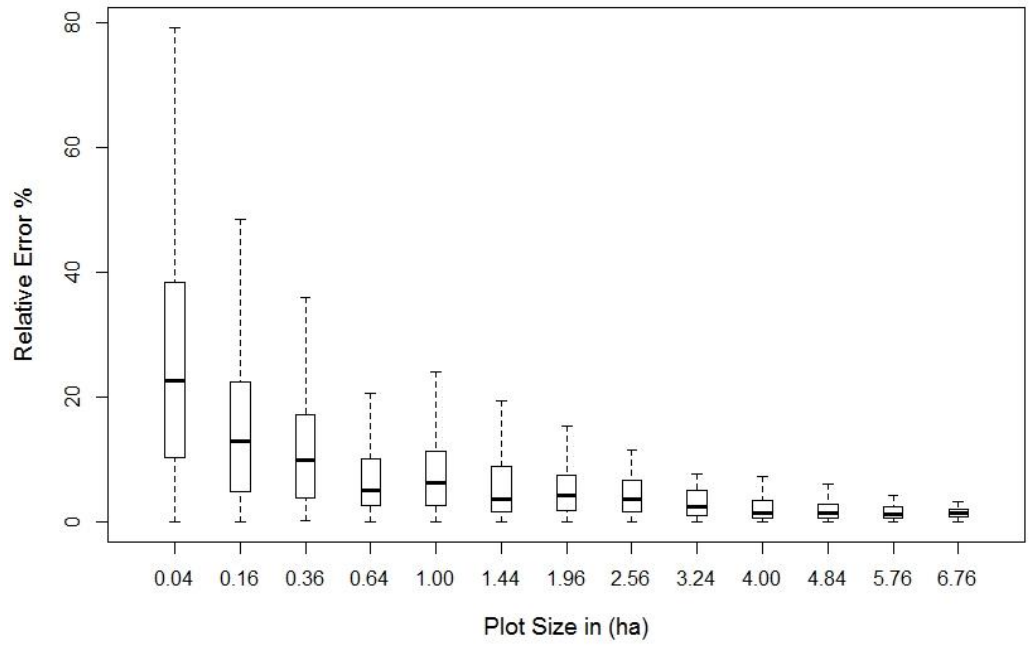


*Figure 3.4. Variability of Relative Error (%) with increase in plot size*

The analysis revealed a clear correlation between plot size and relative error (RE) in AGB estimates. As plot size increased, the RE demonstrably decreased. The average RE started at approximately 24% for plots as small as 0.04 hectares and steadily declined to around 1.5% for the largest plots of 6.76 hectares (Figure 3). This trend is particularly pronounced for plots smaller than 0.5 hectares, where the overall variability of RE was significantly higher.

The interquartile range (IQR), which quantifies the dispersion of the data, further corroborated this diminishing uncertainty with increasing plot size. The IQR decreased from a maximum of 25% for the smallest plots to around 1% for the largest ones. Importantly, both the mean RE and IQR fell below 10% for plots exceeding 0.5 hectares. These findings align with the notion that larger plot sizes

(up to 0.64 hectares or more) can effectively mitigate uncertainty in AGB estimates by incorporating a more representative sample of the inherent heterogeneity within the forest stand (Chave et al., 2004).

To further explore the influence of plot shape on AGB estimation accuracy, we conducted an analysis while maintaining a constant plot area. Synthetic circular plots of seven distinct sizes were generated, ranging from 0.08 hectares to 0.5 hectares. Additionally, corresponding square plots with equivalent areas were created (Figure 3.5). The boxplot in Figure 3.5 depicts the RE for both circular and square plots across various plot sizes, based on 100 simulations.

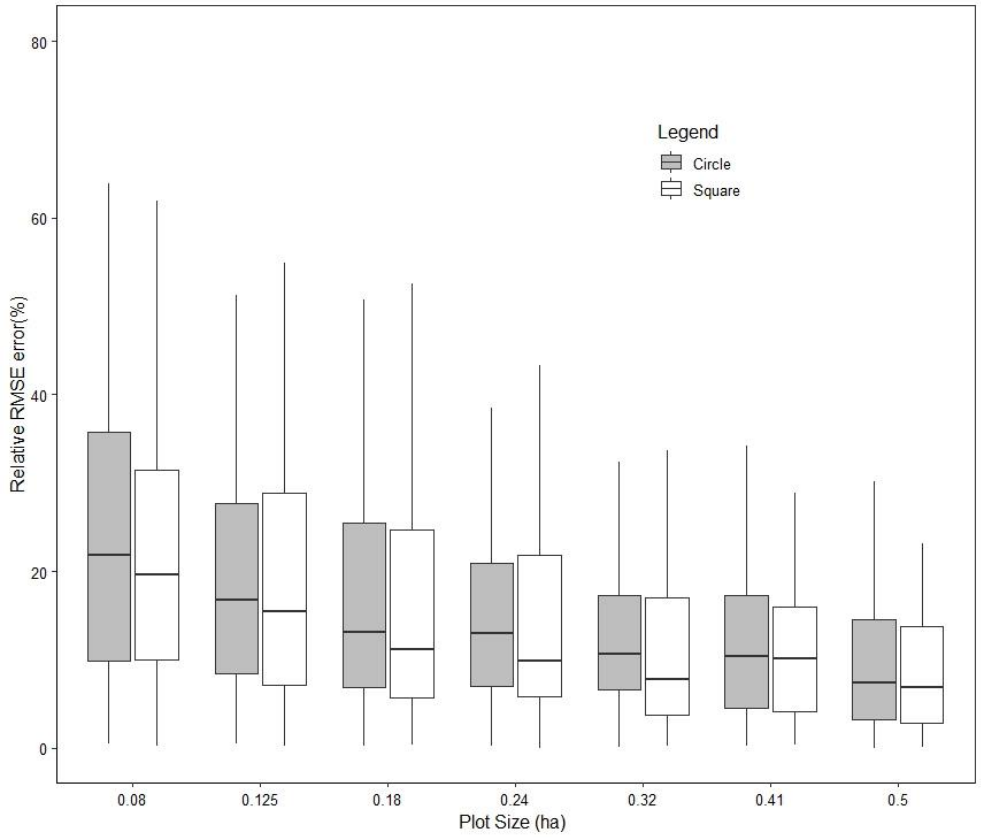


*Figure 3.5. Relative error with respect to plot shape keeping the plot area as constant for Betul (BTL).*

The results interestingly indicate that irrespective of plot shape (circular or square), the average RE exhibited a consistent decreasing trend with increasing plot size. The RE dropped from approximately 20% for the smallest plots (0.08 hectares) to about 8% for the largest (0.5 hectares). However, it is noteworthy that for a given area, a circular plot necessitates a smaller perimeter compared to a square plot. This suggests that when aiming for similar plot areas, circular plots might introduce marginally higher uncertainty.

Our analysis demonstrates a clear advantage of increasing plot size for mitigating uncertainty in AGB estimates. However, practical considerations necessitate a balance between accuracy and feasibility. While plots exceeding 0.64

hectares effectively capture stand heterogeneity, logistical constraints like challenging terrain and accessibility often render larger plots (>2 ha) impractical. Additionally, the diminishing reduction in error (7.2% at 0.64 ha to 4.2% at 2 ha) suggests a limited benefit for significantly larger plots. Therefore, considering the trade-off between accuracy and implementation challenges, plots exceeding 0.64 hectares are recommended for optimal AGB estimation across diverse Indian forest ecosystems.

### 3.3.2 Across All Study Sites

To assess the influence of plot size across diverse forest types, synthetic plots ranging from 0.04 ha (20 x 20 m) to 0.81 ha (90 x 90 m) were generated using data from all three study sites (BTL, ACK, and YLP) (Figure 3.1). Figure 3.6 depicts the trend of relative error (RE) in AGB estimates for each site as plot size increases.

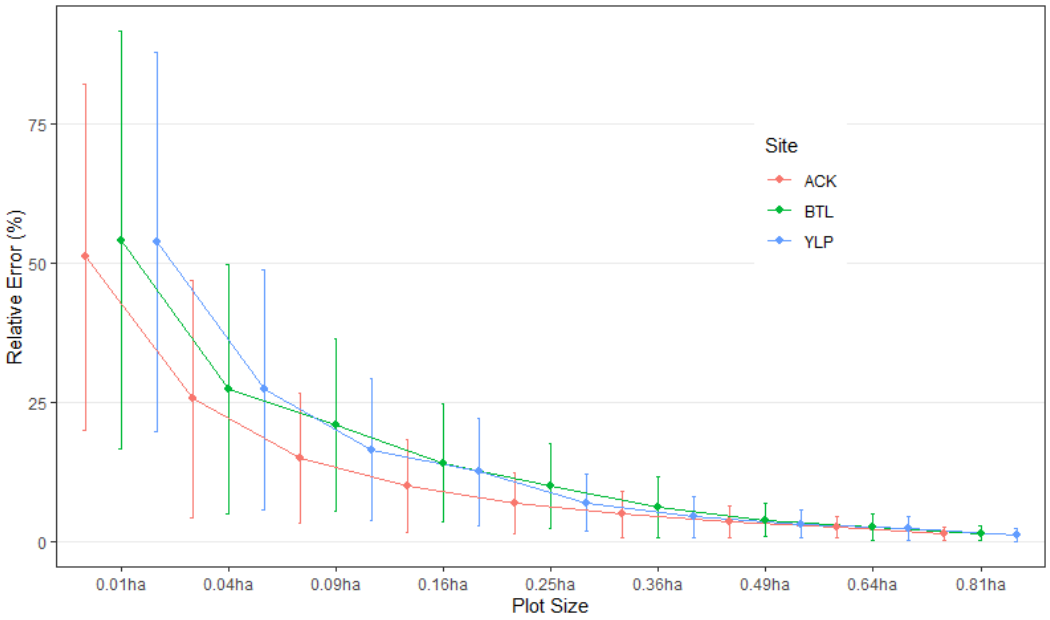


*Figure 3.6. Field measurement error associated with different plot sizes. The straight lines at each point indicate the relative error variability across 100 simulations per plot size.*

As evident in Figure 3.6, RE consistently decreases across all sites with increasing plot size. At the smallest plot size (0.01 ha, 10 x 10 m), RE is around 50%. This value drops significantly to approximately 10% at a plot size of 0.25 ha (50 x 50 m). A further reduction in RE is observed with an increase to 0.49 ha (70 x 70 m), bringing it down to roughly 5%. Notably, expanding plot size beyond this threshold has minimal impact on RE (Figure 3.6).Based on this analysis, we recommend a minimum plot size of 70 x 70 m (0.49 ha) as the optimal reference for AGB estimation in Indian tropical forests. This plot size effectively balances accuracy with feasibility, ensuring reliable estimates while optimizing resource allocation for upscaling models.

As previously discussed (refer to discussion on plot shape in the previous section), plot shape has minimal influence on RE when maintaining a constant area. While both circular and square plots exhibit a decreasing RE trend with increasing size (Figure 3.5), circular plots with a smaller perimeter compared to squares of the same area may introduce slightly higher uncertainty. Therefore, plots exceeding 0.5 ha are recommended for reliable AGB upscaling with minimal uncertainty, regardless of their specific shape (circular or square).

## 3.4 Discussion

This study clearly demonstrates the critical role of plot size in achieving accurate aboveground biomass (AGB) estimates in tropical forests. This accuracy is crucial not only for cost-effective field campaigns but also for effectively linking field measurements with remotely sensed data. Such integration benefits REDD+ initiatives and forest carbon cycle studies.

The current research investigated the impact of plot size and shape on representing field AGB. We generated a wide range of synthetic plots (varying in size and shape) from a network of large plots across central Indian dry deciduous forests. Our findings confirm the positive effect of increasing plot size on field AGB estimates. The mean relative error (RE) decreased significantly with increasing area, highlighting the importance of capturing spatial heterogeneity. Smaller plots (0.04 ha) often miss this variation, leading to over- or underestimation of AGB. In contrast, larger plots (6.76 ha) capture more diverse tree populations and provide a more accurate representation of mean AGB values.

Interestingly, despite the theoretical advantage of circular plots (minimized boundary effects due to lower perimeter-to-area ratio), our study found minimal influence of plot shape (circular vs. square) on field AGB estimates for tropical deciduous forests, as long as the plot area remains constant. This finding emphasizes the primary importance of plot size for accurate AGB estimation in this specific forest type.

While accuracy improves with increasing plot size, there's a diminishing return on investment beyond a certain size threshold. The RE decreased significantly from 0.04 ha (24%) to 0.64 ha (7%), but the decrease was marginal for larger plots (0.64 ha to 1.96 ha and 1.96 ha to 6.76 ha). Considering the rising costs of fieldwork with increase in plot size, a minimum size of 0.64 ha to 1 ha provides a good balance between estimation accuracy and plot establishment cost for reliable AGB estimation in tropical dry deciduous forests of India.

Another key element concerns the influence of plot positioning errors (due to GPS inaccuracies) on AGB estimates derived from remote sensing data. Smaller plots exhibited a higher bias in RMSE% compared to larger plots. This is

likely because larger plots have a greater overlap with remote sensing data, mitigating potential errors associated with imprecise GPS locations (Frazer et al., 2011; Gonçalves et al., 2017; Zhang et al., 2013).

## 3.5 Summary

This study investigated the impact of plot size, shape, and location errors on AGB estimation in a tropical dry deciduous forest, considering both field measurements and remote sensing predictions. Our findings demonstrate the clear advantage of larger plots for achieving accurate and reliable AGB estimates. Larger plots not only capture greater spatial heterogeneity, leading to lower coefficient of variation (CV), but also could significantly improve the fit of models with remote sensing data by minimizing edge effects.

However, a practical balance is crucial. While plots exceeding 1 ha offer marginal gains in precision compared to the additional cost of establishment, a size range of 0.64 ha to 1 ha provides an optimal balance between accuracy and feasibility for this specific study area. This plot size range ensures precise spatial AGB modeling while remaining cost-effective.

Our emphasis on large plots aligns with the findings presented in the previous section, which highlighted the importance of capturing spatial variability for accurate AGB estimation across diverse Indian forest types. Therefore, establishing calibration and validation reference plots exceeding 0.5 ha is recommended for future initiatives focused on large-scale AGB mapping (viz., GEDI, NISAR, BIOMASS etc.,) and the development of globally consistent AGB reference maps.

A critical element in minimizing uncertainty associated with plot-level AGB estimation is the selection of an appropriate allometric model. The next chapter delves into this aspect, evaluating the impact of allometric models on uncertainty and exploring methods to reduce it. Additionally, the chapter examines the errors introduced by allometric models at larger plot sizes, a crucial step in generating high-quality ground data for calibrating and validating Earth Observation (EO) data.

# Chapter 4

# Developing non-destructive volume equations using Terrestrial LiDAR

Estimating Above-Ground Biomass (AGB) across large regions requires ground reference plots for calibrating and validating remote sensing products. Traditionally, reference plot AGB relies on allometric models that estimate tree-level biomass based on diameter and height measurements along with species information. However, recent advancements in Terrestrial Laser Scanning (TLS) offer exciting possibilities for non-destructive tree volume extraction from detailed 3D canopy structure data. In this chapter, we leverage 3D point clouds captured by TLS to develop a framework for constructing local allometric equations tailored to tropical dry deciduous forests in Betul, Madhya Pradesh, India. This framework utilizes TLS data to model individual trees and extract their volumes, ultimately leading to the development of more accurate and localized allometric equations for improved AGB estimations.

## 4.1 Introduction

Forests are crucial to the global carbon cycle as they hold approximately 45% of the terrestrial carbon in live above-ground biomass (Bonan, 2008) and offsets nearly 30% of anthropogenic $CO_2$ emissions through carbon sequestration (Friedlingstein et al., 2020). Nevertheless, quantitative information on the spatial distribution and absolute magnitude of above-ground biomass of forests (hereinafter used interchangeably as "AGB" or "biomass") remain substantially uncertain, in part due to the measurement errors in tree-level biomass estimations (Mitchard et al., 2014a).

Accurately estimating tree-level above-ground biomass (AGB) traditionally involved destructive sampling - felling and weighing entire trees (Brown, 1997). While this method yields the most accurate measurements, it is prohibitively costly, time-consuming, and impractical for large-scale assessments. Consequently, most contemporary AGB estimates rely on indirect methods, primarily allometric equations (Réjou-Méchain et al., 2019). These equations calculate tree volume based on readily measurable parameters like diameter at breast height (DBH) and tree height (Calders et al., 2014; Kearsley et al., 2017).

However, these equations often rely on extensive destructive sampling, propagating potential errors and biases through calculations for entire stands and large-scale biomass mapping (Chen et al., 2015). Therefore, selecting an appropriate allometric model becomes crucial for minimizing these errors in spatial AGB estimates (Picard et al., 2015).

Although generalized allometric models for different forests and tree species exists (Basuki et al., 2009; Chave et al., 2014), the use of these equations can potentially lead to bias in AGB estimation for a particular species since biomass is also dependent on several other underlying factors such as topography, environmental conditions, stand age, species composition, and disturbance history. Therefore, using species- and site-specific allometric models is considered substantial improvement in reducing the uncertainty in upscaling the forest biomass from field to satellite measurements (Brown, 1997; Réjou-Méchain et al., 2019). Nonetheless, the development of such equations is highly limited by numerous factors including the number of samples, cost, labor, restrictions, and complex topography. Hence there is a significant need for alternative, non-destructive approaches to assess forest biomass accurately and sustainably.

Recently, Terrestrial Laser Scanning (TLS) has shown its potential to estimate basic tree attributes (viz., DBH, height, crown width, crown volume etc.) using dense three dimensional (3D) point clouds with a high degree of accuracy and precision (Bienert et al. 2006; Huang et al. 2011; Maas et al. 2008; Reddy et al. 2018). Further, the detailed point clouds from TLS provide the basis for building 3D tree models using quantitative structure models (QSMs). A QSM is a geometric model that describes a complete tree using a hierarchical collection of cylinder objects estimating topological, geometrical and volumetric details of the tree's woody structure (Raumonen et al., 2013). QSMs are reported to be adequate to estimate stem volumes (or biomass) at individual tree levels with very low bias in a non-destructive manner (Hackenberg et al., 2015; Mayamanikandan et al., 2019). The non-destructive tree volume estimation overcomes nearly every complicating factor associated with destructive sampling, including the issue of insufficient sample size in the development of allometric equations (Chave et al., 2004; Stovall et al., 2018). On the other hand, though QSMs could provide tree-level estimations of tree volume and biomass (when combined with wood density information), estimating tree-level biomass for all trees over a given area is still a daunting task. Nevertheless, the non-destructive QSM derived tree volumes would aid in the development and improvement of site-specific (and species-specific) allometric equations for tree volume estimation, which are still a major source of uncertainty in field biomass estimation (Réjou-Méchain et al., 2019).

In this context, the current work aims to assess the use of terrestrial LiDAR data in estimating tree volumes using QSMs and then to develop site-specific allometric models for tropical dry deciduous forests in the Central Indian region. In brief, the objectives of the present work are to (1) generate QSMs for major tree species in the study area using 3D point clouds, (2) estimate QSM based tree volumes for the TLS sampled trees, (3) develop allometric equations for tree volume estimation using tree structure attributes (viz., DBH and height), and (4) assess the accuracy and precision of TLS allometric models.

## 4.2 Materials and Methods

### 4.2.1 Study Area

The study area is located within tropical dry deciduous forests of the Betul district, Madhya Pradesh, India (Figure 4.1). Its geographical coordinates are 21° 49′ to 21° 54′ N latitude and 77° 21′ to 77° 27′ E longitude, with an elevation of approximately 507 meters above sea level. The climate experiences distinct seasons: hot summers and cool winters.

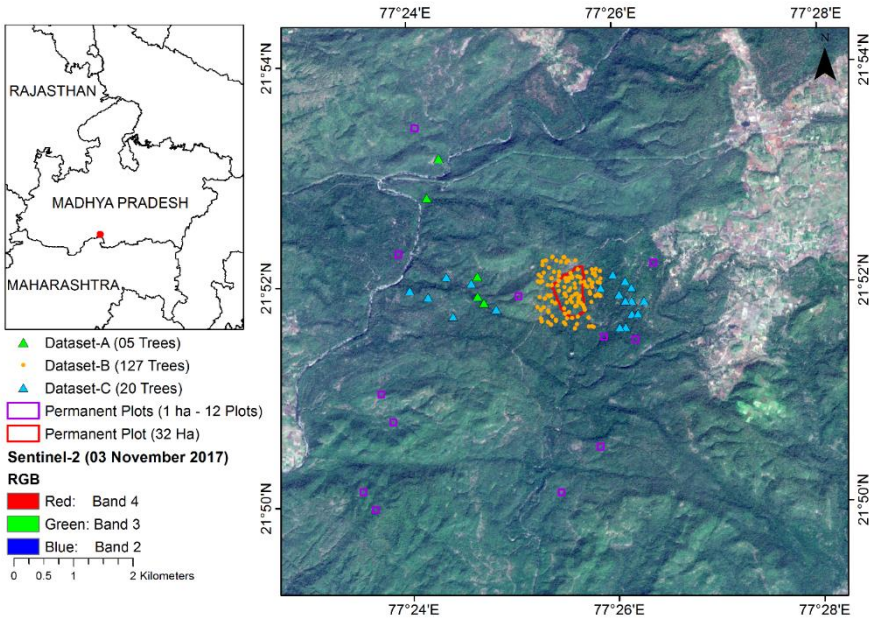


*Figure 4.1. Study site location map showing 13 permanent plots (12- 1ha and 1 -32ha) in tropical dry deciduous forests of Betul, Madhya Pradesh (India). Digital elevation model in the background. The TLS samples and the destructive tree measurements are also shown as dots and triangles respectively.*

The average annual rainfall is around 1100 mm, while the mean temperature hovers around 27°C (Rodda et al., 2021). The landscape is predominantly flat, characterized by a dominant presence of Teak (*Tectona grandis*) with an average stand height of ~22 meters. Other prominent tree species

include *Lagerstroemia parviflora, Anogeissus latifolia, Madhuca indica, Terminalia alata*, and *Diospyros melanoxylon*. The vegetation exhibits strong seasonality, with peak canopy cover occurring from September to October, followed by a complete leaf-off period between April and May.

### 4.2.2 Identification of dominant species

To generate robust allometric models, we focused on volumetric measurements from the most influential tree species within the study area. We identified these dominant and co-dominant species using the Importance Value Index (IVI) analysis. IVI considers a species' relative frequency, relative density, and relative dominance within the tree community structure (Curtis and McIntosh, 1951; Kent, 2011). For the IVI analysis, we utilized tree-level data collected during field inventory work described in Chapter 3 (Objective-I) related to plot size and its impact on overall field biomass accuracy. This data encompassed 13 spatially distributed plots across the study area (Figure 4.1). Twelve plots were 1-hectare squares arranged in a regular grid (100 x 100 meters each), with one additional irregular plot covering a continuous area of 32 hectares.

*Table 4.1. Detailed information on the number of TLS samples used for individual species and their associated importance value index (IVI) based on the available permanent plot data at the study site. DBHmin and DBHmax are minimum and maximum diameter at breast height values respectively. IQR indicates the inter quartile range of the diameter distribution of tree species between 5th and 95th Quartiles*

| | **Large Permanent Plot Data** | | | | **TLS Tree Samples** | | |
|---|---|---|---|---|---|---|---|
| **Species Name** | **IVI** | **$DBH_{min}$ [cm]** | **$DBH_{max}$ [cm]** | **IQR [cm]** | **$DBH_{min}$ [cm]** | **$DBH_{max}$ [cm]** | **N** |
| *Tectona grandis* | 103.9 | 7.6 | 68.9 | 10 - 38 | 11.1 | 46.2 | 33 |
| *Phyllanthus emblica* | 20.9 | 8.0 | 38.5 | 10 - 27 | 9.7 | 26.6 | 11 |
| *Lagerstroemia parviflora* | 19.7 | 9.5 | 44.6 | 9 - 25 | 9.9 | 25.1 | 12 |
| *Terminalia paniculata* | 15.4 | 9.8 | 60.5 | 10 - 39 | 10.0 | 28.8 | 13 |
| *Diospyros melanoxylon* | 14.4 | 7.6 | 52.4 | 9 - 34 | 10.2 | 35.8 | 16 |
| *Terminalia alata* | 12.9 | 9.5 | 78.6 | 12 - 51 | 10.2 | 43.3 | 19 |
| *Miscellaneous (Other Species)* | | 8.9 | 118.7 | 10 - 42 | 9.7 | 41.8 | 23 |

Using the BiodiversityR package (Kindt and Kindt, 2015), we calculated the IVI for each species within the region by summing its relative frequency, relative density, and relative dominance. Based on this analysis, the identified dominant and co-dominant species with their corresponding IVI values were: *Tectona grandis* (103.9), *Diospyros melanoxylon* (20.9), *Phyllanthus emblica* (19.7), *Terminalia alata* (15.4), *Lagerstroemia parviflora* (14.4), *Terminalia Paniculata* (12.9) (Table 4.1).

### 4.2.3 TLS Data Collection and Processing

Following the identification of dominant and co-dominant species through IVI analysis, we employed a RIEGL VZ-1000 terrestrial laser scanner (RIEGL Laser Measurement Systems GmbH) to capture detailed 3D point cloud data during the 2019-2020 period. The sampling strategy focused on ensuring variability in both tree species and diameter range within the selected dominant and co-dominant groups to create more robust and generalizable allometric models (Table 4.1). Each TLS scan was acquired in high-speed mode with a pulse repetition rate of 300 kHz, with a resolution of 0.05-degree vertical and horizontal angles. To guarantee complete 3D tree structure capture, each tree was scanned from a minimum of three different directions, resulting in a comprehensive 360° coverage. In conjunction with the TLS data acquisition, basic tree inventory measurements of diameter at breast height (DBH) and tree height were collected in the field using a measuring tape and hypsometer.

*Table 4.2. Description of the datasets used for the current study*

| | **No. of samples** | **Measurement Type** | **Usage** |
|---|---|---|---|
| **Dataset-A** | 5 | TLS and ground Destructive samples | For standardization of parameters in TreeQSM approach |
| **Dataset-B** | 127 | TLS only | For development of allometric models |
| **Dataset-C** | 25 | 20 trees with only ground destructive samples and 5 trees from Dataset-A | For validation of generated allometric models using Dataset-B |

A total of 132 trees across various species, with a focus on dominant and co-dominant species identified through IVI analysis (Table 4.1), were scanned using the TLS system to capture their 3D structure. The sampling strategy aimed for variability in both tree species and diameter range to ensure the development of robust allometric models. Due to regulations protecting trees at the study site, destructive sampling (complete felling and weighing) was only possible for five Tectona grandis trees (referred to as Dataset-A; N=5 trees). This destructive sampling involved felling the trees, sectioning them into smaller parts for weight measurement using a weighing machine and metal chains in the field. Wood density for each tree was then determined in the lab using the oven-drying method (105°C for 48 hours) and the water displacement method. The limited destructive samples (Dataset-A; N = 5 trees; destructive samples + TLS data) were used to establish a standardized method for estimating tree volume from 3D point cloud data captured by TLS. The remaining 127 TLS scans (Dataset-B; N=127 trees; TLS tree samples only) were then employed to estimate tree volume

using this standardized 3D tree modeling approach. Ultimately, these volume estimates from Dataset-B will be used to develop non-destructive allometric equations based on TLS data.

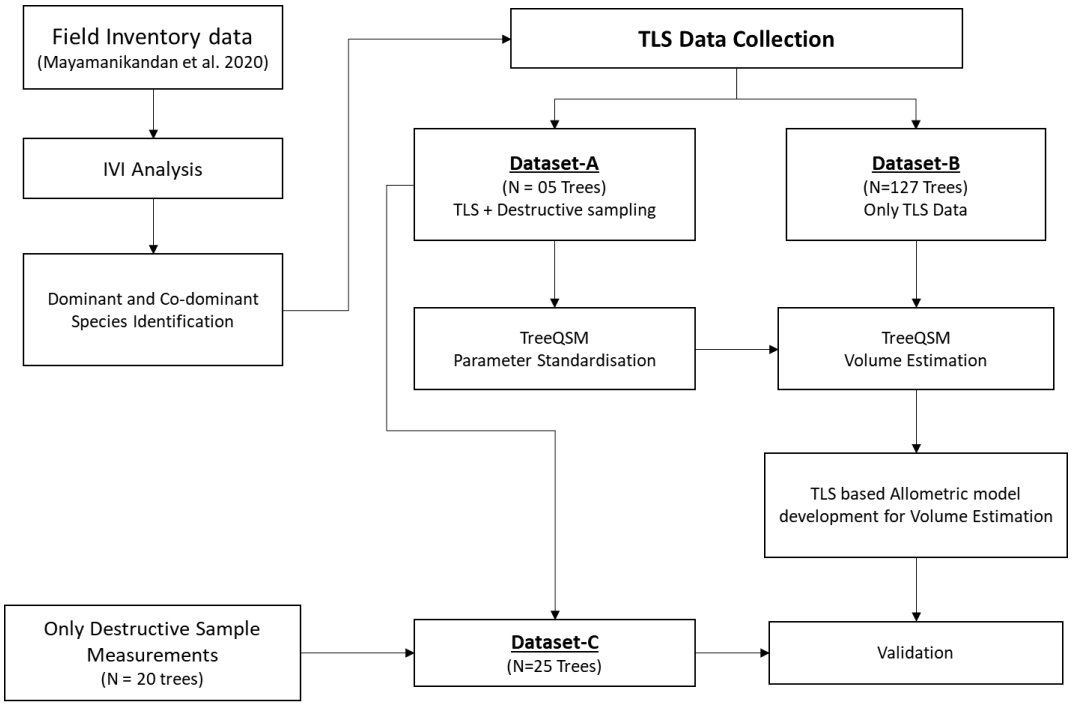


*Figure 4.2. Flowchart describing the methodology used in the study*

The allometric models developed using Dataset-B were validated using an independent destructive sampling measurement. This measurement involved 20 trees collected during 2019. These trees were not scanned with TLS prior to felling, making them an ideal validation dataset. Since the destructive sampling measurements from Dataset-A were not used in allometric model development, they also serve as a valuable validation resource. Therefore, a combined dataset of 25 trees (referred to as Dataset-C in this manuscript; N = 25; destructive samples only) was used to validate the non-destructive allometric models developed for tree volume estimation using Dataset-B. The summary details of the three datasets used in the current study are provided in Table 4.2. A detailed flowchart explaining the overall workflow is shown in Figure 4.2.

The TLS data collected for each tree was processed in accordance with the standard protocols. The three TLS scans captured for each tree from different directions were co-registered and merged. This process involved using artificially placed targets within the scans to facilitate alignment and achieve a comprehensive 3D coverage of each tree for virtual reconstruction. Automatic co-registration was performed using the triangulation algorithms available in the RiSCAN PRO 2.0 software (RIEGL Laser Measurement Systems GmbH). For this process, reflective stickers were placed on the tree of interest in each scan to aid identification during co-registration. Following co-registration and merging, the merged scan data was normalized by estimating a ground model using the

minimum elevation approach (Reddy et al. 2018). This step helps to remove any ground level variations from the data. Next, the tree of interest was manually segmented from the background using CloudCompare software. This segmentation process essentially isolates the tree points from the surrounding environment.

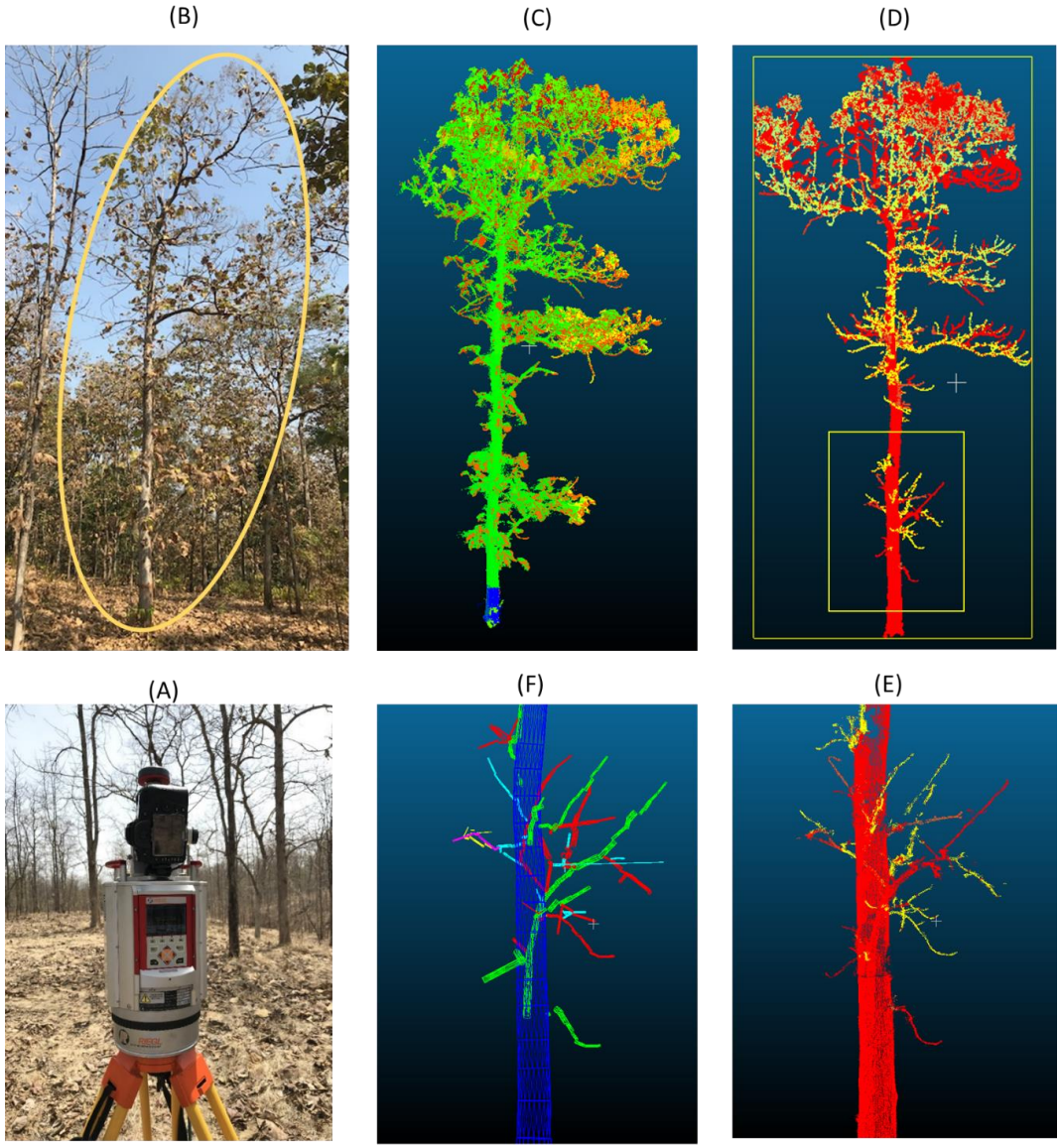


*Figure 4.3. The step-by-step procedure of (A) TLS-scanning of a (B) selected tree, (C) isolating using visual techniques in cloudcompare, (D) leaf removal. The final data is then used for QSM modeling using TreeQSM library. (E) and (F) shows the zoomed portion of the selected tree and respective 3D model as wireframe.*

Finally, individual tree point clouds were manually filtered to remove leaves, leaving only the tree stem and branches. The resulting tree skeletons were then processed for volume estimation using the QSM approach. This method involves fitting cylinders in a hierarchical order to approximate the tree structure

(Raumonen et al., 2013). Figure 4.3 visually illustrates the steps involved in the QSM approach for tree volume estimation.

### 4.2.4 Tree Volume Estimation using TreeQSM

A QSM, or Quantitative Structure Model, of a tree is a model representing a woody tree structure from detailed TLS point clouds using geometric primitives (cylinders in this case) and can quantitatively describe the basic topological (branching structure), geometric and volumetric properties of a tree (Raumonen et al., 2013). In this work, we utilized the open-source TreeQSM library (Version 2.4.0; https://github.com/InverseTampere/TreeQSM ) to construct QSMs of trees (Raumonen et al., 2013). Several studies have demonstrated the effectiveness of TreeQSM in generating 3D tree models and estimating tree volume, ultimately leading to accurate AGB estimations with minimal bias (Calders et al., 2014; de Tanago et al., 2018).

The TreeQSM algorithm reconstructs trees in a two-step process: segmentation and cylinder fitting. First, segmentation separates the point cloud data into the main trunk and individual branches. This process utilizes a "cover set" approach, working in two phases with small subsets of points called patches. In the first phase, large constant size patches with radius (defined as Patch Diameter 1; PD1) are used across the tree to identify a coarse tree architecture and branches. In the second phase, a finer cover with patch size varying from a minimum Patch Diameter 2 (PD2min) to a maximum Patch Diameter 2 (PD2max). This step refines the branch topology by capturing finer details.

Following segmentation, individual branch elements are reconstructed by fitting cylinders to the point cloud data using a least squares approach. A critical parameter in TreeQSM is the selection of PD2min, as it determines the smallest branch feature that can be resolved with the given data and ultimately affects volume estimation accuracy. Additionally, due to the randomness of patch generation, TreeQSM recommends generating multiple models with the same parameter settings. This allows for the calculation of modeling confidence based on the resulting estimates from these 3D tree models (de Tanago et al., 2018; Raumonen et al., 2015).

To optimize the TreeQSM parameters (PD1, PD2min, and PD2max) for this study, we adopted a literature review-based approach (Brede et al., 2019; Calders et al., 2015; de Tanago et al., 2018). PD1 was set to a constant value of 10 cm for all trees. PD2min values were varied from 4 cm to 6 cm with increments of 0.5 cm, and PD2max values ranged from 10 cm to 14 cm with intervals of 1 cm. We then generated ten TreeQSM models for each possible parameter combination. The optimum settings for PD2min and PD2max were selected based on mean point-cylinder model distances from the trunk and first order branches (Calders et al., 2018). The final 3D tree model was generated using the optimized

parameter settings. Tree volume was then estimated by summing the volumes of all cylinders within the 3D model. To evaluate the reconstruction method's robustness and assess tree volume uncertainty, we created 20 models for each tree using the optimal parameter combination. The mean tree volume and standard deviation were calculated from these 20 models. We employed MATLAB (Version 9.5) for the TreeQSM reconstruction process and R Statistical software (Version 4.1.3) for further data analysis.

### 4.2.5 Development and Evaluation of TLS based Allometric Models

Allometric models were developed to estimate tree volume using the non-destructive volumes derived from the TLS measurements. We opted for log-linear regression as it has been one of the accurate descriptors of tree allometry in Central Indian forests compared to other descriptors (Chaturvedi and Raghubanshi, 2015). Also, this is the most common form of equation to describe tree allometry in biomass studies (Jenkins et al., 2003; Lin et al., 2017; Moussa and Mahamane, 2018). We employed a combination of DBH (diameter at breast height) and tree height as predictor variables to explain tree volume using two different equation forms, as shown below:

$$V = \exp\big(a + b\ (ln(DBH))\big) \quad (1)$$

$$V = \exp\big(a + b\ (ln(DBH^2 \times Height\ ))\big) \quad (2)$$

Where V represents tree volume, a and b are the estimated parameters of the fitted model, DBH is measured in meters, and H represents tree height in meters.

The log-linear relationships with high RMSE tend to underestimate the predicted value when back-transformed to log units, so a correction factor has been proposed to remove this bias. During the back transformation of log-linear relationships, a correction factor has been proposed to reduce the systematic bias in the predicted value (Baskerville, 1972). The correction factor (CF) uses the mean square error (MSE) and is calculated as:

$$CF = e^{\frac{MSE}{2}} \quad (3)$$

Where MSE is the mean square error of the modelled equation as per Equations (1) and (2).

The allometric equations were evaluated in terms of the coefficient of determination ($R^2$) and uncertainty using the destructive sampled tree measurement data as the reference value. The model uncertainty was calculated in terms of Bias, rBias%, root mean square error (RMSE) and rRMSE% of the estimated value and the corresponding reference measurement value.

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ri}) \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_{ri})^2}{n}} \quad (5)$$

$$rBias\% = \frac{Bias}{\bar{y_r}} \times 100\% \quad (6)$$

$$rRMSE\% = \frac{RMSE}{\bar{y_r}} \times 100\% \quad (7)$$

Where $y_i$ is the estimated value of the $i^{th}$ tree using the TreeQSM approach, $y_{ri}$ represents the reference measurement value of that tree, $\bar{y_r}$ represents the average value of all the reference measurements, and n represents the total number of tree samples used.

Apart from the model uncertainty, the validation uncertainty was computed using the independent destructive sampling dataset (Dataset-C). Also, the concordance correlation coefficient (CCC; Lin 1989) was calculated to compare the agreement of TLS derived tree volumes with reference measurements. In addition, we have also performed an uncertainty analysis of the generated allometric models with respect to change in sample size.

### 4.2.6 Comparison with other Conventional Allometric models over Indian forests

We also compared our models with the commonly used generic models in tropical forests of India (Table 4.3; derived from FSI (1996) and Chave et al., (2014)). A tree-level comparison was used to evaluate the TLS-based and the generic allometric models using the independent destructive sampling dataset (Dataset-C). Absolute and relative RMSE and bias were computed using the equations mentioned in the previous section. Further, we computed the average bias at plot-level in terms of AGB (by employing the reported wood density values at the species level; Table 4.3) for the 13 permanent plots at the study site due to choice of tree-level allometry using the generic models versus the newly developed models using TLS.

*Table 4.3. Traditional allometric models used at the study site to predict tree volumes and AGB using diameter and height as predictive variables. The wood density (WD) values for Chave et al., 2014 model are used as per species information.*

| | Species | Model | WD |
|---|---|---|---|
| **FSI (1996)** | *Tectona grandis* | $V = 0.04346 - 0.26352\sqrt{DBH} + 8.79334\ DBH^2$ | 0.55 |
| | *Phyllanthus emblica* | -NA- Miscellaneous equation is used | *0.63* |
| | *Lagerstroemia parviflora* | $V = 0.01617 - 0.66446\ DBH + 9.71038\ DBH^2$ | *0.62* |
| | *Terminalia paniculata* | $V = 0.131 - 1.87132\ DBH + 9.47681\ DBH^2$ | *0.63* |
| | *Diospyros melanoxylon* | $V = 0.0333 - 0.93267\ DBH + 8.15911\ DBH^2 + 1.30093\ DBH^3$ | *0.68* |
| | *Terminalia alata* | $V = 0.131 - 1.87132\ DBH + 9.47681\ DBH^2$ | *0.63* |
| | *Miscellaneous (Other Species)* | $V = 0.0697 - 1.4597\ DBH + 11.79933\ DBH^2 - 2.35397\ DBH^3$ | *0.63* |
| | | | |
| **Chave et al., 2014** | General Equation (for all species) | $V = 0.0559\ (WD\ x\ DBH^2 x\ H)$ | - |

## 4.3 Results and Discussion

### 4.3.1 DBH and Tree Height

To ensure robust allometric model development, TLS sampling focused on dominant and co-dominant species identified through permanent plot data (Table 4.1). This strategy was aimed to capture a representative range of species and diameter classes across the study site (Figure 4.1). The diameter range of the sampled trees using TLS closely matched the 5$^{th}$ and 95$^{th}$ percentiles of the diameter distribution for each species based on the permanent plot data (Table 4.1).

The accuracy of the TLS-derived DBH and height estimates was validated for all 132 trees (Dataset-A & Dataset-B) by comparing them with reference field measurements collected during data acquisition (Figure 4.4). The estimated DBH ranged from 8.1 cm to 42.73 cm, with reference values spanning from 9.7 cm to 46.2 cm. A strong positive correlation ($R^2 = 0.96$; $n = 132$; $p < 0.01$) was observed between the estimated and reference DBH values (Figure 4.4A). The root mean squared error (RMSE) was 1.9 cm, translating to a relative RMSE (rRMSE) of 8.9%. Notably, the majority of residuals fell within ±3.0 cm of the reference

values. The relative RMSE was slightly higher in the lower DBH range (<20 cm) at 11.8%, decreasing to 3.3% for trees with DBH greater than 40 cm.

Similarly, tree height estimates using the TLS-QSM method exhibited a strong positive correlation ($R^2$ = 0.98; n = 132; p < 0.01) with reference field measurements (Figure 4.4B). The RMSE for height estimation was 0.54 m, corresponding to a rRMSE of 4.1%. The uncertainty in tree height estimation has not changed substantially with the increase in tree height (Figure 4.4B).

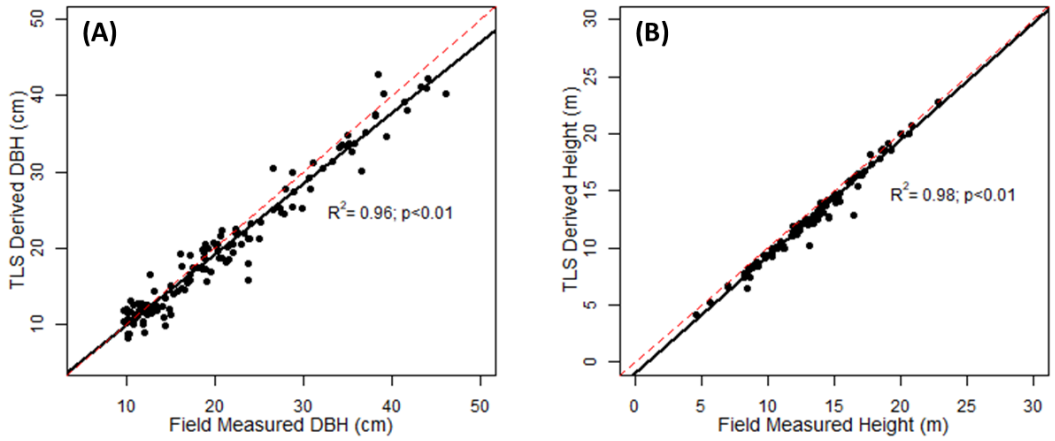


*Figure 4.4. Scatterplot showing comparison of TLS-derived parameters with field measurements for (A) DBH and (B) height for the TLS scanned trees).*

### 4.3.2 Tree Volume

The accuracy of the QSM-derived tree volume estimates from TLS data was evaluated by comparing them with reference volumes obtained from destructive sampling measurements (Dataset-A). The estimated TLS volumes showed a close agreement with the reference values, exhibiting a bias of < 6% (Table 4.4). A linear regression analysis between the estimated and reference volumes revealed a strong positive correlation ($R^2$ = 0.99) with a slope of 1.09, indicating a slight overestimation of tree volume for larger trees (Figure 4.5). The standard deviation in the estimated tree volume due to the random variation of patch diameters within the TreeQSM method was found to be less than 5%, indicating a high level of reliability for non-destructive volume estimations using TLS. Additionally, the root mean squared error (RMSE) was 0.012 $m^3$, and the concordance correlation coefficient (CCC) was 0.98 (95% Confidence Interval: 0.95 – 0.99). Given the high level of agreement between the TLS-derived tree volumes and the destructive samples, the parameterization used for Dataset-A was applied to the remaining 127 trees (Dataset-B) to compute non-destructive tree volumes for allometric model development.

*Table 4.4. Non-destructive tree volume estimates in comparison with the destructive measurements for 05 teak (Tectona grandis) trees as part of Dataset-A. Where $X_{vol}$ denotes the volume with respect to various methods (X) employed in the study.*

| | **Field Height (m)** | **Field DBH (cm)** | **Weight (kg)** | **Des$_{vol}$ (m$^3$)** | **TLS$_{vol}$ (m$^3$)** | **Bias (%)** | **FSI$_{vol}$ (m$^3$)** | **Chave$_{vol}$ (m$^3$)** |
|---|---|---|---|---|---|---|---|---|
| **Tree1** | 19.30 | 43.93 | 1041.85 | 1.847 | 1.926 | 4.3% | 2.466 | 2.082 |
| **Tree2** | 17.87 | 23.87 | 283.18 | 0.487 | 0.459 | -5.7% | 0.655 | 0.569 |
| **Tree3** | 20.20 | 40.43 | 1076.84 | 1.783 | 1.895 | 6.3% | 2.068 | 1.845 |
| **Tree4** | 24.76 | 47.75 | 1383.15 | 2.188 | 2.296 | 4.9% | 2.939 | 3.155 |
| **Tree5** | 25.46 | 52.20 | 1449.56 | 2.593 | 2.758 | 6.4% | 3.543 | 3.878 |

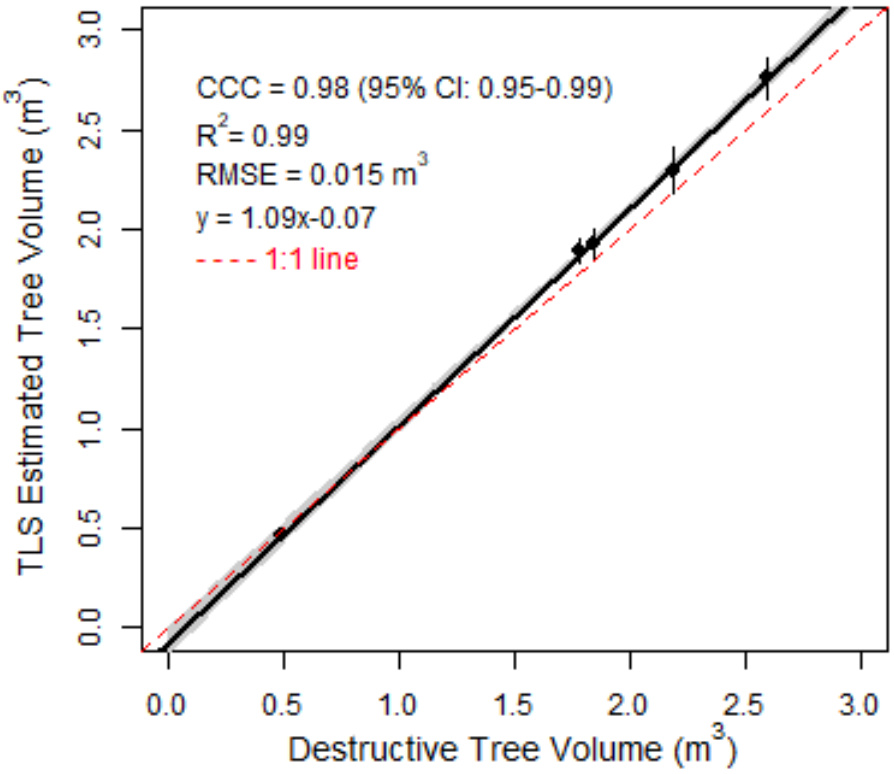


*Figure 4.5. Comparison of destructive tree volume estimates for Tectona grandis trees with the TLS-QSM derived tree volumes for Dataset-A.*

### 4.3.3 Allometric Models

The 127 trees in Dataset-B encompassed a broad diameter range (9.7 cm to 46.2 cm) and represented various species commonly found in tropical dry deciduous forests of India (Figure 4.6). This dataset provided the foundation for developing

local, species-specific allometric models. Subsequently, we utilized the non-destructive tree volumes derived from the TLS-QSM method (refer to section 4.2.4 for details on TLS-QSM) as the dependent variable in allometric model development. Diameter at breast height (DBH) and tree height were employed as the independent predictor variables in two separate log-linear model forms, as shown in Equations (1) and (2).

Following model fitting and bias correction, the final allometric models for the entire site (including all species data from Dataset-B; n = 127) are presented in their back-transformed log-linear forms below:

$$V = 9.543 * (D)^{1.93} \quad (8)$$

$$V = 0.681 * (D^2H)^{0.739} \quad (9)$$

Where V is the estimated tree volume ($m^3$), D is the diameter at breast height (m), and H is the tree height (m).

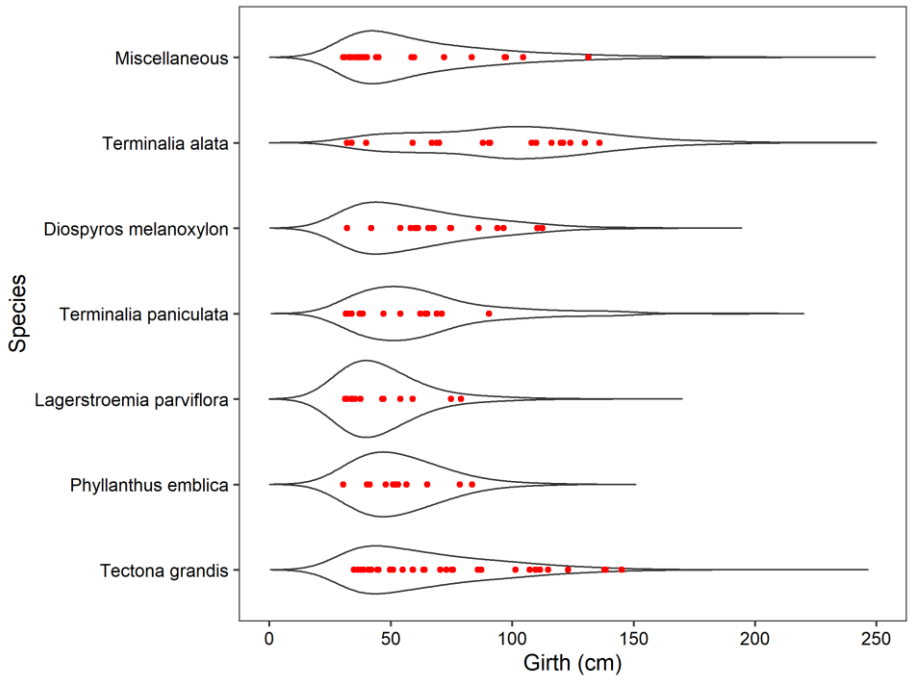


*Figure 4.6. Species-level information of all trees at the study site as violin plot indicating the density distribution and range using all the tree samples data from 55-ha permanent plots. The red dots indicate the girth values of 127 acquired TLS samples for allometric model development.*

The allometric model that incorporates tree height as a predictor variable (height-inclusive model) exhibited superior performance compared to the model using only diameter (diameter-exclusive model). The height-inclusive model achieved a higher coefficient of determination ($R^2$ = 0.94) and a lower root mean

squared error (RMSE) of 21.1% when compared to the diameter exclusive allometric model (R2 = 0.93; RMSE = 24.4%) (Figure 4.7).

Table 4.5 presents detailed information about species-specific allometric equations developed in both model forms (with and without height) using the tree volume from TLS data in a non-destructive way. The RMSE for these models ranged from 15.2% to 21.4% for the height-inclusive model and 17.8% to 28.2% for the diameter-exclusive model. This trend confirms that including tree height as a predictor variable consistently reduces the overall error in volume estimation across all species (Table 4.5). Notably, for the mixed-species category (miscellaneous), incorporating height substantially reduced the RMSE from 28.2% to 19.3%.

To assess model uncertainty beyond the fitting data, relative RMSE errors were computed using the independent data (Dataset-C; n = 25 trees). Since the species-specific samples were few, the errors were computed using all species information. With RMSE of 10.9% compared to diameter-exclusive allometric model (RMSE = 13.5%), the all-species model showed better performance with inclusion of height as predictor variable. Using the species-specific models, the RMSE is 15.4% and 21.6% for diameter-height and diameter models respectively, for the independent dataset. However, the samples were sparse for testing species-specific models.

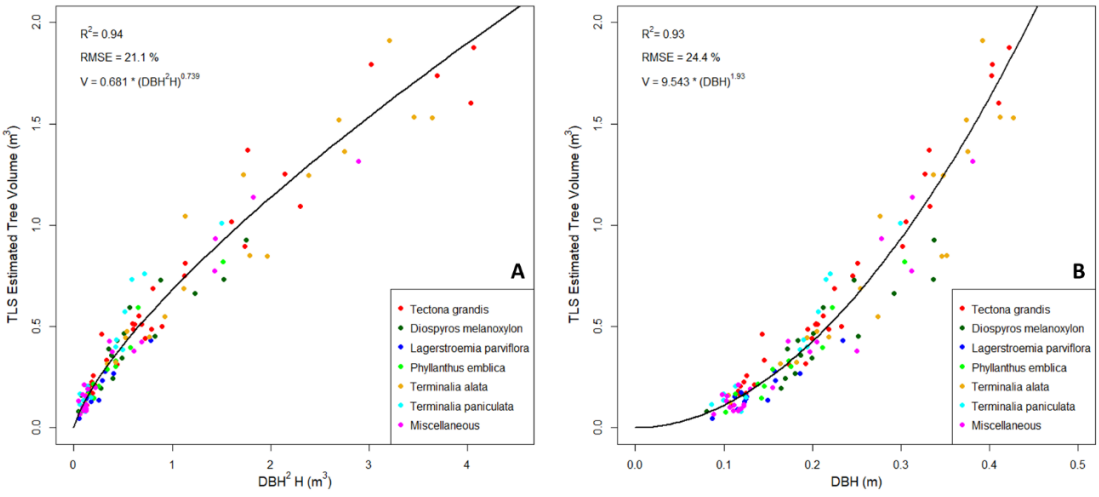


*Figure 4.7. Site-specific allometry derived from TLS-QSM tree volumes with (A) Diameter and Height (B) Diameter as predictive models.*

*Table 4.5. General and species-specific allometric model coefficients (a & b as defined in Equations (1) & (2)) and respective standard errors (se) for the dominant and co-dominant tree species at the study site using TLS-QSM volume estimates. R2 is based on the fit of the species-specific log-transformed observations using the model coefficients after bias removal. Relative RMSE (%) was calculated as the average of all relative RMSE estimates across all diameter classes. RMSE-Valid indicates the validation RMSE error using Dataset-C evaluated using general species equation.*

| Equation | Species | n | DBH Range [cm] | a | a [se] | b | b[se] | $R^2$ | RMSE (%) | RMSE - Valid (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| **$DBH^2$ H** | All Species | 127 | 9.7 - 46.2 | -0.4134 | 0.0268 | 0.7387 | 0.0186 | 0.94 | 21.1 | 10.9 |
| | *Tectona grandis* | 33 | 11.1 - 46.2 | -0.3826 | 0.0341 | 0.7400 | 0.0297 | 0.95 | 17.6 | |
| | *Phyllanthus emblica* | 11 | 9.7 - 26.6 | -0.4432 | 0.0775 | 0.7689 | 0.0557 | 0.95 | 15.2 | |
| | *Lagerstroemia parviflora* | 12 | 9.9 - 25.1 | -0.6069 | 0.1780 | 0.7313 | 0.0996 | 0.90 | 17.9 | |
| | *Terminalia paniculata* | 13 | 10.0 - 28.8 | -0.2168 | 0.1492 | 0.7548 | 0.0913 | 0.91 | 21.4 | |
| | *Diospyros melanoxylon* | 16 | 10.2 - 35.8 | -0.4885 | 0.0736 | 0.6323 | 0.0606 | 0.87 | 18.8 | |
| | *Terminalia alata* | 19 | 10.2 - 43.3 | -0.4036 | 0.0375 | 0.7221 | 0.0362 | 0.91 | 17.9 | |
| | Miscellaneous | 23 | 9.7 - 41.8 | -0.4686 | 0.1083 | 0.7221 | 0.0578 | 0.97 | 19.3 | |
| **DBH** | All Species | 127 | 9.7 - 46.2 | 2.2212 | 0.0938 | 1.9280 | 0.0535 | 0.93 | 24.4 | 13.5 |
| | *Tectona grandis* | 33 | 11.1 - 46.2 | 2.2140 | 0.1263 | 1.8389 | 0.0767 | 0.98 | 19.3 | |
| | *Phyllanthus emblica* | 11 | 9.7 - 26.6 | 2.3067 | 0.3845 | 2.0039 | 0.2096 | 0.94 | 17.8 | |
| | *Lagerstroemia parviflora* | 12 | 9.9 - 25.1 | 2.1214 | 0.5777 | 1.9782 | 0.2881 | 0.88 | 20.4 | |
| | *Terminalia paniculata* | 13 | 10.0 - 28.8 | 2.4345 | 0.4389 | 1.9536 | 0.2306 | 0.92 | 20.3 | |
| | *Diospyros melanoxylon* | 16 | 10.2 - 35.8 | 1.7286 | 0.2609 | 1.6710 | 0.1570 | 0.83 | 22.3 | |
| | *Terminalia alata* | 19 | 10.2 - 43.3 | 2.0074 | 0.1459 | 1.7688 | 0.1034 | 0.88 | 21.1 | |
| | Miscellaneous | 23 | 9.7 - 41.8 | 2.1817 | 0.3024 | 1.9631 | 0.1559 | 0.93 | 28.2 | |

### 4.3.4 Comparison with other Allometric Equations

The performance of the TLS-derived allometric models was evaluated against commonly used volume equations from FSI (1996) and Chave et al. (2014) typically applied for AGB estimation across tropical Indian forests (Fararoda et al., 2021; Rajashekar et al., 2018). An independent validation dataset (n = 25; Dataset-C) not used in model development was employed for this comparison. Since the predictor variables differed between the existing equations (FSI using DBH only and Chave et al. including both DBH and height), we compared them to corresponding models developed in this study. For a fair comparison, the species-specific FSI (1996) models were compared with the TLS-derived diameter-only models at the species level, while the Chave et al. (2014) model was compared with the all-species model that incorporates both DBH and height as predictors.

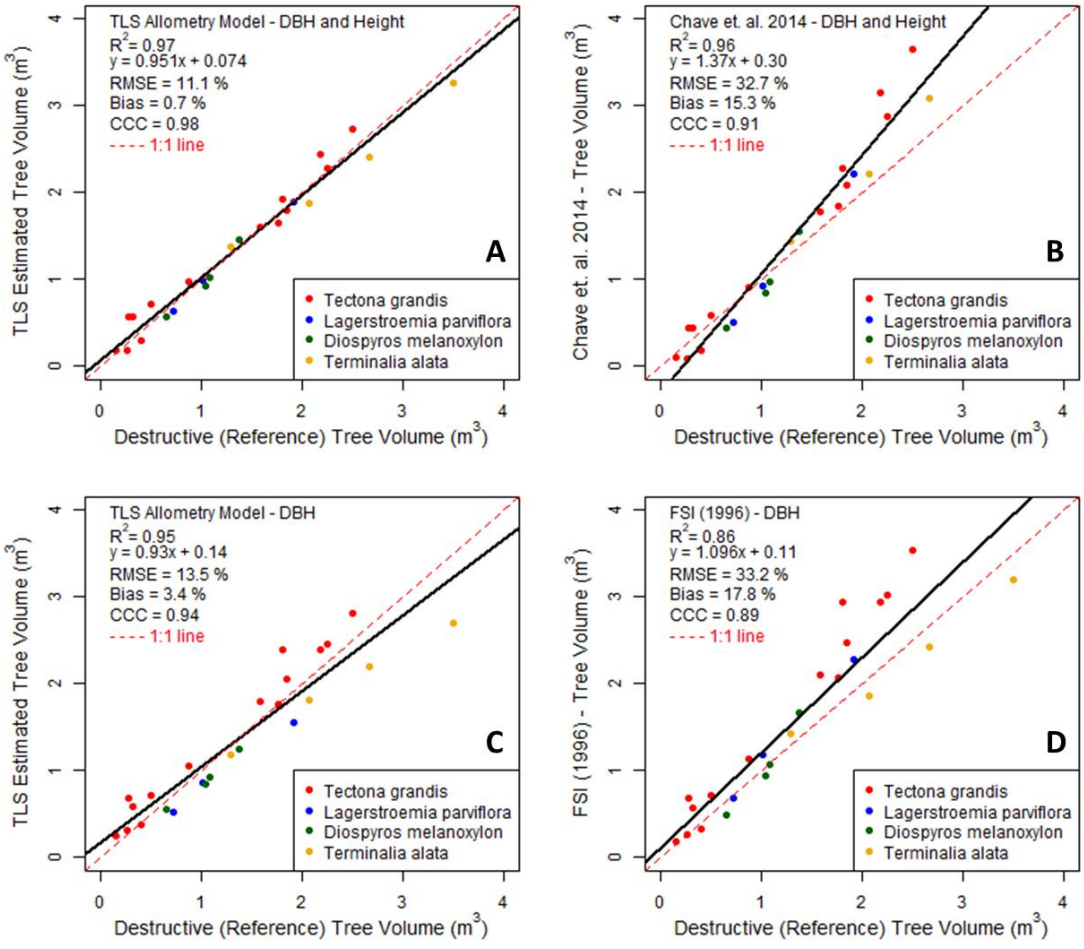


*Figure 4.8. Comparison of destructively measured tree volumes for Dataset-C (n=25 trees) with different allometric models (A) TLS-DBH and Height, (B) Chave et al., 2014, (C) TLS-DBH, and (D) FSI (1996). Generic (all species) TLS allometry models were considered in all cases.*

The locally derived, non-destructive TLS-based allometric models demonstrated significant advantages over the commonly used equations within the study site (Figure 4.8). The TLS model incorporating both diameter and height achieved a substantially lower RMSE (11.1%) and bias (0.7%), along with a high concordance correlation coefficient (CCC) of 0.98, compared to the Chave et al. (2014) model (RMSE = 32.7%, bias = 15.3%, CCC = 0.91). Similarly, the TLS diameter-only model (RMSE = 13.5%, bias = 3.4%, CCC = 0.94) outperformed the FSI (1996) diameter-based models (RMSE = 33.2%, bias = 17.8%, CCC = 0.89).

In summary, the existing volume equations commonly used for tropical dry deciduous forests in India exhibited higher uncertainty and a tendency to overestimate tree volume and consequently, aboveground biomass at the individual tree level. The TLS-derived models offer a more accurate and site-specific alternative for biomass estimation.

## 4.4 Discussion

Accurate allometric models are essential for reducing uncertainty in biomass mapping. This study investigated the potential of Terrestrial Laser Scanning (TLS) data as a promising solution. TLS offers detailed 3D information for non-destructive tree volume estimation. We explored this potential by utilizing TLS data to capture intricate tree structures and integrating it with a limited number of destructive field measurements. These field measurements played a crucial role in validating the accuracy of TLS data and ensuring reliable model development. Our findings convincingly demonstrate the superiority of using TLS data combined with appropriate modelling techniques compared to relying solely on traditional field measurements. The results show that the TLS data combined with QSM modeling is much more capable than just replicating simple field measurements and providing relevant information on complex tree dimensions towards tree volume estimations.

### 4.4.1 Allometric Models

Individual tree AGB is highly influenced by wood density, which exhibits significant variation across species, sites, and even within individual trees due to age and height (Chave et al., 2009). While excluding wood density from biomass allometry offers greater flexibility for large-scale estimations (Stovall et al., 2018), the observed variability in model coefficients underscores the need for region-specific volume equations (Duncanson et al., 2015). In recognition of these factors, this study focused on developing non-destructive, region-specific volume equations for central Indian dry deciduous forests, excluding wood density as the dependent variable and utilizing Terrestrial Laser Scanning (TLS) data.

While QSM offers a effective semi-analytical baseline method for TLS based estimation of tree volume with high degree of precision (Hackenberg et al., 2015; Stovall et al., 2018), the potential for over- or underestimation remains dependent on segmentation parameter selection. In order to reduce the chances of erroneous inferences, we employed a limited number of field-based destructive measurements (Dataset-A) to validate the TLS derived volume. The results are promising, with QSM modeling of 3D point clouds achieving high retrieval accuracy (RMSE = 6%, meanBias = -3.2%). This surpasses the accuracy of other common allometric models for this site (Table 4.4). These findings align with previous studies that utilized QSM-based models for tree volume estimation (Calders et al., 2014; Hackenberg et al., 2015). Hackenberg et al. (2015) reported volume errors within 10% for 36 trees of various species using high-quality TLS data. Calders et al. (2014) observed a slight overestimation (9.68%) of tree volumes in open eucalypt forests (Australia) compared to destructive sampling. Furthermore, the close agreement between TLS-derived diameter and height measurements and field data (Figure 4.4) strengthens the reliability of the TLS-derived volume estimates for developing non-destructive allometric models.

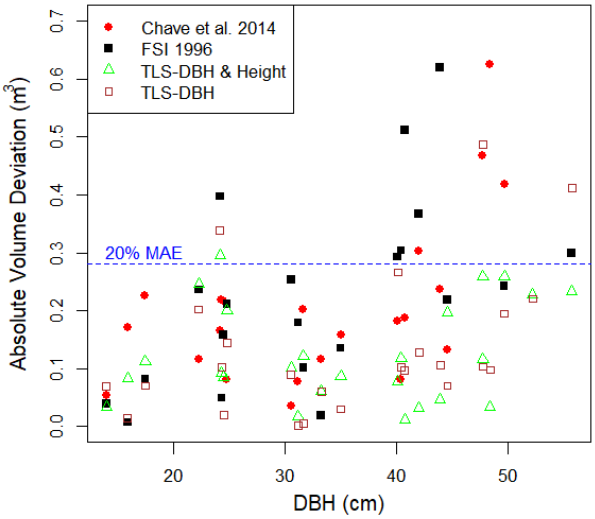


*Figure 4.9. Mean Absolute error distribution in tree volume with respect to tree diameter for the independent destructive sampling dataset (Dataset-C) due to varying choice of allometric model. The blue line indicates a 20% mean absolute error with respect to mean tree volume of the dataset*

Our site-level, generic TLS-QSM allometric models (Table 4.5) demonstrated high accuracy when validated against an independent destructive measurements (Figure 4.8). Including tree height (H) as a predictor variable alongside diameter at breast height (DBH) significantly reduced modeling error, achieving the highest coefficient of concordance (CCC) close to 1, indicating excellent model reproducibility. Notably, while the root mean square error (RMSE) difference between the DBH+H and DBH-only models was minimal, the

absolute volume deviations for large diameter trees were considerably higher using the DBH-only model (Figure 4.9). Conversely, the DBH+H model exhibited consistent volume estimations across the entire diameter range (Figure 4.9), highlighting the importance of incorporating height for accurate volume prediction, particularly for larger trees. Tree height is an integral part of tree growth since trees tend to optimize growth strategies and align physiological processes driving tree growth to fill the available 3D space in the canopy in order to maximize photosynthetic input and reduce competition (West et al., 1999).

The inclusion of both DBH and height (H) in the model aligns closely with the concept of cylindrical volume estimation, suggesting a near-linear relationship (Equations (1) and (2)). The model coefficients likely represent variations in wood allocation patterns relative to a standardized diameter and height. However, the DBH+H model doesn't entirely replace the need for species-specific models. Due to differing wood allocation strategies across species, the species-level models achieved lower errors compared to the single, combined model across all species (Table 4.5). However, these species-specific models were not tested with validation measurements (Dataset-C) due to the limited availability of destructive tree measurements at the species-level.

The significant benefit of the TLS in non-destructive volume estimation could be ascertained in both large tree volume estimation and increasing the sampling size, which are major constraints in the existing allometric models. Large trees are often under-represented since they are not easy to be measure due to practical limitations. Using allometric models which lack higher diameter range would possibly predict the tree volumes with large absolute errors and could be potentially unrealistic (Calders et al., 2014). Considering that the large trees hold approximately 40% of stand-level biomass (Brown, 1997), increasing sampling size particularly for large trees, could potentially alter our understanding in terms of forest-wide volume and biomass allocation.

### 4.4.2 Uncertainty due to Sample Size

Traditional volume equations derived from destructive measurements are often hampered by limited calibration samples. This can lead to systematic bias when applied to estimate tree volumes across diverse diameter and height classes (Clark and Kellner, 2012). To assess this limitation, we conducted a sample size analysis on model coefficients and uncertainties. We performed iterative random sampling from Dataset-B, ranging from 5 to 125 samples in increments of 5. Each sample set was used to develop allometric models, which were then evaluated against independent destructive measurements (25 trees; Dataset-C). Root Mean Square Error (RMSE) and bias were computed for each model. This process was repeated 1000 times at each sample size to analyze parameter variations.

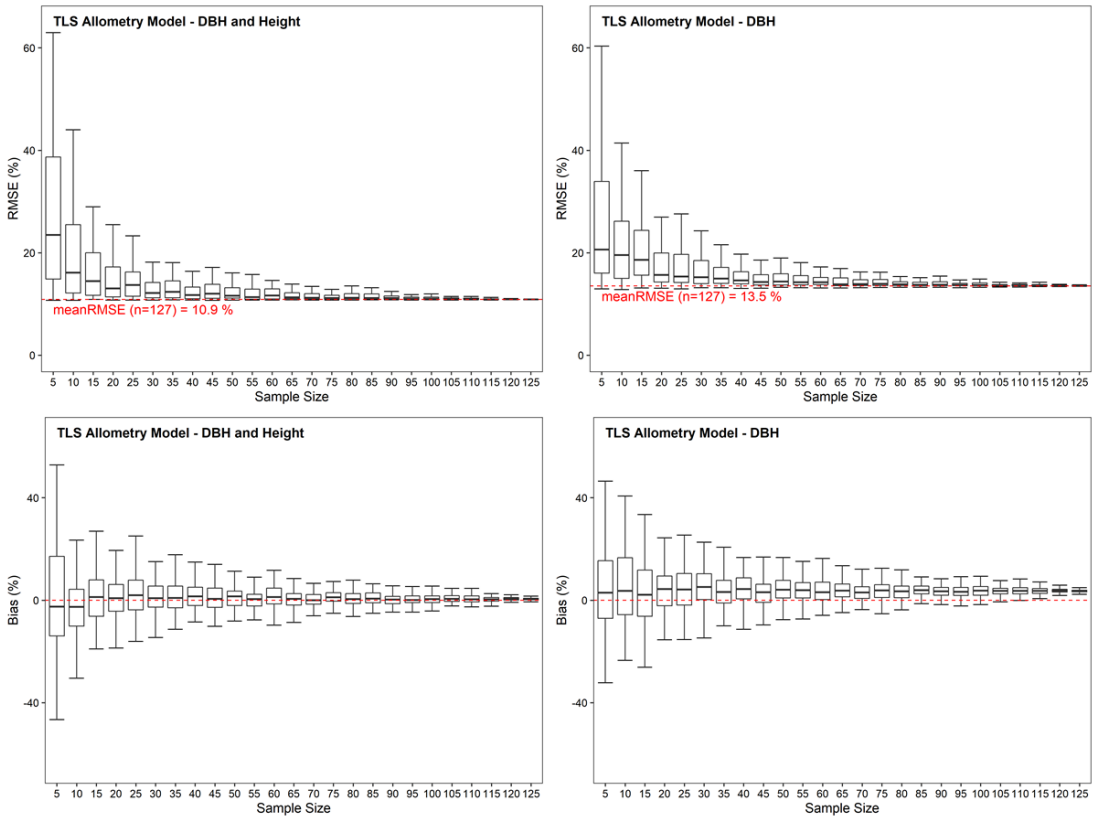


*Figure 4.10. Box-plot variations of TLS-QSM allometry in terms of RMSE and bias using [left] diameter-height and [right] diameter as predictor variables with respect to varying sample size. The red dotted line indicates the mean-RMSE at the maximum sample size.*

Figure 4.10 illustrates the impact of sample size on model performance through boxplots of RMSE and bias variations across 1000 simulations. Both diameter-only and diameter-height models exhibit a clear trend of stabilizing RMSE and bias errors as sample size increases beyond 50-55 samples. With over 100 samples, the median RMSE across simulations falls below 1% of the mean RMSE obtained using the entire dataset, with an interquartile range (IQR) of less than 0.5%. Notably, the inclusion of height alongside diameter (diameter-height models) does not significantly alter this stabilization pattern.

These findings differ from Stovall et al. (2018), where diameter-height models achieved stability at 100 samples compared to 200 samples for diameter-only models. This discrepancy might be attributed to inherent co-linearity between height and diameter, with diameter capturing most of the volume variation. However, a consistent, systematic median bias is observed in the diameter-only models with increasing sample size, while the bias in diameter-height models remains close to zero. Consequently, diameter-height models demonstrate comparatively lower RMSE and are preferable for minimizing uncertainty in plot-level biomass predictions.

Our findings emphasize the importance of adequate sample size to minimize prediction errors and bias in tree volume and subsequent biomass estimations. These results align with prior studies demonstrating high variability and consistent bias at low sample sizes, using both non-destructive TLS sampling (Stovall et al., 2018) and destructive measurements (Roxburgh et al., 2015). Given the efficiency of TLS for non-destructive tree volume acquisition, this technology offers the potential to develop or improve numerous allometric equations, particularly through targeted measurements stratified by specific geographic regions.

In practice, diameter-only volume equations are often preferred due to the readily available nature of diameter measurements in forestry practices. Height measurements can be less reliable and often unavailable in existing plots due to field measurement challenges (Anderson-Teixeira et al., 2015). However, our analysis clearly demonstrates the value of height data in improving overall prediction quality. A recent study by Sullivan et al. (2018) highlights that including even a relatively small number of individual tree heights (as few as 20) in tropical allometry can significantly improve predictive power compared to regional or climate-based equations. Thus, TLS-based tree volume equations offer a significant advantage by overcoming traditional limitations associated with achieving high sample sizes and overcoming difficulties in measuring basic parameters like diameter and height. This technology facilitates efficient and non-destructive tree volume measurements.

### 4.4.3 Comparison with Other Allometric Equations

TLS-derived allometric models outperformed traditional volume equations when applied at the individual tree level (Figure 4.9). Traditional models significantly overestimated biomass and exhibited higher uncertainty, particularly for large diameter trees (>35-40 cm) in our study region (Figure 4.9). This highlights a key limitation of applying these equations at a fine scale. Most national/pan-tropical allometric equations are built using data from diverse locations and are not intended for small-scale application. In our study site, commonly used allometric equations for the Indian region performed adequately for smaller diameter trees, but both variability and bias increased significantly for larger diameter trees. This discrepancy might be attributed to the limited availability of samples within the higher biomass range during model development, potentially leading to reduced model accuracy in these classes.

The observed systematic bias in applying national-scale allometry at the local level, particularly for large diameter trees, warrants further investigations. Future studies could explore how incorporating regional data or stratifying models by diameter class can improve the accuracy of biomass estimates. This highlights the advantage of TLS-based approaches, which can efficiently collect

data on a large number of trees within a specific region, potentially leading to more accurate and regionally-specific biomass models.

## 4.5 Summary

This study successfully evaluated the effectiveness of TLS point clouds for generating accurate, non-destructive tree volume estimates at the individual tree level. We established a framework for developing local allometric models, applicable to both site-level and species-level analyses in central Indian dry deciduous forests. These models utilize diameter and diameter-height as predictor variables for tree volume computation. Validation demonstrated the superiority of TLS allometric models compared to traditional methods. Notably, the TLS models achieved lower RMSE (11.1% with diameter and height; 13.5% with diameter only) and higher CCC (0.98 and 0.94, respectively) for tree volume estimation. Traditional models, on the other hand, exhibited significantly higher RMSE (>30%). This suggests that locally derived allometric models based on TLS data outperform traditional approaches for individual tree volume predictions. The higher errors observed in traditional models for larger diameter classes likely stem from potential under-sampling of these trees during model development.

A key limitation of the current TLS-based volume estimation method lies in the parameterization of the TreeQSM approach. This method relies on coincident destructive sampling measurements, and in this study, it was specifically standardized using teak trees within the study area. While the independent validation dataset helps mitigate substantial bias or errors in the estimated volume, potential biases might still be introduced due to variations in tree architecture across different species. Ideally, future studies would incorporate coincident TLS and destructive samples from a wider range of species to optimize TreeQSM parameterization for diverse tree morphologies and eliminate potential errors associated with this approach.

This study highlights a significant advantage of TLS-based QSM for tree volume estimation. Traditional allometric models often suffer from limitations related to sample size requirements. Many models developed with less than 100 samples exhibit substantial bias in tree-level volumes and aboveground biomass (AGB) estimations. The strength of the TLS-based QSM approach lies in its demonstrated robustness for volume estimation without prior knowledge of tree structure. This eliminates a major constraint of traditional allometric models and facilitates the non-destructive inclusion of large trees, which are often under-represented due to practical limitations associated with traditional measurement techniques. By overcoming these limitations, TLS-based QSM paves the way for the development of improved, site-specific allometric models with reduced uncertainty in biomass mapping. This could lead to more accurate estimations of current and past forest carbon stocks.

### 4.5.1 Influence of allometric model at 1-ha plot sizes

While allometric models might exhibit some error at the individual tree level, these errors tend to cancel out as the number of trees in a plot increases (Chave et al., 2004; Réjou-Méchain et al., 2019). This is because the errors are often random and independent of each other. Our study confirms this by analyzing above-ground biomass (AGB) estimates at large field plots (1-hectare and greater) (Figure 4.11).

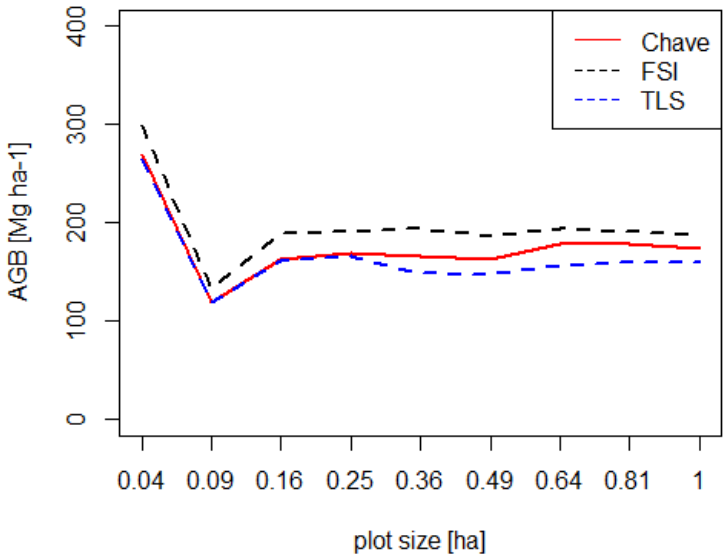


*Figure 4.11. Variability in Plot-level AGB for one plot with increase in plot size as per the choice of allometric model for the study site - Betul, Madhya Pradesh*

We have carried out bias estimation in AGB (by using reported wood density values; Table 4.3) with respect to the traditionally used volume equations (Table 4.3) and the TLS based allometric models at 13 large field plots (1-ha and greater sizes; see Section 3.2 for more details on field plots at Betul). Since the species-specific TLS models were developed using sample sizes of less than 30 for most cases, only all species based TLS models were employed in this analysis. The observations make it clear that the TLS allometric models (both DBH only and DBH+H) are very close with each other at plot level. The FSI allometric based estimates were observed to be in ±10% bias with the TLS-allometry based estimates (Figure 4.12). However, larger biases at plot-level estimations are observed with the pan-tropical model (Chave et al., 2014) compared to TLS-allometry based estimates (Figure 4.12). The plot-level biases using Chave model is strongly linked to the diameter distributions. This highlights a key advantage of using larger plots (~1 hectare): the random errors associated with individual tree models tend to average out, providing more consistent estimates. Further, our findings support that the bias between allometric models was significantly lower for the 32-ha plot compared to smaller plots. However, they impractical to establish due to logistical challenges and costs.

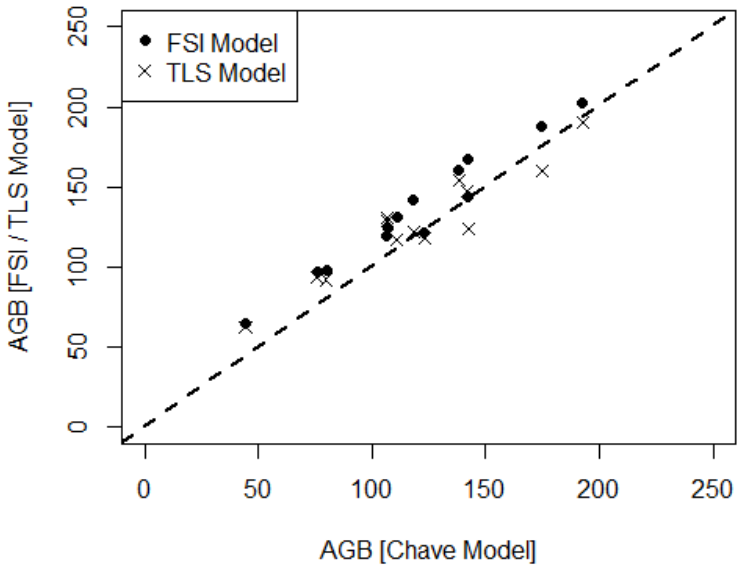


*Figure 4.12. Comparison of Plot-level AGB for all plots using different allometric models available for Betul, Madhya Pradesh*

For other study sites (Achanakmar and Yellapur; see Section 3.2), we have made a similar analysis comparing the plot-level AGB estimations between FSI model and the Chave model. At 1-ha plot size, the plot-level AGB showed minimal variability between the two traditionally available models over Indian tropical forests.

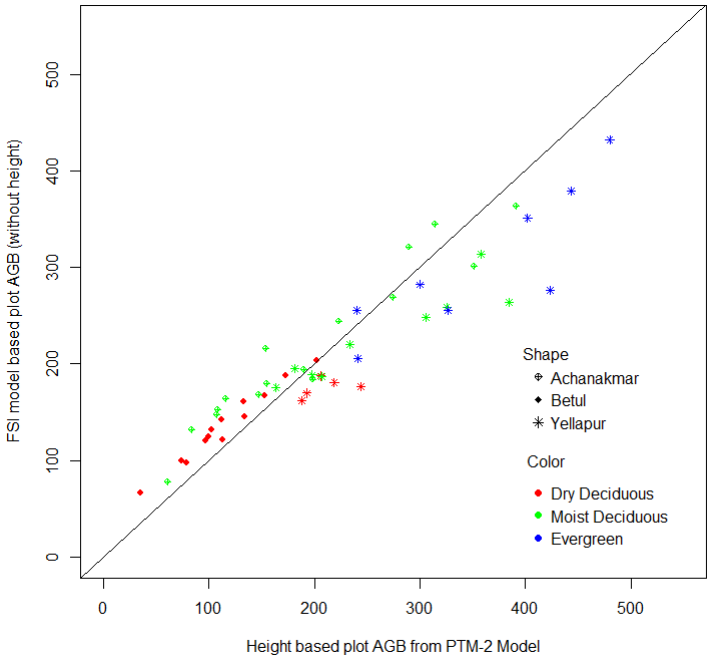


*Figure 4.13. Plot-level (1-ha) comparison of AGB estimates across distinct forest types and sites using traditionally available allometric models*

### 4.5.2 Suitability of Allometric Models for Spatial AGB Estimates

Allometric models, while imperfect for individual trees, become more reliable for estimating plot-level AGB with increase in plot sizes (typically 1 hectare and above) due to error cancellation. This section guides the selection of the most suitable model for spatial AGB assessments across India's diverse forests. Three primary options exist (Figure 4.14):

- Existing FSI Volume Equations: Locally-developed at species level. They are readily available. However, these equations lack uncertainty estimates and focus on merchantable tree volume, introducing inaccuracies when converted to total biomass.
- Pan-Tropical Chave Model (2014): When locally-derived models are unavailable, Chave's model offers a strong alternative. It provides a good balance between accuracy and applicability across various forest types. Additionally, it incorporates uncertainty estimates, critical for generating reliable spatial AGB maps.
- Locally Developed TLS-based Models: Ideally, non-destructive volume equations derived from Terrestrial Laser Scanning (TLS) point clouds should be prioritized. These models offer superior accuracy due to: (a) Site-specific calibration and (b) Ability to capture large trees, often underestimated by traditional methods and (c) detailed uncertainty estimates.
    - However, their current limitation lies in availability. Locally developed TLS models are not yet extensive, restricting their large-scale implementation.

| FSI Model | Chave Model | Local TLS Model |
|---|---|---|
| ✓ Generated using Local species level destructive sample datasets<br>✓ Available across different districts of India<br><br>X No uncertainty estimates<br>X Designed only for merchantable tree volume | ✓ Generated using destructive sample datasets across various species (n>4000) in tropics<br>✓ Uncertainty estimates<br><br>X Needs a Height measurements or H-D model | ✓ Generated Locally using TLS non-destructive volume estimation<br>✓ Uncertainty estimates<br><br>X Not extensively available across various locations |
| Though variations exist, the AGB of plots with 1-ha was found to be in error range of the model used across tropical forests | | |
| Not preferred due to lack of uncertainty estimates for propagating spatial uncertainty | Preferred as consistent with global maps | Highly preferred but not available |

*Figure 4.14. Suitability Assessment of Allometric Models for Spatial Above-Ground Biomass (AGB) Estimation in Indian Forests*

**Recommendations for selection of allometric models towards improved spatial AGB estimation over Tropical Indian forests.**

A. Prioritize locally developed TLS models when available for superior accuracy.
B. Use the Pan-Tropical Chave Model (2014) for large-scale applications due to its global consistency, uncertainty estimates, and broad applicability.
C. The use of FSI models could be limited due to limitations in scope and lack of uncertainty estimates.

Prioritizing locally developed TLS models offers the most accurate approach. However, their current limitations necessitate alternative models like Chave's for large-scale applications. FSI models, while convenient, are less suitable due to missing uncertainty estimates and limited scope. Future efforts should focus on expanding the availability of locally developed TLS models for broader implementation and enhanced spatial AGB assessments across India.

Statistically significant differences were observed in individual tree biomass estimates based on the chosen model, highlighting the importance of selection. Plot size also demonstrated a substantial effect, especially for plots smaller than 0.5 hectares, due to ground measurement errors. At the 1-hectare scale, the Chave model emerged as the preferred option due to its advantages. Locally developed TLS models hold promise but require wider implementation.

# Chapter 5

# Generating High-quality reference LiDAR AGB maps with uncertainty estimates

Understanding the global carbon cycle and combating climate change hinges on accurately measuring the carbon stored in tropical forests. However, existing satellite-derived AGB maps produce markedly divergent estimates and are afflicted with uncertainties. While promising new AGB mapping satellite missions like GEDI and BIOMASS are on the horizon, their accuracy relies on high-quality reference data (Duncanson et al., 2019).

This chapter addresses this challenge to develop high-quality reference AGB data for improving large-scale Spatial AGB estimates for accuracy. Our method combines two key data sources: detailed three-dimensional information of the forest canopy captured by Aerial LiDAR, and ground-truth measurements of AGB within specific areas obtained from well-characterized field plots. Additionally, our approach incorporates Monte Carlo simulations to generate uncertainty maps alongside each AGB map, to help calibrating EO based AGB mapping missions. Initially focused on three Indian sites, collaboration expanded the research to encompass a total of 13 sites across South Asia and Central Africa. The resulting LiDAR-derived AGB reference maps are openly available at both 100m and 40m resolutions, covering individual airborne LiDAR footprints ranging from 100 to 40,000 hectares.

## 5.1 Introduction

Tropical forests play a vital role in the Earth's carbon cycle and contribute largely to uncertainties in the global carbon budget (Mitchard et al., 2013). Methods to accurately map and monitor tropical forest carbon – or aboveground biomass (AGB) – are thus urgently needed to improve Earth system models and to help design carbon emission mitigation strategies in the context of Reducing Emissions from Deforestation and forest Degradation (REDD+) (Herold et al., 2019; Schimel et al., 2015). In the last decade, spaceborne Earth Observation (EO) data in combination with forest inventory measurements have been extensively used to generate spatially continuous AGB maps at pan-tropical scale using different modelling strategies (Dubayah et al., 2020; Herold et al., 2019; Mitchard

et al., 2014a; Réjou-Méchain et al., 2019). However, existing broad-scale maps show divergent estimates among themselves and differ from field-derived forest AGB stocks at different spatial scales (Fararoda et al., 2021; Mitchard et al., 2013, 2014a; Réjou-Méchain et al., 2019), indicating the presence of high uncertainties in prediction maps. To improve the accuracy and reliability of AGB maps over the tropics, several ongoing and upcoming EO missions (NASA's GEDI, ESA's BIOMASS, NASA-ISRO's NISAR and JAXA's ALOS-4 missions, notably) have been specifically designed to collect satellite data sensitive to forest structure, hence to forest AGB (Amelung and others, 2019; Dubayah et al., 2020; Motohka et al., 2021; Quegan et al., 2019). While these new spaceborne datasets will undoubtedly revolutionise broad-scale forest AGB mapping, a network of high-quality reference data is needed to calibrate and validate the mapping algorithms (Chave et al., 2019; Labrière et al., 2023). Besides, using the same sets of reference data across different EO missions would vastly improve the comparability and confidence in the derived AGB maps, enabling their use in a wide range of science, policy, and management applications (Duncanson et al., 2019).

The establishment and long-term maintenance of a network of reference forest AGB observatories across the tropics entails a myriad of challenges, particularly concerning the representativeness of the network (Labrière et al., 2023). Ideally, the network should be relatively evenly distributed in space and cover the main environmental gradients. While scientific discussions on site selection are on-going (Labrière et al., 2023), the Global Ecosystem Dynamics Investigation (GEDI) sensor on-board the International Space Station has already acquired data for a longer period than its initially projected lifetime. Data users would benefit from open-access reference AGB data, particularly in Asia where large geographic regions are not represented in the calibration/validation dataset of GEDI biomass mapping algorithm (Duncanson et al., 2022; Rodda et al., 2023). Besides the notion of spatial representativeness, hurdles related to the temporal mismatch between reference AGB and EO data should not be neglected. Rapid growth in regenerating forests or forest clearing/degradation – which notably characterise rural landscapes around central African cities, where slash-and-burn agriculture induces relatively fast dynamics – could rapidly make tens of thousands of GEDI data shots unusable. We argue that airborne LiDAR data acquired during GEDI lifetime over rapidly changing landscapes are invaluable and should be utilized to improve GEDI biomass mapping algorithms, notably on the lower-end of the forest biomass gradient to best capture forest degradation and regeneration gradients.

In this context, we aim to generate reference biomass datasets over the tropics for eight sites in Central Africa and five sites in South Asia (Figure 5.1) by calibrating airborne LiDAR data with locally established field plots. This chapter briefly describes (1) the details of the study sites and the datasets used, (2) the methodology used to generate the reference AGB maps (Figure 5.2), and (3) the

Monte Carlo simulation workflow used to generate uncertainty maps along with each reference AGB map. Finally, this paper provides access to these reference AGB datasets generated at 100 m and 40 m spatial resolutions over airborne LiDAR footprints ranging from 100 to 40,000 ha.

## 5.2 Materials and Methods

### 5.2.1 Sampling sites and associated inventory and LiDAR datasets

The co-located forest inventory and LiDAR datasets were compiled from 13 sampling sites in Central Africa and South Asia encompassing an array of abiotic conditions, forest types and structures (Figure 5.1, Table 5.1 and Table 5.2). Forest inventories were carried out at each site, and LiDAR datasets were obtained with an absolute temporal difference of 2.2 ± 1.9 years (range: 0 – 6.2 years) from the field measurements.

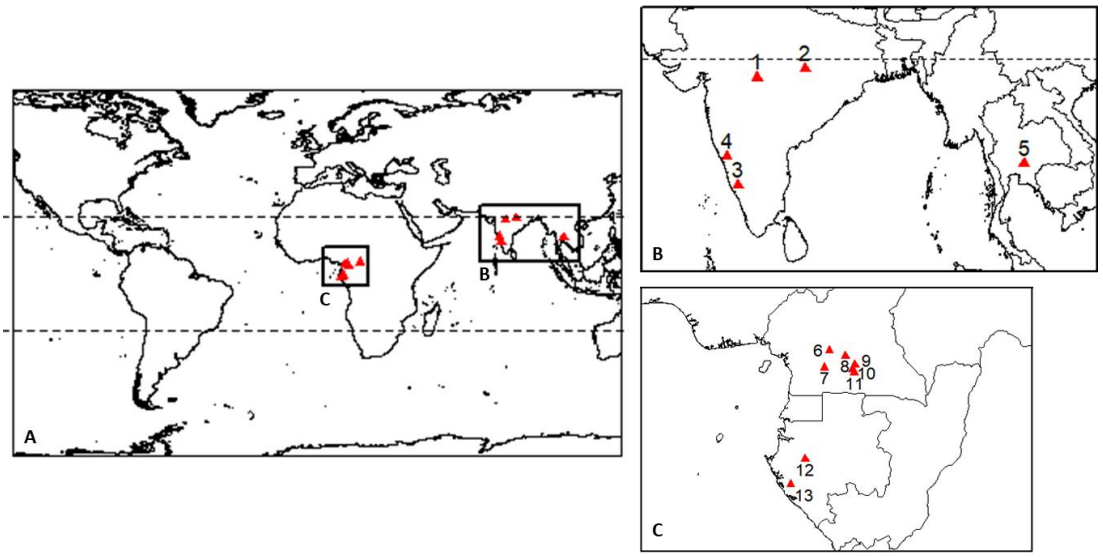


*Figure 5.1. (A) Overview map showing the locations of sampling sites (n = 13) used in the current study. Outlined regions are expanded in (B): South Asian region and in (C): Central African region.*

Forest inventories were conducted by different teams but followed similar protocols. In each plot, the diameter at breast height (DBH or referred to as D in this study, with D ≥ 10 cm) and the taxonomic identification of each tree were recorded. Tree relative coordinates within the plots were measured either at the individual or at the 20 x 20 m quadrat level. For a subsample of trees within the plots, tree height (H) was measured using a laser rangefinder device. Finally, plot geographic coordinates were determined using points measured every 20 m along the plot borders using a combination of differential GPS measurement system and electronic total station (in Asia) or a regular GPS system (in Africa), to warrant an accurate link between ground and remote-sensing data.

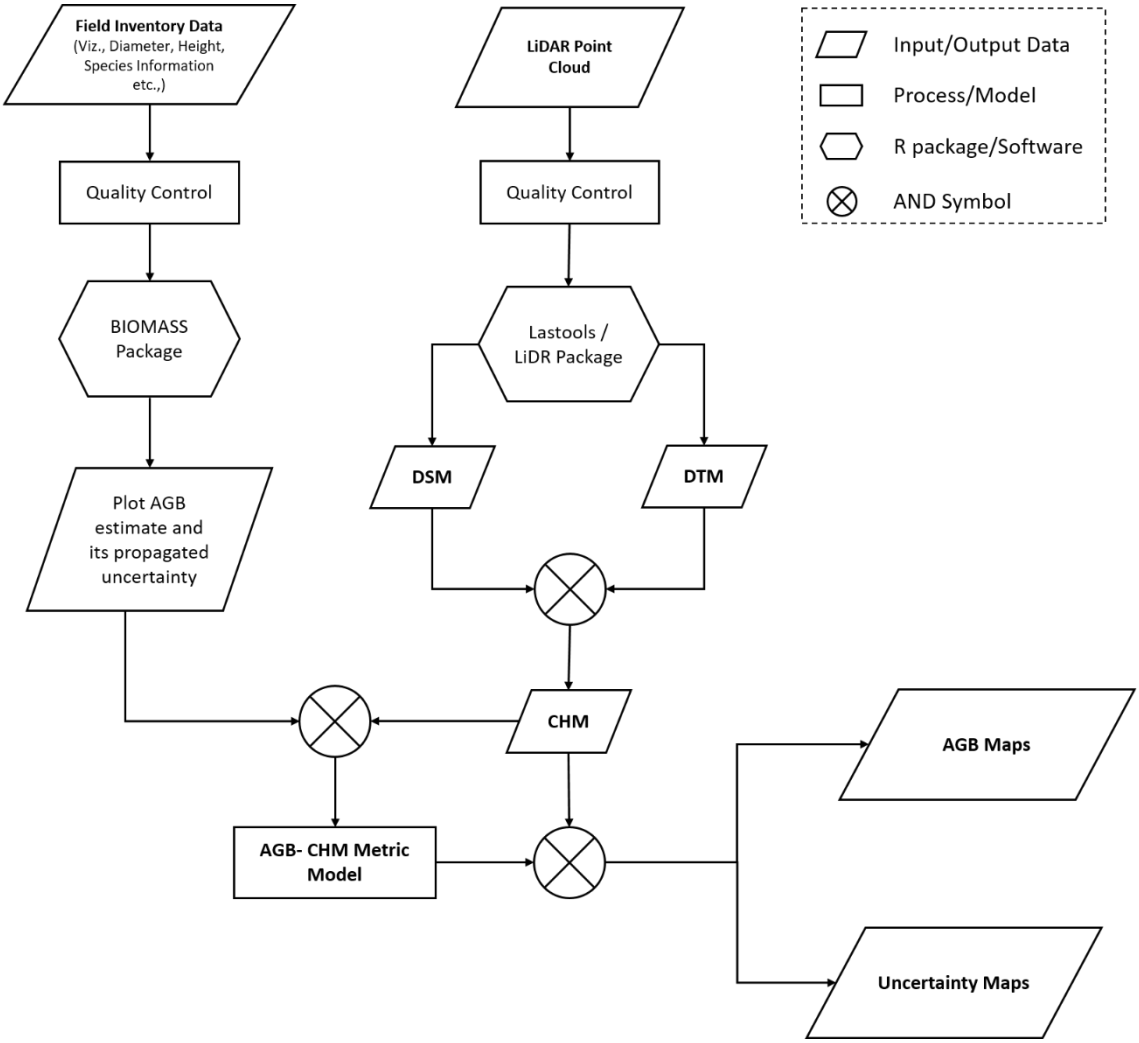


*Figure 5.2. Flowchart depicting workflow of the data analysis procedure to generate reference AGB datasets.*

The complete inventory dataset includes information on D and H measurements for respectively 97,251 and 13,303 trees, and identification rates of 89% at the species level and 92% at the genus level (8% of the trees were left unidentified). The number, size and layout of the inventory plots are uneven across sampling sites with, e.g., a single large 25-ha plot in the Forest-Geo "Rabi" site, a large 30-ha plot and smaller plots of 1-ha and 0.48-ha in the "Khao Yai" site, or a varying number of scattered 1-ha plots (ranging from 2 to 16 in the "Atout" and "Achanakmar" site, respectively). In general, the inventoried extent per site is smaller in Africa (9 ± 8 hectares) than in Asia (27 ± 13 hectares). For a breakdown of plot number, size, tree measurements and identification rates per sampling site, please refer to Table 5.3.

The LiDAR data at each sampling site were acquired between 2012 and 2022 using either aircraft or unmanned aerial vehicles (UAV, Table 5.2).

*Table 5.1. Environmental conditions across sampling sites. Over the LiDAR acquisition area, the statistics (mean ± standard deviation) of elevation and slope are computed using SRTM at 30-m spatial resolution (V3 product), Mean Annual Temperature (MAT) and Mean Annual Precipitation (MAP) are computed using WorldClim Version 2.1 data.*

| S No. | Site name | Region | Elevation | Slope | MAT | MAP |
|---|---|---|---|---|---|---|
| 1 | Betul | S-As | 487 ± 48 | 7.8 ± 5.7 | 25.6 ± 0.3 | 1266 ± 27 |
| 2 | Achanakmar | S-As | 813 ± 165 | 8.8 ± 6.8 | 23.3 ± 0.7 | 1328 ± 42 |
| 3 | Yellapur | S-As | 459 ± 110 | 8.7 ± 7.8 | 24.4 ± 0.5 | 2383 ± 421 |
| 4 | Uppangala | S-As | 377 ± 122 | 18.6 ± 8.5 | 25.1 ± 0.5 | 3789 ± 63 |
| 5 | Khaoyai | S-As | 757 ± 34 | 6.8 ± 4.2 | 23.3 ± 0.1 | 1127 ± 7 |
| 6 | Nachtigal | C-Af | 527 ± 50 | 4.5 ± 3.4 | 24.4 ± 0.5 | 1588 ± 17 |
| 7 | Mbalmayo | C-Af | 662 ± 11 | 5.2 ± 3.3 | 23.6 ± 0.1 | 1706 ± 1 |
| 8 | Atout | C-Af | 715 ± 11 | 5.2 ± 3.0 | 23.7 ± 0.1 | 1572 ± 1 |
| 9 | Kompia | C-Af | 705 ± 20 | 5.7 ± 3.3 | 23.5 ± 0.1 | 1606 ± 2 |
| 10 | Somalomo | C-Af | 656 ± 13 | 4.7 ± 3.0 | 23.7 ± 0.1 | 1602 ± 3 |
| 11 | Bouamir | C-Af | 696 ± 14 | 5.3 ± 3.1 | 23.5 ± 0.1 | 1616 ± 2 |
| 12 | Mabounié | C-Af | 88 ± 31 | 7.3 ± 4.6 | 26.1 ± 0.1 | 2034 ± 16 |
| 13 | Rabi | C-Af | 68 ± 15 | 6.3 ± 3.8 | 25.6 ± 0.1 | 1826 ± 4 |

*Table 5.2. Sampling site details on forest types, inventory statistics and characteristics of the LiDAR acquisitions. AreaINV indicates the total area of field inventories, LiDARDate indicates the month and year of acquisition of LiDAR data and LiDARArea indicates the total area covered by LiDAR data over the site. Nrange and BArange indicate the range in number of trees and basal area per hectare across the inventoried area, respectively. The associated plant functional types (PFT's) for each site are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type product (MCD12Q1) which follows Land Cover Type 5 Classification Scheme; a similar strategy is adopted by GEDI Mission.*

| Sno | Site name | Region | PFT | $INV_{Date}$ | $Area_{INV}$ | $N_{range}$ | $BA_{range}$ | $LiDAR_{Date}$ | $LiDAR_{Area}$ | Platform |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Betul | S-As | DBT, GSW | 02-2016 to 03-2016 | 46 | 150 - 437 | 7.2 - 20.2 | 04-2014 | 10000 | Aircraft |
| 2 | Achanakmar | S-As | EBT, DBT | 12-2017 to 01-2018 | 16 | 186 - 509 | 10.0 - 31.3 | 10-2015 | 11000 | Aircraft |
| 3 | Yellapur | S-As | EBT, DBT, GSW | 01-2015 & 01-2016 | 15 | 140 - 749 | 20.3 - 43.6 | 11-2017 | 40000 | Aircraft |
| 4 | Uppangala | S-As | EBT | 03-2013 & 03-2014 | 23 | 370 - 838 | 10.3 - 62.4 | 11-2013 | 900 | Aircraft |
| 5 | Khaoyai | S-As | EBT | 11-2015 to 04-2018 | 34.84 | 394 - 1015 | 20.8 - 38.1 | 04-2017 | 4500 | Aircraft |
| 6 | Nachtigal | C-Af | DBT, GSW | 02-2018 to 05-2018 | 13.16 | 216 - 538 | 23.4 - 35.0 | 01-2012 | 25000 | Aircraft |
| 7 | Mbalmayo | C-Af | EBT | 04-2021 to 07-2021 | 9 | 442 - 596 | 20.1 - 32.5 | 02-2023 | 400 | UAV |
| 8 | Atout | C-Af | EBT | 07-2017 | 2 | 475 - 574 | 30.8 - 32.0 | 11-2021 | 150 | UAV |
| 9 | Kompia | C-Af | EBT | 12-2018 | 2 | 423 - 519 | 29.0 - 32.3 | 12-2019 | 400 | UAV |
| 10 | Somalomo | C-Af | EBT | 09-2022 | 8 | 325 - 508 | 25.4 - 33.6 | 04-2022 | 800 | UAV |
| 11 | Bouamir | C-Af | EBT | 12-2018 | 4 | 362 - 552 | 24.9 - 35.8 | 12-2018 | 1000 | UAV |
| 12 | Mabounié | C-Af | EBT | 04-2012 to 10-2012 | 11 | 222 - 492 | 17.7 - 31.4 | 11-2007 | 16000 | Aircraft |
| 13 | Rabi | C-Af | EBT | 2015 | 25 | 392 - 532 | 20.5 - 36.6 | 10-2012 | 1000 | Aircraft |

*Table 5.3. Site-level details on field plot layout description, the number of compiled plots at 1-ha (N1ha) and 0.16-ha (N0.16ha), number of total trees across all plots (NTrees), number of trees measured for height (NTree_hts). Species, Genus and Family (%) stands for the identification rate (in %) at the given taxomonic level.*

| No | Site | Field Plots description | $N_{1ha}$ | $N_{0.16ha}$ | $N_{Trees}$ | $N_{Tree\_hts}$ | Species [%] | Genus [%] | Family [%] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Betul | Single plot of 34-ha and 12 distributed plots of 1-ha | 34 | 227 | 15672 | 677 | 95% | 96% | 96% |
| 2 | Achanakmar | 16 distributed plots of 1-ha | 16 | 64 | 5750 | 1546 | 88% | 88% | 88% |
| 3 | Yellapur | 15 distributed plots of 1-ha | 15 | 60 | 8519 | 4932 | 97% | 97% | 97% |
| 4 | Uppangala | Single plot of 10-ha and 13 distributed plots of 1-ha | 23 | 112 | 14967 | 1729 | 61% | 61% | 61% |
| 5 | Khaoyai | Single plot of 30-ha, 1 plot of 1-ha and 8 plots of 0.48ha | 31 | 192 | 19461 | 517 | 100% | 100% | 100% |
| 6 | Nachtigal | 11 distributed plots of 1-ha & 18 distributed plots of 0.16 ha | 11 | 62 | 4315 | 703 | 86% | 96% | 97% |
| 7 | Mbalmayo | 9 distributed plots of 1-ha | 9 | 36 | 4378 | 532 | 74% | 97% | 99% |
| 8 | Atout | 2 distributed plots of 1-ha | 2 | 8 | 1049 |  | 96% | 99% | 100% |
| 9 | Kompia | 2 distributed plots of 1-ha | 2 | 8 | 942 | 49 | 95% | 98% | 98% |
| 10 | Somalomo | 8 distributed plots of 1-ha | 8 | 32 | 4564 | 215 | 93% | 99% | 99% |
| 11 | Bouamir | 4 distributed plots of 1-ha | 4 | 16 | 1784 | 171 | 92% | 98% | 98% |
| 12 | Mabounié | 11 distributed plots of 1-ha | 11 | 44 | 4425 | 570 | 93% | 100% | 100% |
| 13 | Rabi | Single plot of 25-ha | 25 | 144 | 11425 | 1662 | 94% | 100% | 100% |

### 5.2.2 Inventory data processing: computation of reference AGB predictions

Forest inventories were first split into 1-ha (i.e. 100 x 100 m) and 0.16-ha (i.e. 40 x 40 m) plots, using information on tree location recorded in the field (i.e. either individual tree location or quadrat number). The two plot sizes correspond to the two mapping resolutions considered in this study. The 40-m resolution was chosen to account for plots where individual tree locations were only recorded at 20 x 20 m quadrat-level. In cases where the original plot size was not a multiple of the desired output size (typically when splitting 100 x 100 m plots into 40 x 40 m plots), subplots of the desired outputs size were selected at the edges of the original plot, thus leaving-out parts of the original inventory dataset (20 m wide bands in the center as per the previous example). The resulting number of 1-ha and 0.16-ha plots compiled at each sampling site is provided in Table 5.3.

*Table 5.4. H-D Model coefficients (a, b, c) of the 2nd order log-log polynomial model form ($ln(H) = a + b \times (ln(D)) + c \times ln(D^2) + \varepsilon$), where H is the height of the tree and D is the tree diameter. $\varepsilon$ is the normally distributed error to be used during back-transformation for Baskerville correction*

| Sno | Site | a | b | c | sigma [$\varepsilon$] | $R^2$ | RMSE [m] |
|---|---|---|---|---|---|---|---|
| 1 | Betul | -1.163 | 1.857 | -0.194 | 0.148 | 0.71 | 2.37 |
| 2 | Achanakmar | -1.314 | 1.770 | -0.159 | 0.250 | 0.63 | 4.06 |
| 3 | Yellapur | 0.563 | 0.851 | -0.050 | 0.231 | 0.60 | 3.75 |
| 4 | Uppangala | 0.164 | 1.123 | -0.077 | 0.249 | 0.67 | 4.33 |
| 5 | Khaoyai | 1.094 | 0.479 | 0.025 | 0.329 | 0.66 | 4.90 |
| 6 | Nachtigal | -0.446 | 1.457 | -0.127 | 0.223 | 0.72 | 5.06 |
| 7 | Mbalmayo | -0.081 | 1.286 | -0.104 | 0.243 | 0.71 | 4.61 |
| 8 | Atout | | | | | | |
| 9 | Kompia | | | | | | |
| 10 | Somalomo | | | | | | |
| 11 | Bouamir | | | | | | |
| 12 | Mabounié | 0.283 | 1.102 | -0.083 | 0.236 | 0.66 | 5.47 |
| 13 | Rabi | 1.234 | 0.614 | -0.022 | 0.224 | 0.58 | 4.46 |

Subsequently, the BIOMASS R package (Réjou-Méchain et al., 2017) (version 2.1.8) within the R statistical platform (version 4.1.3) was used to compute reference AGB predictions for forest inventory plots at the two spatial resolutions (1-ha and 0.16-ha). To that end, we differentiated sites with a cumulated forest inventory area of 10 ha or more (i.e., 8 out of 13 sites, Table 5.2 and Table 5.3) from those with less than 10 ha of cumulated forest inventory area (i.e., 5 sites). In the former case, we developed site-specific tree height-diameter (H-D) allometric models using second-order polynomials on log-transformed data (modelHD function in the BIOMASS package) and these models were used to predict the height of trees without H measurements in each respective site. In the latter case, which pertained to sites located in moist dense forests of Cameroon (SiteIDs 7 to 11 in Table 5.2), all inventory data from that country and biome were pooled into a single training dataset and the same H-D modelling procedure was applied. The resulting country- and biome-specific model was then used for predicting tree height at those sites. The H-D model coefficients for these site-level and Cameroon level model are presented in Table 5.4. Next, a wood density (WD) estimate was attributed to each tree based on its taxonomic identification using the getWoodDensity function.

Considering that tree AGB prediction is associated with various sources of uncertainty (including measurement errors of the independent variables such as tree diameter, height, and wood density, as well as prediction errors of the H-D models and the AGB allometric model) (Duncanson et al., 2021; Réjou-Méchain et al., 2019), we used a Monte Carlo approach for uncertainty propagation. Specifically, we employed the AGBmonteCarlo function of the BIOMASS package (Réjou-Méchain et al., 2017), which allows propagating the above-mentioned sources of uncertainty and outputs 1000 tree-level and subsequently plot-level AGB predictions. Tree AGB predictions were made using the pantropical AGB allometric model (i.e., Equation-4 in Chave et al., (2014)). For each plot, the 1000 AGB predictions were (i) averaged to obtain a reference plot-level AGB density (hereafter $AGB_{REF}$) for the development of LiDAR-AGB models and (ii) used for the propagation of uncertainties to the final AGB maps (see section "mapping forest AGB and prediction uncertainty").

### 5.2.3 LiDAR data processing: computation of canopy height metrics

LiDAR data from African and Asian sites were processed using LAStools (version 201124) and the lidR R package (version 4.0.1), respectively. The same processing chain was applied to generate the canopy metrics in both cases. First, a digital surface model (DSM) free of pits and spikes was generated at a 1-m resolution by interpolating the highest points on a 1-m grid. Second, a ground point classification was performed on the point cloud and a digital terrain model (DTM) was interpolated from ground-points. The canopy height model (CHM) was then derived by subtracting the DTM from the DSM. Finally, the 1-m CHM

was used to compute 15 canopy metrics for each plot (Table 5.5) as candidate predictors of forest AGB.

*Table 5.5. List of Canopy Metrics derived from LiDAR-derived CHMs over the forest plots extent*

| **LiDAR Canopy Metric (LCM)** | **Description** |
|---|---|
| H40 | Percentile of CHM values (ex. H98 for the 98th percentile, in m) |
| H50 | |
| H60 | |
| H70 | |
| H80 | |
| H90 | |
| H98 | |
| meanTCH or meanH | Mean of CHM values (in m) |
| sdH | Standard deviation of CHM values (in m) |
| CV | Coefficient of variation of CHM values (meanTCH divided by sdH) |
| QMCH | Quadratic mean of CHM values |
| CCF2 | Percentage of CHM values above 2, 5 and 10 m (in %) |
| CCF5 | |
| CCF10 | |
| rumple | Roughness of CHM surface (rumple_index function in lidR R package) |

### 5.2.4 Specification of a general AGB model form

While LiDAR-based AGB mapping models were trained at the site or regional level (for some Cameroon sites), to minimise local bias in model predictions (Asner et al., 2012; Duncanson et al., 2021), we privileged the use of (i) a single AGB model form across all sites to facilitate sites inter-comparison and the subsequent use of AGB predictions for spaceborne products calibration/validation and (ii) a simple, parametric modelling approach, keeping the number of predictors to a minimum to avoid overfitting and multicollinearity issues. To specify the AGB model form, we used linear mixed-effects models to identify the most predictive LiDAR-derived canopy height metrics (LCMs) on $AGB_{REF}$ variation while accounting for the hierarchical spatial structure of the data. In practice, 15 linear mixed-effects models (one for each LCM) were built on the log-transformed variables of $AGB_{REF}$ and LCM (Eq. (10):

$$\log(AGB_{REF}) = a + b \times \log(LCM) + RE_{site} + \varepsilon \quad (10)$$

where a and b are the model's coefficients, LCM represents the Lidar-derived Canopy Metric, $AGB_{REF}$ corresponds to the field-derived AGB prediction at a given spatial resolution (i.e. 0.16- or 1-ha), REsite denotes the random site effect used in linear mixed-effects modelling and $\varepsilon$ is the error term, assumed to follow a normally distribution with a mean of zero and a standard error $\sigma$. Based on the AIC criterion, the meanTCH metric (i.e. the mean of all CHM values in the plot area) emerged as the best predictor of $AGB_{REF}$ variation at both 1-ha and 0.16-ha spatial resolutions (Table 5.6).

A similar procedure was run on $AGB_{REF}$ prediction models combining each pair of LCMs rather than a single predictor. At both spatial resolutions, the best two-predictor model resulted in a modest improvement in relative RMSE (i.e., <0.2%, Table 5.7) compared to the model based on meanTCH only. The latter model form was thus selected for biomass mapping. In line with the H:D modelling procedure, LiDAR-based AGB mapping models were either trained at the site-level (for sites with a cumulated forest inventory area of 10 ha more) or on a pooled training dataset containing all inventory data from Cameroonian moist dense forests (for sites with a cumulated forest inventory area smaller than 10 ha), henceforth referred to as the "regional" AGB model. It is noteworthy that including sites as an additional fixed-effect covariate in the regional model did not yield significant effects for this variable at a 5% risk (neither in terms of site-level intercepts nor in terms of interactions between sites and the meanTCH predictor), suggesting a minimal site effect on the regional model's predictions, if any.

*Table 5.6. LiDAR-AGB Linear Mixed Effects Model performance statistics at 1-ha and 0.16-ha plot sizes. The table is sorted in ascending order based on the column "AIC" (Akaike information criterion) when the respective LiDAR Canopy Metric (LCM) is used for Eq. (10. $R^2$ and RMSEs (in Mg $ha^{-1}$ and in %) are computed on back-transformed predictions.*

| Plot Size (1-ha) | | | | | Plot Size (0.16-ha) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **LCM** | **AIC*** | **$R^2$** | **RMSE [Mg $ha^{-1}$]** | **RMSE [%]** | **LCM** | **AIC** | **$R^2$** | **RMSE [Mg $ha^{-1}$]** | **RMSE [%]** |
| meanH | -180.42 | 0.89 | 43.62 | 15.2% | meanH | 149.94 | 0.71 | 88.55 | 32.4% |
| RH60 | -169.22 | 0.88 | 44.75 | 15.6% | QMCH_chm | 194.11 | 0.70 | 89.31 | 32.7% |
| RH50 | -167.22 | 0.88 | 45.27 | 15.8% | RH90 | 379.89 | 0.65 | 96.81 | 35.4% |
| RH70 | -151.53 | 0.87 | 45.83 | 16.0% | RH98 | 642.43 | 0.59 | 105.29 | 38.5% |
| QMCH_chm | -148.49 | 0.88 | 45.30 | 15.8% | RH40 | 869.25 | 0.60 | 106.49 | 39.0% |
| RH80 | -133.52 | 0.87 | 47.14 | 16.5% | RH50 | 889.42 | 0.59 | 106.84 | 39.1% |
| RH40 | -114.58 | 0.86 | 48.46 | 16.9% | CCF10 | 928.45 | 0.50 | 116.11 | 42.5% |
| RH90 | -102.74 | 0.85 | 50.54 | 17.7% | RH60 | 975.37 | 0.59 | 106.68 | 39.0% |
| RH98 | -23.33 | 0.79 | 59.26 | 20.7% | RH80 | 1033.34 | 0.62 | 103.08 | 37.7% |
| CCF10 | 15.74 | 0.72 | 70.11 | 24.5% | RH70 | 1036.05 | 0.59 | 106.41 | 38.9% |
| CCF5 | 66.87 | 0.63 | 79.56 | 27.8% | CCF5 | 1101.04 | 0.46 | 120.43 | 44.1% |
| CV | 85.22 | 0.61 | 81.18 | 28.4% | CCF2 | 1105.63 | 0.45 | 121.80 | 44.6% |
| CCF2 | 90.21 | 0.60 | 82.20 | 28.7% | CV | 1265.00 | 0.41 | 132.29 | 48.4% |
| sdH | 114.75 | 0.62 | 79.52 | 27.8% | sdH | 1365.98 | 0.41 | 128.62 | 47.1% |
| rumple | 131.76 | 0.58 | 83.77 | 29.3% | rumple | 1473.62 | 0.39 | 130.34 | 47.7% |

*Table 5.7. LiDAR-AGB Linear Mixed Effects Model performance statistics at 1-ha and 0.16-ha plot sizes using two LCMs as predictive variables. The table is sorted in ascending order based on the column "AIC" (Akaike information criterion) when the respective LiDAR Canopy Metrics (LCM) are used in Eq. (10. $R^2$ and RMSEs (in Mg $ha^{-1}$ and in %) are computed on back-transformed predictions.*

| **Plot Size (1-ha)** | | | | | **Plot Size (0.16-ha)** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **LCMs** | **AIC** | **$R^2$** | **RMSE [Mg $ha^{-1}$]** | **RMSE [%]** | **LCMs** | **AIC** | **$R^2$** | **RMSE [Mg $ha^{-1}$]** | **RMSE [%]** |
| RH50 , RH90 | -192.69 | 0.89 | 43.75 | 15.3% | meanH , CCF2 | 114.03 | 0.71 | 88.00 | 32.2% |
| RH50 , RH80 | -191.69 | 0.88 | 43.84 | 15.3% | RH40 , meanH | 131.70 | 0.71 | 88.09 | 32.2% |
| RH50 , QMCH_chm | -188.18 | 0.88 | 43.87 | 15.3% | RH98 , QMCH_chm | 134.13 | 0.71 | 87.66 | 32.1% |
| RH50 , RH70 | -188.15 | 0.88 | 44.18 | 15.4% | meanH , CCF10 | 136.58 | 0.71 | 88.07 | 32.2% |
| RH50 , meanH | -186.09 | 0.89 | 43.67 | 15.3% | RH50 , meanH | 137.09 | 0.71 | 88.22 | 32.3% |
| RH40 , RH80 | -185.00 | 0.88 | 43.96 | 15.4% | meanH , QMCH_chm | 139.23 | 0.71 | 88.29 | 32.3% |
| RH40 , RH50 | -183.00 | 0.88 | 44.67 | 15.6% | RH70 , meanH | 139.36 | 0.71 | 88.44 | 32.4% |
| RH50 , RH60 | -182.72 | 0.88 | 44.37 | 15.5% | RH60 , meanH | 139.42 | 0.71 | 88.36 | 32.3% |
| RH40 , RH90 | -182.32 | 0.88 | 44.15 | 15.4% | meanH , CCF5 | 141.40 | 0.71 | 88.14 | 32.3% |
| RH40 , RH70 | -181.96 | 0.88 | 44.26 | 15.5% | RH90 , meanH | 147.02 | 0.71 | 88.54 | 32.4% |

*Table 5.8. Model coefficients along with standard errors (in brackets) for site-wise level models at 1-ha and 0.16-ha resolution. For Cameroon sites listed from 7-11 in column "Sno", a single regional model is employed. Sigma is the model residual standard error in log-transformed units. $R^2$ and RMSEs (in Mg $ha^{-1}$ and in %) are computed on back-transformed predictions.*

| Sno | Site | Plot Size (1-ha) | | | | | | Plot Size (0.16-ha) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | a (se) | b (se) | sigma | $R^2$ | RMSE [Mg $ha^{-1}$] | RMSE [%] | a (se) | b (se) | sigma | R2 | RMSE [Mg $ha^{-1}$] | RMSE [%] |
| **1** | **Betul** | 2.043 (0.191) | 1.247 (0.087) | 0.095 | 0.87 | 12.04 | 9.9 | 2.605 (0.107) | 0.988 (0.049) | 0.162 | 0.65 | 21.08 | 17.3 |
| **2** | **Achanakmar** | 2.046 (0.219) | 1.173 (0.080) | 0.113 | 0.94 | 23.51 | 11.7 | 2.058 (0.175) | 1.165 (0.064) | 0.185 | 0.84 | 36.58 | 18.3 |
| **3** | **Yellapur** | 0.500 (0.478) | 1.691 (0.158) | 0.111 | 0.90 | 31.07 | 11.0 | 0.684 (0.380) | 1.632 (0.126) | 0.188 | 0.74 | 66.65 | 23.1 |
| **4** | **Uppangala** | 0.969 (0.442) | 1.523 (0.134) | 0.192 | 0.86 | 82.16 | 18.9 | 1.087 (0.335) | 1.464 (0.100) | 0.328 | 0.66 | 145.85 | 32.6 |
| **5** | **Khaoyai** | 1.934 (0.445) | 1.236 (0.144) | 0.109 | 0.72 | 32.27 | 9.9 | 1.360 (0.189) | 1.411 (0.061) | 0.219 | 0.74 | 74.08 | 23.1 |
| **6** | **Nachtigal** | 2.009 (1.004) | 1.037 (0.314) | 0.233 | 0.55 | 43.21 | 18.6 | 1.902 (0.081) | 1.083 (0.029) | 0.284 | 0.82 | 48.28 | 27.9 |
| **7** | **Mbalmayo** | 1.721 (0.517) | 1.253 (0.155) | 0.135 | 0.74 | 52.62 | 14.1 | 1.230 (0.396) | 1.39 (0.119) | 0.282 | 0.59 | 108.54 | 29.1 |
| **8** | **Atout** | | | | | | | | | | | | |
| **9** | **Kompia** | | | | | | | | | | | | |
| **10** | **Somalomo** | | | | | | | | | | | | |
| **11** | **Bouamir** | | | | | | | | | | | | |
| **12** | **Mabounié** | 2.471 (0.447) | 1.015 (0.136) | 0.100 | 0.86 | 32.26 | 9.4 | 2.159 (0.498) | 1.098 (0.152) | 0.265 | 0.55 | 94.23 | 28.0 |
| **13** | **Rabi** | 1.267 (0.671) | 1.397 (0.213) | 0.123 | 0.65 | 38.78 | 13.1 | 1.386 (0.407) | 1.344 (0.130) | 0.312 | 0.43 | 99.63 | 33.8 |

### 5.2.5 Mapping forest AGB and prediction uncertainty

Forest AGB and prediction uncertainty were mapped over the extent of airborne LiDAR data at each site using a Monte Carlo approach similar to that used to compute plot-level $AGB_{REF}$. More specifically, we used the 1000 plot-level AGB predictions generated at the first modelling level (i.e., from tree to plot) to build 1000 LiDAR-based models per site (or at "regional" level for Cameroonian sites with less than 10 ha of cumulated forest inventory area). At the second modelling level (i.e., from plot to landscape), pixel AGB predictions derived from LiDAR-based models suffer from additional uncertainty associated to the LiDAR-based models themselves.

To propagate this additional uncertainty, we mimicked the procedure used in BIOMASS to propagate the uncertainty associated to the tree-level AGB allometric model (see Appendix S1 of Réjou-Méchain et al., (2017) for codes and details), which entailed using a Markov chain Monte Carlo algorithm to infer the uncertainty on Lidar-based models' parameters (i.e., models' coefficients and associated RSE). The Markov chain outputted 1000 sets of model parameters per model. For each of the 1000 LiDAR-based model at each site, we then (1) randomly selected a set of parameters among the 1000 available sets, (2) used the model coefficient selected in (1) to predict pixels AGB and (3) added to all pixels an error term randomly drawn from a normal distribution N(0, RSEi) where RSEi is the model RSE selected in (1). This procedure led to 1000 predictions of pixels AGB embedding the prediction uncertainty from both the first and second modelling levels. Finally, reference AGB maps and associated spatial uncertainty maps were generated as the mean and standard deviation of the 1000 pixel AGB predictions, respectively. Hereafter, we refer to pixels mean AGB prediction as $AGB_{PRED}$.

## 5.3 Results

### 5.3.1 LiDAR-AGB models and Site-Level AGB Maps

Using LiDAR data, we aimed to predict above-ground biomass (AGB) with minimal local bias for diverse forest sites. Out of the tested model across 15 canopy height metrics, a single, parsimonious model (involving meanH as the predictor variable) across all sites was found to perform better facilitating comparisons and future applications in EO data validation. Additionally, we kept the number of predictors low to avoid overfitting. The coefficients and calibration statistics of LiDAR-based AGB mapping models at site-level are provided in Table 5.8, while Figure 5.3 shows scatterplots of 'reference' against predicted AGB values.

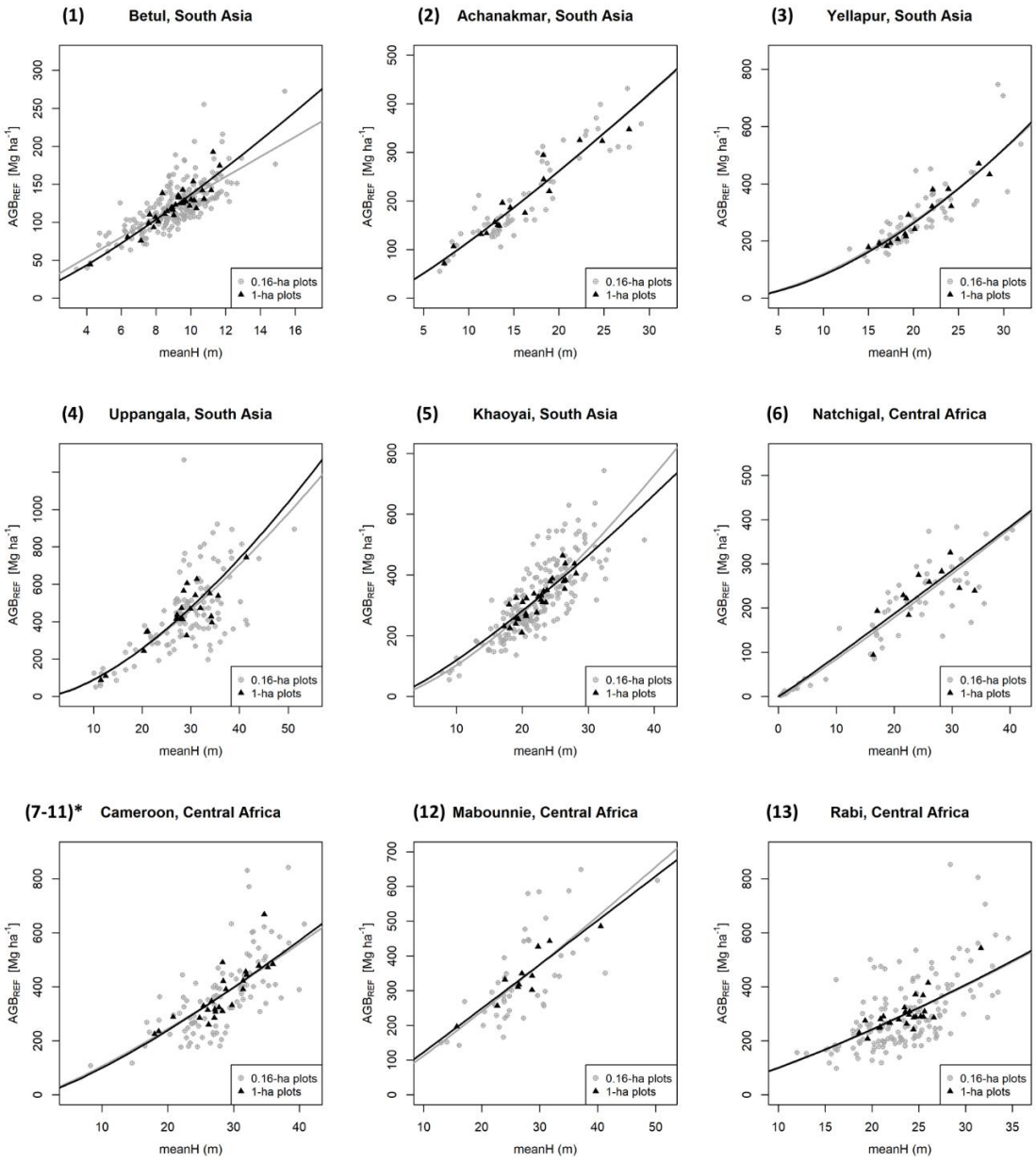


*Figure 5.3. LiDAR-AGB models of Asian and African sites at 1-ha and 0.16-ha resolutions. (7-11)* refers to the regional model established over moist dense forests of Cameroon.*

Through detailed analysis, we identified mean top canopy height as the most powerful predictor of AGB at both 1-ha and 0.16-ha resolutions, outperforming combinations of other LiDAR metrics (Figure 5.3). Similar to our tree height-diameter modeling, LiDAR-based AGB models were trained either at the site level (for larger areas) or regionally (for smaller areas). Notably, incorporating site information in the regional model had minimal impact on predictions, suggesting a consistent model performance across regions.

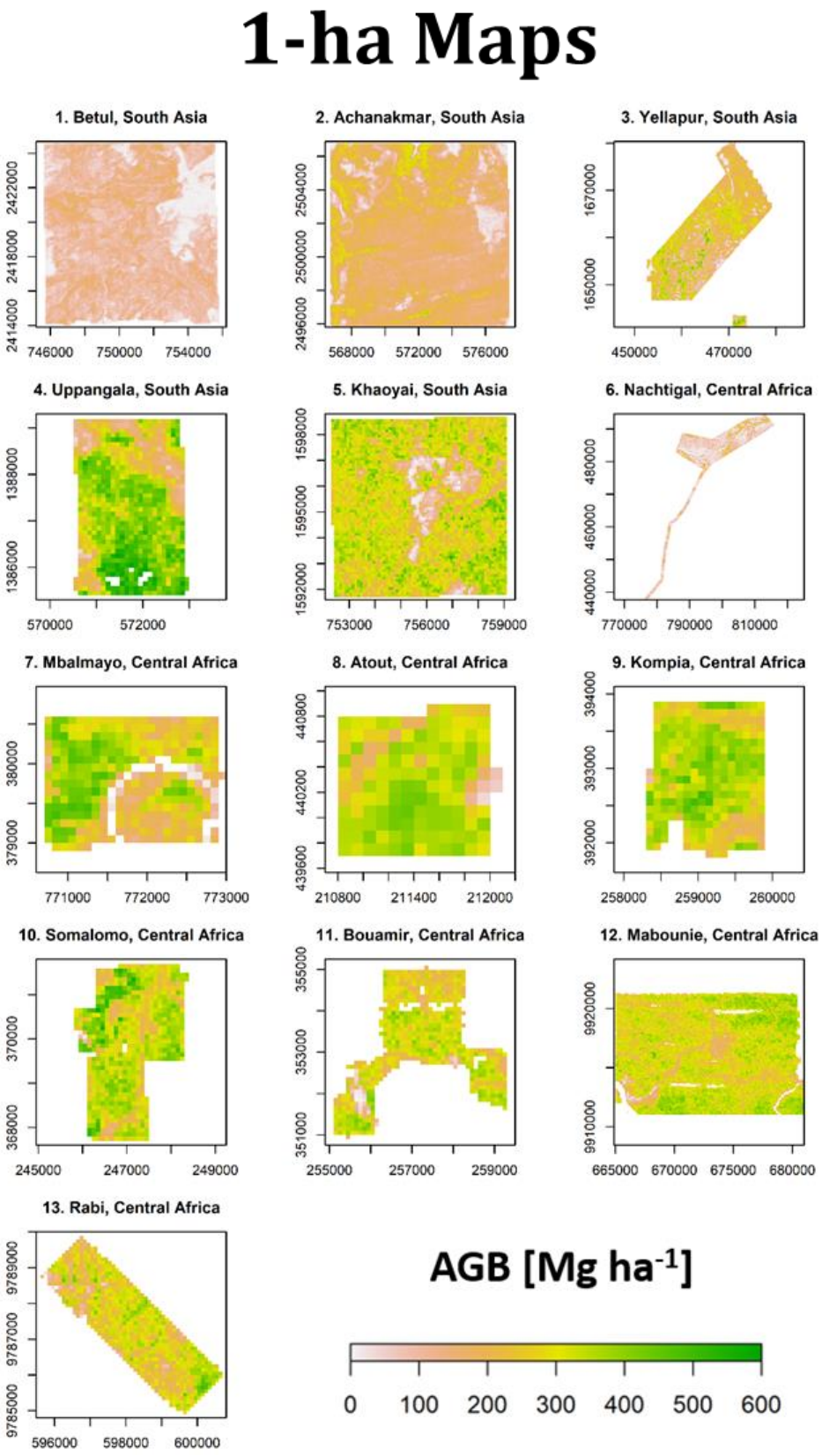


*Figure 5.4. Reference AGB maps of Asian and African sites at 1-ha spatial resolution*

Finally, the models are used to generate reference AGB maps for the LiDAR sites along with uncertainty estimates. Reference AGB maps at 1-ha resolution are shown in Figure 5.4 and the density distributions of 1-ha AGB maps are represented in Figure 5.5-A along with uncertainty levels in Figure 5.5-B expressed as a coefficient of variation (CV, in % of mean AGB) (see Figure 5.6 and Figure 5.7 for AGB maps and respective density distributions at 0.16-ha resolution). Figure 5.5-B shows that the mean uncertainty across sites is 15.4%, with site-level mean uncertainty ranging from 10.8 to 31%. It can be observed that Nachtigal and Uppangala sites have larger mean uncertainties than other sites, with 31% and 20.1%, respectively. This can be explained by larger LiDAR-AGB model uncertainties at these sites and mapping resolution.

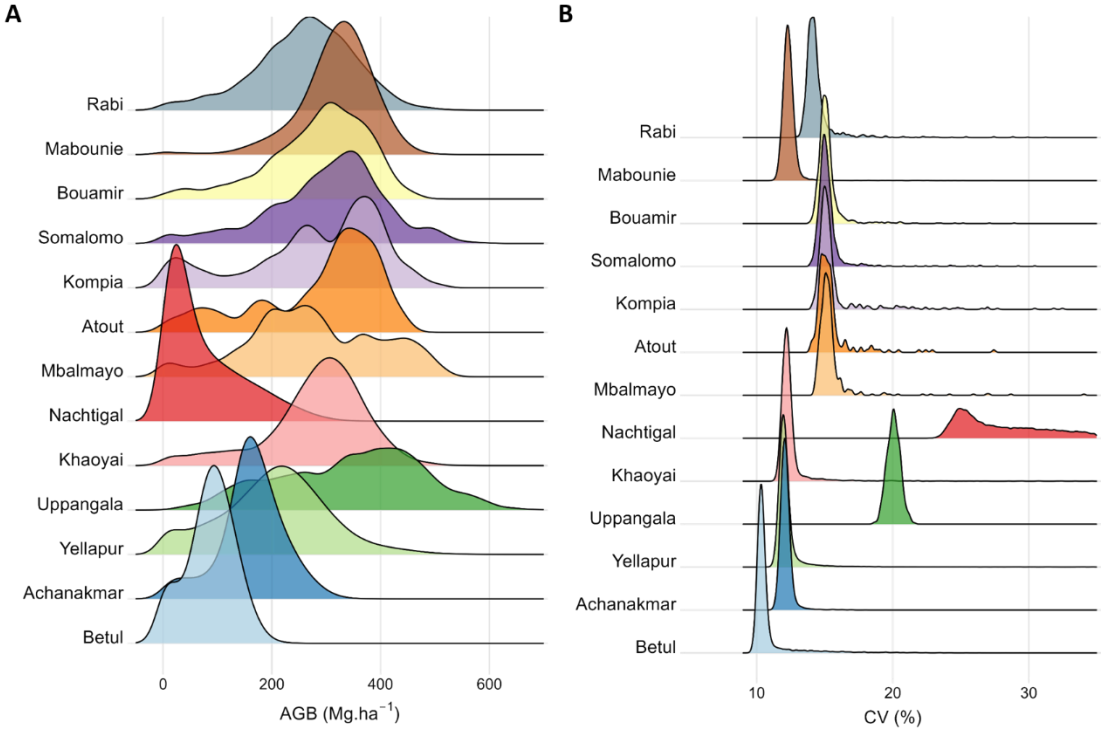


*Figure 5.5. Density distributions of (A) mean pixel AGB and (B) AGB uncertainty, expressed as a coefficient of variation (CV, in %), at 1-ha resolution across sites.*

For each site, AGB and uncertainty maps are distributed as a single GeoTiff file at the two spatial resolutions (1- and 0.16-ha) through Dataverse (Rodda et al., 2024). Each file comprises three individual layers. The two first layers named meanAGB and sdAGB correspond to the mean and standard deviation of AGB predictions over the 1000 Monte Carlo simulations, respectively. The file projection system is Universal Transverse Mercator. The third layer named Nbin corresponds to the bin number each map pixel is associated with in the binning approach proposed by McRoberts et al., (2022) to allow users reconstituting a matrix of pairwise population unit covariances estimates (See Section 5.3.2 for more details).

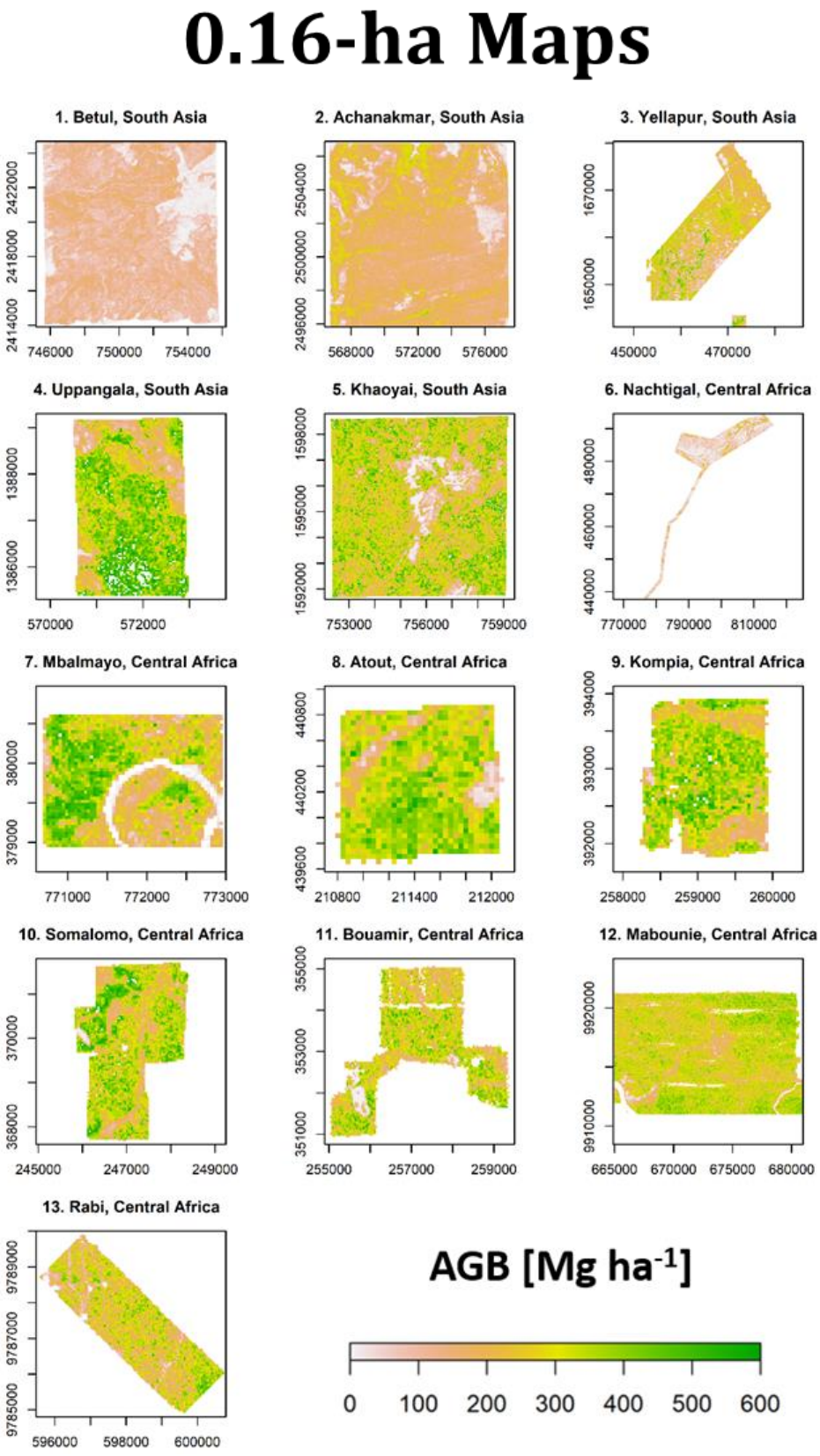


*Figure 5.6. Reference AGB maps of Asian and African sites at 1-ha spatial resolution*

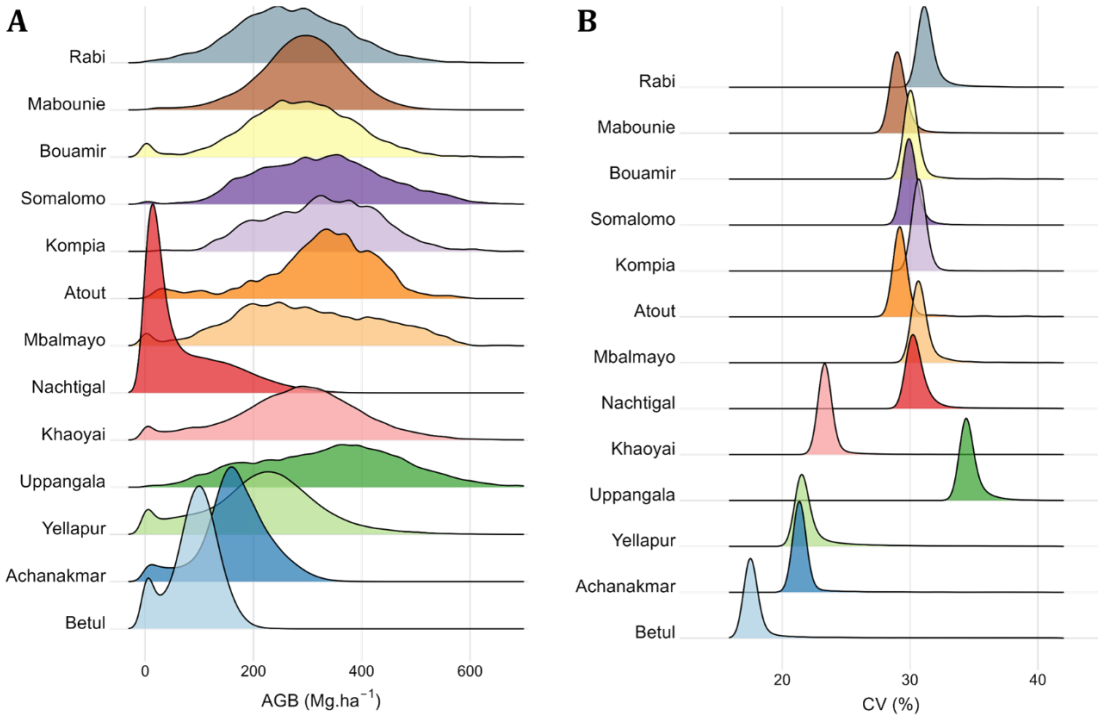


*Figure 5.7. Density distributions of (A) mean pixel AGB and (B) AGB uncertainty, expressed as a coefficient of variation (CV, in %), at 0.16-ha resolution across sites.*

### 5.3.2 Additional Error Metrics for AGB maps

LiDAR-based AGB maps produced in the present study are intended to support calibration and validation efforts of spaceborne data. To maximise their usefulness, we provide additional information that users may require – depending on their study's objective and methodological choices – to facilitate their integration with spaceborne data and/or develop comprehensive uncertainty propagation schemes up to the final, spaceborne-derived AGB map.

A first challenge users may face relates to the computation of the uncertainty associated with the mean AGB of arbitrary subregions of LiDAR AGB maps. Such subregions could for instance correspond to the footprints of spaceborne data unit pixels. Estimating the total mean squared error associated with a map (sub) population mean requires access to the matrix of pairwise population unit covariances, which is rarely communicated by map makers to users because of its large size. Yet, McRoberts et al., (2022) recently showed that pairwise population unit covariances could largely contribute to total mean squared error, and proposed an averaging and binning approach to drastically reduce the matrix size, thus facilitating its publication along with AGB maps. While we refer interested readers to McRoberts et al., (2022) for methodological details, we provided in Supplementary data all information recommended by the authors to allow map users to comply with IPCC good practice guidelines for greenhouse gas inventories (Rodda et al., 2024). We note that for each pixel of the

LiDAR-based AGB maps provided in this study, a bin number is available in the third map layer.

Another challenge lies in the propagation of uncertainties in multi-level hierarchical modelling, which is a likely use-case of the LiDAR-based maps we produced. These maps were generated by applying two hierarchically nested models: a tree allometric model linking field measurements to tree AGB, and a mapping model linking plot AGB to LiDAR data. LiDAR-based AGB maps users may employ a three-steps hierarchical modelling approach and add as a third step a model linking high resolution AGB predictions from the LiDAR-based maps to the coarser resolution of spaceborne data. An example of such an approach is presented in detail in Saarela et al., (2023) and referred to as "three-phase hierarchical model-based inference". The uncertainty assessment in such a nested modelling approach requires information at the two first modelling steps that goes beyond the results of the Monte Carlo simulation we used to produce pixel-level uncertainty estimates. While we refer interested readers to Saarela et al., (2023) for methodological details, we provide in Supplementary data all information allowing users to assess uncertainty as described in Saarela et al., (2023). This information notably includes the variance-covariance matrix of model parameters for each sampling site as well as statistics on parameters (DBH, AGB, pixels' height from CHM, etc.,) used at various levels in the chain of hierarchical models.

## 5.4 Discussion

This study offers valuable LiDAR-based forest above-ground biomass (AGB) maps for the remote sensing community, particularly those focused on carbon stock assessment. These maps are expected to be highly useful for calibrating and validating next-generation biomass mapping models based on upcoming spaceborne missions like NASA's GEDI, NASA-ISRO's NISAR, and ESA's BIOMASS. Additionally, they can improve the accuracy of existing AGB maps, especially in data-poor regions with unreliable estimates.

Beyond immediate applications, the study identifies the potential of "super-sites" across the tropics. These super-sites combine extensive forest inventory data (ideally with multiple censuses) over large areas (≥ 10 ha) with airborne LiDAR data. Such data, collected through long-term, multi-organizational efforts, is crucial for monitoring the impacts of global changes on forest ecosystems.

In addition to the per pixel estimates of uncertainty accompanying AGB maps, we hereafter provide (i) an assessment of mapping model predictive performances using a spatial model cross-validation technique (Ploton et al., 2020), to provide additional insights into the reliability of AGB predictions on each map and (ii) an assessment of mapping models extrapolation at each

sampling site, which may be useful to help users for filtering-out pixels where extrapolation occurred and only retaining the highest quality AGB predictions for spaceborne products calibration/validation.

### 5.4.1 LiDAR-AGB Model Spatial Cross-Validation

Model calibration statistics in Table 5.8 likely overestimate model predictive performance on pixels that are not used for model training, that is, on most maps' pixels. We performed a cross-validation (CV) of each model to provide more reliable insights into model predictive performance. Field plots at each site are iteratively split into training and test data and model CV statistics are built on the set of test data predictions. Regarding CV design, we selected a buffered leave-one-out cross-validation (LOO-CV; Parmentier et al., 2011) where a spatial buffer around test data is used to exclude from model training dataset observations located at the neighbourhood of test data, thus avoiding inflation in CV statistics due to spatial autocorrelation in forest AGB (Ploton et al., 2020).

*Table 5.9. Error statistics of modified LOO-CV procedure at site-level for 1-ha and 0.16-ha plots.*

| Sno | Site | 100m | | | 40m | | |
|---|---|---|---|---|---|---|---|
| | | R2 | LOOCV - RMSE [Mg ha$^{-1}$] | LOOCV - RMSE [%] | R2 | LOOCV - RMSE [Mg ha$^{-1}$] | LOOCV - RMSE [%] |
| 1 | Betul | 0.76 | 12.85 | 10.5 | 0.55 | 21.47 | 17.6 |
| 2 | Achanakmar | 0.90 | 27.05 | 13.5 | 0.81 | 39.00 | 19.5 |
| 3 | Yellapur | 0.85 | 36.27 | 12.9 | 0.62 | 74.21 | 25.8 |
| 4 | Uppangala | 0.64 | 100.44 | 23.2 | 0.37 | 171.68 | 38.4 |
| 5 | Khaoyai | 0.65 | 37.40 | 11.5 | 0.62 | 75.05 | 23.4 |
| 6 | Nachtigal | 0.22 | 60.49 | 26.1 | 0.81 | 49.85 | 28.8 |
| 7 | Mbalmayo | 0.66 | 56.36 | 15.1 | 0.46 | 110.39 | 29.6 |
| 8 | Atout | | | | | | |
| 9 | Kompia | | | | | | |
| 10 | Somalomo | | | | | | |
| 11 | Bouamir | | | | | | |
| 12 | Mabounié | 0.79 | 37.47 | 11.0 | 0.46 | 96.32 | 28.7 |
| 13 | Rabi | 0.54 | 46.11 | 15.5 | 0.39 | 101.74 | 34.5 |

As a compromise between the diversity in terms of number and spatial arrangement of field data across sites (e.g. multiple individual 1-ha plots vs. single large plot), the consistency of the CV approach across sites, as well as our expectation for a relatively weak spatial autocorrelation in forest AGB at the high resolution of the maps (< 100 m$^2$) (Réjou-Méchain et al., 2014), we selected a LOO-CV with a 100 m buffer radius for all sites and mapping resolutions (i.e. 100 x 100

m and 40 x 40 m). This CV design notably implies that (i) when a test observation came from a large field plot (i.e. > 1-ha, e.g. the 25-ha plot at Rabi), subplots at its direct neighbourhood were not used for model training (i.e., all subplots intersecting a 100 m circular buffer around the center of a test subplot were excluded from the training set, regardless of the mapping resolution), and (ii) at the 40 x 40 m mapping resolution, when a test observation came from a 1-ha field plot, the remaining three subplots of that 1-ha plot were not used for model training. The results of the buffered LOO-CV are presented in Table 5.9. They show that the predictive performances of mapping models developed in this study are comparable to those found in the literature (i.e. 15-20 % on average for the tropical forest biome (Zolkos et al., 2013) with relative RMSEs ranging from 10.6 to 20.1 % (mean across sites: 14.1 %) at 1-ha and 17.7 to 33.7 % (mean across sites: 25.7 %) at 0.16-ha.

### 5.4.2 Model Extrapolation in the predictor Space

Uncertainty maps, AGB maps, and model CV results provide insights into the reliability of AGB predictions within the calibration domain of mapping models. It is however likely that the entire gradient of forest structure sampled by LiDAR data was not fully sampled in the model's training set, thus leading to situations of predictive extrapolation where prediction uncertainty is unknown. To investigate this issue, we compared the range of vegetation height (i.e., meanTCH) sampled by the training set of each mapping model to the full range found in the LiDAR data, restricting the analysis to pixels considered as vegetated, i.e., with meanTCH ≥ 2 m. We found that the proportion of pixels affected by predictive extrapolation strongly varied across sites and at the two mapping resolutions. Generally, the upper range of meanTCH (and thus of AGBPRED) found at a landscape scale in the LiDAR data were sampled in the training set (Figure 5.8 A-B), which probably is a reflection of the "majestic forest bias" (Malhi et al., 2002) – that is, the tendency for researchers to preferentially establish sample plots where forest stands appear the less disturbed (e.g. tallest canopy height, the highest abundance of large trees, etc.). However, a varying and often substantial proportion of maps on the lower end of the meanTCH gradient was outside the model's calibration domain. For instance, in the Nachtigal site predictive extrapolation occurred on about 83% of the vegetated pixels on the 1-ha AGB map. This can be explained by the nature of this site, a forest-savanna mosaic, where the meanTCH of all herbaceous and shrubby savannas is lower than the height of the smallest 1-ha forest stand (ie., 16.4 m) found in model training set (Figure 5.8A). However, this proportion dropped to 0% at the 0.16-ha mapping resolution thanks to the inclusion into the model training set of 18 additional 0.16-ha plots established in savannas-dominated areas (Figure 5.8B, Table 5.3).

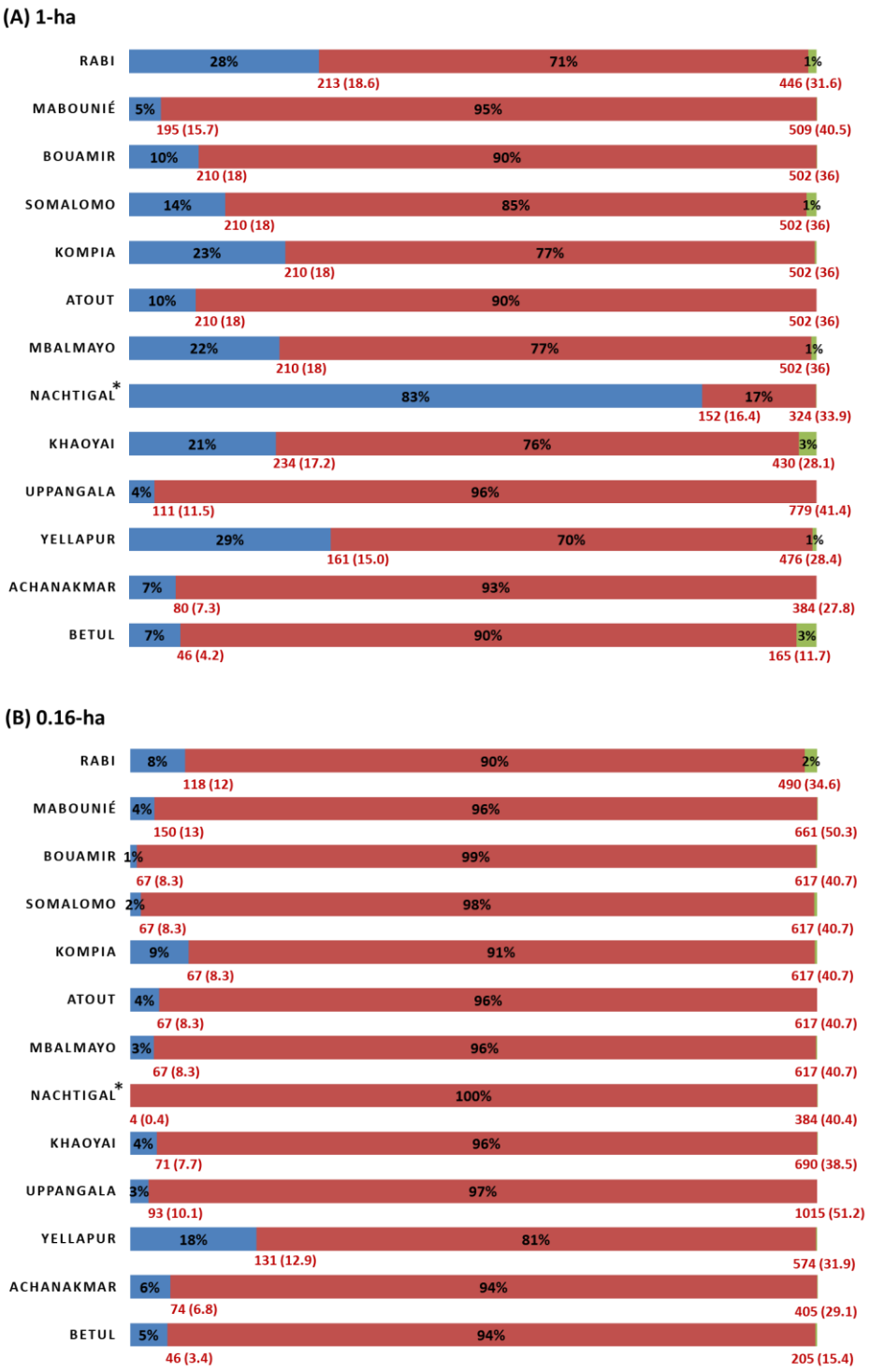


*Figure 5.8. Proportion (in %) of map pixels outside and inside models calibration domains at 1-ha (panel A) and 0.16-ha (panel B) mapping resolutions. The proportions are computed with respect to the total number of map pixels with CHM > 2 m at the exception of the Natchigal site where a 0.4 m threshold is used so as to account for the nature of the site i.e., a forest-savanna mosaic. The proportion of map pixels within model calibration domains is represented in red. Map pixels below and above the range of model calibration domains are represented in blue and green, respectively.*

## 5.5 Summary

Existing satellite-based AGB maps are unreliable due to high uncertainties. Improving these maps is crucial for understanding the global carbon cycle and combating climate change. This chapter presents a novel method to address this challenge. We combine high-resolution airborne LiDAR data, which captures the 3D structure of the forest canopy, with high quality ground-based AGB estimates from established field plots. This approach also incorporates Monte Carlo simulations to generate uncertainty maps alongside each AGB map, promoting transparency and aiding in the calibration of ongoing or upcoming spaceborne missions (viz., NASA"s GEDI, NASA-ISRO's NISAR and ESA's BIOMASS missions).

We applied this method at 13 strategically chosen sites across South Asia and Central Africa, focusing on regions with limited data for calibrating satellite missions, particularly in Asia (Duncanson et al., 2022) and are marked by notable uncertainties in AGB estimates (Rodda et al., 2023). Establishing a long-term network of such reference sites across the tropics remains a challenge, but airborne LiDAR data is a valuable tool, especially for capturing the dynamics of rapidly changing landscapes.

The resulting high-quality LiDAR-derived AGB maps are a significant contribution to the remote sensing community studying forest carbon. These maps offer two key benefits. First, they can be used to calibrate and validate next-generation biomass mapping models from upcoming spaceborne missions, leading to more accurate large-scale AGB estimates (eg. Avitabile et al., 2016; Mitchard et al., 2014b). Second, they can improve the accuracy of existing AGB maps, especially in data-poor regions with unreliable estimates. Our study sites are located in areas with known data scarcity and significant uncertainties in existing AGB estimates. Therefore, these new maps provide a valuable resource for researchers reevaluating existing maps or developing new models.

These openly available AGB reference maps, with resolutions of 100m and 40m, cover individual airborne LiDAR footprints ranging from 100 to 40,000 hectares. This chapter highlights the importance of establishing a long-term network of reference sites across the tropics. Such a network would be crucial for improved monitoring of forest AGB and the calibration of future satellite missions, ultimately leading to a more comprehensive understanding of the role of tropical forests in the global carbon cycle.

# Chapter 6

# Calibration and Validation of Existing EO-Based AGB products

Airborne LiDAR (ALS) and terrestrial forest inventories have established themselves as robust methods for estimating above-ground biomass density (AGBD or AGB used interchangeably in this chapter) at local scales. However, their spatial and temporal coverage limitations necessitate alternative approaches for regional and global biomass assessments. Space-based LiDAR missions like Global Ecosystem Dynamics Investigation (GEDI) offer promising solutions, providing sample-based vertical vegetation structure data but requiring validation due to their novelty and limited geographic coverage (Dubayah et al., 2020). The current chapter uses the data from the new-era spaceborne LiDAR systems ICESat-2 and GEDI and validates the canopy height metrics and subsequently the above-ground biomass density (AGBD) product (L4A) from GEDI over one site – tropical deciduous forests of Betul, Madhya Pradesh. Building on the reference LiDAR-derived AGB maps developed in Chapter 5 , the chapter further explores the possibility of recalibrating biomass mapping models for one specific plant functional type. This was developed using data from five additional sites across South Asia, aiming to improve the accuracy of GEDI's biomass estimates. Overall, the research demonstrates the value of reference LiDAR data in calibrating satellite-based mapping missions for more reliable large-scale biomass assessments.

## 6.1 Introduction

Light Detection and Ranging (LiDAR) systems have enabled the retrieval of three-dimensional (3D) forest canopy structure information to measure canopy height, monitor forest degradation or restoration stages, estimate above-ground biomass density (AGBD), and model other key ecosystem variables such as primary productivity and biodiversity at local to regional scales (Coops et al., 2021). The airborne LiDAR systems (ALS), LiDAR systems mounted on aircraft or drones, provide spatially continuous and accurate measurements of forest canopy structure of an area of interest but are often limited in terms of spatial and temporal coverage due to high costs (Wulder et al., 2012). On the other hand,

space-based LiDAR systems can capture information on vertical vegetation structure as samples and, when combined with wall-to-wall data from optical or radar sensors, allow large-area estimation of 3D forest structure parameters in both space and time (Potapov et al., 2021; Silva et al., 2021).

The two new space-based LiDAR missions (a) Ice, Cloud and land Elevation Satellite – ICESat-2 (Neumann et al., 2019) and (b) Global Ecosystem Dynamics Investigation (GEDI) (Dubayah et al., 2020) are providing information on elevation and canopy height metrics since 2018 and 2019 respectively. The ICESat-2, or ATLAS, is a photon counting LiDAR that fires 532 nm (green) lasers and collects information from overlapping circular laser footprints (of diameter 14-17m) at every 0.7m in the along-track direction (Neuenschwander and Pitts 2019). ICESat-2 offers wide spatial coverage (88°N–88°S latitude) but with reduced sampling densities over tropical and temperate forests. In contrast, GEDI operating on the International Space Station offers specific coverage (viz., between 51.6° N and 51.6° S) emphasizing measurements of vertical 3D structure from temperate and tropical forests. GEDI is a full waveform recording LiDAR that uses laser beam at 1064 nm to illuminate a circular footprint of ~25m diameter on the earth's surface at every 60m interval in the along track direction and stores the reflected continuous full waveform by recording the amount of laser energy reflected by plant material at different heights from the ground (Dubayah et al., 2020).While a direct comparison of parameters from ICESat-2 and GEDI would be inappropriate due to the differences in laser altimetry techniques (photon-counting vs. full-waveform), both systems offer unique advantages in the retrieval of forest canopy height, a key parameter for the estimation of AGBD and carbon balance studies (Duncanson et al., 2020; Narine et al., 2020). However, validating parameters retrieved from these novel sensors under different conditions (viz., geographical, forest type, canopy conditions, etc.) is critical to the ongoing global research. Further, the space-borne LiDAR measurements are sample-based and are used as input reference data in combination with remote sensing image datasets using statistical or machine learning models to generate wall-to-wall estimates of canopy height or forest AGBD (Ghosh et al., 2022; Lefsky, 2010; Musthafa and Singh, 2022; Nandy et al., 2021; Simard et al., 2011). Hence, a consistent validation of these space-borne LiDAR measurements are also critical to generate/validate such canopy height or AGBD maps at large spatial scales.

The accuracy of the terrain and canopy heights from the space-borne LiDAR systems are generally assessed using similar measurements from ALS data as a reference. Studies have reported accuracies of ICESat-2 for terrain height retrieval with RMSE of 0.73 – 1.89 m in various forest types, namely boreal forests of Finland (Neuenschwander and Magruder 2019), forests of interior Alaska (Wang et al., 2019), and forests of South Carolina (USA) (Xing et al., 2020). The accuracy of GEDI in terrain height retrievals was observed to be lower than

ICESat-2 with RMSE of 1.67 – 4.48 m in temperate (Adam et al., 2020) and Mediterranean forests (Dorado-Roda et al., 2021). Similar results of ICESat-2 (RMSE = 2.24 m) outperforming GEDI (RMSE = 4.03 m) in terrain height retrievals were observed by Liu et al. (2021), when evaluating ATLAS and GEDI metrics over 40 different Land cover sites in and around the USA. It was also observed that over these sites, GEDI is better performing than ICESat-2 in canopy height retrievals with RMSE of 3.56 m and 5.02 m respectively. The errors in terrain and canopy height retrievals are found to be strongly related to laser intensity, slope, canopy cover and vegetation height. The strong (or power) beams in both GEDI and ICESat-2 seem to offer better accuracy than weak (or coverage) beams in height retrievals (Liu et al. 2021; Neuenschwander et al. 2020). Nevertheless, most of the studies mentioned above are conducted mostly in boreal, temperate, and sub-tropical forests, with very few samples in tropical savannas and forests. A recent study using ICESat-2 ATLAS for height retrievals in tropical Mesoamerican forests reported large RMSE of 10.58 m and 13.35m in terrain height and canopy height respectively (Fernandez-Diaz et al., 2022). Hence, the need to evaluate the performance of GEDI and ICESat-2 in various forest types in different latitudinal bands for broader usage.

While the spaceborne LiDAR mission GEDI offers above-ground biomass density (AGBD) estimates through its L4A data products, these are derived from models using relative height metrics in L2A data (Duncanson et al., 2022). These models, although globally calibrated across plant functional types, lack comprehensive reference data, leading to uncertainties in regional and national biomass estimates. Therefore, validating GEDI's AGBD products is crucial for understanding and improving their accuracy.

This study addresses this need by performing a systematic accuracy assessment of GEDI and ICESat-2 data products in tropical dry deciduous forests of Central India. We compared the accuracy of terrain and canopy height retrievals from both satellites (ATL08 for ICESat-2 and L2A for GEDI) against reference airborne LiDAR data. Additionally, we validated GEDI's derived AGBD estimates (L4A) using a reference biomass map generated from airborne LiDAR data and field biomass plots. Next, the chapter explores a possibility to demonstrate re-calibration of GEDI AGBD estimates for one plant functional type of South Asia using the reference LiDAR-AGB maps generated in Chapter 5 .

## 6.2 Materials and Methods

### 6.2.1 Aerial LiDAR Sample Sites

The study utilizes the Aerial LIDAR data captured over tropical forests of Asia as described in Chapter 5 (Figure 6.1). More specifically, the study sites include – Betul, Achanakmar, Uppangala, Yellapur and Khaoyai over tropical Asian

region. These sites cover wide range of plant functional types (PFTs), which are key parameter for generating GEDI L4A (AGBD) product from relative height metrics generated from GEDI L2A products. The detailed descriptions of study area characteristics, inventory data, LiDAR data processing and generation of AGB maps are given in Section 5.2.

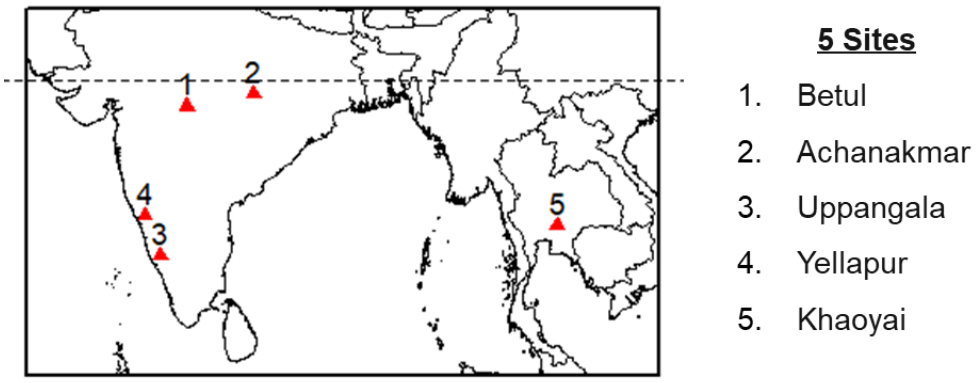


*Figure 6.1. Reference LiDAR-AGB Sites over Tropical Asia*

### 6.2.2 Space-borne LIDAR Products

The ICESat-2 ATL08 product (Version 5 released November 2021; Neuenschwander et al., 2021) provides information about terrain height, canopy height, canopy cover, and other parameters at a fixed step length of 100m along the ground track direction. The terrain and canopy height estimates are based on the distribution of signal photons and their subsequent classification as either noise, ground, canopy or top-of-canopy photons. The parameters from the ATL08 product used in the current study are h_te_best_fit (terrain height best fit), h_canopy (indicates 98th canopy height percentile), night_flag (information about night or day), gt (ground track beam information), and sc_orient (when combined with parameter "gt" distinguishes strong beam and weak beam). These metrics from ATL08 product refer to the ground segment 100 x 14 m polygon.

The GEDI Level-2A (L2A; version 2 released April 2021; Dubayah et al., 2021) data product provides information on geolocation, ground elevation (or elev_lowestmode), canopy height, and relative height (RH) metrics from the received waveform at footprint-level (Dubayah et al., 2020). The RH metrics (viz., rh0, rh1, rh2, ….., rh98, rh99, rh100) indicate the height above the ground at each energy quartile in the received waveform (Dubayah et al., 2020). The GEDI Level-4A (L4A; version 2.1 released March 2022; Dubayah et al., 2022) is footprint-level above-ground biomass density (AGBD, Mg $ha^{-1}$) product derived through parametric linear models that relate GEDI L2A waveform RH metrics to AGBD. The models are generated by combined stratification of plant functional type and world regions and applying natural logarithm or square root transformation on

the response and predictor variables to stabilize the prediction interval and residual variance (Duncanson et al., 2022). Though the models are generated and validated in more than 27 countries across six continents, the important regions such as continental Asia and Africa are still under-represented (Duncanson et al., 2022). In this study, in addition to the RH metrics and elev_lowestmode (from L2A) and AGBD values (from L4A), elevation_auxillary information of each GEDI shot such as beam_flag (to differentiate power and coverage beam), date and time (to differentiate day or night), beam sensitivity (defined as the maximum canopy cover through which GEDI waveform can detect ground with 90% probability) and quality_flag are used for further analysis. Each GEDI shot refers to a ground footprint of 25m.

*Table 6.1. List of GEDI passes over Betul, Madhya Pradesh (till August 2022). The days with greater than 90% of the shots effected due to quality issues are not used.*

| Sno | Date | Remark | N_total | N_L2A [QF = 1] | N_L4A [QF = 1] |
|---|---|---|---|---|---|
| 1 | 2019.06.22 | *Leaf-on; cloudy; not used* | 1625 | 20 | 0 |
| 2 | 2019.10.19 | *Leaf-on; cloudy; not used* | 362 | 0 | 0 |
| 3 | 2019.11.09 | *Leaf-on* | 998 | 970 | 238 |
| 4 | 2020.01.09 | *Leaf-on* | 30 | 30 | 27 |
| 5 | 2020.04.14 | *Leaf-off; cloudy; not used* | 1581 | 157 | - |
| 6 | 2020.04.20 | *Leaf-off* | 316 | 265 | - |
| 7 | 2020.04.30 | *Leaf-off* | 1553 | 548 | - |
| 8 | 2020.07.28 | *Leaf-on; cloudy; not used* | 14 | 0 | 0 |
| 9 | 2020.09.29 | *Leaf-on* | 1002 | 534 | 303 |
| 10 | 2020.10.06 | *Leaf-on* | 280 | 130 | 55 |
| 11 | 2020.10.19 | *Leaf-on* | 1549 | 1354 | 1006 |
| 12 | 2021.03.25 | *Leaf-off* | 300 | 276 | - |
| 13 | 2021.04.02 | *Leaf-off* | 34 | 18 | - |
| 14 | 2021.04.13 | *Leaf-off* | 1084 | 492 | - |
| 15 | 2021.08.14 | *Leaf-on; cloudy; not used* | 1620 | 31 | 0 |
| 16 | 2021.10.02 | *Leaf-on* | 24 | 6 | 3 |
| 17 | 2021.10.13 | *Leaf-on* | 35 | 35 | 35 |
| 18 | 2022.03.14 | *Leaf-off* | 318 | 192 | - |

The detailed evaluations of ICESat-2 and GEDI products were initially carried out over tropical dry deciduous forests of Betul, Madhya Pradesh. All available ICESat-2 (Version 5), GEDI L2A (Version 2), and L4A (Version 2.1) products over Betul, Madhya Pradesh were queried from availability till August 2022 are downloaded from https://search.earthdata.nasa.gov/ website in hdf5

format and are processed in R software environment (Version 4.1.3). Over the study area, 6 passes of ATLAS and 18 passes of GEDI were downloaded. Out of the 6 passes of ATLAS, 3 passes were acquired during the leaf-on season and 3 passes were during the leaf-off season. Of the 18 GEDI passes, 11 passes were during the leaf-on season, 7 passes were during the leaf-off season. The individual observations were filtered using quality flags (cloud_flag_atm <=1 for ATLAS and quality_flag = 1 for GEDI). Also, the passes which are heavily affected by quality issues (i.e. >90% of the shots in the pass are not passing the quality flag) are ignored for this study. The remaining observations are further filtered spatially using a forest mask over the study area. Finally, 777 ATLAS ATL08 observations (181 during the leaf-on season and 596 during the leaf-off season) and 4414 GEDI L2A shots (2830 during the leaf-on season and 1584 during the leaf-off season) were evaluated for terrain and canopy height in the current study. The GEDI L4A AGB product was also similarly filtered and is only evaluated for the leaf-on season since the L4A algorithm is valid only during the leaf-on stage (Duncanson et al., 2022). Finally, a total filtered 1557 GEDI L4A observations were evaluated against the ALS reference AGB map.

*Table 6.2. List of ATLAS passes over Betul, Madhya Pradesh (till August 2022).*

| Sno | Date | Remark | N_Total_Shots | N_ATL03 [Cloud Flag <= 1] |
|---|---|---|---|---|
| 1 | 2018.12.13 | leaf-on | 131 | 91 |
| 2 | 2020.03.11 | leaf-off | 347 | 347 |
| 3 | 2020.09.09 | leaf-on | 31 | 31 |
| 4 | 2021.03.09 | leaf-off | 272 | 272 |
| 5 | 2021.04.07 | leaf-off | 106 | 106 |
| 6 | 2021.12.07 | leaf-on | 78 | 65 |

The total and filtered observations used in this study are listed in Table 6.1 and Table 6.2. Considering the geolocation errors of GEDI (viz., 8 – 10m; Dubayah et al. 2021) and ICESat-2 (viz., 2 – 3m; Neuenschwander et al. 2021), the GEDI metrics are evaluated against the reference ALS metrics extracted using 8m buffer around the GEDI footprint, and the ATLAS metrics are evaluated against reference ALS metrics extracted using 2m buffer around the original ATLAS 100 x 14m ground segments.

### 6.2.3 Evaluation of Canopy Height & AGB Products

For every ATLAS or GEDI footprint that was found to be of good quality, the metrics of terrain height, canopy height and AGB (or AGBD) estimates were compared against the respective reference DTM, CHM and LiDAR-AGB maps.

For instance, the RH98 metric from GEDI or h_canopy from ATLAS are evaluated against the 98th quantile of the CHM raster cells within the extent of the respective buffered sensor footprints. Similarly, the ground elevation values (viz., elev_lowestmode from GEDI or h_te_best_fit) are evaluated against the mean of the DTM raster cells within the extent of the respective buffered sensor footprints. The GEDI-AGBD estimate over the footprint was evaluated against the area-weighted average of the LiDAR-AGB raster cells within the footprint.

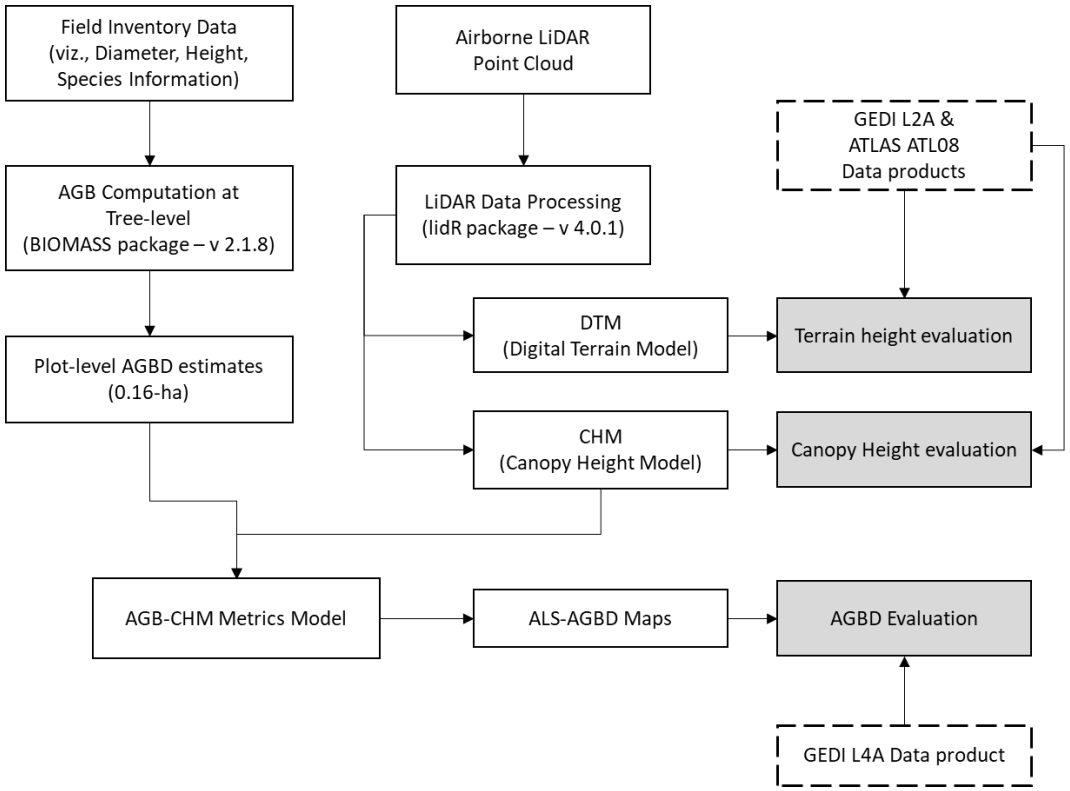


*Figure 6.2. Flowchart describing the methodology used in this study.*

The uncertainty of each of the metric from space-borne LiDARs was expressed in terms of root mean square error (RMSE), relative RMSE (rRMSE), Bias, relative Bias (rBias), mean absolute error (MAE), and pearson's correlation coefficient (r) described by the below equations (Equations (11 – (16).

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{13}$$

$$rBias\ (\%) = \frac{Bias}{\bar{x}} \times 100 \quad (14)$$

$$rRMSE\ (\%) = \frac{RMSE}{\bar{x}} \times 100 \quad (15)$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (16)$$

Where $x_i$ is the reference ALS metric (viz., CHM Height percentile or AGBD), $y_i$ is the GEDI or ATLAS derived height metric or AGBD metric, $\bar{x}$ is the mean value of the reference ALS metric, $\bar{y}$ is the mean value of the GEDI or ATLAS metric and n is the number of samples used for the evaluation.

### 6.2.4 Re-calibration of GEDI AGBD Products

Over the 05 tropical forest sites in tropical South Asia with LiDAR-AGB reference maps, a total of ~60,000 GEDI shots were available. The data from these shots are downloaded, filtered and compiled for investigating accuracy of above-ground biomass density (AGBD) estimates derived from the spaceborne LiDAR mission GEDI in diverse forest ecosystems across South Asia. Five representative sites encompassed the three major plant functional types (PFTs) found in the region - Dry Broadleaf (DBT), Evergreen Broadleaf (EBT), and Grasses Shrubs and Woodlands (GSW).

To facilitate the comparison, all available GEDI AGBD estimates at the footprint level within the study areas were compiled. Subsequently, data processing steps were implemented as follows:

- GEDI Footprint Selection: All GEDI shots intersecting with the reference AGBD maps (generated from airborne LiDAR data) were meticulously identified and extracted.
- Quality Control: Stringent filtering based on GEDI's quality flags ensured inclusion of only reliable AGBD estimates in the analysis.
- Reference Data Extraction: Corresponding reference AGB values were extracted from the LiDAR reference map for each selected GEDI shot, enabling a point-to-point comparison.
- Multi-Resolution Analysis: The comparison was conducted at three spatial resolutions - 100m, 500m, and 1000m grids - to capture potential scale-dependent variations in accuracy.
- Bias Correction: To address systematic biases between GEDI and LiDAR AGBD estimates, a first-order bias correction model was employed for EBT plant functional type.

## 6.3 Results

### 6.3.1 Terrain and Canopy Height Accuracy

The terrain height derived from ATL08 (ATLAS) and L2A (GEDI) is validated against the mean reference ALS-DTM value for the corresponding buffered ground footprint of GEDI and ATLAS. The analysis is carried out independently for leaf-on and leaf-off periods to understand the influence of canopy cover on terrain height retrieval (Figure 6.3). The scatterplot (Figure 6.3) suggests that the ATLAS outperforms GEDI in detecting ground surface during leaf-on and leaf-off periods. Further, the shots acquired during the night for both sensors appeared close to that of the reference DTM. For GEDI, during leaf-off periods, the daytime shots appear to be more distributed with significant uncertainties.

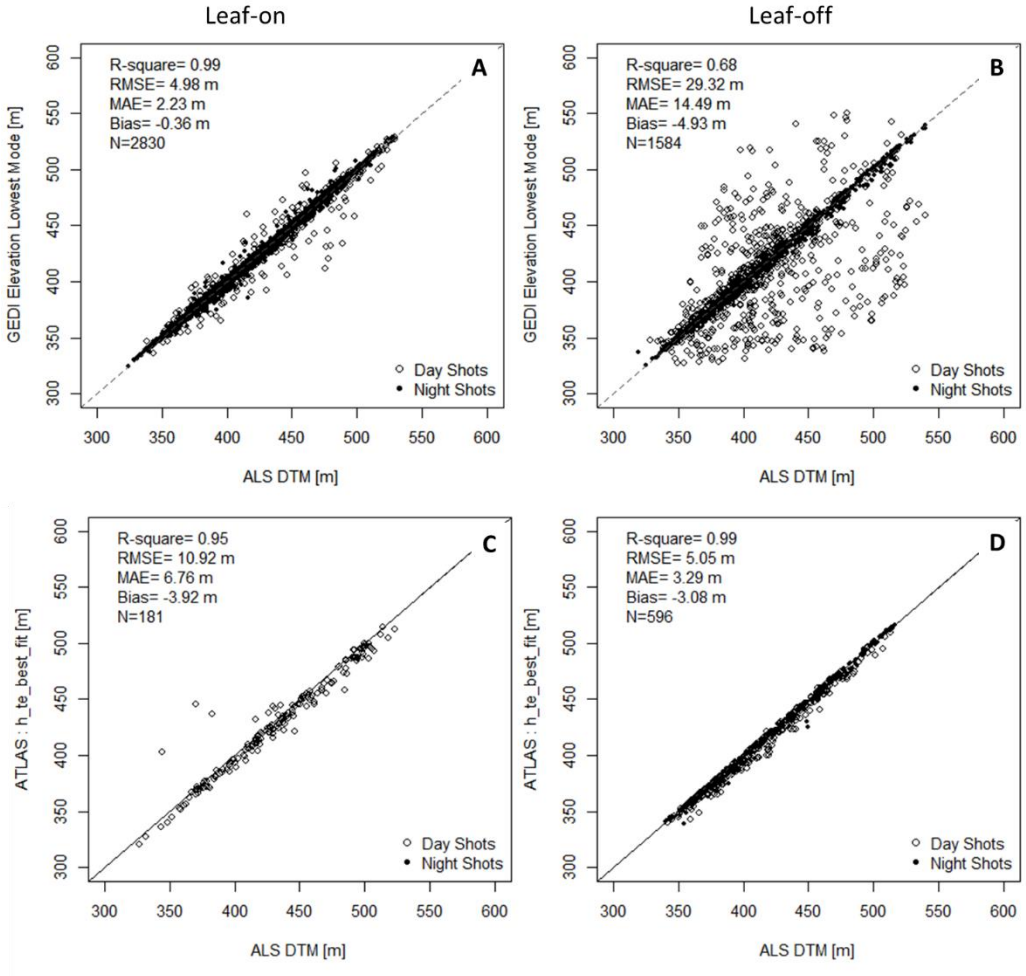


*Figure 6.3. (A) & (C) The scatterplot between reference DTM value from the Aerial LiDAR (ALS) and GEDI and ICESat-2 (or ATLAS) respectively during leaf-on season. (B) & (D) The scatterplot between reference DTM value from the Aerial LiDAR (ALS) and GEDI and ICESat-2 (or ATLAS) respectively during leaf-off season. The daytime shots are shown in hollow circles and night time shots are shown in filled circles.*

*Table 6.3. Error estimates of Terrain height retrievals from ATL03 (ATLAS) and L2A (GEDI) for different scenarios according to beam intensity, data acquisition time and canopy condition.*

| Sensor | Canopy Condition | Beam Flag | Day/Night | Parameter | N | MAE [m] | RMSE [m] | BIAS [m] | R | RMSE [%] |
|---|---|---|---|---|---|---|---|---|---|---|
| GEDI | Leaf-on | Power | Day | elev_lowest_mode | 931 | 2.0 | 5.6 | -0.8 | 0.99 | 1.3% |
| GEDI | | Coverage | Day | elev_lowest_mode | 473 | 3.7 | 7.4 | -0.7 | 0.99 | 1.7% |
| GEDI | | Power | Night | elev_lowest_mode | 867 | 1.7 | 2.8 | -0.2 | 1.00 | 0.7% |
| GEDI | | Coverage | Night | elev_lowest_mode | 559 | 2.1 | 3.8 | 0.4 | 1.00 | 0.9% |
| ATLAS | | Strong | Day | h_te_best_fit | 167 | 6.7 | 11.1 | 3.6 | 0.97 | 2.6% |
| ATLAS | | Weak | Day | h_te_best_fit | 14 | 7.7 | 8.7 | 7.7 | 1.00 | 2.0% |
| ATLAS | | Strong | Night | h_te_best_fit | No data available | | | | | |
| ATLAS | | Weak | Night | h_te_best_fit | No data available | | | | | |
| GEDI | Leaf-off | Power | Day | elev_lowest_mode | 738 | 23.2 | 39.2 | -9.8 | 0.61 | 9.6% |
| GEDI | | Coverage | Day | elev_lowest_mode | 266 | 17.6 | 28.7 | 0.1 | 0.74 | 6.8% |
| GEDI | | Power | Night | elev_lowest_mode | 155 | 1.8 | 2.7 | 0.3 | 1.00 | 0.6% |
| GEDI | | Coverage | Night | elev_lowest_mode | 425 | 2.0 | 3.4 | -1.4 | 1.00 | 0.8% |
| ATLAS | | Strong | Day | h_te_best_fit | 183 | 3.3 | 4.6 | 3.1 | 1.00 | 1.1% |
| ATLAS | | Weak | Day | h_te_best_fit | 148 | 6.2 | 7.6 | 6.2 | 0.99 | 1.8% |
| ATLAS | | Strong | Night | h_te_best_fit | 265 | 1.7 | 3.2 | 1.4 | 1.00 | 0.8% |
| ATLAS | | Weak | Night | h_te_best_fit | No data available | | | | | |

*Table 6.4. Error estimates of Canopy height retrievals from ATL03 (ATLAS) and L2A (GEDI) for different scenarios according to beam intensity, data acquisition time and canopy condition.*

| Sensor | Canopy Condition | Beam Flag | Day/Night | Parameter | N | MAE [m] | RMSE [m] | BIAS [m] | R | RMSE (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| GEDI | Leaf-on | Power | Day | rh98 | 931 | 2.3 | 3.3 | 0.6 | 0.68 | 18% |
| GEDI | | Coverage | Day | rh98 | 473 | 3.0 | 4.3 | -0.9 | 0.47 | 22% |
| GEDI | | Power | Night | rh98 | 867 | 2.7 | 3.9 | 1.1 | 0.63 | 20% |
| GEDI | | Coverage | Night | rh98 | 559 | 2.6 | 3.9 | -0.2 | 0.54 | 20% |
| ATLAS | | Strong | Day | h_canopy | 166 | 9.8 | 12.5 | -9.6 | 0.22 | 65% |
| ATLAS | | Weak | Day | h_canopy | 14 | 11.3 | 12.4 | -10.1 | 0.68 | 61% |
| ATLAS | | Strong | Night | h_canopy | No data available | | | | | |
| ATLAS | | Weak | Night | h_canopy | No data available | | | | | |
| GEDI | Leaf-off | Power | Day | rh98 | 738 | 7.1 | 8.7 | -6.5 | 0.25 | 46% |
| GEDI | | Coverage | Day | rh98 | 266 | 10.2 | 11.6 | -9.7 | 0.02 | 62% |
| GEDI | | Power | Night | rh98 | 155 | 6.3 | 7.6 | -5.8 | 0.68 | 39% |
| GEDI | | Coverage | Night | rh98 | 425 | 7.3 | 8.7 | -7.0 | 0.52 | 47% |
| ATLAS | | Strong | Day | h_canopy | 178 | 4.2 | 5.6 | -0.6 | 0.51 | 28% |
| ATLAS | | Weak | Day | h_canopy | 145 | 6.1 | 7.5 | -4.9 | 0.48 | 38% |
| ATLAS | | Strong | Night | h_canopy | 245 | 2.4 | 3.5 | -0.4 | 0.63 | 18% |
| ATLAS | | Weak | Night | h_canopy | No data available | | | | | |

Table 6.3 presents the uncertainty analysis of GEDI and ATLAS terrain height against different beam strengths and acquisition time. For GEDI night-time shots, during both leaf-on and leaf-off periods, the terrain height RMSE was found to be low and in a similar range (2.7m – 3.8m; Table 6.3). Though a high RMSE was observed for day-time acquisition during leaf-on periods (5.6m – 7.4m; Table 6.3), RMSE recorded for day-time leaf-off periods was unreasonably high (28.7m – 39.2m; Table 6.3).

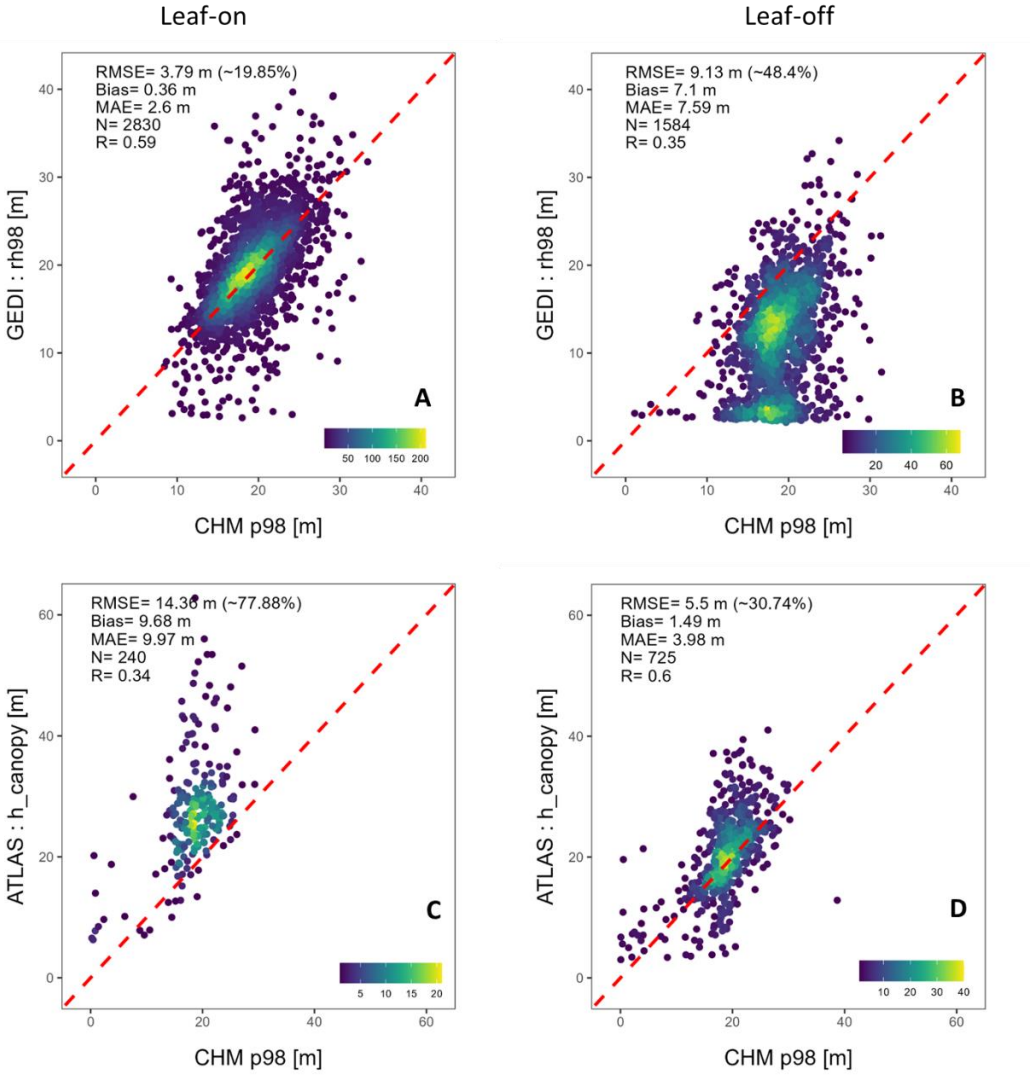


*Figure 6.4. (A) & (C) The density scatterplot between reference 98th percentile of canopy height model (CHM p98) and the 98th height percentile values from GEDI and ICESat-2 (or ATLAS) respectively during leaf-on season. (B) & (D) The density scatterplot between reference 98th percentile of canopy height model (CHM p98) and the 98th height percentile values from GEDI and ICESat-2 (or ATLAS) respectively during leaf-off season. The reference 98th percentile is computed over the respective ground segments of the spaceborne LiDAR systems.*

This unreasonable RMSE during leaf-off was mostly due to a single acquisition (2021-04-13) during the early morning hours (0730 hours IST). The

early morning atmospheric conditions due to the presence of aerosols could possibly attenuate the GEDI waveform signal through scattering (Fayad et al., 2021). The mean aerosol optical depth over the study area, derived from MODIS MCD19A2 product (Lyapustin et al. 2018), for the date (2021-04-13) was observed to be ~0.8, indicating high aerosol content. Nevertheless, a more detailed analysis about the presence of aerosol particles on GEDI signal attenuation could not be carried out due to limited GEDI observations. Similar observations were made over parts of France, Tunisia, and French Guiana that the high aerosol optical depth (>0.8) affected the viability of >50% of the GEDI shots (Fayad et al., 2021). For ATLAS, due to the fewer observations over the study site in leaf-on periods, nighttime data is not captured. However, the night-time data captured during leaf-off periods was observed to be better at detecting terrain height (RMSE = 3.2m) compared to day time accuracy ranging between 4.6m – 11.1m across different scenarios (Table 6.3). Overall, the terrain height retrievals for both sensors was observed to be more accurate during the night time acquisitions and even more consistent with strong beams, in line with the reported studies (Adam et al., 2020; Liu et al., 2021).

Figure 6.4, shows the density scatter plot between rh98 (GEDI) and h_canopy (ATLAS) validated against the CHM-p98 during leaf-on and leaf-off periods. It was observed that the GEDI rh98 metric was strongly linked to the CHM-p98 metric during the leaf-on period and rather weak during the leaf-off period. In contrast, ATLAS was observed to detect the canopy height reasonably well even during the leaf-off periods (Figure 6.4). The Table 6.4 describes a detailed analysis of rh98 and h_canopy under different beam strengths and acquisition times. The RMSE varied minimally between 3.3m – 4.3 m for GEDI beams during the leaf-on period considering beam strength and acquisition time. Also, the mean residuals (or bias) were found to be in the order of ±1m indicating high reliability of GEDI shots for delineation of canopy height during leaf-on seasons. However, significant under-estimation (bias range -5.83 to -9.73m) with high RMSE (7.6m – 11.6m) was observed in canopy height delineation from GEDI shots acquired during the leaf-off season. During the leaf-off season due to the very little or no leaf availability over the study area, detecting the top of the trees against the background noise could be difficult and thus leading to errors in height recovery (Dubayah et al., 2020).

Though night-time ATLAS acquisitions are limited to a leaf-off season, the strong night beams could delineate canopy height with RMSE of 3.53m and a mean bias of -0.4 m, similar to the GEDI shots in the leaf-on season. The day-time ATLAS acquisitions for canopy height retrieval were largely inaccurate with high RMSE (>5m) regardless of beam strength and leaf condition. For these day-time acquisitions, bias of 3.1m – 7.7m was also observed in terrain height retrieval, indicating the inability of ATLAS to either reach or identify ground points due to background solar noise and subsequently leading to potential bias in the

canopy height estimation (Table 6.4). Similar observations of the reliability of the ATLAS system in low canopy conditions were observed in the USA (Wang et al., 2019). In summary, though GEDI outperforms ATLAS in canopy height delineation, strong night beams of ATLAS could add significant information to the integrated canopy height measurements over tropical forests.

### 6.3.2 GEDI L4A AGBD Accuracy

The GEDI footprint-level AGBD (L4A) is derived from the RH metrics of L2A products using calibration equations based on the stratification involving five plant function types and seven geographic regions across the globe (Duncanson et al., 2022). The GEDI AGBD estimates were strongly mismatched against the reference AGBD estimates with an RMSE of ~46% (N = 1557; Figure 6.5). Since the AGBD (L4A) over the study area was calculated from L2A RH metrics using two models (Table 6.5), viz., DBT_SAs (Deciduous Broadleaf Trees - South Asia) and GSW_SAs (Grasses, Shrubs and Woodlands – South Asia), we performed an independent model level error analysis. Nevertheless, performance of both models over the study area was found to be inaccurate with high RMSE of 33.8% for the DBT_SAs model and 64.0% for the GSW_SAs model (Figure 6.5). Further, the AGBD values of DBT_SAs model were underestimated (Bias = -11.7%; N = 1022) compared to a significant over-estimation for the shots using GSW_SAs model (Bias = 42.9%; N = 535). The high RMSE and substantial biases could be attributed to the lack of ground reference datasets over South Asia. More specifically, zero waveforms were used to calibrate GEDI-AGBD models for the DBT and GSW plant functional types over the South Asian region (Duncanson et al., 2022).

*Table 6.5. Error estimates of AGBD (GEDI L4A) with the reference ALS AGBD map. The variables (xvar1 and xvar2) indicates the square transformation of the GEDI L2A relative height metrics rh50 and rh98 respectively. xvar1 = √(100+rh50) and xvar2 = √(100+rh98).*

| GEDI Model | Equation | N | RMSE [Mg $ha^{-1}$] | MAE [Mg $ha^{-1}$] | BIAS [Mg $ha^{-1}$] | RMSE [%] | BIAS [%] |
|---|---|---|---|---|---|---|---|
| DBT_SAs | AGBD = 1.113*(-104.966 + 6.802*xvar1 + 3.955*xvar2) | 1022 | 38.2 | 30.7 | -13.2 | 33.8% | -11.7% |
| GSW_SAs | AGBD = 1.118*(-124.832 + 12.426*xvar2) | 535 | 64.0 | 45.5 | 40.1 | 68.4% | 42.9% |
| DBT_SAs + GSW_SAs | Respective models for respective shots. | 1557 | 48.6 | 35.9 | -5.1 | 45.8% | -4.8% |

*Table 6.6. Error estimates of AGBD (GEDI L4A) estimates with DBT_SAs model for different scenarios according to beam intensity, data acquisition time and beam sensitivity ($Beam_{sen}$)*

| Category | N | $Beam_{sen}$ | MAE [Mg $ha^{-1}$] | RMSE [Mg $ha^{-1}$] | BIAS [Mg $ha^{-1}$] | R | Mean Reference AGBD [Mg $ha^{-1}$] | RMSE [%] |
|---|---|---|---|---|---|---|---|---|
| All Shots | 1022 | 96.9% | 30.7 | 38.2 | -13.2 | 0.61 | 112.9 | 34% |
| Power-Night | 157 | 97.2% | 34.7 | 44.3 | -2.9 | 0.52 | 118.6 | 37% |
| Power-Day | 544 | 97.2% | 30.0 | 37.1 | -17.9 | 0.65 | 109.4 | 34% |
| Coverage-Night | 154 | 96.3% | 28.0 | 34.4 | -4.0 | 0.51 | 121.1 | 34% |
| Coverage-Day | 169 | 96.2% | 31.8 | 38.5 | -15.9 | 0.61 | 111.5 | 39% |

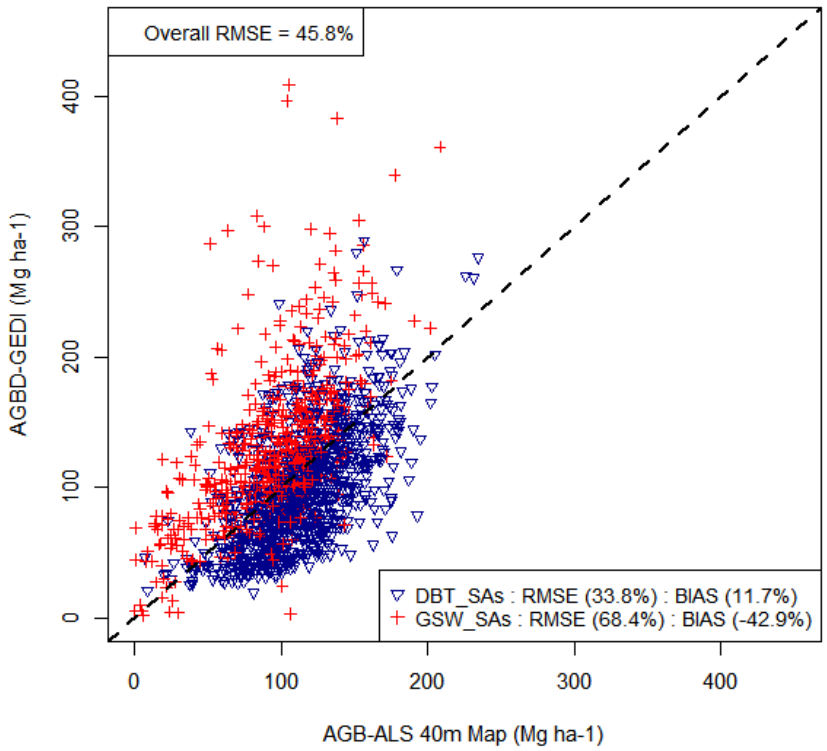


*Figure 6.5. Scatterplot of GEDI L4A AGBD estimates versus the reference ALS AGBD estimates over the study area. The plot distinguishes between the two different GEDI L4A models used over the study area*

To assess the impact of the beam energy (power vs. coverage) and acquisition time (day vs. night), we have only used the GEDI-AGBD estimates from the DBT_SAs model due to the high uncertainty of the GSW_SAs model over the study area (Table 6.6). It was observed that RMSE varied minimally (34% - 39%; Table 6.6) across these scenarios due to similar beam sensitivity. Though

the AGBD was underestimated in four scenarios, the GEDI shots acquired during nights are observed to be with lower bias when compared to day time GEDI shots. In addition, the correlation between GEDI-AGBD and ALS-AGBD was observed to be marginally higher for power-beams than the coverage beams.

### 6.3.3 Re-calibration of GEDI AGBD Product

Following the establishment of a robust comparison framework, a comprehensive accuracy assessment of GEDI's above-ground biomass (AGB) estimates was undertaken across all five South Asian forest sites. This assessment revealed a consistent and concerning trend: a statistically significant bias was observed across all sites, irrespective of the plant functional type (PFT). To address this systematic error, the investigation focused on the Evergreen Broadleaf (EBT) PFT, which exhibited the most pronounced bias.

A first-order bias correction model was subsequently developed and evaluated specifically for GEDI-AGBD estimates in South Asian EBT forests (Figure 6.6). To ensure the generalizability and robustness of this corrective model, a rigorous "Leave-One-Site-Out" cross-validation approach was employed. This technique iteratively constructs and validates the model on four out of the five EBT sites. The remaining site is then used for validation, allowing for the model coefficients to be refined for that specific region. By repeating this process for each site, five unique correction models were effectively created, each tailored to address biases within distinct regions of the EBT biome.

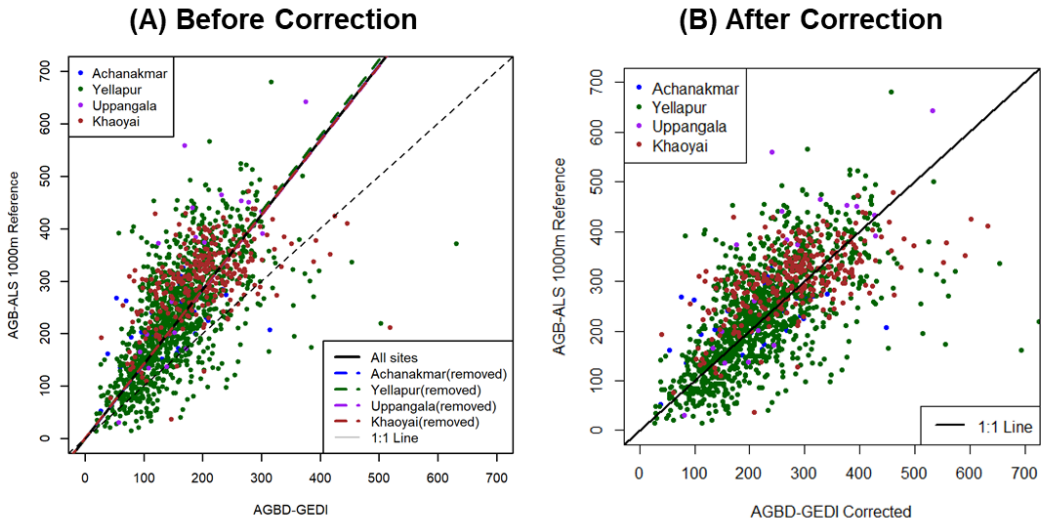


*Figure 6.6. Impact of Bias Correction on GEDI AGBD Accuracy at 100m Grid Resolution. (A) Before and (B) After correction.*

Furthermore, the investigation explored the potential impact of spatial resolution on the efficacy of the bias correction model. The model was evaluated at multiple resolutions (100m, 500m, and 1000m) as illustrated in Figure 6.7. To guarantee reliable calibration at each resolution, minimum thresholds were

established for the number of qualifying GEDI shots within each grid cell: 3 for 100m, 25 for 500m, and 50 for 1000m.

*Table 6.7. Site-level comparison between LiDAR-AGB and GEDI AGB Estimates at 100m resolution before and after correction. Individual sites indicates estimates during Leave-One-Site-Out Validation.*

| | | After Correction | | Before Correction | |
|---|---|---|---|---|---|
| **Sites** | **N** | **RMSE** | **$R^2$** | **RMSE** | **$R^2$** |
| All Sites | 1449 | 85.8 | 0.42 | 114 | 0.42 |
| Achanakmar (test) | 46 | 72.7 | 0.25 | 96.2 | 0.25 |
| Yellapur (test) | 1097 | 85.2 | 0.42 | 110 | 0.42 |
| Uppangala (test) | 19 | 116 | 0.63 | 177 | 0.63 |
| Khaoyai (test) | 287 | 87.8 | 0.28 | 126 | 0.28 |

The results yielded promising insights. The model demonstrably reduced bias, as evidenced by the model output (Figure 6.7). However, at the 100m resolution, a residual degree of data dispersion persisted, suggesting limitations in capturing fine-scale variability of biomass. Interestingly, the Root Mean Square Error (RMSE) exhibited a trend of improvement with increasing grid resolution. This underscores the potential advantages of employing coarser scales to mitigate inherent uncertainties associated with GEDI data. At the 1km grid resolution, the overall RMSE decreased to approximately 20%, signifying a substantial improvement.

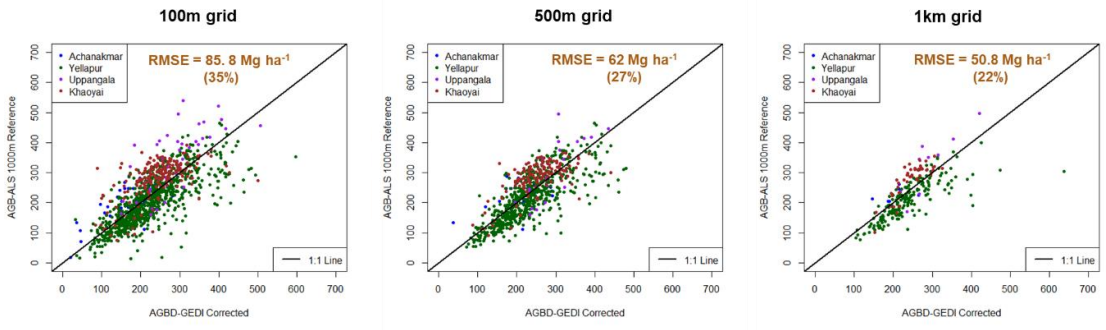


*Figure 6.7. Scatterplots Illustrating GEDI AGBD Accuracy Improvement with Increasing Grid Resolution after Bias Correction for EBT plant function type.*

These findings offer valuable and generalizable knowledge. While current data limitations restrict analysis of other biomes, the success achieved with EBT

forests suggests broader potential applicability of this approach. The LiDAR reference biomass maps employed within this study convincingly demonstrate their utility as calibration tools for Earth Observation (EO) missions. This paves the way for the development of more accurate, large-scale assessments of forest biomass across diverse regions.

## 6.4 Discussion

The advent of space-borne LiDAR systems, GEDI and ATLAS, presents a unique opportunity to analyze 3D vegetation structure globally. Though both systems are widely different in sensor technologies, primary science goals and intended geographical coverage, the results suggest they can complement each other in given circumstances. Since the study site can be characterized largely as open forests with considerable canopy gaps, the ground detection under favorable (i.e. cloud-free) night-time conditions was not found to be problematic in both leaf-on and leaf-off conditions for strong beams of both the sensors.

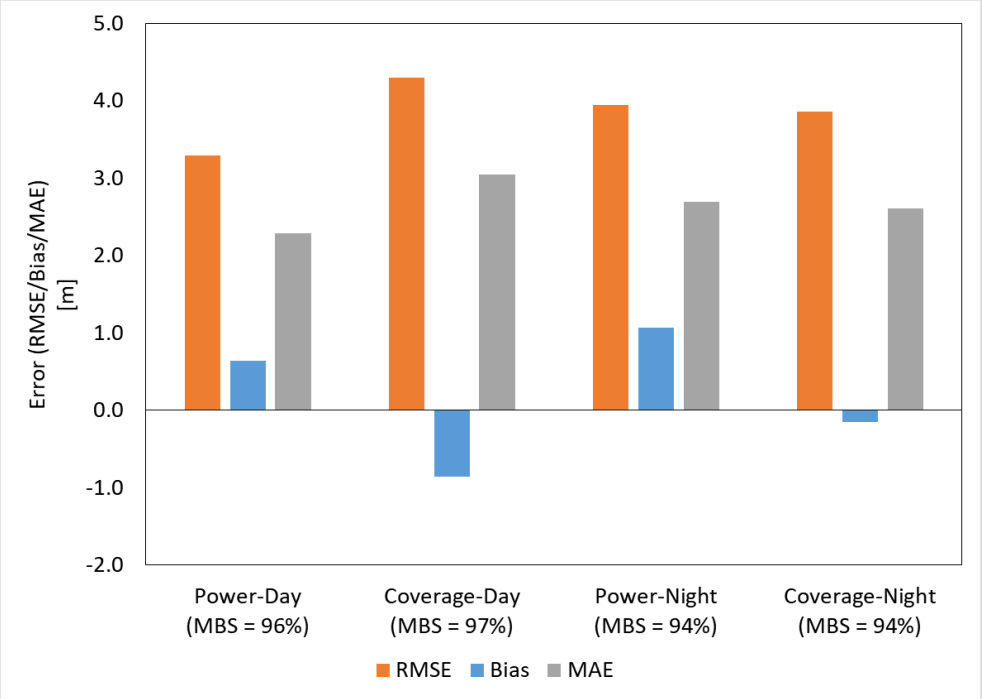


*Figure 6.8. GEDI canopy height retrieval errors during leaf-on period across various beam strengths and acquisition types. MBS indicates the mean beam sensitivity of all the shots in the associated category*

The errors in canopy height retrievals from GEDI during leaf-on season acquisition and beam types were observed to be in similar, contrasting to the results over forests in and around USA, where the strong beams of GEDI were found to be more accurate in detecting canopy tops than weak beams (Liu et al., 2021). This could be attributed to the minimal variability in mean beam sensitivity (94% – 97%) of the GEDI shots in varying beam strength and

acquisition time conditions (Figure 6.8). Considering canopy openings over the study site, both laser types (power and coverage beam) of GEDI are capable of penetrating canopies with high accuracy and thus yield comparable RMSE and bias errors. Similar results were shown by a study over tropical dense forests of French Guiana, where the GEDI shots with high-beam sensitivity (≥ 98%) tend were found to have greater penetration and can detect ground and canopy top accurately irrespective of the laser type (Fayad et al., 2022).

The RMSE in canopy height retrievals over this tropical forest site during the ideal acquisition conditions (night-time and high power beams) was found to be 3.9m for GEDI during leaf-on season and 3.5m for ATLAS during leaf-off season. The errors obtained in this study are higher than the reported RMSE of 1.1m for canopy height retrievals using ICESat-2 ATLAS over sub-tropical forests of India in northwestern Himalayas (Nandy et al., 2021). However, the number of sample sites used for validation are very few (N = 6) and the sample sites are of 32m x 32m in shape which are very different from the footprint of ATLAS products at 100m x 14m. A better approach is employed by Musthafa et al. (2023) to use field plots of 32m x 32m to validate GEDI and ICESat-2 products by using interpolation of the footprint-level canopy height retrievals. The study reports an RMSE in the range 2.62m – 3.99m for GEDI and 5.08 – 5.71m for ICESat-2 over tropical and sub-tropical forests of India (Musthafa et al., 2023). Though these studies have used field-measured canopy heights for validation of space-borne LiDAR metrics, the tree heights measured from ground are often error-prone depending up on the visibility of the tree-top in dense canopy systems (Réjou-Méchain et al., 2019). Hence for an accurate reference measurement of tree height to validate space-borne LiDAR metrics and LiDAR data from airborne platforms (both aircrafts/drones) is chosen as the ground reference across various studies to assess the uncertainty of these space-borne canopy height retrievals (Dorado-Roda et al. 2021; Neuenschwander and Magruder 2019).

Nevertheless, the time-difference between reference ALS data and the space-borne LiDAR acquisitions play a crucial role in such evaluations in different tree species. Though it is ideal to have both acquisitions in the similar time frame, it is often not possible considering the costs and planning involved during Aerial LiDAR missions. On the other hand, evaluation of space-borne LiDAR missions through field plots acquiring canopy height measurements over large areas is extremely difficult, expensive and time-consuming. Hence, though slightly apart ALS serves as the best alternative of the reference canopy height and was used by several studies to validate space-borne forest canopy height retrievals with time-differences up to 1 – 10 years between the acquisitions (Adam et al., 2020; Lahssini et al., 2022; Sothe et al., 2022). However, this assumption may be invalid in fast growing tree plantations (viz., Eucalyptus or *Prunus avium* etc.) or areas with very young trees (Musthafa et al., 2023). In the current study, although there is a notable difference between ALS acquisition and

Space-borne LiDAR acquisitions (~5 years), considering the natural forest system in the study area, the increments in forest canopy height over an area are assumed to be minimal although other forest canopy structures may change. Further, we have ensured no major land-cover changes over the forests which impact forest heights drastically are observed using temporal Landsat-8 images.

The accuracy of canopy height and terrain height retrievals from GEDI during day-time was observed to be better in terms of bias and RMSE when compared to ATLAS retrievals in both leaf-off and leaf-on periods (Table 6.3 and Table 6.4). This could be largely attributed to the difference in wavelengths of ATLAS (532 nm) and GEDI (1064 nm), the day time observations of ATLAS are more susceptible to scattering during the day and hence could lead to such potential uncertainties. During the leaf-on period irrespective of time of acquisition, GEDI was able to successfully retrieve canopy and terrain heights (Table 6.3 and Table 6.4). The full waveform retrieved from the GEDI was able to successfully notify the ground and canopy structures so as to enable accurate ground detection owing to the high canopy penetration of GEDI signal. In comparison, the photon-counting technology of ATLAS would be tested over the dense canopies due to lack of penetration. But, this could not be evaluated over the current study site due to lack of data during leaf-on season when the canopy cover at the study site is maximum. However, the underestimation of canopy height retrieval from ATLAS was observed by Liu et al. (2021) with increasing canopy cover and canopy heights. Consecutively, considering low canopy cover during leaf-off season, the performance of ATLAS was similar to GEDI during leaf-on season. However, during leaf-off season GEDI was not able to differentiate ground and canopy structure efficiently leading to large underestimations.

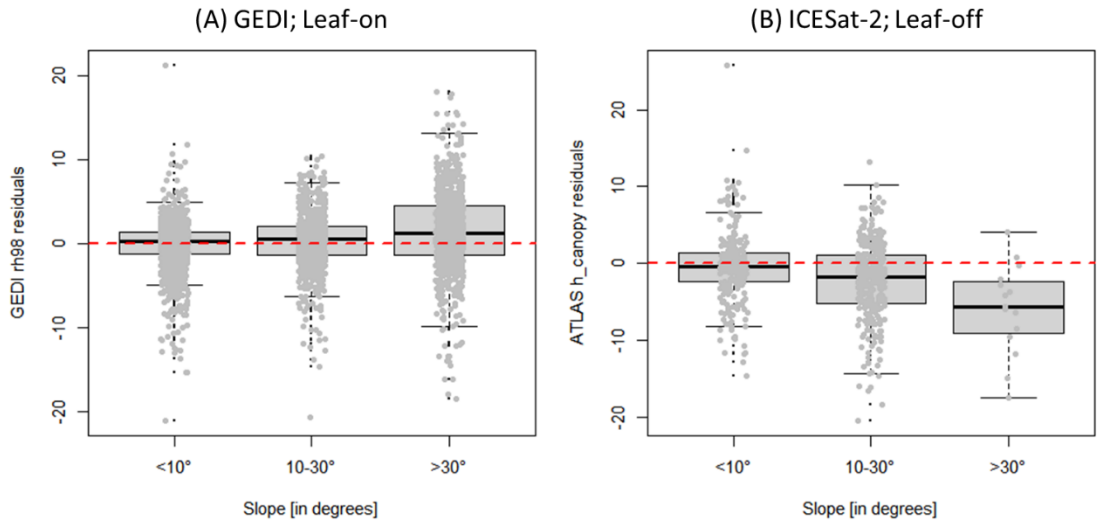


*Figure 6.9. Influence of terrain slope on the canopy height residuals as retrieved from (A) GEDI during leaf-on season and (B) ICESat-2 (ATLAS) during leaf-off season.*

The GEDI-AGBD estimates over the study area were found to be largely underestimated and biased (with RMSE of 46%) owing to the fact that the parametric models used to generate AGBD estimates from RH metrics for the South Asian region are not calibrated locally and use models calibrated over other plant functional types and regions. For example, over the study area, the DBT_SAs model is calibrated over the Evergreen Broadleaf Trees in Africa and the GSW_SAs model is calibrated using data from all over the globe. In this regard, the use of the GEDI-AGBD product over Indian (and South Asian region) is still notional and needs significant improvement and understanding.

The topographic slope is a crucial factor affecting signal processing and influencing the outputs derived from the space-borne LIDARs (Dorado-Roda et al., 2021; Xing et al., 2020). In the study area, the RMSE error was observed to be low for the flat areas (slopes < 10°) when compared to steep slopes (> 30°) for the canopy heights derived from both GEDI and ICESat-2. Since the GEDI AGBD was also derived from the relative height metrics from the GEDI L2A data, the terrain slope strongly affected the AGBD accuracy. The steeper slopes were observed to be strongly overestimating both canopy height and AGBD during leaf-on season. During leaf-off season, variability in slope least effected the retrieval of GEDI based canopy heights. Similar observations of steeper slopes (>30°) affecting the accuracy of ICESat-2 and GEDI in predicting terrain and canopy height are reported by Liu et al. (2021) and Wang et al. (2019). For GEDI, the best rh98 predictions were made for flat areas during the leaf-on season with an RMSE of 2.95m and a mean bias of -0.18 m (Figure 6.9). Similarly, for ATLAS, the h_canopy predictions over flat areas yielded the lowest RMSE of 4.85m (bias = 0.38 m) during the leaf-off season (Figure 6.9).

## 6.5 Summary

This study evaluated the performance of two spaceborne altimeters, GEDI and ICESat-2, in retrieving terrain and canopy heights in tropical dry deciduous forests of India. The goal was to understand how well these satellites performed under varying leaf cover conditions in comparison to the reference data from airborne LiDAR during both leaf-on and leaf-off seasons. During the leaf-on season, GEDI accurately measured canopy heights under various acquisition conditions. However, its performance suffered during the leaf-off season. GEDI struggled to differentiate between ground and canopy, leading to underestimates of canopy height. In contrast, ICESat-2 performed better for leaf-off canopy height retrieval, particularly using strong nighttime beams. This is likely because the laser beams could penetrate more effectively through the sparse canopy during this time. Overall, GEDI seems better suited for canopy height retrieval during leaf-on seasons due to its superior ability to penetrate through leaves. However, further evaluation across diverse forest types is necessary to solidify this conclusion.

While assessing GEDI's effectiveness for above-ground biomass density (AGBD) estimation, the study also highlighted challenges associated with GEDI AGBD estimates. These estimates exhibited significant discrepancies when compared to reference data specifically in South Asia. This underlines the need for more calibration sites and robust models, particularly for South Asia.

To address GEDI's AGBD estimation limitations, we compared its estimates with reference maps derived from airborne LiDAR data and field plots across diverse South Asian forests. The study examined GEDI's accuracy across various locations and forest types, it's potential for comprehensive AGBD mapping when combined with airborne LiDAR data, and the uncertainties associated with GEDI measurements. This analysis aimed to identify areas for improvement and ultimately enable more reliable use of GEDI data for robust AGBD estimation.

A key finding of the research is a novel bias correction method specifically designed for GEDI AGBD estimates in South Asian Evergreen Broadleaf forests. This method leverages leave-one-site-out cross-validation for robustness and considers the unique characteristics of these forests. The correction model demonstrates improved accuracy with increasing spatial resolution, suggesting promise for regional-scale biomass mapping using GEDI data. However, challenges remain at finer resolutions. Further research is warranted to explore the applicability of this methodology to other biomes and larger areas. The initial findings are promising, indicating that GEDI data, combined with robust correction models and high-quality reference data, has the potential to deliver reliable and efficient regional biomass mapping, ultimately contributing to improved carbon monitoring and forest management practices.

The research emphasizes the importance of generating high-quality reference datasets with well-understood uncertainties for improved GEDI AGBD estimates. Overall, this research provides valuable insights for utilizing GEDI and ICESat-2 measurements under various conditions to obtain accurate canopy heights. These findings can inform the development of improved GEDI models that translate into more reliable above-ground biomass estimates at regional and national scales.

# Chapter 7

# Conclusions and Future Scope

## 7.1 Conclusions

This thesis comprehensively investigated and addressed critical aspects of improving spatial Above-Ground Biomass (AGB) estimations for tropical forest ecosystems using Earth Observation (EO) data. The research focused on integrating Light Detection and Ranging (LiDAR) technology within a hierarchical framework to enhance accuracy, reliability, and uncertainty quantification. The key achievements of this thesis are as below.

- Defined an optimal field plot size for generating high-quality plot-level AGB data. This finding provides valuable guidance for future forest assessments, ensuring reliable data suitable for upscaling to remote sensing applications.
- Established a novel workflow utilizing Terrestrial LiDAR (TLS) for non-destructive tree allometry. This approach reduces uncertainties in plot-level AGB estimates, improving accuracy and efficiency compared to traditional methods.
- Generated comprehensive reference AGB maps using Airborne LiDAR data across diverse tropical forest sites. These open-access maps serve as a valuable resource for the remote sensing community, providing ground-truth data for calibrating and validating large-scale biomass mapping models.
- Developed detailed uncertainty estimates for reference LiDAR AGB maps, facilitating the implementation of a hierarchical modeling approach for regional AGB mapping as recommended by the IPCC. This approach ensures reliable regional AGB maps with accurate error representations.
- Demonstrated the applicability of the developed methods by creating a correction model for Indian forests that improves the accuracy of NASA's

GEDI spaceborne LiDAR mission. This showcases the real-world impact of the research, contributing to more accurate global biomass assessments.

Traditional methods for spatial AGB estimates exhibited significant discrepancies across India. These discrepancies were traced back to limitations in ground data collection. Plot size limitations resulted in insufficient sampling of tree size variations, leading to underestimates of total biomass. Additionally, the accuracy of allometric models used to convert field measurements to biomass estimates was found to be a crucial factor. Existing models may not adequately capture the specific growth patterns and biomass allocation of tropical tree species.

This research not only provided a standardized protocol for field data collection but also explored the development of TLS-derived tree volume equations, offering a non-destructive alternative to traditional methods. TLS technology allows for the creation of highly detailed 3D tree models, enabling accurate volume estimations without the need for destructive sampling. Additionally, a methodology for generating reference LiDAR-AGB maps applicable across diverse tropical forest ecosystems was established. This methodology considers variations in forest structure and species composition, ensuring the generalizability of the approach.

Integrating high-resolution Aerial LiDAR data offered substantial improvements in AGB estimations. The detailed 3D structural information captured by LiDAR facilitated the development of more accurate reference LiDAR-AGB maps. These maps consider the spatial distribution of vegetation within a forest stand, providing a more comprehensive picture of biomass compared to traditional plot-based approaches. The reference maps serve as valuable calibration and validation datasets for EO missions like GEDI, promoting more reliable biomass assessments from space.

The thesis emphasized the importance of a hierarchical modeling approach that incorporates error propagation throughout the estimation process. This approach, recommended by the IPCC, strengthens the estimation process across all stages. It allows for the quantification of uncertainties associated with ground measurements, plot size selection, allometric model choice, and the upscaling process using EO data. By accounting for these uncertainties, the final regional AGB estimates become more robust and defensible.

This thesis contributes significantly to advancing the accuracy and reliability of EO-based spatial AGB estimates for tropical forests. The open-access reference LiDAR-AGB maps and established methodologies pave the way for improved monitoring of biomass change dynamics and carbon stocks. These resources can be utilized by researchers, policymakers, and environmental agencies to develop more effective strategies for forest conservation and carbon sequestration efforts. The findings highlight the crucial role of LiDAR

technology, not just in generating high-quality reference data but also in calibrating and validating upcoming EO missions dedicated to biomass assessments. This research lays the groundwork for more robust and defensible biomass estimates at regional and global scales, enabling a more comprehensive understanding of forest carbon pools and their role in the global carbon cycle.

## 7.2 Future Research Directions

Further refinement of allometric models, particularly for temperate Indian forests, is necessary to enhance regional AGB estimations. The development of ecosystem-specific models that account for variations in tree species composition, growth patterns, and wood density would improve the accuracy of biomass estimates across diverse forest types. Additionally, exploring the potential of machine learning techniques for allometric model development holds promise for further advancements in this area.

Expanding the network of reference LiDAR-AGB maps across a broader range of tropical forest types would strengthen the calibration and validation of EO data. A wider geographical scope would improve the generalizability of the findings and enhance the applicability of the established methodologies. This could involve collaborating with research institutions and forestry agencies in other tropical regions to establish additional reference sites.

Investigating the integration of additional remote sensing data sources, such as optical and radar imagery, with the established LiDAR-based approach could offer further improvements in spatial AGB estimations. By leveraging the complementary strengths of different sensors, a more comprehensive understanding of forest structure and biomass could be achieved. For example, optical imagery can provide information on leaf area index and vegetation health, while radar imagery can penetrate through the forest canopy to capture information on underlying structural elements.

In conclusion, this thesis successfully demonstrated the effectiveness of LiDAR technology within a hierarchical framework for enhancing spatial AGB assessments in tropical forests. The developed methodologies and open-access data resources hold significant promise for improving the accuracy and reliability of EO-based biomass monitoring efforts. This research contributes to a more comprehensive understanding of global carbon dynamics, which is crucial for informing effective climate change mitigation strategies and ensuring the sustainable management of our vital forest ecosystems.

# Chapter 8

# Bibliography

Adam, M., Urbazaev, M., Dubois, C., Schmullius, C., 2020. Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters. Remote Sens. 12, 3948.

Amelung, F., others, 2019. NASA-ISRO SAR (NISAR) Mission Science Users' Handbook. Jet Propuls. Lab.

Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Joseph Wright, S., Abu Salim, K., Almeyda Zambrano, A.M., Alonso, A., Baltzer, J.L., others, 2015. CTFS-Forest GEO: a worldwide network monitoring forests in an era of global change. Glob. Chang. Biol. 21, 528–549.

Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J.S., van Breugel, M., 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia 168, 1147–1160.

Avitabile, V., Herold, M., Heuvelink, G.B.M., Lewis, S.L., Phillips, O.L., Asner, G.P., Armston, J., Ashton, P.S., Banin, L., Bayol, N., others, 2016. An integrated pan-tropical biomass map using multiple reference datasets. Glob. Chang. Biol. 22, 1406–1420.

Baccini, A., Friedl, M.A., Woodcock, C.E., Warbington, R., 2004. Forest biomass estimation over regional scales using multisource data. Geophys. Res. Lett. 31.

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., Samanta, S., Houghton, R.A., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nat. Clim. Chang.

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. Environ. Res. Lett. 3, 45011.

Baskerville, G.L., 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. For. Res. 2, 49–53.

Basuki, T.M., Van Laake, P.E., Skidmore, A.K., Hussin, Y.A., 2009. Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. For. Ecol. Manage. 257, 1684–1694.

Bienert, A., Maas, H.-G., Scheller, S., 2006. Analysis of the information content of terrestrial laserscanner point clouds for the automatic determination of forest inventory parameters, in: Workshop on 3D Remote Sensing in Forestry. p. 15th.

Biradar, A.S., Roy Chowdhury, D., Garg, S., Zhang, J., Zhu, Z., Anavaya, H.M., 2020. Spaceborne LiDAR for 3D vegetation structure retrieval over tropical forest: Accounting for footprint variability and canopy gaps. Remote Sens. 12, 2202. https://doi.org/10.3390/rs12142202

Bonan, G.B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science (80-. ). 320, 1444–1449.

Borsah, A.A., Nazeer, M., Wong, M.S., 2023. LIDAR-Based Forest Biomass Remote Sensing: A Review of Metrics, Methods, and Assessment Criteria for the Selection of Allometric Equations. Forests 14, 2095.

Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS J. Photogramm. Remote Sens. 61, 325–340.

Brede, B., Calders, K., Lau, A., Raumonen, P., Bartholomeus, H.M., Herold, M., Kooistra, L., 2019. Non-destructive tree volume estimation through quantitative structure modelling: Comparing UAV laser scanning with terrestrial LIDAR. Remote Sens. Environ. 233, 111355.

Brown, S., 1997. Estimating biomass and biomass change of tropical forests: a primer.

Brown, S., Iverson, L.R., Prasad, A., UT-Battelle, L.L.C., 2001. Geographical Distribution of Biomass Carbon in Tropical Southeast Asian Forests: A Database (NDP-068).

Cairns, M.A., Brown, S., Helmer, E.H., Baumgardner, G.A., 1997. Root biomass allocation in the world's upland forests. Oecologia 111, 1–11.

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., Culvenor, D., Avitabile, V., Disney, M., Armston, J., 2015. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. Methods Ecol. Evol. 6, 198–208.

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., Culvenor, D., Avitabile, V., Disney, M., Armston, J., others, 2014.

Nondestructive estimates of above-ground biomass using terrestrial laser scanning. Methods Ecol. Evol.

Calders, K., Origo, N., Burt, A., Disney, M., Nightingale, J., Raumonen, P., Åkerblom, M., Malhi, Y., Lewis, P., 2018. Realistic forest stand reconstruction from terrestrial LiDAR for radiative transfer modelling. Remote Sens. 10, 933.

Chaturvedi, R.K., Raghubanshi, A.S., 2015. Allometric models for accurate estimation of aboveground biomass of teak in tropical dry forests of India. For. Sci. 61, 938–949.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and scaling for tropical forest biomass estimates. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 359, 409–420.

Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G., Zanne, A.E., 2009. Towards a worldwide wood economics spectrum. Ecol. Lett. 12, 351–366.

Chave, J., Davies, S.J., Phillips, O.L., Lewis, S.L., Sist, P., Schepaschenko, D., Armston, J., Baker, T.R., Coomes, D., Disney, M., others, 2019. Ground data are essential for biomass remote sensing missions. Surv. Geophys. 40, 863–880.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., others, 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. 20, 3177–3190.

Chen, Q., Laurin, G.V., Valentini, R., 2015. Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. Remote Sens. Environ. 160, 134–143.

Clark, D.B., Kellner, J.R., 2012. Tropical forest biomass estimation and the fallacy of misplaced concreteness. J. Veg. Sci. 23, 1191–1196.

Coops, N.C., Tompalski, P., Goodbody, T.R.H., Queinnec, M., Luther, J.E., Bolton, D.K., White, J.C., Wulder, M.A., van Lier, O.R., Hermosilla, T., 2021. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. Remote Sens. Environ. 260, 112477.

Curtis, J.T., McIntosh, R.P., 1951. An upland forest continuum in the prairie-forest border region of Wisconsin. Ecology 32, 476–496.

de Tanago, J., Lau, A., Bartholomeus, H., Herold, M., Avitabile, V., Raumonen, P., Martius, C., Goodman, R.C., Disney, M., Manuri, S., others, 2018. Estimation of above-ground biomass of large tropical trees with terrestrial LiDAR. Methods Ecol. Evol. 9, 223–234.

Dorado-Roda, I., Pascual, A., Godinho, S., Silva, C.A., Botequim, B., Rodr\'\iguez-Gonzálvez, P., González-Ferreiro, E., Guerra-Hernández, J., 2021. Assessing the accuracy of GEDI data for canopy height and aboveground biomass estimates in Mediterranean forests. Remote Sens. 13, 2279.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., others, 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci. Remote Sens. 1, 100002.

Dubayah, R.O., Drake, J.B., 2000. Lidar remote sensing for forestry. J. For. 98, 44–46.

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Carter, S., Chave, J., Herold, M., Crowther, T.W., others, 2019. The importance of consistent global forest aboveground biomass product validation. Surv. Geophys. 40, 979–999.

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Carter, S., Chave, J., Herold, M., MacBean, N., others, 2021. Aboveground Woody biomass product validation good practices protocol.

Duncanson, L., Kellner, J.R., Armston, J., Dubayah, R., Minor, D.M., Hancock, S., Healey, S.P., Patterson, P.L., Saarela, S., Marselis, S., others, 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. Remote Sens. Environ. 270, 112845.

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C.A., Armston, J., Luthcke, S.B., Hofton, M., others, 2020. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. Remote Sens. Environ. 242, 111779.

Duncanson, L.I., Dubayah, R.O., Enquist, B.J., 2015. Assessing the general patterns of forest structure: quantifying tree and forest allometric scaling relationships in the U nited S tates. Glob. Ecol. Biogeogr. 24, 1465–1475.

Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., 2006. 2006 IPCC guidelines for national greenhouse gas inventories.

Falkowski, M.J., Smith, A.M.S., Hudak, A.T., Gessler, P.E., Vierling, L.A., Crookston, N.L., 2006. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. Can. J. Remote Sens. 32, 153–161.

Fararoda, R., Reddy, R.S., Rajashekar, G., Chand, T.R.K., Jha, C.S., Dadhwal, V.K., 2021. Improving forest above ground biomass estimates over Indian forests

using multi source data sets with machine learning algorithm. Ecol. Inform. 101392.

Fayad, I., Baghdadi, N., Lahssini, K., 2022. An Assessment of the GEDI Lasers' Capabilities in Detecting Canopy Tops and Their Penetration in a Densely Vegetated, Tropical Area. Remote Sens. 14, 2969.

Fayad, I., Baghdadi, N., Riédi, J., 2021. Quality assessment of acquired gedi waveforms: Case study over france, tunisia and french guiana. Remote Sens. 13, 3144.

Fernandez-Diaz, J.C., Velikova, M., Glennie, C.L., 2022. Validation of ICESat-2 ATL08 Terrain and Canopy Height Retrievals in Tropical Mesoamerican Forests. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 2956–2970.

Foody, G.M., Boyd, D.S., Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sens. Environ. 85, 463–474.

Franklin, J., 1986. Thematic Mapper analysis of coniferous forest structure and composition. Int. J. Remote Sens. 7, 1287–1301.

Frazer, G.W., Magnussen, S., Wulder, M.A., Niemann, K.O., 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sens. Environ. 115, 636–649.

Friedlingstein, P., O'Sullivan, M., Jones, M.W., Andrew, R.M., Hauck, J., Olsen, A., Peters, G.P., Peters, W., Pongratz, J., Sitch, S., others, 2020. Global carbon budget 2020. Earth Syst. Sci. Data 12, 3269–3340.

FSI, 2021. Indian State of Forest Report 2021. Forest Survey of India, Ministry of Environment and Forests, Dehra Dun.

FSI, 1996. Volume Equations for Forests of India, Nepal and Bhutan.Forest Survey of India. Ministry of Environment and Forests, Govt. of India, Dehradun.

Garcia, M., Riaño, D., Chuvieco, E., Danson, F.M., 2010. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. Remote Sens. Environ. 114, 816–830.

Ghosh, S.M., Behera, M.D., Kumar, S., Das, P., Prakash, A.J., Bhaskaran, P.K., Roy, P.S., Barik, S.K., Jeganathan, C., Srivastava, P.K., others, 2022. Predicting the Forest Canopy Height from LiDAR and Multi-Sensor Data Using Machine Learning over India. Remote Sens. 14, 5968.

Gobakken, T., Næsset, E., 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. Can. J. For. Res. 39, 1036–1052.

Gonçalves, F., Treuhaft, R., Law, B., Almeida, A., Walker, W., Baccini, A., Dos Santos, J.R., Graça, P., 2017. Estimating aboveground biomass in tropical forests: Field methods and error analysis for the calibration of remote sensing observations. Remote Sens. 9, 47.

Hackenberg, J., Wassenberg, M., Spiecker, H., Sun, D., 2015. Non Destructive Method for Biomass Prediction Combining TLS Derived Tree Volume and Wood Density. Forests 6, 1274–1300.

Harris, N.L., Brown, S., Hagen, S.C., Saatchi, S.S., Petrova, S., Salas, W., Hansen, M.C., Potapov, P. V., Lotsch, A., 2012. Baseline Map of Carbon Emissions from Deforestation in Tropical Regions. Science (80-. ).

Herold, M., Carter, S., Avitabile, V., Espejo, A.B., Jonckheere, I., Lucas, R., McRoberts, R.E., Næsset, E., Nightingale, J., Petersen, R., others, 2019. The role and need for space-based forest biomass-related measurements in environmental management and policy. Surv. Geophys. 40, 757–778.

Hofton, M.A., Rocchio, L., Blair, J.B., 2006. Comparison of lidar flight line planning strategies for variable density forests. Can. J. Remote Sens. 32, 142–150.

Houghton, R.A., Hall, F., Goetz, S.J., 2009. Importance of biomass in the global carbon cycle. J. Geophys. Res. Biogeosciences 114.

Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J., Guo, Q., 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. Remote Sens. 8, 565.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., Wang, L., 2011. Automated methods for measuring DBH and tree heights with a commercial scanning lidar. Photogramm. Eng. Remote Sens. 77, 219–227.

Hyyppa, J., Kelle, O., Lehikoinen, M., Inkinen, M., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Trans. Geosci. Remote Sens. 39, 969–975.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49, 12–35.

Jennings, S.B., Brown, N.D., Sheil, D., 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry 72, 59–74.

Kearsley, E., Moonen, P.C.J., Hufkens, K., Doetterl, S., Lisingo, J., Bosela, F.B., Boeckx, P., Beeckman, H., Verbeeck, H., 2017. Model performance of tree height-diameter relationships in the central Congo Basin. Ann. For. Sci. 74, 1–13.

Kent, M., 2011. Vegetation description and data analysis: a practical approach.

John Wiley \& Sons.

Kindt, R., Kindt, M.R., 2015. Package 'BiodiversityR.' R Proj.

Kobayashi, H., Iwabuchi, H., 2008. A coupled 1-D atmosphere and 3-D canopy radiative transfer model for canopy reflectance, light environment, and photosynthesis simulation in a heterogeneous landscape. Remote Sens. Environ. 112, 173–185.

Koch, B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS J. Photogramm. Remote Sens. 65, 581–590.

Labrière, N., Davies, S.J., Disney, M.I., Duncanson, L.I., Herold, M., Lewis, S.L., Phillips, O.L., Quegan, S., Saatchi, S.S., Schepaschenko, D.G., others, 2023. Toward a forest biomass reference measurement system for remote sensing applications. Glob. Chang. Biol. 29, 827–840.

Labrière, N., Davies, S.J., Disney, M.I., Duncanson, L.I., Herold, M., Lewis, S.L., Phillips, O.L., Quegan, S., Saatchi, S.S., Schepaschenko, D.G., others, 2022. Toward a forest biomass reference measurement system for remote sensing applications. Glob. Chang. Biol.

Lahssini, K., Baghdadi, N., Le Maire, G., Fayad, I., 2022. Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests. Remote Sens. 14, 6264.

Latifi, H., 2012. Characterizing forest structure by means of remote sensing: a review. Remote Sensing-Advanced Tech. Platforms 953–978.

Lefsky, M., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., 1999a. Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. Remote Sens. Environ. 70, 339–361. https://doi.org/https://doi.org/10.1016/S0034-4257(99)00052-8

Lefsky, M., Harding, D., Cohen, W.B., Parker, G., Shugart, H.H., 1999b. Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. Remote Sens. Environ. 67, 83–98.

Lefsky, M.A., 2010. A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. Geophys. Res. Lett. 37.

Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular inter. Bioscience 52, 19–30.

Liang, X., Li, X., Sun, Y., Li, D., Wang, Z., Zeng, S., 2018. Forest inventory using

terrestrial laser scanning: a comparison of plot-based and area-based inventory methods. Remote Sens. 10, 70. https://doi.org/10.3390/rs10010070

Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. Prog. Phys. Geogr. 27, 88–106.

Lim, K.S., Treitz, P.M., 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scand. J. For. Res. 19, 558–570.

Lin, K., Lyu, M., Jiang, M., Chen, Y., Li, Y., Chen, G., Xie, J., Yang, Y., 2017. Improved allometric equations for estimating biomass of the three Castanopsis carlesii H. forest types in subtropical China. New For. 48, 115–135.

Lin, L.I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 255–268.

Liu, A., Cheng, X., Chen, Z., 2021. Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. Remote Sens. Environ. 264, 112571.

Liu, Y.Y., Van Dijk, A.I.J.M., De Jeu, R.A.M., Canadell, J.G., McCabe, M.F., Evans, J.P., Wang, G., 2015. Recent reversal in loss of global terrestrial biomass. Nat. Clim. Chang. 5, 470–474.

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27, 1297–1328.

Maas, H.-G., Bienert, A., Scheller, S., Keane, E., 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. Int. J. Remote Sens. 29, 1579–1593.

Malhi, Y., Phillips, O.L., Lloyd, J., Baker, T., Wright, J., Almeida, S., Arroyo, L., Frederiksen, T., Grace, J., Higuchi, N., others, 2002. An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). J. Veg. Sci. 13, 439–450.

Maltamo, M., Mustonen, K., Hyyppä, J., Pitkänen, J., Yu, X., 2004. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. Can. J. For. Res. 34, 1791–1801.

Mayamanikandan, T., Reddy, S., Fararoda, R., Chand, K., Soma, M., Praveen, S., Rajashekar, G., Jha, C.S., Das, I.C., Gummapu, J., 2020. Quantifying the influence of plot-level uncertainty in above ground biomass up scaling using remote sensing data in central Indian dry deciduous forest. Geocarto Int. 0, 1–15. https://doi.org/10.1080/10106049.2020.1864029

Mayamanikandan, T., Suraj Reddy, R., Jha, C.S., 2019. Non-Destructive Tree Volume Estimation Using Terrestrial Lidar Data in Teak Dominated Central

Indian Forests, in: IEEE Geoscience and Remote Sensing Conference.

McRoberts, R.E., Næsset, E., Saatchi, S., Quegan, S., 2022. Statistically rigorous, model-based inferences from maps. Remote Sens. Environ. 279, 113028.

McRoberts, R.E., Tomppo, E., Næsset, E., Rautiainen, A., 2010. Remote sensing derived estimates of forest biomass: a review. Scand. J. For. Res. 25, 70–88. https://doi.org/10.1080/02827580903419651

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J., others, 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogramm. Eng. Remote Sensing 66, 1367–1372.

Mitchard, E.T.A., Feldpausch, T.R., Brienen, R.J.W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T.R., Lewis, S.L., Lloyd, J., Quesada, C.A., Gloor, M., others, 2014a. Markedly divergent estimates of A mazon forest carbon density from ground plots and satellites. Glob. Ecol. Biogeogr. 23, 935–946.

Mitchard, E.T.A., Feldpausch, T.R., Brienen, R.J.W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T.R., Lewis, S.L., Lloyd, J., Quesada, C.A., Gloor, M., others, 2014b. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. Glob. Ecol. Biogeogr. 23, 935–946.

Mitchard, E.T.A., Saatchi, S.S., Baccini, A., Asner, G.P., Goetz, S.J., Harris, N., Brown, S., 2013. Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. Carb Bal Manag.

Mitchard, E.T.A., Saatchi, S.S., Lewis, S.L., Feldpausch, T.R., Woodhouse, I.H., Sonké, B., Rowland, C., Meir, P., 2011. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest--savanna boundary region of central Africa using multi-temporal L-band radar backscatter. Remote Sens. Environ. 115, 2861–2873.

Morsdorf, F., Ewel, J., Rogerts, G., McFadden, J.P., 2009. Estimating tree heights and biomass in a boreal forest using a combination of spectral reflectance and lidar data. Remote Sens. Environ. 113, 1000–1012. https://doi.org/10.1016/j.rse.2008.09.004

Morsdorf, F., Kötz, B., Meier, E., Itten, K.I., Allgöwer, B., 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. Remote Sens. Environ. 104, 50–61.

Motohka, T., Kankaku, Y., Miura, S., Suzuki, S., 2021. Overview of ALOS-2 and ALOS-4 L-band SAR, in: 2021 IEEE Radar Conference (RadarConf21). pp. 1–4.

Moussa, M., Mahamane, L., 2018. Allometric models for estimating aboveground biomass and carbon in Faidherbia albida and Prosopis africana under agroforestry parklands in drylands of Niger. J. For. Res. 29, 1703–1717.

Musthafa, M., Singh, G., 2022. Forest above-ground woody biomass estimation using multi-temporal space-borne LiDAR data in a managed forest at Haldwani, India. Adv. Sp. Res. 69, 3245–3257.

Musthafa, M., Singh, G., Kumar, P., 2023. Comparison of forest stand height interpolation of GEDI and ICESat-2 LiDAR measurements over tropical and sub-tropical forests in India. Environ. Monit. Assess. 195, 1–17.

Naesset, E., 1997. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS J. Photogramm. Remote Sens. 52, 49–56.

Næsset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scand. J. For. Res. 19, 164–179.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99.

Nandy, S., Srinet, R., Padalia, H., 2021. Mapping forest height and aboveground biomass by integrating ICESat-2, Sentinel-1 and Sentinel-2 data using Random Forest algorithm in northwest Himalayan foothills of India. Geophys. Res. Lett. 48, e2021GL093799.

Narine, L.L., Popescu, S.C., Malambo, L., 2020. Using ICESat-2 to estimate and map forest aboveground biomass: A first example. Remote Sens. 12, 1824.

Neuenschwander, A., Guenther, E., White, J.C., Duncanson, L., Montesano, P., 2020. Validation of ICESat-2 terrain and canopy heights in boreal forests. Remote Sens. Environ. 251, 112110.

Neuenschwander, A., Pitts, K., 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. Remote Sens. Environ. 221, 247–259.

Neuenschwander, A.L., Magruder, L.A., 2019. Canopy and terrain height retrievals with ICESat-2: A first look. Remote Sens. 11, 1721.

Neuenschwander, A.L., Pitts, K.L., Jelley, B.P., Robbins, J., Klotz, B., Popescu, S.C., Nelson, R.F., Harding, D., Pederson, D., Sheridan, R., 2021. ATLAS/ICESat-2 L3A land and vegetation height, version 3. Boulder, Color. USA. NASA Natl. Snow Ice Data Cent. Distrib. Act. Arch. Cent.

Neumann, T.A., Martino, A.J., Markus, T., Bae, S., Bock, M.R., Brenner, A.C., Brunt, K.M., Cavanaugh, J., Fernandes, S.T., Hancock, D.W., others, 2019. The Ice, Cloud, and Land Elevation Satellite--2 Mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. Remote Sens. Environ. 233, 111325.

NOAA, 2012. Lidar 101: An introduction to lidar technology, data, and applications. Natl. Ocean. Atmos. Adm.

Oliver, Chadwick Dearing, Larson, B.C., Oliver, C D, 1996. Forest stand

dynamics. Wiley New York.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A large and persistent carbon sink in the world's forests. Science 333, 988–993.

Parmentier, I., Harrigan, R.J., Buermann, W., Mitchard, E.T.A., Saatchi, S., Malhi, Y., Bongers, F., Hawthorne, W.D., Leal, M.E., Lewis, S.L., others, 2011. Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. J. Biogeogr. 38, 1164–1176.

Parrish, C.E., Scarpace, F.L., 2007. Investigating full-waveform lidar data for detection and recognition of vertical objects, in: American Society of Photogrammetry and Remote Sensing (ASPRS) Annual Conference.

Pelletier, J., Ramankutty, N., Potvin, C., 2011. Diagnosing the uncertainty and detectability of emission reductions for REDD+ under current capabilities: an example for Panama. Environ. Res. Lett. 6, 24005.

Picard, N., Bosela, F.B., Rossi, V., 2015. Reducing the error in biomass estimates strongly depends on model selection. Ann. For. Sci. 72, 811–823.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., others, 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540.

Ploton, P., Pélissier, R., Proisy, C., Flavenot, T., Barbier, N., Rai, S.N., Couteron, P., 2012. Assessing aboveground tropical forest biomass using Google Earth canopy images. Ecol. Appl. 22, 993–1003.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646–655.

Popescu, S.C., Zhao, K., Warner, B.A., Malkinson, D., 2007. Assessing variability in forest structure with ground-based lidar: Validation of scanning methodology and consequences for remote sensing applications. Remote Sens. Environ. 108, 375–383. https://doi.org/10.1016/j.rse.2006.10.010

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., others, 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sens. Environ. 253, 112165.

Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D.H.T., Lomas, M., D'alessandro, M.M., Paillou, P., Papathanassiou, K., others, 2019. The European Space Agency BIOMASS mission: Measuring forest above-ground

biomass from space. Remote Sens. Environ. 227, 44–60.

Rajashekar, G., Fararoda, R., Reddy, R.S., Jha, C.S., Ganeshaiah, K.N., Singh, J.S., Dadhwal, V.K., 2018. Spatial distribution of forest biomass carbon (Above and below ground) in Indian forests. Ecol. Indic. 85, 742–752.

Raumonen, P., Åkerblom, M., Kaasalainen, M., Casella, E., Calders, K., Murphy, S., 2015. Massive-scale tree modelling from TLS data. ISPRS Ann. Photogramm. Remote Sens. \& Spat. Inf. Sci. 2.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M., Lewis, P., 2013. Fast automatic precision tree models from terrestrial laser scanner data. Remote Sens. 5, 491–520.

Reddy, R Suraj, Fararoda, R., Jha, C.S., Rajan, K.S., 2018. Automatic estimation of tree stem attributes using terrestrial laser scanning in central Indian dry deciduous forests. Curr. Sci. 114.

Reddy, R. Suraj, Jha, C.S., Rajan, K.S., 2018. Automatic Tree Identification and Diameter Estimation Using Single Scan Terrestrial Laser Scanner Data in Central Indian Forests. J. Indian Soc. Remote Sens. 46, 937–943. https://doi.org/10.1007/s12524-018-0753-7

Reddy, S.R., Rajashekar, G., Jha, C.S., Dadhwal, V.K., Pelissier, R., Couteron, P., 2016. Estimation of Above Ground Biomass Using Texture Metrics Derived from IRS Cartosat-1 Panchromatic Data in Evergreen Forests of Western Ghats, India. J. Indian Soc. Remote Sens. 1–9. https://doi.org/10.1007/s12524-016-0630-1

Réjou-Méchain, M., Barbier, N., Couteron, P., Ploton, P., Vincent, G., Herold, M., Mermoz, S., Saatchi, S., Chave, J., de Boissieu, F., others, 2019. Upscaling Forest biomass from field to satellite measurements: sources of errors and ways to reduce them. Surv. Geophys. 40, 881–911.

Réjou-Méchain, M., Muller-Landau, H.C., Detto, M., Thomas, S.C., Toan, T. Le, Saatchi, S.S., Barreto-Silvia, J.S., Bourg, N.A., Bunyavejchewin, S., Butt, N., others, 2014. Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. Biogeosciences Discuss. 11, 5711.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., Hérault, B., 2017. biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. Methods Ecol. Evol. 8, 1163–1167.

Reutebuch, S.E., Andersen, H.-E., McGaughey, R.J., 2005. Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. J. For. 103, 286–292.

Richards, J.F., Flint, E.P., Daniels, R.C., 1994. Historic land use and carbon estimates for South and Southeast Asia 1880-1980. Carbon Dioxide

Information Analysis Center, Oak Ridge National Laboratory Oak Ridge.

Rodda, S.R., Fararoda, R., Jha, N., Réjou-Méchain, M., Couteron, P., Gopalakrishnan, R., Barbier, N., Alfonso, A., Bako, O., Bassama, P., Behera, D., Bissiengou, P., Biyiha, H., Brockelman Y., W., Chanthorn, W., Chauhan, P., Dadhwal, V.K., Dauby, G., Deblauwe, V., Dongmo, N., Droissart, V., Jeyakumar, S., Jha, C.S., Kandem, N.G., Katembo, J., Kougue, R., Leblanc, H., Lewis, S., Libalah, M., Manikandan, M., Martin-Ducup, O., Mbock, G., Memiaghe, H., Mofack, G., Mutyala, P., Narayanan, A., Nathalang, A., Oum Ndjock, G., Ngoula, F., Nidamanuri, R.R., Pélissier, R., Saatchi, S., Sagang, L.B., Salla, P., Simo-Droissart, M., B. Smith, T., Sonké, B., Stevart, T., Tjomb, D., Zebaze, D., Zemagho, L., Ploton, P., 2024. South Asian and Central African maps from: LiDAR-based reference aboveground biomass maps for tropical forests of South Asia and Central Africa. https://doi.org/10.23708/H2MHXF

Rodda, S.R., Nidamanuri, R.R., Fararoda, R., Mayamanikandan, T., Rajashekar, G., 2023. Evaluation of Height Metrics and Above-Ground Biomass Density from GEDI and ICESat-2 Over Indian Tropical Dry Forests using Airborne LiDAR Data. J. Indian Soc. Remote Sens. 1–16.

Rodda, S.R., Thumaty, K.C., Praveen, M.S.S., Jha, C.S., Dadhwal, V.K., 2021. Multi-year eddy covariance measurements of net ecosystem exchange in tropical dry deciduous forest of India. Agric. For. Meteorol. 301–302. https://doi.org/10.1016/j.agrformet.2021.108351

Roxburgh, S.H., Paul, K.I., Clifford, D., England, J.R., Raison, R.J., 2015. Guidelines for constructing allometric models for the prediction of woody biomass: how many individuals to harvest? Ecosphere 6, 1–27.

Saarela, S., Varvia, P., Korhonen, L., Yang, Z., Patterson, P.L., Gobakken, T., Næsset, E., Healey, S.P., Ståhl, G., 2023. Three-phase hierarchical model-based and hybrid inference. MethodsX 11, 102321.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., others, 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. 108, 9899–9904.

Santoro, M., Cartus, O., 2023. ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2010 2017 2018 2019 and 2020. NERC EDS Cent. Environ. Data Anal.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabile, V., Araza, A., De Bruin, S., Herold, M., Quegan, S., Rodr\'\iguez-Veiga, P., others, 2021. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth Syst. Sci. Data 13, 3927–3950.

Schimel, D., Pavlick, R., Fisher, J.B., Asner, G.P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., Cox, P., 2015. Observing terrestrial ecosystems and the carbon cycle from space. Glob. Chang. Biol. 21, 1762–1776.

Seidel, D., Fleck, S., Leuschner, C., Hammett, T., 2011. Review of ground-based methods to measure the distribution of biomass in forest canopies. Ann. For. Sci. 68, 225–244.

Silva, C.A., Duncanson, L., Hancock, S., Neuenschwander, A., Thomas, N., Hofton, M., Fatoyinbo, L., Simard, M., Marshak, C.Z., Armston, J., others, 2021. Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. Remote Sens. Environ. 253, 112234.

Simard, M., Pinto, N., Fisher, J.B., Baccini, A., 2011. Mapping forest canopy height globally with spaceborne lidar. J. Geophys. Res. Biogeosciences 116.

Sothe, C., Gonsamo, A., Lourenço, R.B., Kurz, W.A., Snider, J., 2022. Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel. Remote Sens. 14, 5158.

Spawn, S.A., Gibbs, H.K., 2020. Global aboveground and belowground biomass carbon density maps for the year 2010. Ornl Daac.

Stovall, A., Anderson-Teixeira, K., Shugart, H., 2018. Assessing terrestrial laser scanning for developing non-destructive biomass allometry. For. Ecol. Manage. 427, 217–229.

Sullivan, M.J.P., Lewis, S.L., Hubau, W., Qie, L., Baker, T.R., Banin, L.F., Chave, J., Cuni-Sanchez, A., Feldpausch, T.R., Lopez-Gonzalez, G., others, 2018. Field methods for sampling tree height for tropical forest biomass estimation. Methods Ecol. Evol. 9, 1179–1189.

Thumaty, K.C., Fararoda, R., Middinti, S., Gopalakrishnan, R., Jha, C.S., Dadhwal, V.K., 2016. Estimation of Above Ground Biomass for Central Indian Deciduous Forests Using ALOS PALSAR L-Band Data. J. Indian Soc. Remote Sens. 44, 31–39.

Van Aardt, J.A.N., Wynne, R.H., Scrivani, J.A., 2008. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. Photogramm. Eng. Remote Sens. 74, 1033–1044.

Véga, C., Vepakomma, U., Morel, J., Bader, J.-L., Rajashekar, G., Jha, C.S., Ferêt, J., Proisy, C., Pélissier, R., Dadhwal, V.K., 2015. Aboveground-Biomass Estimation of a Complex Tropical Forest in India Using Lidar. Remote Sens. 7, 10607–10625.

Wagner, W., Hollaus, M., Briese, C., Ducic, V., 2008. 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. Int. J. Remote Sens. 29, 1433–1452.

Wang, C., Zhu, X., Nie, S., Xi, X., Li, D., Zheng, W., Chen, S., 2019. Ground elevation accuracy verification of ICESat-2 data: A case study in Alaska, USA. Opt. Express 27, 38168–38179.

Weishampel, J.F., Ranson, K.J., Harding, D.J., 1996. Remote sensing of forest canopies. Selbyana 6–14.

West, G.B., Brown, J.H., Enquist, B.J., 1999. A general model for the structure and allometry of plant vascular systems. Nature 400, 664–667.

Wulder, M.A., White, J.C., Nelson, R.F., N\aesset, E., \\Orka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. Lidar sampling for large-area forest characterization: A review. Remote Sens. Environ. 121, 196–209.

Xing, Y., Huang, J., Gruen, A., Qin, L., 2020. Assessing the performance of ICESat-2/ATLAS multi-channel photon data for estimating ground topography in forested terrain. Remote Sens. 12, 2084.

Zhang, G., Ganguly, S., Nemani, R., White, M., Milesi, C., Wang, W., Saatchi, S., Yu, Y., Myneni, R.B., 2013. A simple parametric estimation of live forest aboveground biomass in California using satellite derived metrics of canopy height and Leaf Area Index. Geophys. Res. Lett.

Zhang, Y., Liang, S., 2020. Fusion of multiple gridded biomass datasets for generating a global forest aboveground biomass map. Remote Sens. 12, 2559.

Zolkos, S.G., Goetz, S.J., Dubayah, R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sens. Environ. 128, 289–298.

# Chapter 9

# List of Publications

1. **Rodda, S. R**., Nidamanuri, R. R., Mayamanikandan, T., Rajashekar, G., Jha, C. S., & Dadhwal, V. K. (2023). Non-Destructive Allometric Modeling for Tree Volume Estimation in Tropical Dry Deciduous Forests of India Using Terrestrial Laser Scanner. **Journal of the Indian Society of Remote Sensing, 1–15**.

2. **Rodda, S. R.,** Nidamanuri, R. R., Fararoda, R., Mayamanikandan, T., & Rajashekar, G. (2023). Evaluation of Height Metrics and Above-Ground Biomass Density from GEDI and ICESat-2 Over Indian Tropical Dry Forests using Airborne LiDAR Data. **Journal of the Indian Society of Remote Sensing, 1–16**.

3. **Rodda, S. R.,** Fararoda, R., Jha, N., Réjou-Méchain, M., Couteron, P., Gopalakrishnan, R., et al. (2024). South Asian and Central African maps from: LiDAR-based reference aboveground biomass maps for tropical forests of South Asia and Central Africa. **Nature Scientific Data**. **Accepted 19-Feb-24.**