THE UNIVERSITY
of EDINBURGH

Doctor of Philosophy in Atmospheric and Environmental Sciences

Mapping forest aboveground biomass and its change in China

by

Wenquan Dong

2024

Declaration

I declare that this thesis was composed by myself, that the work contained therein is my own, except where explicitly stated otherwise in the text, and that it has not been submitted, in whole or in part, for any other degree or professional qualification.

Wenquan Dong

This thesis was conducted under the supervision of Prof Casey Ryan and Prof Edward Mitchard.

Dedicated to my wife,

Man Chen,

and to my lovely family.

Abstract

Accurate quantification of forest aboveground biomass (AGB) and its changes is crucial for understanding the role of forests in the global carbon cycle and developing effective climate change mitigation strategies. This thesis presents a comprehensive assessment of forest AGB in China using a combination of field data, remote sensing observations, and machine learning algorithms. The research is divided into four chapters, each addressing specific aspects of forest AGB estimation and mapping.

Chapter 2 used established methods, but novel focused datasets, to create the first high resolution map of China's biomass for the mid-2000s. I generated a wall-to-wall AGB map of China for the year 2007 at a 50 m resolution using Ice, Cloud, and land Elevation Satellite (ICESAT) Geoscience Laser Altimeter System (GLAS) Lorey's height data, L-band SAR from Advanced Land Observing Satellite (ALOS) Phased Array L-band SAR (PALSAR), C-band SAR from Environmental Satellite (Envisat) Advanced Synthetic Aperture Radar (ASAR), and optical satellite data from Landsat-5 and field data. I averaged GLAS data within 0.01 degrees × 0.01 degrees grid cells, obtaining 8,981 cells with at least two GLAS footprints. Lorey's height was converted to AGB using allometric equations developed independently for northern and southern China using field data. Random forest (RF) regression was then employed to extrapolate AGB grid cells from GLAS data to a continuous map at 50 m resolution, using variables derived from Earth observation datasets and layers for training. Our estimates of total carbon stored in the forest in China was 9.52 Pg C, with an average forest AGB of 104 Mg ha^{-1}.

Chapter 3 looked in detail at two regions where I conducted fieldwork, to assess more accurately how well satellite LiDAR from the GEDI sensor can estimate tree height and biomass, and how well these isolated footprints can be spatially extrapolated using other datasets. It involved generating 25 m resolution AGB maps for two 500 km x 500 km

regions in northeastern and southwestern China for the year 2021, using GEDI data, field measurements, and Sentinel-1, ALOS-2 PALSAR-2, and Sentinel-2 data. We measured 26 plots (24 under GEDI footprints) in the northeastern region and 16 plots (12 under GEDI footprints) in the southwestern region. In the northeastern region, the closest relationship was observed between field AGB and RH98, while in the mountainous southwestern region, field AGB exhibited a stronger correlation with RH80. The fitted relationships were used to convert RH98 and RH80 to AGB for both regions, respectively. We found many GEDI footprints had errors not detected by their quality flags; filtering using remote sensing data to remove low-quality footprints improved results. In both the northeastern and southwestern regions, an inverse correlation between slope steepness and model accuracy was observed. Specifically, in the northeastern region, coefficient of determination (R^2) values exhibited a decline from 0.93 in areas of minimal slope to 0.42 in locales exhibiting slopes greater than 30 degrees. Concurrently, the root mean square error (RMSE) values escalated from 14 to 38 Mg ha^{-1}. Similarly, in the southwestern region, R^2 values decreased from 0.75 in relatively flat terrains to 0.50 in areas with slopes exceeding 30 degrees, alongside an increase in RMSE values from 20 to 44 Mg ha^{-1}.

In Chapter 4 I examine whether the latest deep learning methods can improve biomass estimation, beyond methods using relatively simple tree-based methods (Random Forest) as used in chapters 2 and 3. In particular I explore the use of the attention UNet (AU) deep learning model for estimating forest AGB in Guangdong Province. I converted GEDI relative height (RH) metrics to AGB using the allometric equations established in Chapter 3 and then employed attention UNet to extrapolate GEDI footprints to a wall-to-wall AGB map. The AU model demonstrated superior performance in biomass estimation accuracy compared to the traditional RF method, though computational requirements were considerable. I managed to produce a 2019 10-meter resolution AGB map for the whole of Guangdong using the AU model. This involved a novel approach to reduce boundary artifacts in patch-based predictions, which are prevalent in deep learning applications for image analysis. By overlapping patches and excluding edge pixels, the method improved spatial consistency and accuracy at the edges of predictions, leading to more dependable AGB estimates, without patch artefacts. Although AGB distributions from both AU and RF models were closely aligned, with similar mean values, the AU-derived AGB map offered more realistic spatial details upon visual assessment.

Chapter 5 employs a multi-step approach to estimate forest AGB in China for 2021 and analyze its changes since 2007. It employs the approaches explored in Chapter 3, at the same scale as Chapter 2, and without using the advanced ML methods of Chapter 4 as the computational burden was not considered. First, GEDI L2A RH metrics were converted to AGB estimates using region-specific allometric equations developed in Chapter 2. Next, random forest models were trained for each first-level administrative unit in China to predict AGB using GEDI-derived AGB estimates and predictor variables from Sentinel-1, PALSAR-2, and Sentinel-2. The trained models were then applied to generate 25 m resolution wall-to-wall AGB maps for each first-level unit and then merged into a national-scale AGB map. Our results show that the total carbon stored in the forest in China for 2021 was 13.06 Pg C, with a mean AGB density of 108.75 Mg ha^{-1}. Comparing the 2021 AGB map with the 2007 map, we observed an overall increase in carbon storage, with a net gain of 3.54 Pg C and varying spatial patterns of AGB changes across the country. Spatial variations in AGB alterations were observed across the nation, exhibiting the most significant increments in the northeastern and south-central regions of China.

The high-resolution AGB maps and quantified biomass changes generated in this thesis provide valuable insights into the spatial distribution and dynamics of forest biomass resources in China, supporting decision-making in forest management and climate change mitigation efforts.

Lay Summary

Forests play a vital role in regulating Earth's climate by absorbing carbon dioxide from the atmosphere, and trapping it in trees, their roots, and the soil. Accurately measuring the amount of carbon stored in forest biomass and understanding how it changes over time is essential for developing strategies to mitigate climate change, and monitoring projects that protect or restore forests. One area where big changes in forest cover and the density of carbon stored within that forest is known to be China, where rapid industrialisation, the development of new forests and plantations, recovery from past disturbance, and climate change, are all causing rapid changes. This thesis focuses on estimating forest AGB in China using a combination of field measurements, satellite data, and advanced computer algorithms.

The research is divided into an introductory chapter and four results chapters, each addressing different aspects of forest AGB estimation and mapping. Chapter 2 uses satellite data and field measurements to create a detailed map of forest AGB in China for the year 2007. Chapter 3 generates AGB maps for two smaller regions in China for the year 2021, using more advanced satellite data and computer models, and a set of field data I collected by measuring hundreds of trees in inventory plots in each region. Chapter 4 explores the use of deep learning, a type of artificial intelligence, to estimate forest AGB in Guangdong Province in 2021, achieving higher accuracy compared to traditional methods.

Building upon the findings from Chapters 3 and 4, Chapter 5 presents the creation of a nationwide AGB map for China in 2021. The results show a total forest AGB of 27.79 billion tons in China for 2021, with an average AGB density of 108.75 tons per hectare. By comparing the 2021 AGB map with the 2007 map from Chapter 2, we observed an overall increase in carbon storage, with variations in AGB changes across the country.

This thesis contributes to our understanding of forest biomass distribution and changes in

China, providing valuable information for forest management and climate change mitigation efforts. The detailed AGB maps and quantified biomass changes serve as essential tools for decision-makers and researchers working to protect forests and combat climate change.

In summary, this research employs an innovative multi-step approach, combining multiple data sources such as GEDI, multi-source satellite imagery, and ground survey data, along with machine learning algorithms to generate high-resolution forest aboveground biomass maps for China in 2007 and 2021, revealing an overall increasing trend in forest carbon storage across the country during this period. These findings can provide important evidence for monitoring forest resources nationwide and valuable information for optimizing forest management and climate change mitigation policies.

Acknowledgements

Time flies, and the years have gone by in the blink of an eye. The four years I spent in Edinburgh have passed swiftly, with a significant portion of that time marked by the COVID-19 pandemic. Throughout this incredible journey, I have been fortunate to have the support, guidance, and love of so many remarkable individuals. It is with a deep sense of gratitude that I acknowledge the contributions of those who have played a vital role in shaping my academic career and personal growth.

First and foremost, I extend my sincerest gratitude to Professor Edward Mitchard, who served as my principal supervisor during the initial three years of my research. His mentorship was instrumental in shaping my approach to scientific research, guiding me through the intricacies of field data collection, research methodology, and academic writing. His patience, expertise, and encouragement have been invaluable to my development as a researcher.

I would also like to extend my sincere thanks to Professor Casey Ryan, who took over as my principal supervisor in July 2023 after Edward departure from the university. Casey's enthusiasm and keen involvement in my research have been valuable. His insightful guidance and warm welcome into his research group have been crucial in the latter stages of my PhD, and I am deeply appreciative of his support.

To my wife, Man Chen, I owe a debt of gratitude that words alone cannot convey. Her support, understanding, and care during the demanding periods of my studies have been a source of strength and comfort. Man, I am eternally thankful for your companionship on this journey.

I would also like to acknowledge the invaluable contributions of several individuals who have supported my research in various capacities. Special thanks to Steven Hancock (University of Edinburgh), Maurizio Santoro (Gamma Remote Sensing), and Yuwei Chen (Finnish Geospatial Research Institute) for their assistance, insights, and encouragement.

I would like to express my sincere gratitude to my colleagues and friends who have been instrumental in my academic growth and have provided me with valuable insights and inspiration throughout my PhD journey. Yu Hao, Charlotte Wheeler, Selena Georgiou, Harry Carstairs, Iain McNicol, Chiara Aquino, Shona Jenkins, and Adam Hastie have been constant sources of knowledge, support, and motivation. Our discussions, collaborations, and shared experiences have greatly enriched my understanding of the field and have helped me to push the boundaries of my research.

Lastly, but certainly not least, I must express my heartfelt appreciation to my family. My family has always been my pillars of strength, ensuring that I could focus on my studies without worrying about matters beyond academia. I am blessed to have grown up in such a loving and supportive family. To my parents, Changyin Dong and Xiangrong Kong; my siblings, Bin Fang and Baocheng Xu; and all other family members - your faith in me has been a source of motivation and joy.

To all those mentioned above and the many others not listed who have contributed to my success in ways both big and small, I offer my sincere appreciation. This thesis would not have been possible without your collective support, guidance, and belief in me. Thank you for being a part of this incredible journey.

Contents

2 2007 AGB map for China 21

3 Mapping AGB in two contrasting regions 47

List of Figures

List of Tables

Abbreviations

AGB Aboveground Biomass.

ALOS Advanced Land Observing Satellite.

ALS Airborne laser scanning.

ANN Artificial Neural Network.

APFNet Asia-Pacific Network for Sustainable Forest Management and Rehabilitation.

ASAR Advanced Synthetic Aperture Radar.

ASTER Advanced Spaceborne Thermal Emission and Reflection Radiometer.

AU Attention UNet.

AU-FC UNet Fully Connected.

BN Batch Normalization.

CNNs Convolutional Neural Networks.

DBH Diameter at Breast Height.

DEM Digital Elevation Model.

DN Digital Number.

Envisat Environmental Satellite.

EO Earth Observation.

ESA European Space Agency.

FAO Food and Agriculture Organization.

FC Fully Connected.

GEDI Global Ecosystem Dynamics Investigation.

GEE Google Earth Engine.

GLAS Geoscience Laser Altimeter System.

GRD Ground Range Detected.

ICESat Ice, Cloud, and land Elevation Satellite.

ISS Residual Network.

IW Interferometric Wide swath.

JAXA Japan Aerospace Exploration Agency.

kNDVI Kernel Normalized Difference Vegetation Index.

L2A Level 2A.

L2B Level 2B.

LiDAR Laser-Ranging.

LightGBM Light Gradient Boosting Machine.

MaxEnt Maximum Entropy Modeling.

MSI MultiSpectral Instrument.

NASADEM NASA Digital Elevation Model.

NDMI Normalized Difference Moisture Index.

NDVI Normalized Difference Vegetation Index.

NSTI National Science & Technology Infrastructure.

PALSAR Phased Array L-band SAR.

QA Quality Assessment.

R^2 The coefficient of determination.

ReLU Rectified Linear Unit.

ResNet Residual Network.

RF Random Forest.

RH Relative Height.

RMSE Root Mean Square Error.

RNNs Recurrent Neural Networks.

SAR Synthetic Aperture Radar.

SARs Special Administrative Regions.

SR Surface Reflectance.

SRTM Shuttle Radar Topography Mission.

SVM Support Vector Machines.

SWIR Short Wave Infrared.

TM Thematic Mapper.

VH Vertical Transmit, Horizontal Receive.

VNIR Visible and Near Infrared.

VV Vertical Transmit, Vertical Receive.

Chapter 1

Introduction

1.1 Overview

This thesis presents the results of research into the use of multi-source remote sensing data, field measurements, and machine learning algorithms for estimating and mapping forest AGB and AGB change in China. The research spans a range of spatial scales, from regional to national, and employs various remote sensing datasets, including spaceborne LiDAR, SAR, and optical sensors. The thesis is organized into four main chapters, each focusing on a specific aspect of forest AGB estimation and mapping, with the overarching goal of developing a framework for accurate and efficient monitoring of forest AGB change in China.

The research is divided into four chapters, each addressing specific aspects of forest AGB estimation and mapping. Chapter 2 focuses on generating a high-resolution AGB map of China for the year 2007 using ICESat GLAS, ALOS PALSAR, Envisat ASAR, and Landsat-5 data, along with field measurements. Chapter 3 investigates the potential of using the Global Ecosystem Dynamics Investigation (GEDI) LiDAR data for estimating forest AGB in two contrasting regions of China in 2021, and explores the spatial extrapolation of GEDI-derived AGB estimates using Sentinel-1, ALOS-2 PALSAR-2, and Sentinel-2 data. Chapter 4 evaluates the performance of a deep learning approach, the attention UNet (AU) model, for estimating forest AGB in Guangdong Province in 2019, and compares it with the traditional random forest (RF) method. Finally, Chapter 5 presents a national-scale assessment of forest AGB in China for the year 2021 and analyzes its changes since 2007, using a multi-step approach that combines GEDI data, multi-source satellite imagery, and machine learning algorithms.

This chapter provides a general overview of the importance of forest AGB estimation, the current state-of-the-art methods, and the challenges associated with large-scale AGB mapping. It discusses the role of forests in the global carbon cycle, the significance of China's forests, and the need for accurate estimation of forest AGB change in China. The chapter also reviews the progress in forest AGB estimation methods, including traditional field-based inventory methods, remote sensing techniques, and machine learning algorithms.

1.2 Why map forest AGB and AGB change in China

1.2.1 Forest

Forests cover about 30% of the earth's land surface globally, and play an important role in terrestrial ecosystems, including holding the majority of the world's species, and contributing to an outsized degree to the global water and carbon cycles (Beer et al., 2010). Forests can help to reduce the effects of climate change, which is likely the greatest environmental challenge of the 21st century, largely by capturing and storing carbon. When forests grow, trees absorb carbon from the atmosphere and store carbon in wood, leaves and soil. Carbon stored in forests is estimated to be 861±66 Pg C, with 44% in soil to 1 meter depth, and 42% in above and below ground biomass (Pan et al., 2011). When destroyed, overused or degraded, carbon is released from forests into the atmosphere mostly in the form of carbon dioxide. Carbon dioxide is a greenhouse gas, warming the Earth by absorbing energy and slowing the process that energy escapes to space. However, intact and recovering forests are also a carbon sink, absorbing a similar or greater amount of carbon dioxide each year than is released by deforestation (Mitchard, 2018), and therefore reducing the climate change that should result from our emissions from burning fossil fuels and damaging forests. As one of the most important characters of forests, and given their importance to the global carbon cycle, it is of great significance to study the distribution and change of forest aboveground biomass (AGB).

1.2.2 China's forests and their importance

More than half of the world's forests are found in only five countries (the Russian Federation, Brazil, Canada, the United States of America and China), and the forest area of China is the fifth largest in the world (FAO, 2020). In 1949, the forested area in China encompassed 102 million hectares (ha) (Fang et al., 2001), which is believed to have expanded to 220 million ha by 2020 (FAO, 2020) (Fig. 1.1), representing 23% of the country's total land area.

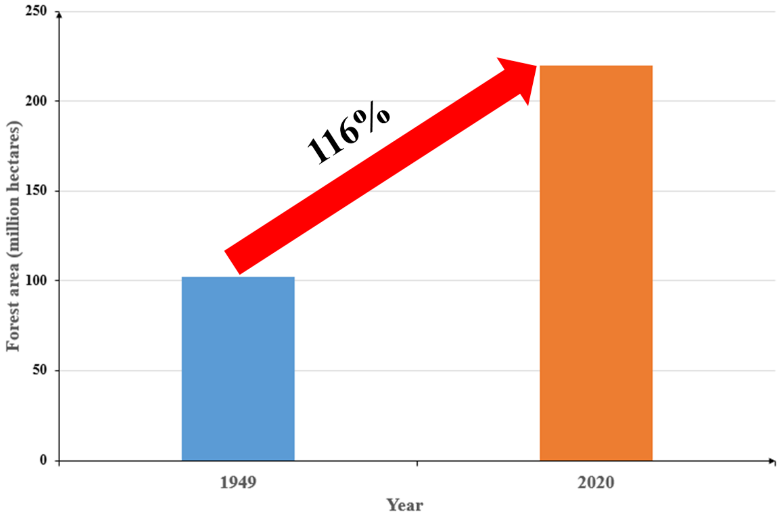


Figure 1.1: Increase in China's forested area from 1949 to 2020, showing a 116% Growth.

China's forests are diverse, ranging from boreal forests in the northeast to tropical rainforests in the south (Xu, Xiao, Zhang, Yang, Zhang, Hull, Wang, Zheng, Liu, Polasky et al., 2017). Forests in China locate over a wide span of latitude, resulting in different forest types, with extensive plantation forests supplmenting natural forests. Forests in China are in 8 vegetation zones, including cold temperate deciduous coniferous forest, temperate conifer-broadleaf forest, warm-temperate deciduous broadleaf forest, Subtropical evergreen broadleaf forest, tropical monsoon forest and rain forest, temperate steppe, temperate desert and Tibetan Plateau alpine vegetation (Fig. 1.2).

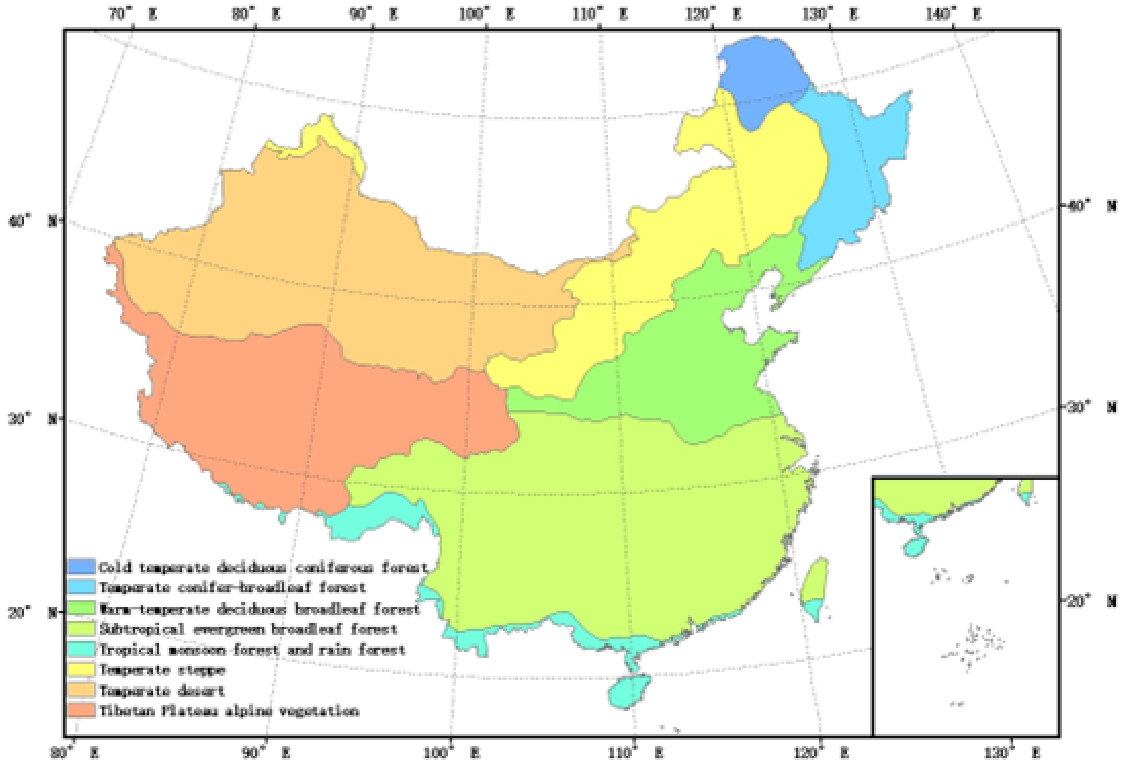


Figure 1.2: Vegetation zones in China.

In recent decades, China has implemented extensive afforestation and reforestation programs, such as the Grain for Green Project, Three-North Shelter Forest Program and the Natural Forest Protection Program, to restore degraded landscapes and enhance forest cover (Delang and Yuan, 2015). Not only the government but also some private companies plant trees in China. For example, a payment and lifestyle company Alipay has planted almost 122 million trees ("Green energy points" are allotted to users each time they opt for an eco-friendly option, and once sufficient points are accumulated, a real tree is planted by Alipay in northwest China). In 2020, the area of planted forests in China was approximately 77 million ha, representing 31.3% of the world's total planted forests (Cheng et al., 2023).

1.2.3 The importance of accurate estimation of forest AGB change in China

Accurate estimation of changes in forest aboveground biomass (AGB) is essential for understanding the dynamics of China's forest ecosystems and their role in the global carbon cycle. Monitoring forest AGB changes over time provides valuable insights into the effectiveness of forest management practices, afforestation and reforestation efforts, and the impact of natural disturbances such as fires, pests, and diseases (Fang et al., 2014).

Moreover, accurate estimation of forest AGB changes is necessary for China to fulfill its international commitments to mitigate climate change. Under the Paris Agreement, China has pledged to peak its carbon emissions by 2030 and achieve carbon neutrality by 2060 (Chen, Chen, Ciais, Zhang, Lü, Wang, Chevallier, Liu, Yuan and Peters, 2022). Forests could play a role in achieving these targets, as they act as carbon sinks that can offset anthropogenic greenhouse gas emissions (Zhang, Song, Band and Sun, 2019), though it is noted that the total carbon locked up in carbon in China's forest (10-15 Pg C) is only 3-4 times China's annual emissions of 3.3 Pg C (REF). Quantifying the changes in forest AGB over time allows policymakers to evaluate the contribution of forests to China's climate change mitigation efforts and develop effective strategies for enhancing forest carbon sequestration (Wang, Zhang, Ju, Chen, Ciais, Cescatti, Sardans, Janssens, Wu, Berry et al., 2020).

1.3 Methods for forest AGB estimation

Forest AGB estimation methods can be broadly categorized into field measurement and remote sensing approaches. This section provides an overview of the progress in forest AGB estimation methods, highlighting their strengths and limitations.

1.3.1 Traditional field-based inventory methods and their limitations

Traditionally, forest AGB has been estimated through field-based inventory methods, which involve destructive sampling or allometric equations based on measured tree parameters such as diameter at breast height (DBH), tree height and wood density (Chave et al., 2005).

Once established, allometric models for tree species enable non-destructive biomass assessments for forest stands. These models are formulated from various integrations of the previously mentioned three parameters, using linear or more normally nonlinear regression analyses (Chave et al., 2014; Overman et al., 1994; Seidel et al., 2011; Feldpausch et al., 2011). While field-based methods normally provide low uncertainty and relatively unbiased AGB estimates at the plot level, they are time-consuming, labor-intensive, and costly, making them impractical for large-scale applications (Zolkos et al., 2013). Moreover, field-based methods are limited in their ability to capture the spatial variability of forest AGB across landscapes and regions (Saatchi et al., 2011).

1.3.2 Remote sensing techniques for forest AGB estimation

Remote sensing has emerged as a powerful tool for estimating forest AGB at various spatial scales. Satellite and airborne sensors can capture information about forest structure and composition, which can be related to AGB using empirical models or physically-based approaches (Lu et al., 2016). The main remote sensing techniques used for forest AGB estimation include optical remote sensing, synthetic aperture radar (SAR), and light detection and ranging (LiDAR).

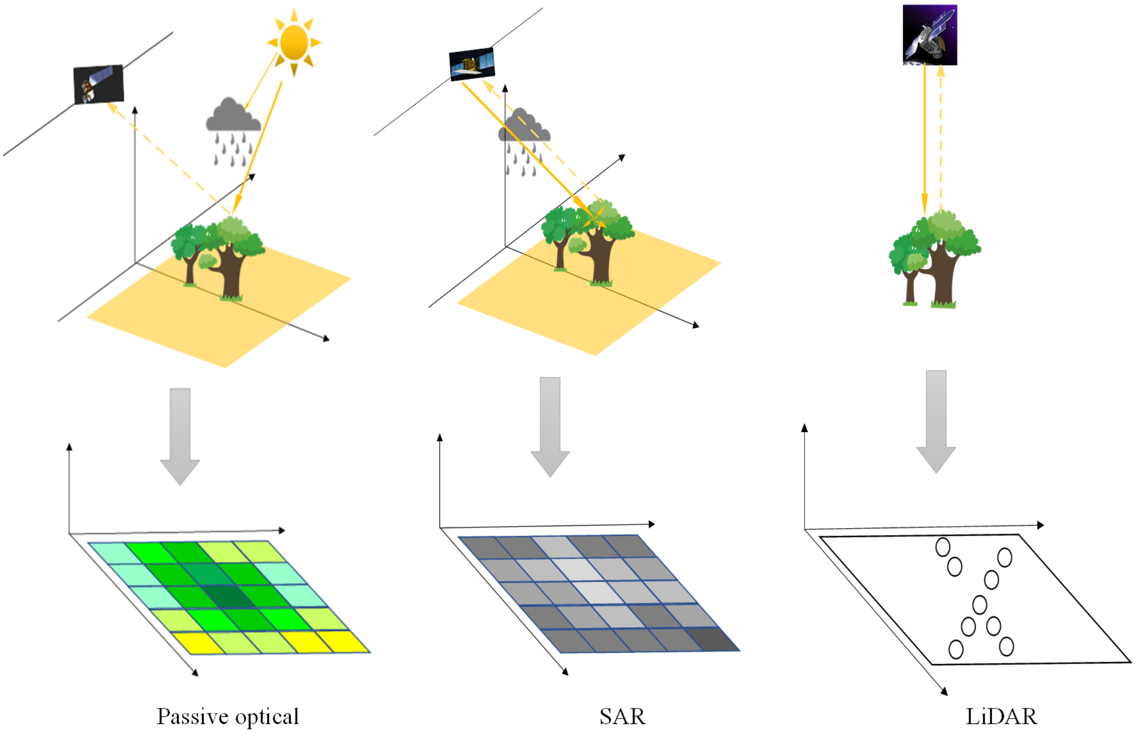


Figure 1.3: Three different types of remote sensing senors.

Passive optical remote sensing

Optical data is the most commonly used data, and there are a number of open access platforms providing data from the 1970s to present, from satellites such as Landsat-1 to Landsat-9 (8 and 9 are in orbit currently), Sentinel-2 and the Moderate Resolution Imaging Spectroradiometer (MODIS) data. Optical remote sensing sensors are passive, making use of natural radiation reflected by the Earth's surface. Due to easy accessibility, ease of interprestation, and frequent revisits, these data have been widely used to monitor forest. The estimation of forest AGB using optical data is usually by establishing relationships between AGB and spectral responses and/or vegetation index derived from multispectral data. However, optical data is not always the ideal data for forest monitoring because of three major factors: optical sensor can only see canopy surface, which means cannot get information on tree structures; it cannot see through clouds Soenen et al. (2010), which leads the observations highly depend on the proportion of time the study area is not under cloud; and the relationship between optical data and biomass often saturates at a very low point, equivalent to canopy closure. In fact, optical data is normally better suited to saying where forest is and is not, and tracking forest loss (Hansen et al. 2013), rather than mapping AGB.

SAR

SAR sensors emit microwave energy and measure the backscattered signal from forest canopies, which is sensitive to forest structure and biomass (Woodhouse, 2017). SAR has the advantage of penetrating through clouds and operating independently of solar illumination, making it suitable for monitoring forests in regions with frequent cloud cover (Ling et al., 2021). Theoretically, the backscatter values increase as the number and/or size of trees in an area increases (Mitchard et al., 2009). Therefore, radar data has been used to provide an accurate estimation of forest AGB. However, radar data also has limitations such as saturation and complex interaction with forests. Radar data which has longer wavelength tends to have a higher saturation with simpler forest structure (Sinha et al., 2015), and typically L-band, the longest wavelength currently available from space, saturates at about 100-150 Mg ha^{-1} (Shugart et al., 2010). In addition, microwave energy transmitted by SAR sensors mainly interacts with forest components comparable in size to the wavelengths (Joshi et al., 2017). Hence, radar backscatter value may differ for the forests with same biomass value, as there are multiple structures of forests that can have the same biomass.

Radar data can be acquired in X-, C-, S-, L-, and P-band as the wavelength of the radar increases. These bands all have their own characteristics, penetrating into and interacting with various depth of canopy. The different backscatter responses mainly contain information on leaves, branches and trunks according to wavelength. P-band sensors can penetrate through canopy and interact with large branches and trunk, and thus more sensitive to forest

AGB. While there has never been a P-band sensor based on satellite, the European Space Agency (ESA) is planning to launch BIOMASS, a satellite with a fully polarimetric P-band sensor, in 2025. This mission is expected to revolutionize biomass mapping by providing unprecedented sensitivity to forest structure and biomass, especially in dense tropical forests where current sensors often saturate (Quegan et al., 2019).

Currently, L-band data is the most sensitive to biomass among the available satellite-based sensors, with data provided by the ALOS-2 satellite. The upcoming NASA-ISRO SAR (NISAR) mission, a collaboration between NASA and the Indian Space Research Organisation (ISRO), is set to launch in 2024. NISAR will carry both L-band and S-band SAR instruments, offering global coverage with a 12-day repeat cycle (Rosen et al., 2017). The L-band component of NISAR is expected to significantly enhance our ability to map forest biomass globally, providing high-resolution, frequent observations that will enable more accurate and timely biomass change detection.

These upcoming missions, BIOMASS and NISAR, represent a new era in biomass mapping. They will provide complementary data at different wavelengths (P-band and L-band, respectively), allowing for a more comprehensive understanding of forest structure and biomass across various forest types and environmental conditions. The combination of these new data sources with existing ones (such as those from Sentinel-1 and ALOS-2) is expected to greatly improve our ability to estimate and monitor forest biomass at global scales, potentially reducing uncertainties in carbon stock estimates and enhancing our understanding of the global carbon cycle (Quegan et al., 2019).

Polarisation is also a parameter that can affect the information getting from forests and the polarisations are usually controlled between horizontal and vertical, and can be detected in either polarisation too. SAR signal is normally sent in an individual polarisation, and that 4 possible combinations of send and receives exist (HH: Horizontal transmit, Horizontal receive; VV: Vertical transmit, Vertical receive; HV: Horizontal transmit, Vertical receive; VH: Vertical transmit, Horizontal receive). Muilt-polarisation can not only detect scattering elements that change the polarisation of transmit signal such as trees, but also provide information of elements that surface scattering takes place such as soil moisture (Kasischke et al., 2011). The interaction with forest can be divide into three scattering categories: surface scattering, double-bounce scattering and volume scattering 1.4. While normally only

1-2 combinations are used, due to power constraints and the close correlation between the single and cross-polarised returns, all four can be at once (as will happen with BIOMASS), giving rich information on scattering mechanisms. Even with dual-polarisation, the absolute backscatter and ratio between the two can provide more information, helping with biomass mapping. For example, a ratio between the power of the HH and HV polarisations can reveal forest structural components as the double-bounce component is largely absent from HV, and volume scattering is in both; it can also reduce the impact of terrain effects compared to the absolute backscatter values (Ranson et al., 2001; Sarker et al., 2012). But if going to use only one, HV polarisation has been proven to be mostly sensitive to forest AGB (Luckman et al., 1997; Mitchard et al., 2011), likely as less impacted by soil scattering.

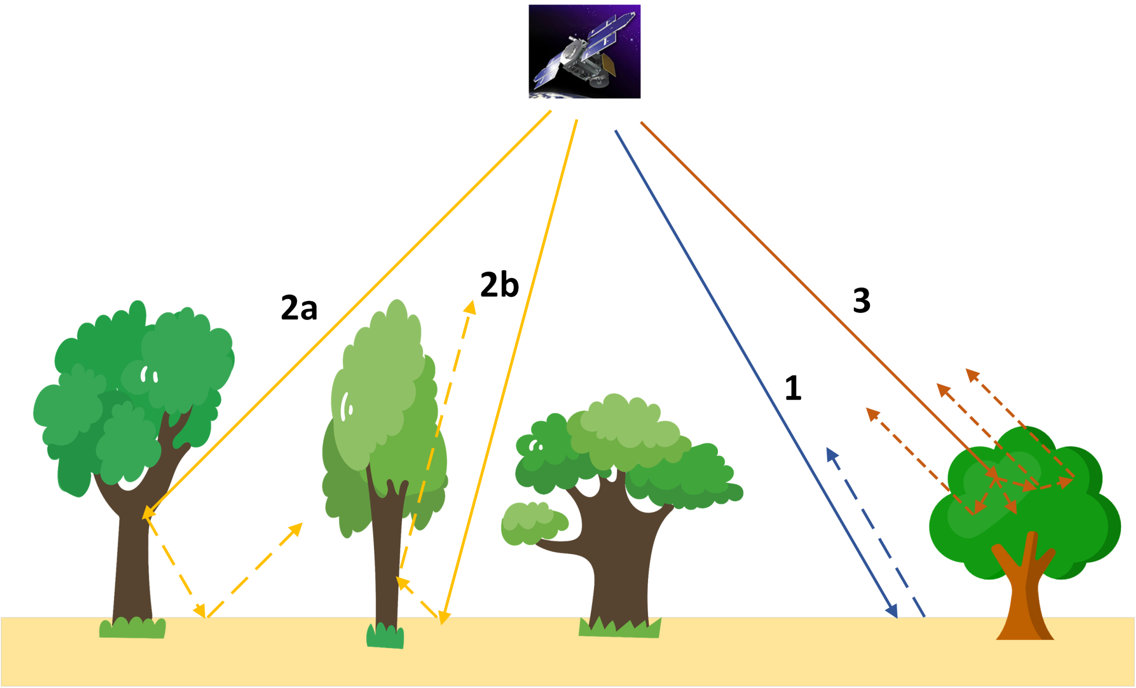


Figure 1.4: SAR backscattering in forests. (1) Surface scattering, (2) double-bounce scattering, (3) volume scattering.

LiDAR

LiDAR is an active remote sensing technology that provides detailed three-dimensional information about forest structure by measuring the time of return of emitted laser pulses, typically at wavelengths between 900 and 1064 nm towards the Earth (Lefsky et al., 2002). While aircraft and Unmanned Aerial System (AES) collected data is commonly collected and highly useful, only satellite data has relatively uniform coverage, and is thus of most use for creating and updating large scale maps. Both platforms collect information in the same way however. The time interval is used to calculate the distance between the sensor and object, and thus canopy height can be derived from waveform. Relative Height (RH) metrics, derived from LiDAR waveforms, are also closely related to forest structure and canopy height. RH metrics represent the height at which a certain percentage of the waveform energy is returned from the canopy (Drake et al., 2002). For example, RH98 represents the height below which 98% of the LiDAR pulse returns (or the total energy of the returns) are found. RH98 effectively represents the canopy top or the highest vegetation within the LiDAR survey area, while being less prone to outliers and noise compared to RH100 (Li et al., 2023). This makes it a more stable indicator for assessing the uppermost vegetation structure.

LiDAR sensors are classified into full-waveform sensors and discrete-return sensors. Full-waveform sensors record a continuous height distribution illuminated by the entire reflected energy from a return, while discrete-return sensors record single or multiple returns (Wulder et al., 2012). This allows full-waveform sensors to capture the detailed vertical structure of forest canopies, as the laser pulse interacts with different canopy components such as leaves, branches, and the ground (Wagner et al., 2006). The resulting waveform data contains rich information about the forest structure, which can be used to derive canopy height, canopy density, and ground elevation (Mallet and Bretar, 2009). However, processing full-waveform data requires sophisticated algorithms to decompose the waveforms and extract the relevant features, which can be computationally intensive, especially in high density aircraft LiDAR data. In contrast, discrete-return LiDAR sensors record one or multiple returns as discrete points, each representing a significant reflection of the laser pulse from an object's surface. These sensors typically record the first and last returns, while some sensors can also capture intermediate returns. The first return usually corresponds to the top of the canopy, while the last return represents the ground or understory vegetation. Discrete-return LiDAR data is commonly used to generate Digital Terrain Models (DTMs) and Digital Surface Models (DSMs), from which Canopy Height Models (CHMs) and other forest structural parameters can be derived (Liu, 2008). Compared to full-waveform data, discrete-return data is less complex and requires less processing time, but will not capture information on the detailed vertical structure of the forest.

LiDAR systems can be classified based on the platform from which they operate, namely airborne and spaceborne LiDAR. Airborne LiDAR systems are mounted on aircraft and can acquire high-density point clouds over targeted areas. Airborne LiDAR campaigns can be flexibly designed to meet specific research or management objectives, with the ability to adjust flight parameters such as altitude, speed, and scan angle to optimize data collection (White et al., 2013). While airborne LiDAR provides high-resolution AGB estimates, its application is limited by high costs and restricted spatial coverage. Spaceborne LiDAR sensors are mounted on satellites orbiting the Earth, providing global coverage and repeated measurements at regular intervals. One of the pioneering spaceborne LiDAR missions was the Ice, Cloud, and land Elevation Satellite (ICESat), which operated from 2003 to 2009 (Schutz et al., 2005). ICESat Geoscience Laser Altimeter System (GLAS) data has been widely used for estimating forest canopy height and biomass, demonstrating the potential of spaceborne LiDAR for large-scale forest monitoring (Lefsky et al., 2005; Simard et al., 2011).

A notable example of a spaceborne LiDAR mission is GEDI, which was launched in 2018 and operates from the International Space Station (ISS) (Dubayah et al., 2020). GEDI is a full-waveform LiDAR system that aims to provide high-resolution observations of forest vertical structure and canopy characteristics on a global scale. While GEDI represents a significant advancement in spaceborne LiDAR technology, it's important to note its limitations. The sensor has a footprint size of approximately 25 m, which results in a relatively coarse spatial resolution compared to airborne LiDAR systems (Hancock et al., 2019). Despite this limitation, GEDI can penetrate through dense forest canopies to capture the vertical distribution of vegetation and the underlying topography, offering valuable insights into forest structure (Dubayah et al., 2020). However, it's crucial to acknowledge that GEDI data, like all remote sensing products, can contain errors and uncertainties. These may arise from factors such as the sampling design, atmospheric conditions, and terrain effects (Duncanson et al., 2022; Potapov et al., 2021). Nonetheless, spaceborne LiDAR data from missions like GEDI, despite their limitations, offer unprecedented opportunities for large-scale forest biomass estimation and carbon monitoring, providing valuable data in areas where field measurements or airborne LiDAR are not feasible (Duncanson et al., 2022).

Multi-source remote sensing data fusion and its advantages

The fusion of multi-source remote sensing data has emerged as a promising approach to improve forest AGB estimation accuracy. By combining data from optical, SAR, and LiDAR sensors, the limitations of individual sensors can be overcome, and the complementary information can enhance AGB estimation performance (Saatchi et al., 2011; Baccini et al., 2012). Each data source offers unique advantages and captures different aspects of forest structure and composition.

Optical missions, such as Landsat and Sentinel-2, provide information on the spectral properties of forest canopies. Vegetation indices, which are formulated by contrasting reflectance measurements in specific bands, are correlated with vegetation health and productivity. However, optical sensors are limited by their inability to penetrate cloud cover and their saturation in dense forest canopies (Lu et al., 2016). SAR offers critical advantages in biomass estimation, primarily due to its capability to penetrate cloud cover and, to some extent, vegetation canopies. However, SAR signals can be affected by factors such as terrain roughness, soil moisture, and vegetation water content, which can introduce large uncertainties and biases in biomass estimation (Mitchard et al., 2009). Many studies have explored the synergistic use of optical and SAR remote sensing data to estimate forest AGB. While these sensor combinations have shown promise, both optical and SAR data provide indirect measurements of AGB, relying on the relationships between spectral or backscatter characteristics and forest structural properties (Lu et al., 2016; Mitchard et al., 2009). In contrast, LiDAR sensors offer a more direct approach to estimating AGB by measuring forest height and vertical structure, which are strongly correlated with biomass (Lefsky et al., 2002).

The integration of spaceborne LiDAR with optical and SAR data has several advantages for wall-to-wall AGB mapping. First, the LiDAR data provide direct measurements of forest height and structure, which can improve the accuracy and robustness of AGB estimation models (Duncanson et al., 2022). Second, the optical and SAR data offer spatially continuous coverage, enabling the extrapolation of LiDAR-based models to larger areas (Xu, Saatchi, Shapiro, Meyer, Ferraz, Yang, Bastin, Banks, Boeckx, Verbeeck et al., 2017). Third, the fusion of multiple sensor types can help to reduce the uncertainties associated with individual datasets, such as the saturation of optical data in dense forests or the sensitivity of SAR data to moisture and terrain conditions (Montesano et al., 2009). Numerous studies have demonstrated the effectiveness of fusing spaceborne LiDAR, optical, and SAR data for wall-to-wall AGB mapping (Huang et al., 2019; Santoro et al., 2020; Schwartz et al., 2024; Fayad et al., 2024). These studies highlight the potential of multi-sensor data fusion for improving the accuracy and spatial resolution of AGB estimates at regional to global scales.

1.3.3 How to convert spaceborne LiDAR data to AGB

When generating wall-to-wall AGB maps using spaceborne LiDAR data and remote sensing imagery, it is typically necessary to first calculate the AGB of the LiDAR footprints. This requires the extraction of relevant metrics from the LiDAR waveforms or point clouds. These metrics typically include canopy height, canopy cover, and vertical structure indices, such as RH metrics (Lefsky et al., 2005; Drake et al., 2002). These metrics are then used to develop models that relate the LiDAR-derived variables to field-measured AGB, often using regression techniques such as linear, nonlinear, or machine learning algorithms (Boudreau et al., 2008; Margolis et al., 2015).

The development of robust AGB estimation models requires a representative sample of field plots that cover a range of forest types, structures, and biomass densities. These plots are typically established using a stratified random sampling approach, with strata defined by factors such as forest type, elevation, or disturbance history (Duncanson et al., 2015). The field plots are then co-located with the spaceborne LiDAR footprints, and the LiDAR metrics are extracted for each plot. The field-measured AGB values are then regressed against the LiDAR metrics to develop the estimation models (Saatchi et al., 2011).

1.3.4 Field-measured AGB

Forest inventory plots, areas where every tree above a certain threshold are 'inventoried', are essential for calibrating and validating remote sensing-based AGB estimation models (Clark and Kellner, 2012). Field-based AGB estimates are typically derived from measurements of individual tree attributes, such as Diameter at Breast Height (DBH), tree height, and wood density, which are then converted to biomass using allometric equations (Chave et al., 2005). Allometric equations are equations relating one measurement of an organism to another, for example the diameter of a tree to its biomass. They are useful because they enable a prediction of useful but hard-, expensive- or destructive-to-measure things such as biomass, from easy to measure things like DBH and height.

DBH is one of the most commonly measured tree attributes in forest inventory plots, as it is strongly correlated with tree biomass and can be easily measured using a diameter tape (Fig. 1.5a) or caliper (Chave et al., 2005). DBH is typically measured at a standard height of 1.3 m above the ground (Fig. 1.5b), although this height may vary depending on the forest type and the presence of buttresses or other stem irregularities (Cushman et al., 2014).

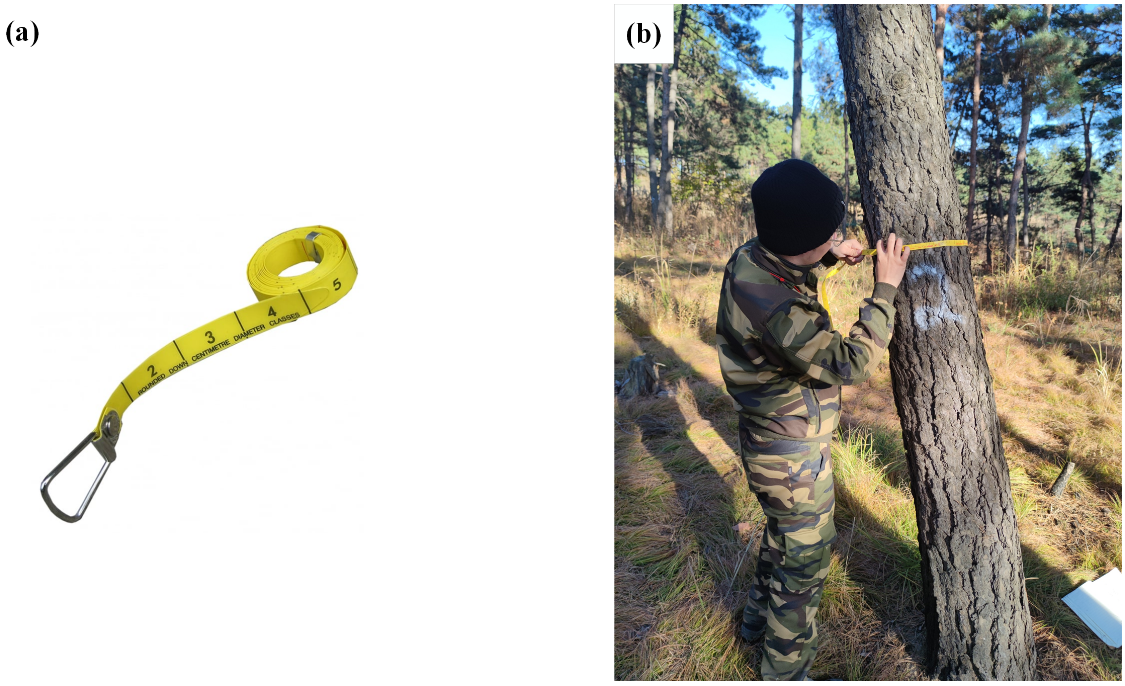


Figure 1.5: Measurement of DBH at a forest plot using a diameter tape. (a)Diameter tape, (b) Measurement of DBH at a field plot.

In addition to DBH, tree height is another important variable for estimating AGB, as it provides information on the vertical structure and volume of the tree (Hunter et al., 2013). Tree height can be measured using a variety of tools, such as clinometers, laser rangefinders, or hypsometers, depending on the forest conditions and the desired level of accuracy (Larjavaara and Muller-Landau, 2013).

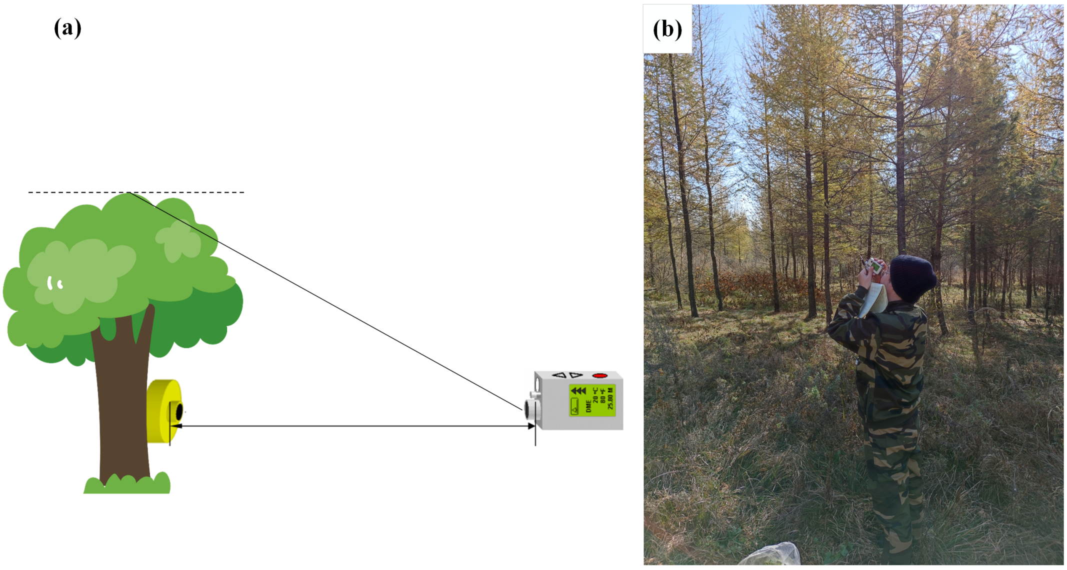


Figure 1.6: Measurement of tree height using laser rangefinders and transponder. (a)Laser rangefinder and transponder, (b)Measurement of tree height at a field plot.

Wood density is another critical parameter for converting tree volume to biomass, as it accounts for the variation in wood mass per unit volume across different tree species and environmental conditions (Chave et al., 2006). Wood density is typically measured by extracting wood cores from a sample of trees and determining their oven-dry mass and fresh volume (Williamson and Wiemann, 2010). Alternatively, wood density can be estimated using global or regional databases that compile wood density measurements for different tree species (Chave et al., 2009).

Once the individual tree measurements have been collected, they are used to estimate tree-level AGB using allometric equations. Allometric equations relate the measured tree attributes to biomass using regression models that are developed from destructive sampling of trees across a range of sizes and species (Chave et al., 2014). The choice of allometric equation is critical for accurate AGB estimation, as different equations can yield widely varying biomass estimates depending on the forest type, region, and size class of the trees (Duncanson et al., 2015). In general, species-specific and locally developed allometric equations are preferred over generalized equations, as they can better capture the variation in tree form and wood density across different forest types (Ngomanda et al., 2014).

After estimating tree-level AGB using allometric equations, the individual tree biomass estimates are summed to obtain plot-level AGB estimates, which are typically expressed in units of megagrams per hectare (Mg ha^{-1}) (Réjou-Méchain et al., 2017). Plot-level AGB estimates are then used to convert LiDAR metrics to AGB and validate AGB maps.

1.3.5 Machine learning algorithms for forest AGB estimation

Machine learning algorithms have emerged as powerful tools for estimating forest AGB by leveraging the complex relationships between remotely sensed data and field-measured biomass. These algorithms can handle high-dimensional datasets, capture non-linear relationships, and integrate data from multiple sources, making them well-suited for forest AGB estimation across diverse landscapes (Lary et al., 2016). In recent years, various machine learning techniques, including parametric and non-parametric methods, have been applied to forest AGB estimation, often outperforming traditional regression approaches (Lu et al., 2016).

Parametric machine learning methods, such as linear regression, logistic regression, and stepwise regression, assume a predefined functional form between the input variables (e.g., remote sensing data) and the target variable (e.g., AGB) (Fassnacht et al., 2014). These methods are relatively simple to implement and interpret, but may not fully capture the complex relationships between remotely sensed data and forest biomass (Lu et al., 2016). However, when the relationship between the input variables and AGB is well-approximated by a linear function, parametric methods can provide reliable and computationally efficient estimates of forest AGB (McNicol et al., 2018; Urbazaev et al., 2018).

Non-parametric machine learning algorithms, such as RF, support vector machines (SVM), and artificial neural network (ANN), do not assume a specific functional form between the input variables and the target variable, making them more flexible and adaptable to complex data structures (Fassnacht et al., 2014). Random forests, an ensemble learning method that combines multiple decision trees, have gained popularity in forest AGB estimation due to their ability to handle high-dimensional data, reduce overfitting, and provide variable importance measures (Baccini et al., 2012; Baghdadi et al., 2013). Random forests have been successfully applied to AGB estimation using a variety of remote sensing datasets, including optical, radar, and LiDAR data, often achieving higher accuracies than parametric methods (Gao et al., 2018; Wai et al., 2022).

More recently, deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been explored for forest AGB estimation (Lang et al., 2022). Deep learning algorithms are a subset of ANNs that consist of multiple hidden layers and can learn hierarchical representations of the input data (Schwartz et al., 2024). CNNs, in particular, have been successful in extracting spatial and contextual features from high-resolution remote sensing imagery, leading to improved AGB estimation accuracies compared to traditional machine learning methods.

The choice of machine learning algorithm for forest AGB estimation depends on several factors, including the characteristics of the input data, the complexity of the relationships between the variables, the sample size, and the computational resources available (Lu et al., 2016). In many cases, an ensemble approach that combines multiple machine learning algorithms can provide more robust and accurate AGB estimates than individual models (Forkuor et al., 2020). Additionally, the incorporation of feature selection and hyperparameter tuning techniques can further improve the performance of machine learning algorithms for forest AGB estimation (Pandit et al., 2018).

In summary, machine learning algorithms offer a powerful and flexible approach to forest AGB estimation using remotely sensed data. Both parametric and non-parametric methods have been successfully applied to AGB estimation across a range of forest types and data sources. As new remote sensing datasets and computational tools become available, the application of advanced machine learning techniques, such as deep learning, is expected to further improve the accuracy and efficiency of forest AGB estimation, supporting better forest management and conservation decisions.

1.3.6 Uncertainty of forest AGB estimation

Estimating biomass using remote sensing data and machine learning methods involves various sources of uncertainty that need to be carefully considered and addressed to improve the accuracy and reliability of the results. In the context of our study, which utilizes multiple remote sensing data sources (ICESat, GEDI, Sentinel-1/-2, Landsat, and ALOS PALSAR) in conjunction with machine learning approaches, the propagation of errors through the estimation process becomes particularly complex. The uncertainty in our biomass estimates stems from several interconnected sources, each contributing to the overall error in unique ways.

One of the primary sources of uncertainty in our approach lies in the field measurements used for model calibration and validation. Tree height and diameter at breast height (DBH) measurements, while fundamental to biomass estimation, are subject to both random and systematic errors. These errors can arise from instrument limitations, observer bias, and the inherent variability in tree form and structure (Calders et al., 2015).

The allometric models used to convert these field measurements into biomass estimates introduce another layer of uncertainty. The selection of appropriate allometric equations for specific forest types and regions is critical, as the application of generalized models can lead to substantial biases in biomass estimates (Chave et al., 2014). Moreover, the parameters of these models carry their own uncertainties which propagate through to the final biomass estimates.

In the realm of remote sensing data, uncertainties arise from various factors including sensor calibration, atmospheric effects, and spatial resolution limitations. The fusion of multiple data sources, while potentially improving overall accuracy, also introduces complexities in error propagation. For instance, the integration of optical (e.g., Sentinel-2, Landsat) and radar (e.g., Sentinel-1, ALOS PALSAR) data requires careful consideration of their respective error characteristics and how they interact in the biomass estimation process (Réjou-Méchain et al., 2017).

The use of machine learning models for biomass estimation introduces another layer of uncertainty. Model uncertainty can arise from the selection of input features, model architecture, and training data quality. Overfitting and underfitting are common issues that can affect model performance. Furthermore, the transferability of models across different regions and forest types can be limited, leading to potential biases in biomass predictions. Error propagation in modeling involves the compounding of uncertainties from input data and model parameters, which can significantly affect the final biomass estimates.

To address these multifaceted uncertainties, we employed a 5-fold cross-validation approach in Chapters 3 and 4 of this thesis. This method provides a robust means of assessing model performance and stability across different subsets of our data. However, it's important to note that while cross-validation is valuable for model evaluation, it may not fully capture all sources of uncertainty, particularly those arising from systematic biases in input data or limitations in the spatial representativeness of our training samples.

In our final national-scale AGB estimation, we recognize the limitations of relying solely on cross-validation for uncertainty quantification. The complexity of error propagation in large-scale biomass mapping, coupled with the potential for unquantified sources of uncertainty, suggests that a more comprehensive approach to uncertainty analysis may be necessary for future research.

1.4 Thesis objectives and key questions

The primary objective of this thesis is to develop robust and accurate methods for estimating aboveground biomass (AGB) in forest ecosystems using a combination of remote sensing data (ICESAT, GEDI, Sentinel-1/-2, Landsat, ALOS PALSAR) and machine learning techniques. To achieve the objective, this thesis addresses the following key research questions:

1. Research Question: Can a combination of ICESat GLAS data, SAR, and optical satellite imagery be used to create a high-resolution, wall-to-wall aboveground biomass (AGB) map of China?

   Hypothesis: The integration of ICESat GLAS-derived Lorey's height data with SAR and optical satellite imagery, coupled with random forest regression, will enable the generation of a accurate, high-resolution (50 m) AGB map for China.

2. Research Question: How accurately can GEDI LiDAR data estimate tree height and biomass in diverse forest ecosystems of China, and how effectively can these estimates be spatially extrapolated using other remote sensing datasets?

   Hypothesis: GEDI LiDAR data will provide accurate estimates of tree height and biomass across different forest types in China, but the accuracy of spatial extrapolation will vary with terrain complexity, with higher accuracy in flat areas compared to steep slopes.

3. Research Question: Can deep learning methods, particularly the attention UNet (AU) model, improve AGB estimation accuracy compared to traditional machine learning approaches like Random Forest?

   Hypothesis: The attention UNet model will outperform the Random Forest method in AGB estimation accuracy and spatial detail, particularly in areas with complex forest structures, albeit at a higher computational cost.

4. Research Question: How has the forest AGB in China changed between 2007 and 2021, and what are the spatial patterns of these changes?

   Hypothesis: There will be a significant increase in total forest AGB in China from 2007 to 2021, with spatial variations in AGB changes reflecting regional differences in forest management practices and environmental factors.

Chapter 2

2007 AGB map for China

Abstract

The forest area of China is the fifth largest of any country, and unlike in many other countries, in recent decades its area has been increasing. However, there are substantial differences in estimates of the amount of carbon this forest contains, ranging from 3.92 to 17.02 Pg C for circa 2007. This makes it unclear how the changes in China's forest area contribute to the global carbon cycle. We generate a circa 2007 AGB map at a resolution of 50 m using optical, radar and LiDAR satellite data. Our estimates of total carbon stored in the forest in China was 9.52 Pg C, with an average forest AGB of 104 Mg ha^{-1}. Compared with three existing AGB maps, our AGB map showed better correlation with a distributed set of forest inventory plots. In addition, our high resolution AGB map provided more details on spatial distribution of forest AGB, and is likely to help understand the carbon storage changes in China's forest.

2.1 Background & Summary

Forests reduce the impact of climate change, which is the greatest environmental challenge of the 21st century (Portner et al., 2022), by capturing CO_2 from the atmosphere, and acting as guardians of a carbon store in their wood, roots, and soils. However, there are major uncertainties as to the amount of carbon stored in forests, and how this is changing through time (Mitchard, 2018). The carbon stored in forests changes as a combination of the processes of deforestation, degradation and growth driven by both natural and anthropogenic disturbances, and climate change itself. This leads to fluxes in stored carbon, making forest carbon stock one of the greatest sources of uncertainty in the global carbon cycle (Mitchard, 2018; Grace et al., 2014; McDowell et al., 2020). It is important to reduce these uncertainties to improve our understanding of the carbon cycle and thus various elements of Earth system and climate models. This reduction is also crucial to enable policies to conserve and increase forest carbon storage to be designed and monitored.

Modern remote sensing data, cloud computing, and machine learning capacity, enable the production of high quality maps of AGB, the largest carbon pool in most forests, and a carbon pool that can change substantially over time (Besnard et al., 2021) . This increasing capacity to generate high quality maps of AGB has been shown by the recent release of a large number of global products (Avitabile et al., 2014; Yang et al., 2020; Spawn et al., 2020; Santoro and Cartus, 2021*a*). However, such global products can have high uncertainties at a local or regional level, and often perform poorly when compared to field plots (Santoro and Cartus, 2021*a*; Mitchard et al., 2014; Santoro et al., 2021).Given this, there is clearly a place for regional or country-specific maps where datasets and methods are tuned to local conditions, and that perform well against independent validation datasets (Rodríguez-Veiga et al., 2019).

China covers 6.3% of the world's land surface, has 18% of its people (The World Bank, 2021), and an estimated 5.4% of its forest (FAO, 2020). As forests play such an important role in the global climate system and in terrestrial ecosystems, what is happening in China's forests is important to the global climate system. The area of China's forest was 102 million hectares (ha) in 1949, and it remained at 106 million ha until 1994–1998 (Fang et al., 2001), and then increased dramatically to 220 million ha in 2020, accounting for 23% of China's area (FAO, 2020). Typically, substantial increases in forest area lead to increases in total AGB. However in China's case, it is possible that higher-biomass, old-growth forests are being replaced with new, lower-biomass forests or plantations (Wenhua, 2004; Viña et al., 2016).

Currently, information of forest AGB can be derived from field inventory data and Earth Observation (EO) data. Conventionally, forest AGB estimation requires inventory forest variables such as tree species, DBH, and tree height, measured in fixed area forest inventory plots, which are converted into estimates of tree biomass through allometric equations

(Chave et al., 2004). Although conventional inventory techniques generally provide accurate estimates of forest AGB at plot level (Henry et al., 2011), it is time-consuming, only possible to collect data covering a very small area, requiring good statistical stratification and placement of a large number of field plots in order for them to represent the wider forest, and sometimes impractical due to inaccessibility. It is obviously impossible in the present time to add sampling plots or locations retroactively for historical data collection. On the other hand, EO data, from satellites or aircraft, has shown promising capabilities for consistent and systematic observations of the dynamics of forest ecosystems (Blackard et al., 2008; Mitchard et al., 2009; Baccini et al., 2012; Araza et al., 2022). However, producing accurate EO-based results relies on consideration of local factors, ideally using local field calibration data as well as appropriate EO data sources.

Some studies of AGB in China have been conducted using satellite data. However, these studies produce very different estimates, due to differences in their prediction models and dataset sources (Table 2.1). The total carbon estimates from the map of Su (Su et al., 2016) are four times higher than that of GEOCARBON (Avitabile et al., 2014). While these maps use different forest cover maps to mask non-forest areas, resulting in variations in forest area, and the maps are for slightly different periods, this disparity in area and time cannot explain much of the observed difference.

Table 2.1. Total carbon stored in forest AGB in China from a number of studies, along with the area covered by the maps. The carbon estimated for GEOCARBON, CCI Biomass, V3 and the Su maps were derived using AGB maps provided by the respective studies. For GEOCARBON and CCI Biomass, V3, non-forested areas were excluded based on Hansen forest cover data. The Su map had already excluded non-forest areas in their AGB map. Estimates for all other sources were obtained directly from the respective publications.

Source	Carbon (Pg C)	Period	Area
GEOCARBON	3.92	2000-2010	Global
Huang, et al 2019	5.44	2006	China
CCI Biomass, v3	5.94	2010	Global
Yin, et al. 2015	8.56	2001-2013	China
Li, et al. 2015	9.87	2010	China
Su, et al. 2016	17.02	2004	China

Therefore, our study aims to produce a reliable 2007 AGB map for China. We chose to map 2007 for a number of reasons, summarised as (i) it is a useful baseline year to compare against forest policy and (ii) a variety of remote sensing and field validation datasets is available around that date. The Asia-Pacific Network for Sustainable Forest Management and Rehabilitation (APFNet) was proposed in 2007, and it further marks the acceleration of the China's reforestation/afforestation programmes, as well as the implementation of forest protection laws (Liu et al., 2017). From the remote sensing standpoint, 2007 is where spaceborne LiDAR and suitable optical and radar satellite data are available. LiDAR data from satellites provide a direct estimate of height, which is closely related to biomass, unlike other satellite sensors which rely on indirect estimation, with the shape of relationships often varying from place to place (Saatchi et al., 2011). The ICESat GLAS operated throughout the 2000s, including collecting lots of data in 2007, and provides a large number of ∼65 m circular footprints with estimates of height, used to train and test the wall-to-wall maps of AGB we produce. In order to extrapolate across the landscape, we need SAR and optical datasets. The C-band Envisat ASAR and ALOS PALSAR were available around 2007. C-band SAR is capable of penetrating into the canopy, and L-band SAR can penetrate deeper into the canopy and thus obtain information of larger element of trees such stems (Naidoo et al., 2015). Landsat-5 TM data was also used in the prediction, as optical data can capture detailed spectral information of forest canopy.

In this study, we introduced a new 2007 AGB map for China using high resolution multisource remote sensing data. Our model was trained specifically for China, considering its unique allometric equation and terrain factors. This dataset can serve as the historical baseline dataset for forest AGB change studies, as it outperforms existing maps when compared to two unseen field validation datasets.

2.2 Methods

We generated point estimates of AGB from ICESat GLAS Lorey's height data (Lefsky, 2010), using local allometric relationships developed independently for northern and southern China using field data we collected in 40 forest inventory plots in China. We then predict AGB spatially to create a wall-to-wall map at a resolution of 50 m using a random forest model with L-band SAR from ALOS PALSAR, C-band SAR from Envisat ASAR, and optical satellite data from Landsat-5, all acquired around 2007, in addition to a Digital Elevation Model from the Shuttle Radar Topography Mission (SRTM) (collected in 2000). Both forest structure and remote sensing responses are different in sloped compared to flat areas (Chen, 2010; Jucker et al., 2018), so we generated continuous AGB maps for flat regions and rugged regions separately, to enable different prediction models for the two and thus reduce the errors in AGB estimates. Due to gaps in the coverage by Envisat ASAR (1.2% of China was missed), we generated a second AGB map with slightly lower accuracy to use in the regions without Envisat ASAR data (Fig. 2.1). As a last step, we combined the maps for flat region (slope $\leq 10°$), rugged region (slope $> 10°$) and region without Envisat data, to produce a final map for validation (Fig. 2.2).

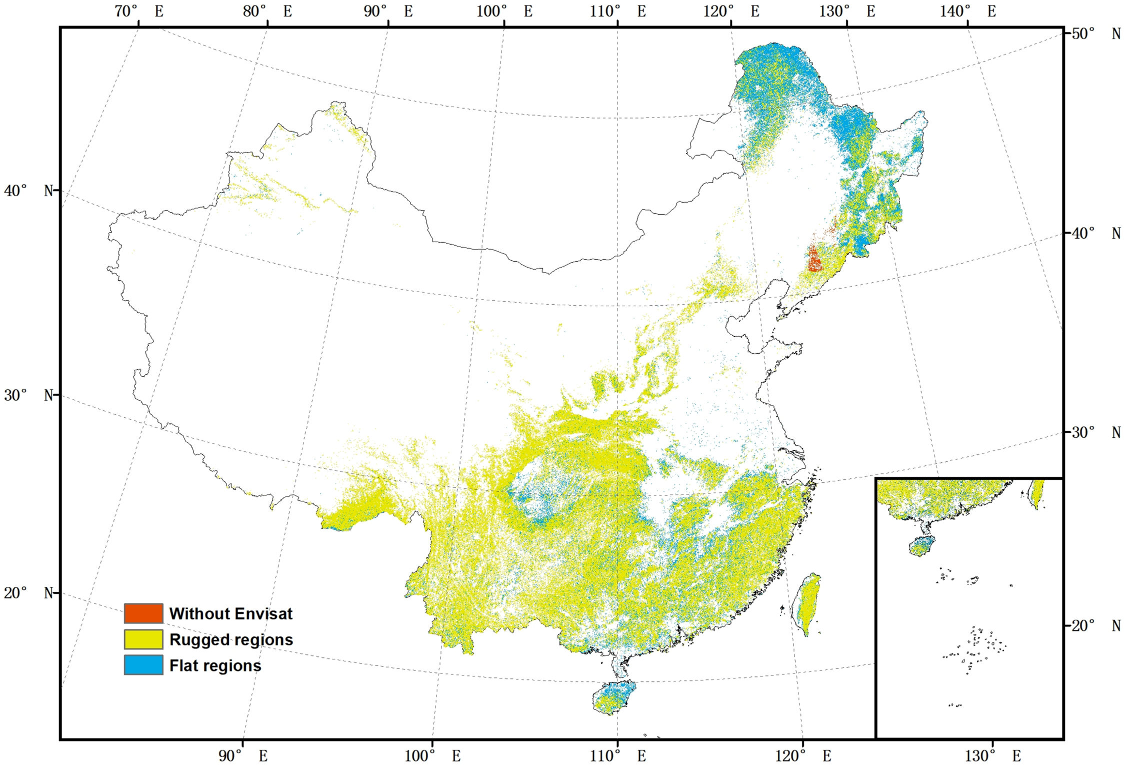


Figure 2.1: The spatial distribution of the three forested regions.

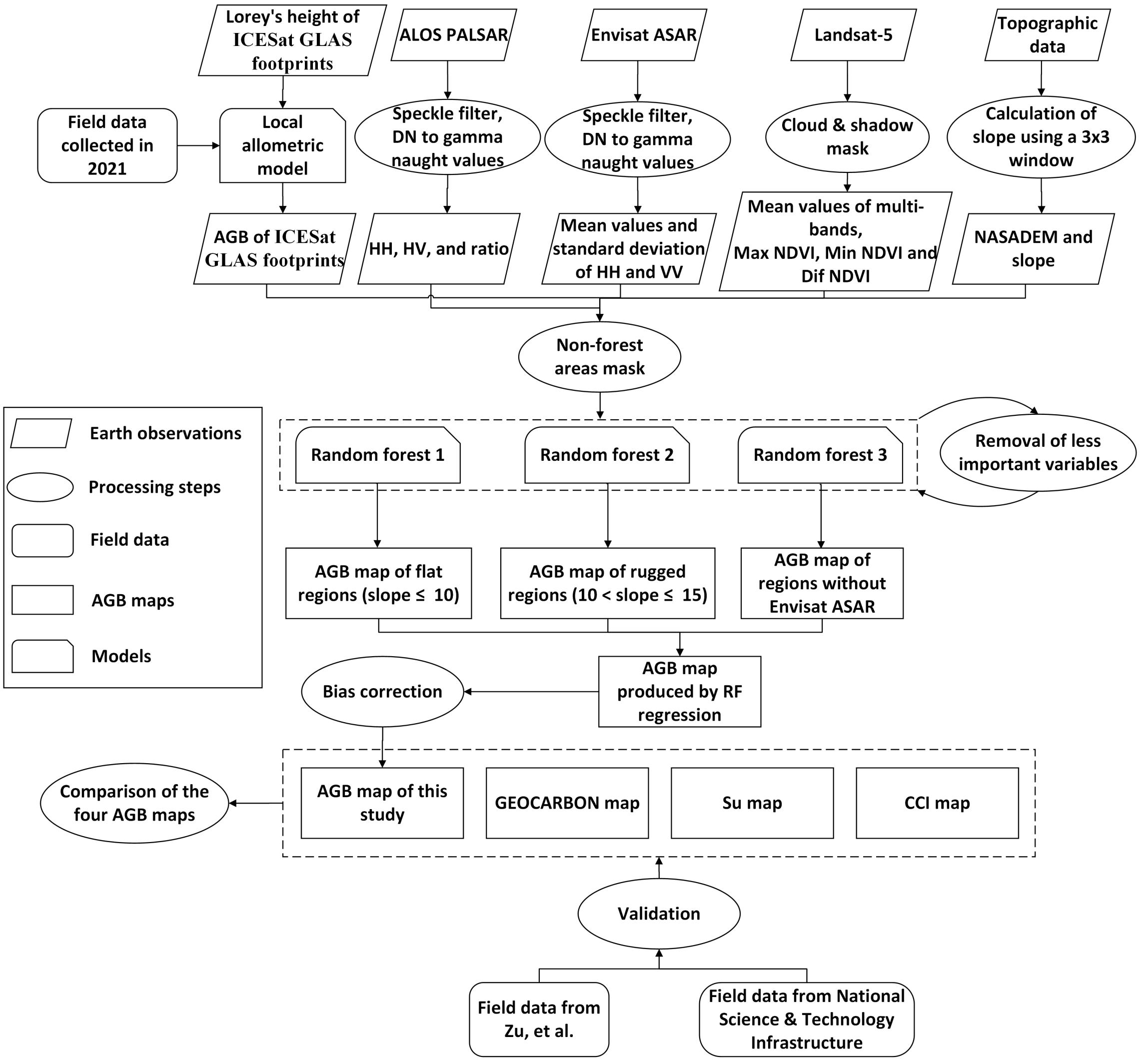


Figure 2.2: Overall workflow of forest AGB estimation and validation.

Further detail about all these steps is given below, and code is available.

2.2.1 Field data

Field data collected in 2021

26 field plots in northeastern China and 14 field plots in Southwestern China were measured in 2021 (Fig.2.3). Each of the field plot was circular with a radius of 12.5 m $\sim 0.05ha$. We measured tree height using Vertex IV and Transponder T3, and measured DBH of all the trees with DBH $\geq$ 5 cm. These measurements were used to estimate the AGB of each tree, utilizing the following biomass equation for China (Luo et al., 2019):

$$W = 0.1355 * \left(D^2 * H\right)^{0.817} \tag{2.1}$$

Where, W is the AGB of each tree (kg), D is DBH (cm) and H is height of the tree (m). We summed the AGB of every tree to estimate total AGB per plot, and converted the total biomass to AGB density at the 1 ha scale (Mg ha^{-1}).

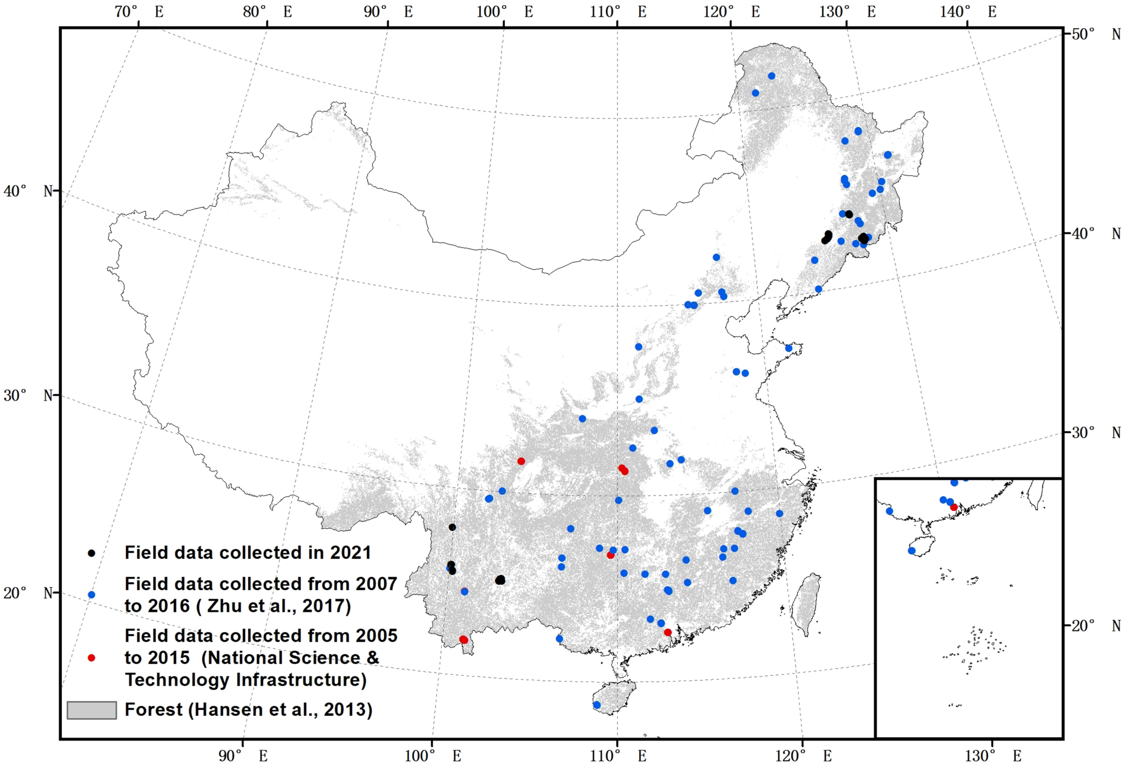


Figure 2.3: Spatial distribution of field data collected for this study in 2021 and compiled from other studies from 2005 to 2015.

We also calculated Lorey's height of each plot using Eq. 2.2. Lorey's height is an index of height whereby individual trees are weighted in proportion to their basal area, which enhances the significance of larger trees in the forest, and is closely correlated with AGB (Saatchi et al., 2011; Liu, Shen, Cao, Wang and Cao, 2018; Cao et al., 2016).

$$\text{Lorey} = \frac{\sum_{i=1}^{N} BA_i h_i}{\sum_{i=1}^{N} BA_i} \tag{2.2}$$

where, BA_i represents the basal areas, and h_i represents the canopy height of the individual trees.

We used these plot-level measurements to develop allometric equations between the field measured Lorey's height and AGB in northern China and southern China independently (2.4).

$$AGB_n = 0.5014 \times \text{ Lorey }^{1.8762} \tag{2.3}$$

$$AGB_s = 0.1237 \times \text{ Lorey }^{2.3281} \tag{2.4}$$

where, AGB_N is AGB density (Mg ha^{-1}) of field plots in northern China, AGB_S is AGB density of field plots in southern China and Lorey is Lorey's height. The allometric equation for northern China showed an R^2 of 0.75 and an RMSE of 35.58 Mg ha^{-1}, while for southern China, the R^2 was 0.50 with an RMSE of 22.96 Mg ha^{-1}.

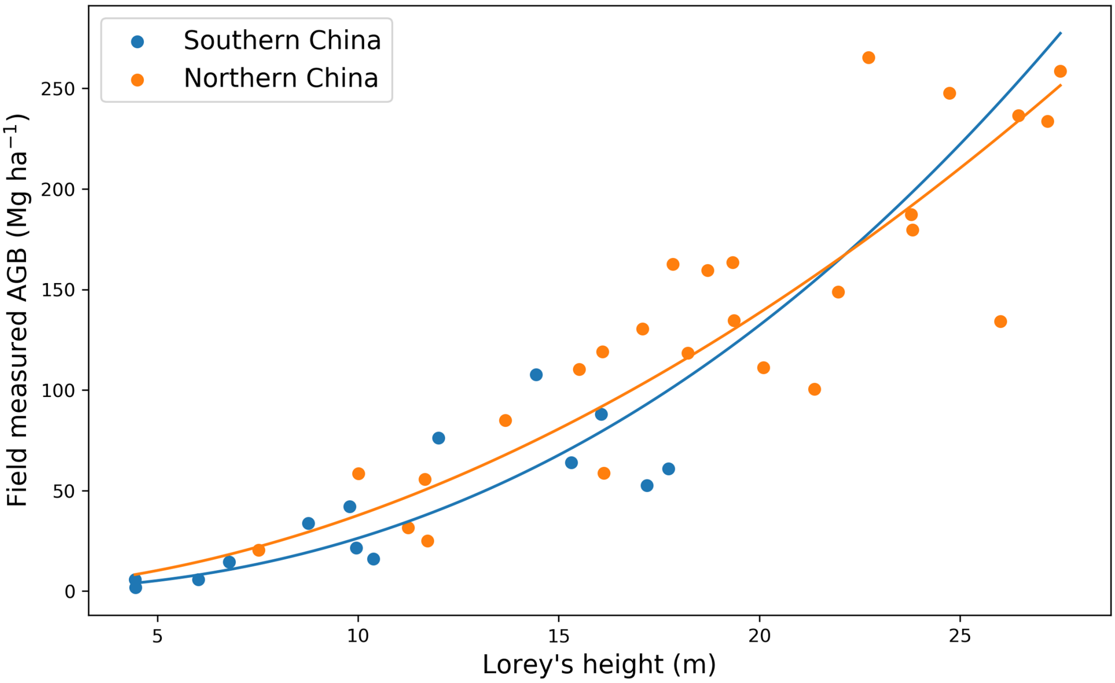


Figure 2.4: Allometric equations between field measured Lorey's height and AGB. The solid lines are fitted lines. Both fitted lines were extended to the full range to provide a more comprehensive representation of the potential trends of the allometric equations.

The field data were collected in 2021, therefore they cannot be used to validate the accuracy of a ∼ 2007 AGB map. Given the diverse forest types and vast geographical expanse of China, using these two equations to convert Lorey's height to forest AGB across China could lead to biases when used in different forest types, and in different areas of China. The fact that the two datasets are very close together does add some confidence that the forest height: biomass relationship of China does not change greatly from location to location. There is precedent in the literature with single equations being used to derive biomass across large regions, with Saatchi et al. (2011) using a single equation per continent in their pantropical biomass map, Baccini et al. (2012) using a single equation across the whole tropics, and the Global Ecosystem Dynamics Investigation (GEDI) LiDAR mission's biomass product itself using a single equation for all of Asia (Duncanson et al., 2020). In order to assess the impact of this decision, and other errors in the biomass modelling, two independent field datasets collected closer to 2007 from across the woody areas of China were used for validation (Fig. 2.3).

Field data from Zhu, et al

One of the datasets consisted of field inventory measurements collected between 2011 and 2016. It included 189 sites with three 20 m × 20 m plots at each site. The longitude and latitude of each site were the averaged values of the three plots within the site (the three plots were usually close to each other, e.g., within 50 m). In order to reduce the uncertainty of biomass change from 2007 to 2016 and the uncertainty of geolocations, 83 field sites were used to validate our map. Field sites were used in analysis if they met the following three criteria (Zhu et al., 2017).

1. At least half of the area in the buffered region 100 m around the field sites were forests according to the Hansen forest cover data (Hansen et al., 2013). We created 100 m buffers to reduce the implication of the uncertainty of the geolocations, considering the geolocations of the sites were averaged values of the three plots in each site.

2. In order to show there had been no major changes in the forest in that area between the time of the creation of the maps and the date of the measurement of the field plots, we confirmed that:

    - The maximum Normalized Difference Vegetation Index (NDVI) of each year calculated from Landsat were stable from 2007 to 2016 (Fluctuations of NDVI less than 0.15).

    - The values of ALOS PALSAR and PALSAR-2 HH and HV polarizations were stable from 2007 to 2016 (Fluctuations of HH band and HV band less than 2 in decibel unit).

3. To exclude the field sites that are highly spatial heterogeneous, which may introduce large errors, we removed field sites that have large ratio of standard deviation of the AGB in the 100 m buffers and the field AGB (ratio > 0.25).

Field data from National Science & Technology Infrastructure (NSTI)

The other dataset from NSTI (`http://www.cnern.org.cn/`) was collected from 2005 to 2015 in 29 sites (Fig.2.3). There were 5 to 100 plots (10 m × 10 m) in each site. The dataset was collected by nine field research stations, and was measured every few years (normally every five years). We used the average values of the measurements of AGB which were collected around 2007. We refrained from implementing the above filters due to several reasons. Firstly, the geolocation uncertainty of each site varied, making it challenging to apply a uniform filtering approach. Secondly, the size of each field site in this dataset was potentially varied, further complicating the filtering process. In addition, the dataset available for analysis was relatively limited, and applying filters could potentially result in insufficient field sites for validation purposes.

2.2.2 Satellite data

ICESat GLAS Lorey's height

GLAS which was launched aboard the NASA Ice Cloud and Land Elevation Satellite (ICESat) satellite on January 12, 2003, was the first laser-ranging (LiDAR) instrument for global observations of the Earth. It was designed to measure ice-sheet topography and associated temporal changes, cloud and atmospheric properties, to obtain information on the height and thickness of radiatively important cloud layers, and as a tertiary aim obtain information on tree heights globally.

GLAS provided data with 65 m diameter circular footprints every 170 m along orbits between -86 to 86° latitude. Since we aimed to produce a forest AGB map in ~ 2007, we selected 287,399 GLAS footprints located on forests (Hansen et al., 2013) acquired during the operating periods L3H and L3I (Table 2.2). In this study, we directly used the Lorey's height values generated by Lefsky (Lefsky, 2010), who developed distinct equations for estimating Lorey's height for needleleaf, broadleaf, and mixed forests globally, based on waveform extent indices from GLAS.

Table 2.2: Reference orbit tracks acquired during ICESat observation periods.

Operation Period	Start Date	End Date	Starting Track	Ending Track
Laser 3H	2007-03-12	2007-04-14	91-day #1279	91-day #426
Laser 3I	2007-10-02	2007-11-05	91-day #1280	91-day #421

We removed the GLAS footprints over hills with slope $> 15°$, because of potential overestimates or underestimates in Lorey's height in high slope areas (Chen, 2010). We also excluded GLAS footprints in forests with slopes between 12° and 15° and where AGB ≤ 40 Mg ha^{-1} (roughly 11–12 m in height). This is because deriving tree height becomes highly uncertain when the topographic relief within a footprint is large compared to the tree height (Fig.2.5), resulting in lower AGB estimates being more susceptible to potential impacts from slope (Harding and Carabajal, 2005).

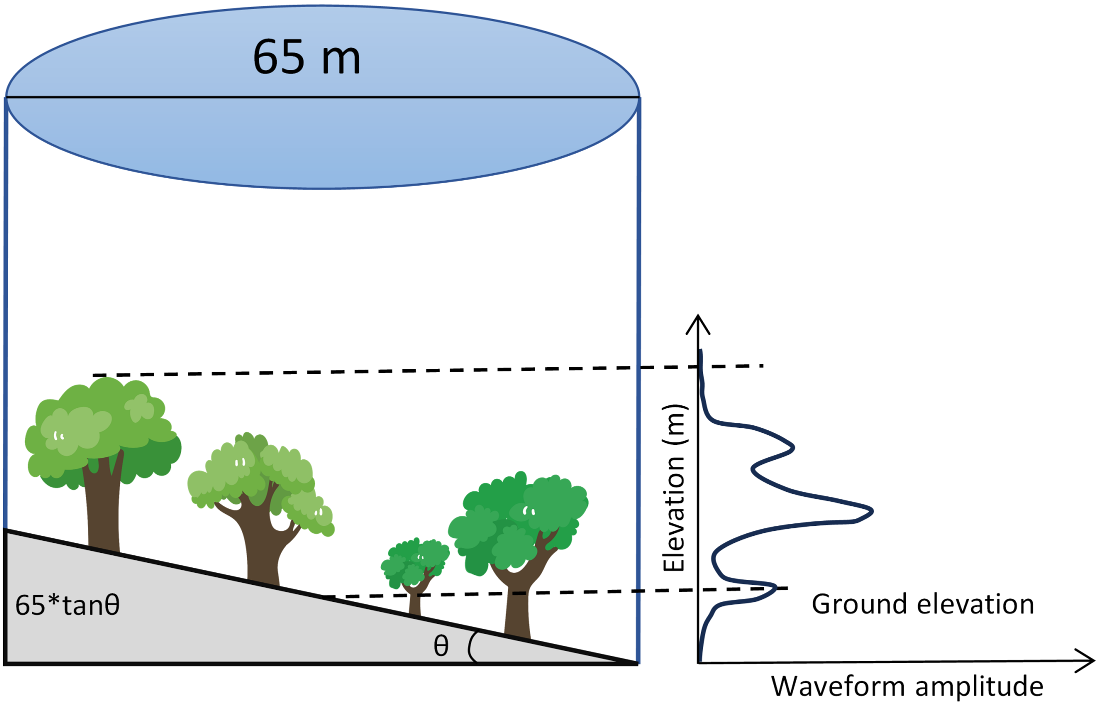


Figure 2.5: The effects of slope on trees of different heights.

ALOS PALSAR annual mosaic data

PALSAR was an L-band SAR instrument on ALOS that operated by the Japan Aerospace Exploration Agency (JAXA) between 2006 and 2011, that has proved very useful in mapping forest biomass across a range of ecosystems(Hamdan et al., 2014; Carreiras et al., 2012; Bouvet et al., 2018; Chen, Wang, Ren, Zhang and Wang, 2019). L-band SAR is the longest wavelength SAR (∼24 cm) currently available from space. JAXA produced annual seamless mosaics of the radar backscattered intensity from the PALSAR observations at HH or HV polarization, with terrain and slope correction already applied (Shimada et al., 2014). We accessed the 2007 mosaic at the resolution of 25 m and did pre-processing through Google Earth Engine (GEE). We applied a focal mean smoothing filter with a radius of 150 m to reduce speckle noise. The digital number (DN) values were converted to gamma naught values in decibel unit (dB) using the equation provided by JAXA:

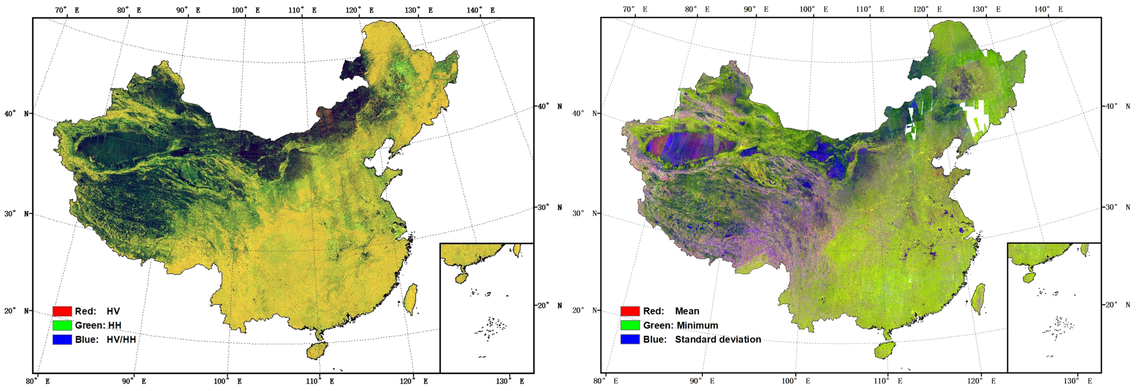
$$\gamma^0 = 10\log_{10}\left(DN^2\right) + CF \tag{2.5}$$

where, γ^0 is the backscattering coefficient in dB, and CF is a calibration factor equal to −83.0 dB (Motohka et al., 2014) for the PALSAR mosaic. HV, HH polarizations and the ratio of HV and HH were used in the estimation model, as the polarization ratio has a higher saturation point (Sarker et al., 2012; Hayashi et al., 2019).

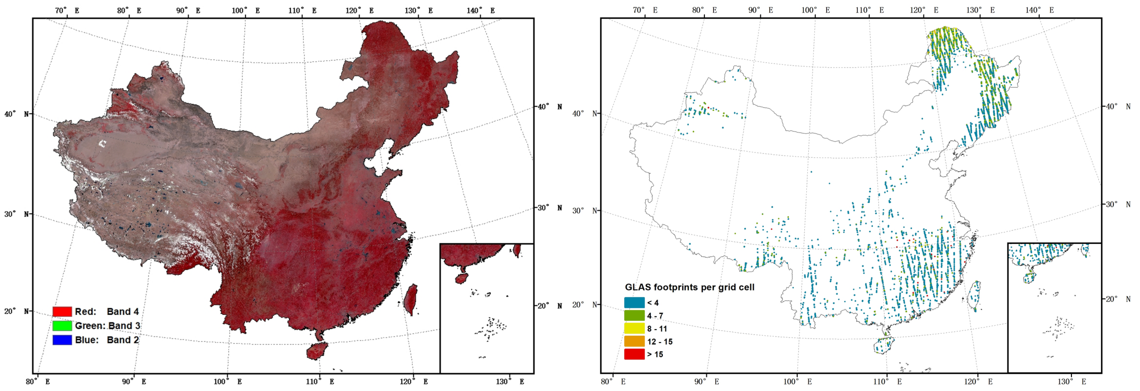
Envisat ASAR

Envisat was a satellite mission operated by the European Space Agency (ESA). The main objective of the Envisat programme was to study and monitor the Earth's environment on various scales, from local through regional to global (Louet and Bruzzi, 1999). Envisat was equipped with ten instruments, including an active C-band sensor ASAR. C-band is a shorter wavelength (∼6 cm) than L-band, making it suitable for assisting with lower biomass areas; additionally its frequent observations in the Wide Swath mode at a resolution of 150 m allow the generation of the standard deviation of the backscatter through the year, which further helps to understand characteristics of the forests that link to biomass (Santoro, Beaudoin, Beer, Cartus, Fransson, Hall, Pathe, Schmullius, Schepaschenko, Shvidenko et al., 2015). Envisat ASAR measured the radar backscatter of the Earth's surface at HH or VV polarization, both of which were collected in our study area.

We calculated the mean values and standard deviation of combined HH and VV images together over 2007 to 2008, as C-band SAR backscatter and forest biomass has similar relationship with both polarizations (Santoro, Beaudoin, Beer, Cartus, Fransson, Hall, Pathe, Schmullius, Schepaschenko, Shvidenko et al., 2015; Castel et al., 2001), and this combination reduced artefacts due to differing look angles and number of observations. In order to calculate the mean values and standard deviation, we only used pixels where there were at least two images of each pixel (gaps correspond to 1.2% of the total area of China, see Fig. 2.6 (b), where a separate AGB retrieval model not based on Envisat data was used (Santoro, Wegmüller, Lamarche, Bontemps, Defourny and Arino, 2015). This method of selecting pixels with sufficient multi-temporal coverage was aimed at ensuring more stable and reliable observations. In total we used 113,698 ASAR images tiled in 1 degree × 1 degree grid cells across China.



(a) ALOS PALSAR (b) Envisat ASAR



(c) Landsat-5 TM (d) ICESAT GLAS

Figure 2.6: Earth observations used in this study. (a) False color composite of ALOS PALSAR annual mosaic data for the year 2007 (HV-polarized backscatter in red, HH-polarized backscatter in green, and the ratio of the backscatter HV/HH in blue). (b) False color composite of the Envisat ASAR dataset covering China acquired over 2007 to 2008 (mean of HH and VV in red, minimum of HH and VV in green, and standard deviation of HH and VV in blue). The white areas are where no observations were collected. (c) False color composite of temporal mean values computed from Landsat-5 TM bands 2, 3 and 4 acquired from 1st January 2006 to 30th December 2008. Band 4 (Near-infrared) in red, Band 3 (Red) in green, Band 2 (Green) in blue. (d) Filtered grid cells of ICESAT GLAS data used to create and test the model.

Landsat-5

Since the first Landsat satellite launched in 1972, Landsat data have contributed to our understanding of Earth, especially in land use and land use change. We used Thematic Mapper (TM) images here, which consist of seven spectral bands with a spatial resolution of 30 meters for Bands 1 to 5 and 7, and a temporal resolution of 16 days. We extracted all the images acquired from 1st January 2006 to 30th December 2008, and used Quality Assessment (QA) band, generated from the CFMASK algorithm (Foga et al., 2017), to mask cloud and cloud shadow. We then stacked the cloud-free Landsat-5 bands and computed the mean value of the spectral reflectance in the stack. The maximum, minimum and the differences of maximum and minimum values of NDVI of each pixel were also calculated because of the strong predictive power when related to AGB (Piao et al., 2005; Dong et al., 2003; Saatchi et al., 2007; Lumbierres et al., 2017).

Topographic data

NASADEM was generated by reprocessing the SRTM data with other DEM data, such as ICESat GLAS, ASTER GDEM and ALOS PRISM DEM data (Crippen et al., 2016). NASADEM is more accurate than SRTM, has no holes, and is globally available at the finer resolution of 30 m (Uuemaa et al., 2020). In this study, we used the elevation in the NASADEM product as well as terrain slope derived from NASADEM.

2.2.3 Forest AGB estimation

We converted Lorey's height derived from GLAS into forest AGB using the allometric equations developed for northern China and southern China independently with the field data. Each individual Lorey's height estimate from GLAS comes with considerable random errors in height estimation (RMSE was 5.9 m and R^2 was 0.67 (Lefsky, 2010)), in addition to geographic uncertainty. In order to reduce the influence of these errors, we averaged the Lorey's height within 0.01 degrees × 0.01 degrees grid cells and obtained 8,981 cells with at least two GLAS footprints (mean 4.14 footprints). We limited Lorey's height values to 25 m, as our review of the literature showed nearly no plots with Lorey's height over 25 m in China (Huang et al., 2019; Luo et al., 2014), especially when averaged in 0.01 degrees × 0.01 degrees grid cells.

All remote sensing imagery was resampled to a spatial resolution of 50 meters in GEE. We then used RF regression (Breiman, 2001)) to extrapolate AGB grid cells from GLAS data to a continuous map at 50 m resolution. We trained the RF regression using the variables derived from Earth Observation datasets listed above as well as layers giving latitude and longitude, to enable the model to take account of the differing relationships across the varied ecosystems of China. As we were focused only on forest AGB, we used the 30 m resolution Hansen et al. (2013) forest cover data to mask all the input layers to forest only, preventing any influence of non-forest areas on the RF regression. The 2007 forest cover map was obtained by removing pixels flagged as lost between 2000 and 2007, from baseline forest canopy cover >10% in the year 2000 canopy cover layer. However, due to the lack of forest gains between 2000 and 2007, the 2007 forest cover map may underestimate the actual extent of forest cover. Since the importance of the variables differs in flat region and rugged region (Fig. 2.7), we modelled the AGB in flat region and rugged region separately to make full use of the remote sensing layers.

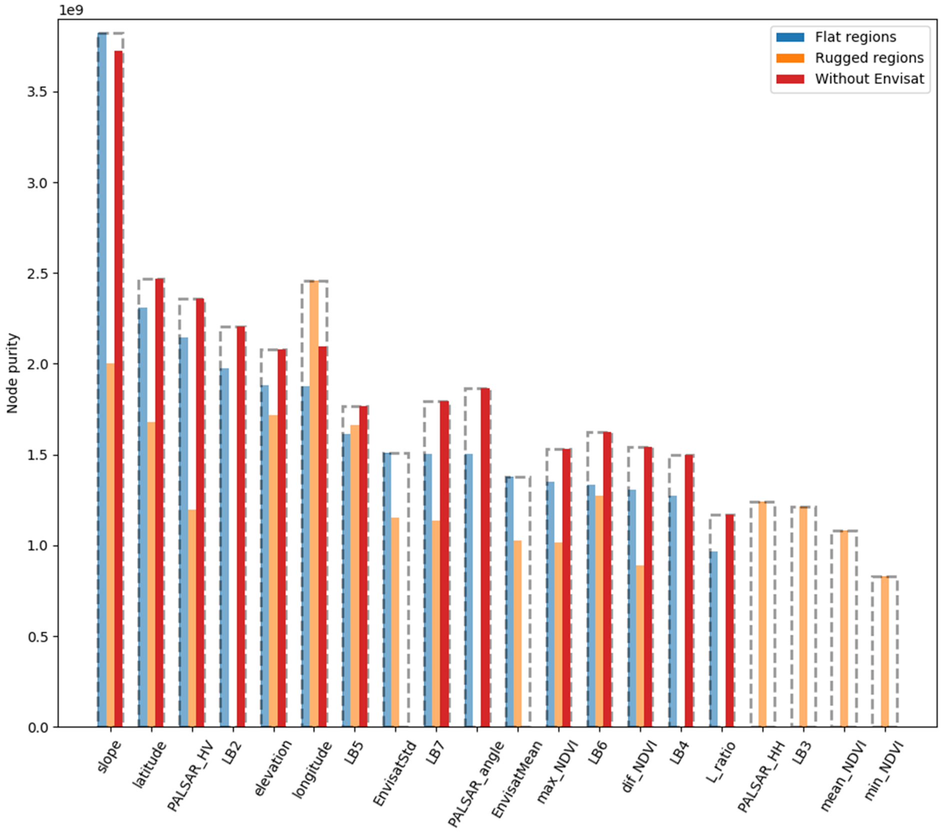


Figure 2.7: Importance rank of selected variables in RF regression (importance of layers in flat regions from high to low). The six least important variables for all the three regions are not shown in the figure. `LBx` represents Landsat-5 Band x, `dif_NDVI` represents the differences of maximum and minimum values of NDVI. `EnvisatMean` and `EnvisatStd` represent the mean values and standard deviation of Envisat data.

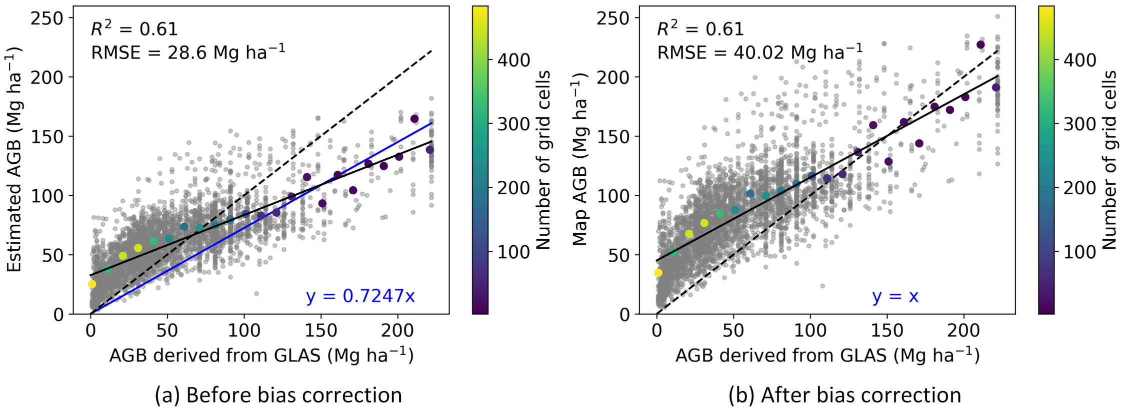
We eliminated redundant information and thus improved the computational efficiency of prediction by analysing the importance of all input layers on the AGB map. We considered the node purity, which is equivalent to Mean Decrease Gini, of each variable (Fig. 2.7). The higher the value of node purity, the higher the importance of the variable to our model.

This is because node purity increased along with a decrease in the residual sum of squares, indicating a decrease in the Gini coefficient in regression analysis (Han et al., 2016; Wang, Song, Du, Wang, Zhou, Du and Li, 2019). We removed the 10 variables with smallest node purity in each of the three models.

2.2.4 Accuracy and bias correction

To evaluate the performance of the AGB estimation model, 8,981 grid cells of 0.01 degrees × 0.01 degrees AGB were randomly split in training set (60% of the data) and validation set (40% of the data). The performance of the model was assessed by aggregating the pixel-wise AGB estimates to the same size of grid cells at 0.01 degrees × 0.01 degrees. The coefficient of determination (R^2) and Root Mean Square Error (RMSE) were used to quantify the performance of the model.

The model estimated AGB corresponded well with the AGB derived from GLAS, and the R^2 and RMSE between the estimated AGB and AGB derived from GLAS were 0.61 and 28.6 Mg ha^{-1}, respectively (Fig. 2.8 a). Our method tended to overestimate forest AGB at the forests with low AGB (AGB $<$ 60 Mg ha^{-1}), and underestimate in forests with large AGB, as is common with RF models. To correct for this bias, a linear regression bias correction was performed to the estimated AGB using the slope of the fitting line of 10 Mg ha^{-1} intervals (Song, 2015; Shendryk, 2022; Liang et al., 2023). This improved, but did not solve, the bias issue, with increased scatter in the output (RMSE = 40 Mg ha^{-1} compared to 28.6 Mg ha^{-1} pre-correction), but with the range of values better matching the range from the GLAS predictions (maximum predicted value 251 Mg ha^{-1} after correction, compared to 182 Mg ha^{-1} pre-correction, compared to the maximum GLAS prediction of 222 Mg ha^{-1}).



(a) Before bias correction

(b) After bias correction

Figure 2.8: Evaluation of the model. (a) Estimated AGB from our model against AGB of averaged grid cells derived from GLAS data (grey points). (b) Bias corrected AGB against AGB of averaged grid cells derived from GLAS data (grey points). The dotted line is the 1:1 line. The black solid line is the regression of the whole set of points. The colour dots are the average values of 10 Mg ha^{-1} interval of AGB derived from GLAS, and the colour bar represents the number of grid cells in a given AGB interval. The blue line is a standard regression of colour dots with the intercept forced to be zero; in the bias corrected figure (b) this line overlaps exactly with the 1:1 line and is thus not visible.

2.3 Data Records

The new forest AGB map was produced using multisource remote sensing datasets, including ICESat GLAS data acquired in 2007, ALOS PALSAR data acquired in 2007, Envisat ASAR data acquired in 2007 and 2008, Landsat-5 data acquired from 2006 to 2008, and NASADEM data. Due to the remote sensing datasets all acquired around 2007, it is a circa 2007 forest AGB map. The AGB map is stored in GeoTIFF (.tif) files on Edinburgh DataShare (Dong, Mitchard, Santoro, Chen and Wheeler, 2023). The map provides the values of AGB in Mg ha^{-1}, and the values have been rounded to integers. It can be converted to carbon by multiplying by 0.47 (Change, 2006).

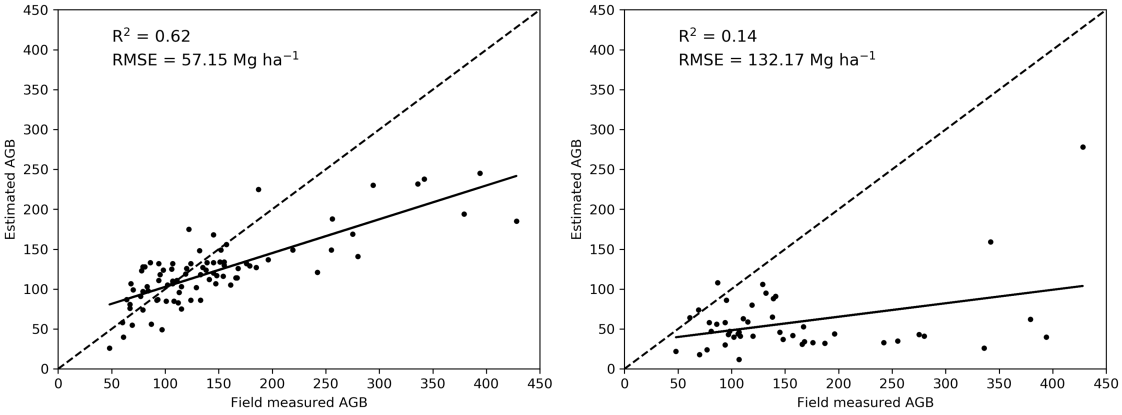
2.4 Technical Validation

Our estimates of total carbon stored in forest AGB in China is 9.52 Pg C. To assess the accuracy of our AGB map, we validated our map using one field datasets from a previous study (Zhu et al., 2017) and the other independent dataset from NSTI. The coefficient of determination (R^2) and RMSE were used to quantify the performance of the maps. In addition, we compared this map to the GEOCARBON map (Avitabile et al., 2014), Su map 2016, and CCI map (Santoro and Cartus, 2021*a*) and attempt to explain the differences.

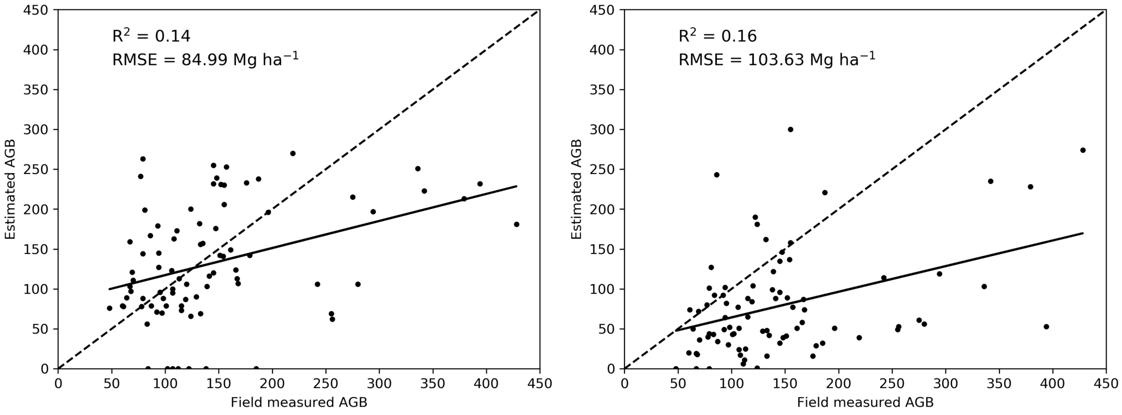
2.4.1 Validation using field data

Comparison to Field data from Zhu, et al

We used 83 sites from Zhu, et al. 2017 to validate our map at a pixel level. Overall, the estimated AGB of our map demonstrated a reasonably good correspondence with field measured AGB, with an R^2 of 0.62 and an RMSE of 57.15 Mg ha^{-1}; this exceeds considerably the metrics obtained from the other datasets (Fig. 2.9). Compared with the field data, all the four maps exhibited a tendency to underestimate AGB values in areas with large AGB (>200 Mg ha^{-1}). The other national scale AGB map, the Su map 2016, exhibited a lower RMSE than the two global AGB maps.



(a) This study (b) GEOCARBON map



(c) Su map (d) CCI map

Figure 2.9: Assessment of AGB map against independent field data from Zhu, et al. The dotted line is the 1:1 line.

To assess the similarities and differences in the AGB distribution, we compared field data from Zhu, et al.(2017) and the pixel values of the four AGB maps within 1 km × 1 km grid cells, which represents the resolution of the coarsest map (Fig. 2.10). The field data hosted the highest median and mean AGB, and half of the field sites had an AGB density of 95–160 Mg ha^{-1}. The map from this study had a similar low end of the interquartile range with the field data, but was more homogeneously distributed with 75% of the AGB below 130 Mg ha^{-1}. GEOCARBON map hosted the smallest median and mean AGB and was most homogeneously distributed. The AGB of Su map had the largest dispersion, with 50% of the AGB between 80 Mg ha^{-1} and 190 Mg ha^{-1}. The CCI map had the second smallest median AGB of 60 Mg ha^{-1}. It can be seen that the two national scale AGB maps are closer match to these field data in this comparison.

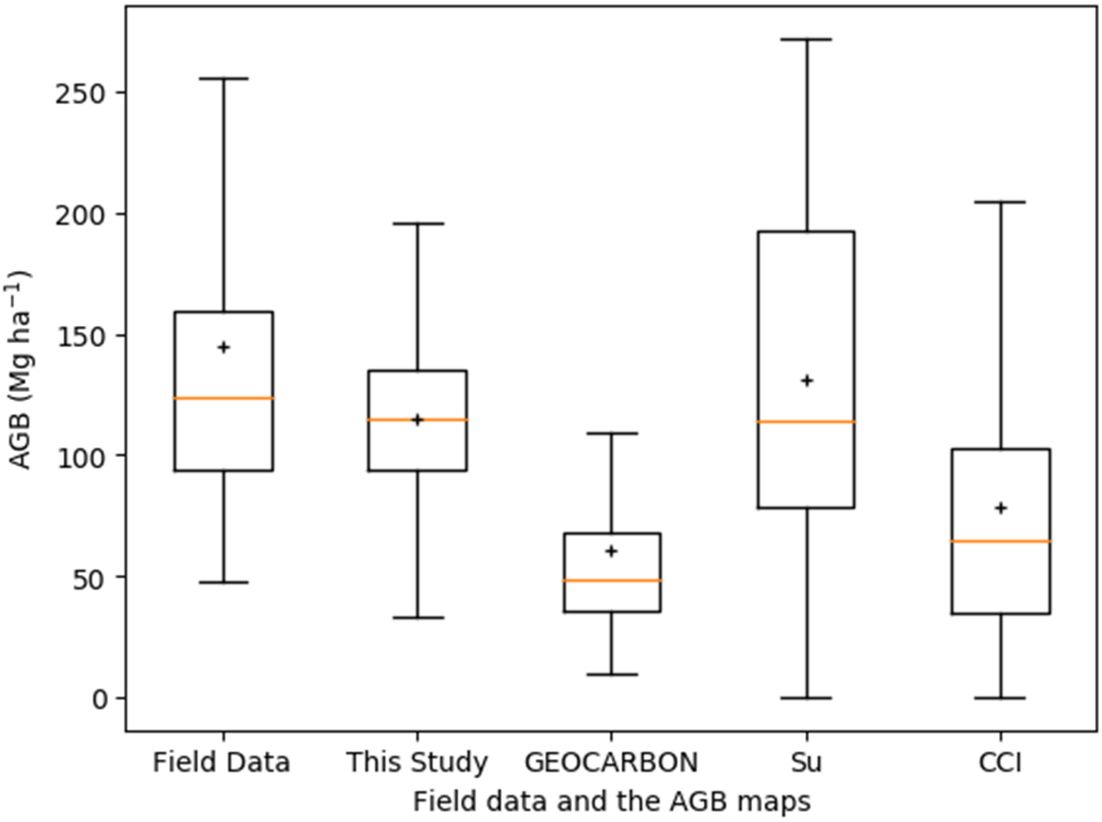
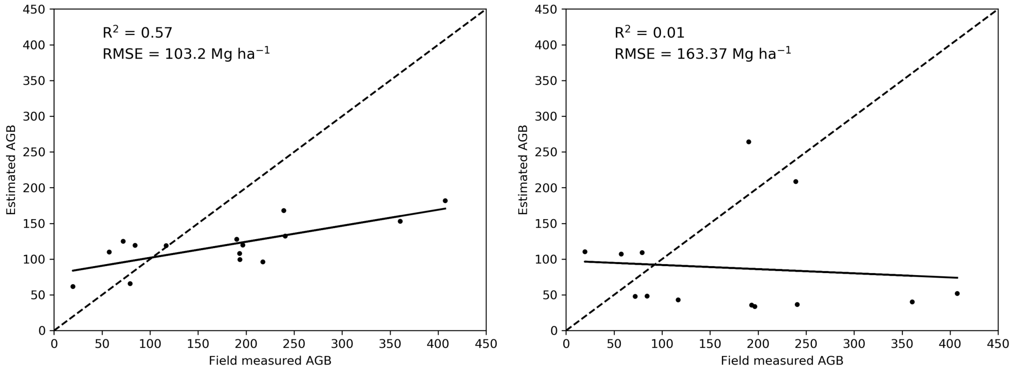


Figure 2.10: Boxplot of field data and forest AGB of the four AGB maps within 1 km × 1 km grid cells of field sites. The orange horizontal bar indicates the median and the boxes show the interquartile range, and the "+" indicates the mean forest AGB.

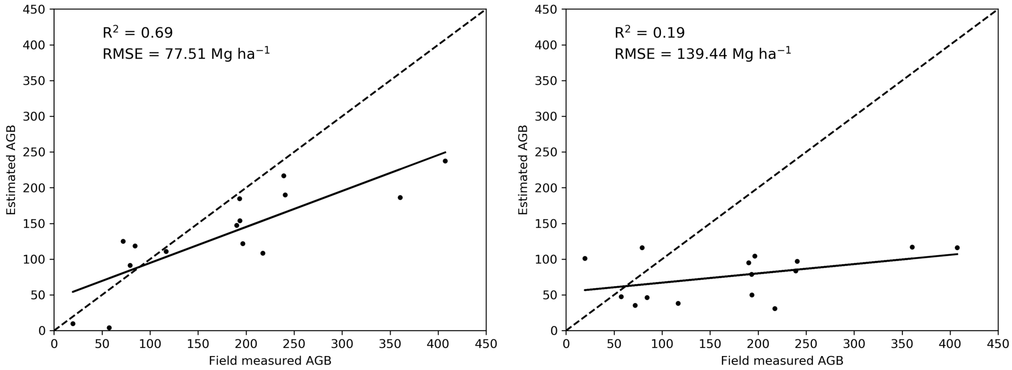
Comparison to Field data from NSTI

We also validated our map, and the GEOCARBON, Su and CCI maps using field data from NSTI. The uncertainty of the geolocations of the data from NSTI were from 0.0001 degrees to 0.01 degrees (∼10 m–1000 m). In order to reduce the effect of errors caused by the coarse geolocation, we aggregated our estimated forest AGB map and the field data to 0.01 degrees × 0.01 degrees grid cells. 15 grid cells aggregated from 29 field sites were used to validate the four AGB maps. The Su map had the strongest correlation with the field data from NSTI, which can be attributed to the incorporation of part of the field data in their training dataset (so this is not an independent test of the Su map). Our map demonstrated the second best correlation with this field data, with an R^2 of 0.57 and an RMSE of 103.2 Mg ha^{-1}. All four maps showed relatively high RMSE values when compared to the field data (Fig. 2.11). This can be attributed to the limited coverage of field measurements within the 0.01 × 0.01 degrees grid cell. In fact, even within the grid cell with the highest number of field plots, the area of the field measurements cover ∼1% of the total area of the grid cell.



(a) This study

(b) GEOCARBON map



(c) Su map

(d) CCI map

Figure 2.11: Assessment of AGB map against independent field data from NSTI. The dotted line is the 1:1 line.

The field data from NSTI had a large range of forest AGB, with 50% of the field data between 80 Mg ha^{-1} and 230 Mg ha^{-1} (Fig. 2.12). The field data had the largest median and mean AGB. 50% of our map had an AGB of 100–150 Mg ha^{-1}, and both of median and mean AGB were around 120 Mg ha^{-1}. Both of GEOCARBON map and CCI map had 50% AGB between 40 Mg ha^{-1} and 100 Mg ha^{-1}. The two maps had similar mean AGB of 80 Mg ha^{-1}, but the median AGB of GEOCARBON map is lower. The Su map had the second largest median AGB and mean AGB of 140 Mg ha^{-1} and 130 Mg ha^{-1}, respectively.

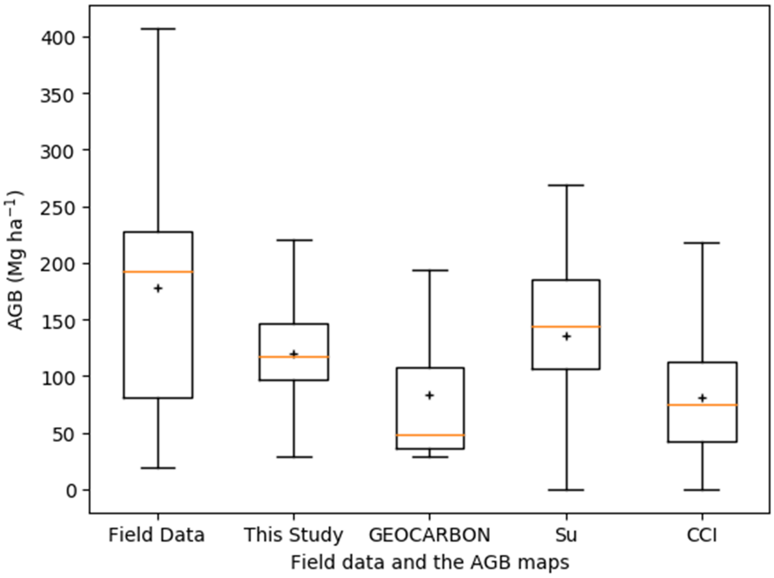


Figure 2.12: Boxplot of field data and forest AGB of the four AGB maps within 1 km × 1 km grid cells of field sites. The orange horizontal bar indicates the median and the boxes show the interquartile range, and the "+" indicates the mean forest AGB.

2.4.2 Comparison to existing AGB maps

The bias-corrected AGB map from this study was compared with AGB maps by Avitabile, et al.(Avitabile et al., 2014), Su, et al.(2016), and Santoro and Cartus (Santoro and Cartus, 2021*a*). Some of the maps are generated globally, while others are for specific smaller regions. We masked the non-forest area of Santoro and Cartus (Santoro and Cartus, 2021*a*) using the Hansen forest cover map in 2007 (Hansen et al., 2013), as this map expresses woody biomass and we focused on forest AGB in this study. In order to compare the maps at pixel level, we reprojected and resampled the AGB maps to the same projection and resolution as our map using a nearest neighbour resampling method, avoiding uncertainty which can be introduced during warping, and appropriate as our map is much higher resolution than the maps compared. Forest AGB density were converted to above-ground carbon using a default conversion factor of 0.47 (Change, 2006). We compared the AGB maps across China and for three specific regions: southern, northeastern and middle part of China (Fig. 2.13). These regions were chosen due to their extensive forest coverage and were derived by combining small vegetation zones (Hou, 2001). Compared to the three AGB maps, our results show large differences in total carbon and the distribution of AGB, in particular in southern China, where there are dense forest and steep slopes which introduce large uncertainties in AGB estimation.

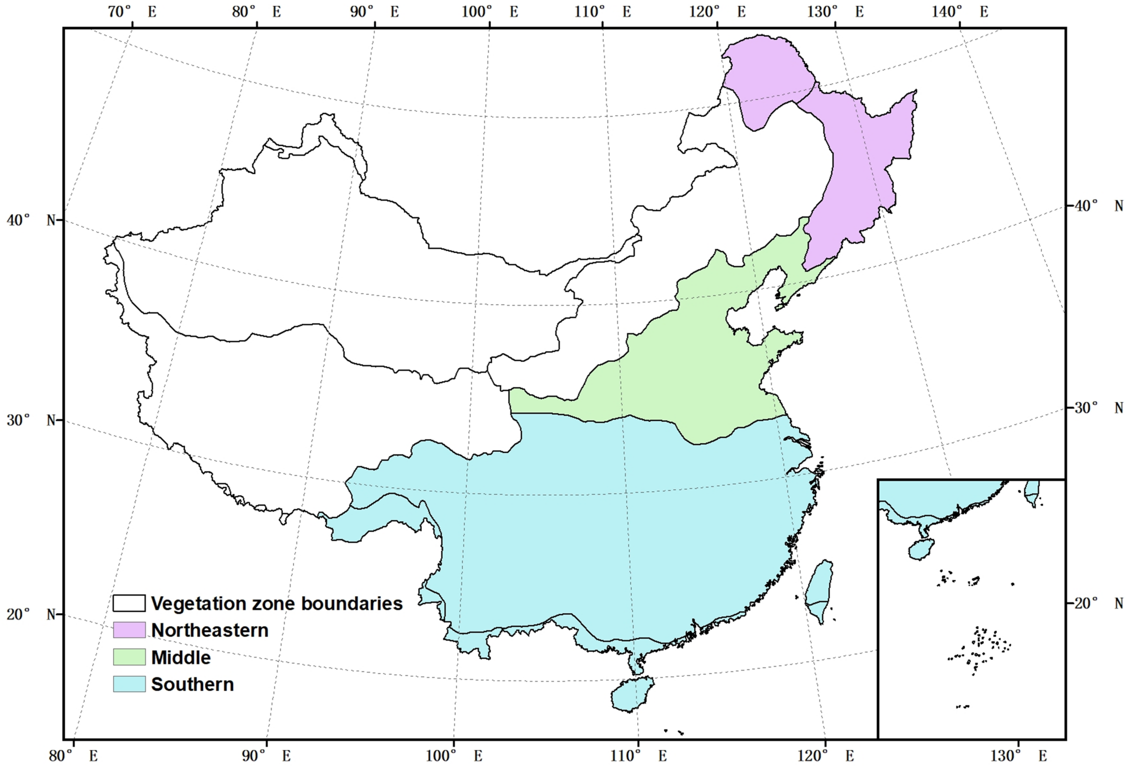
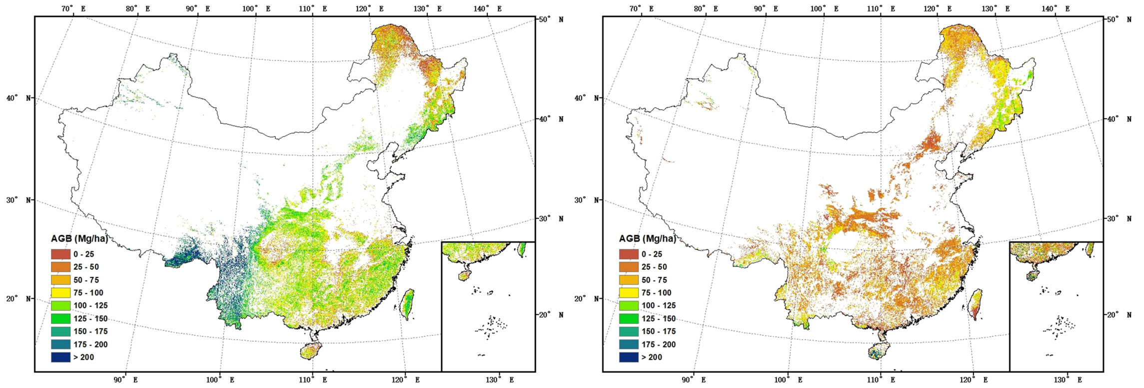


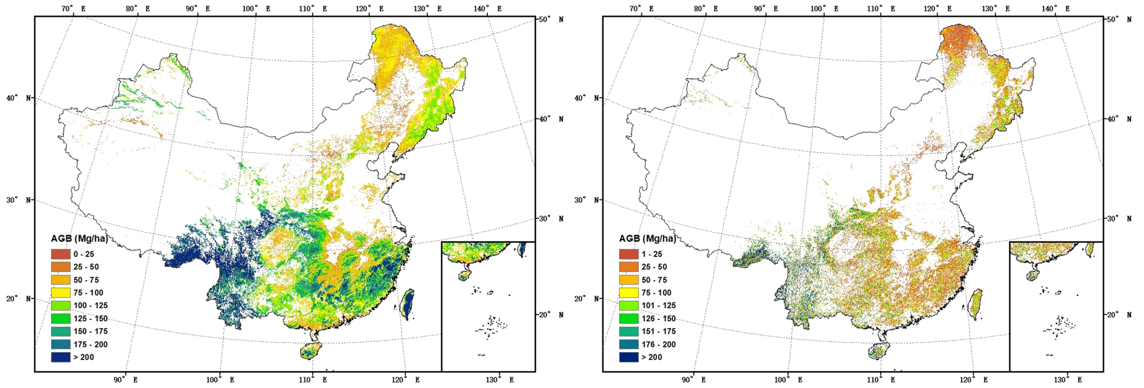
Figure 2.13: Forest zones we used to compare the forest AGB maps.

Across China

All four forest AGB maps showed considerable higher AGB in southwestern China, and the lowest AGB in northern China. However, significant differences were visible when the four AGB maps were compared (Fig. 2.14).



(a) This study (b) GEOCARBON map



(c) Su map (d) CCI map

Figure 2.14: Spatial distribution of forest AGB across China.

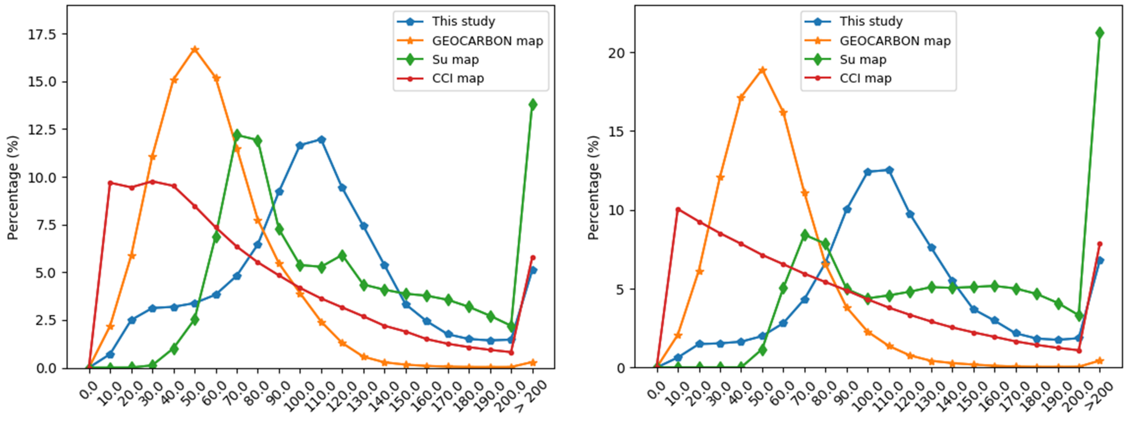
Among the four AGB maps, average AGB from the Su map 2016 was higher than the other three maps, while the average AGB from this study was the second highest and differed greatly from the values from GEOCARBON map (Avitabile et al., 2014) and CCI Biomass map (Santoro and Cartus, 2021*a*).

Table 2.3: Average AGB, forest area and total carbon in China

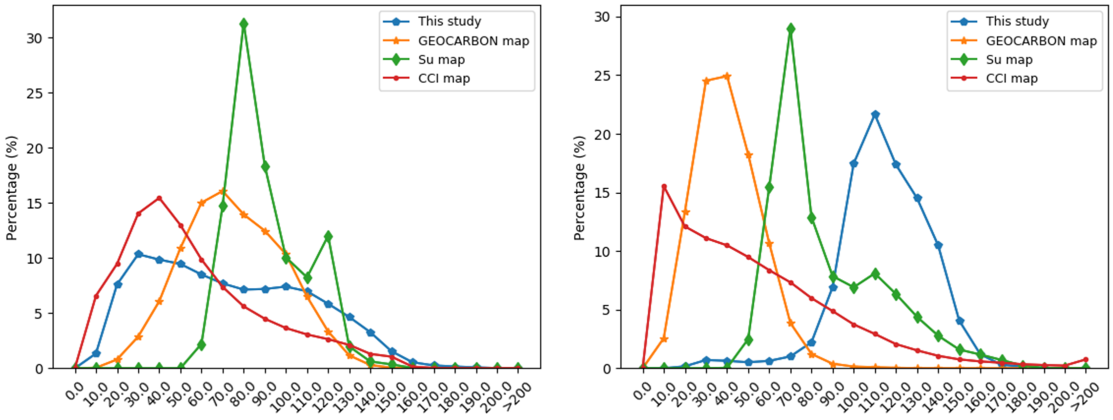
Maps	Year	Resolution	Average AGB ($\mathrm{Mg\ ha^{-1}}$)	Forest area (10^6 ha)	Total carbon (Pg C)
This study	∼ 2007	50 m	104.10	194.53	9.52
GEOCARBON map	∼ 2000 – 2010	0.01 degrees (∼ 1000 m)	53.10	157.03	3.92
Su map	2004	1000 m	121.93	297.03	17.02
CCI map	2010	100 m	73.77	171.24	5.94

Some of the differences in average AGB and total carbon storage in Table 2.3, may relate to the differences in the total area of forest used in the calculations. According to assessment report by the Food and Agriculture Organization (FAO), the forest area in China was 193.0 million ha in 2005 (Food and Agriculture Organization of the United Nations (FAO), 2010). The forest area from this study, which used the Hansen forest cover map (Hansen et al., 2013) to mask the non-forest area, was close to the forest area by FAO. The explanation for a smaller forest area from CCI map (Santoro and Cartus, 2021*a*), which we also used the Hansen forest cover map (Hansen et al., 2013) to mask non-forest area, was that we only counted the pixel values higher than zero as forest, leading a small difference in the forest area of the two maps. The GEOCARBON map (Avitabile et al., 2014) and Su map (2016) showed significant differences in the estimated forest area compared to this study. This is because the two AGB maps used different forest cover maps, the GLC2000 map (Bartholome and Belward, 2005) and the 2000 land use map (Jiyuan et al., 2002). The total above-ground carbon stocks for China were estimated by summing the carbon of all pixels in each AGB map. The total carbon stored in China's forests was estimated 9.52 Pg C in this study, which was almost two and a half times as much as that in GEOCARBON map (Avitabile et al., 2014), 60% higher than that in CCI map (Santoro and Cartus, 2021*a*), but less than 60% of that in Su map (2016) (Table 2.3).

The histograms of the four AGB maps across China in 10 Mg ha^{-1} bins showed that the peak of frequency of AGB in CCI Biomass (Santoro and Cartus, 2021*a*) occurred at a lower AGB of 0 to 40 Mg ha^{-1}, while the peak value in Su map was the highest (Fig. 2.15). The frequency of AGB in this study peaked at 100 to 110 Mg ha^{-1}, and more than half of the AGB ranged from 70 to 130 Mg ha^{-1}. More than 80% of AGB in GEOCARBON map (Avitabile et al., 2014) ranged from 20 to 90 Mg ha^{-1}, with nearly no AGB higher than 150 Mg ha^{-1}. The AGB in CCI Biomass map (Santoro and Cartus, 2021*a*) had no AGB in 10 Mg ha^{-1} bins greater than 10%, and the frequency of AGB decreased from 40 Mg ha^{-1}, and increased at the >200 Mg ha^{-1} interval. The AGB histogram for southern China is similar to that of the entire country, with the exception that the Su map contains a greater frequency of AGB values exceeding 200 Mg ha^{-1}. In the northeastern and middle China, the AGB is predominantly less than 150 Mg ha^{-1}.



(a) Across China (b) Southern China



(c) Northeastern China (d) Middle China

Figure 2.15: Histogram of AGB in 10 Mg ha^{-1} bins, 10 indicts 0–10 bin.

Based on the assessment of the field data above and comparisons with existing AGB maps, we would recommend that this new dataset represents the best estimate of the total carbon storage, and its distribution, for China in this time period.

Chapter 3

Mapping AGB in two contrasting regions

Abstract

Large-scale high spatial resolution AGB maps play a crucial role in determining forest carbon stocks and how they are changing, which is instrumental in understanding the global carbon cycle, and implementing policy to mitigate climate change. The advent of the new spaceborne Light Detection and Ranging (LiDAR) sensor, NASA's Global Ecosystem Dynamics Investigation (GEDI) instrument, provides unparalleled possibilities for the accurate estimation of forest AGB at high resolution, particularly in dense and tall forests, where SAR and optical data exhibit saturation. However, GEDI is a sampling instrument, collecting disperate footprints, and its data must be combined with that from other continuous cover satellites to create maps, using local machine learning methods. In this study, we linked GEDI data with forest AGB using unique field data co-located with GEDI footprints. In the northeast of China, we converted RH98 to field AGB, while in the mountainous southwestern region, a closer relationship was observed between RH80 and field AGB. We applied an ensemble model of decision trees (Light Gradient Boosting Machine, LightGBM), and random forest regression to generate wall-to-wall AGB maps at 25 m resolution, using extensive GEDI footprints as well as Sentinel-1 C-band data, ALOS-2 PALSAR-2 L-band data and Sentinel-2 optical data. We found many GEDI footprints had errors not detected by their quality flags; filtering using remote sensing data to find realistic possible values improved results. Through a 5-fold

cross-validation, LightGBM demonstrated a slightly better performance than Random Forest across two contrasting regions. In the northeastern region, R^2 of the two models is both 0.70 and the root-mean-square-error (RMSE) ranges between 28.44 and 28.77 Mg ha^{-1}. Meanwhile, in the southwestern region, the R^2 values range from 0.60 to 0.62, and the RMSE values range between 37.71 to 38.29 Mg ha^{-1}. However, in both regions, the computation speed of LightGBM is substantially faster than that of the random forest model, requiring roughly one-third of the time to compute on the same hardware. We found in both locations an increase in error as slope increased. In the northeastern region, the R^2 values decreased from 0.93 in flat areas to 0.42 in areas with >30 degrees slope, while the RMSE increased from 14 to 38 Mg ha^{-1}. In the southwestern region, the R^2 values decreased from 0.75 in flat areas to 0.50 in areas with >30 degrees slope, while the RMSE increased from 20 to 44 Mg ha^{-1}. The trained models were tested on nearby but different regions and exhibited good performance. This study not only proposes an approach to map forest AGB using extensive GEDI and continuous SAR, optical data in China but also enables the estimation of AGB changes using the trained model, even in the absence of GEDI but presence of SAR and optical data.

3.1 Introduction

Forests play an essential role in the health and sustainability of the planet's ecosystems (Watson et al., 2018). Accurate information on forest AGB is critical for effective forest management, carbon accounting, and climate change mitigation efforts (Xu et al., 2021; Houghton et al., 2009). Traditional forest inventory methods for estimating forest biomass are labor-intensive and time-consuming, making them impractical for large-scale monitoring and assessment. Approaches based on earth observations have emerged as a promising alternative for estimating forest AGB (Mitchard, 2018; Tucker et al., 2023), offering the potential to rapidly and consistently map AGB over large areas. Optical, SAR, and LiDAR are the three main types of earth observation data.

Optical data measures light reflected from the Earth's surface, and is widely used due to the ability to capture detailed spectral information, which can be used to assess canopy cover and potentially estimates of the density. Optical data provides the longest temporal record, with the series of Landsat that commenced data collection in 1972. It has traditionally been an efficient tool to map wall-to-wall forest AGB by combining with field data through regression models (McNicol et al., 2018; Lu et al., 2004; Zhu and Liu, 2015; H. Nguyen et al., 2019; Zheng et al., 2004). However, optical data can only see the top of the forest canopy, meaning similar canopy with different heights or trunk widths, are not be distinguished well. Additionally, optical data's utility can be significantly hampered by atmospheric conditions such as clouds and haze.

SAR is a type of active sensor that transmits its own energy and subsequently records the quantity of that energy reflected back (Meyer, 2019). The long wavelengths of the SAR sensors used for biomass mapping (C, L, and longer wavelengths if available) enables it to see through clouds, which ensures consistent data acquisition. The continuous, reliable, all-weather, day-and-night imaging capabilities of SAR make it particularly useful for forest AGB estimation (Mitchard et al., 2009). SAR can further penetrate through the canopy and interact with leaves, branches and trunks, and thus SAR is sensitive to forest AGB (Balzter et al., 2007). SAR data which has longer wavelength tends to have a higher saturation point than those with shorter wavelengths (Woodhouse et al., 2012; Sinha et al., 2015). However, even the L-band SAR, which has the longest wavelength among currently available space-borne SAR data, has a relatively low saturation point for AGB at around 100-150 Mg ha^{-1} (Mitchard et al., 2009; Englhart et al., 2011; Mitchard et al., 2012).

LiDAR data employs laser light to estimate tree height and structure, thereby facilitating the construction of a detailed representation of a forest. This technology allows for the direct estimation of 3D structural parameters such as height, stem density, and canopy cover, providing invaluable insights into forest characteristics and dynamics. Therefore, LiDAR suffers little from saturation effects (Wulder et al., 2012; Duncanson et al., 2022). With the launch of GEDI, a LiDAR instrument designed to measure vegetation structure (Dubayah et al., 2020), it is now possible to quantify forest AGB in distributed 25m diameter footprints over very large areas. While this represents an improvement over the previous spaceborne LiDAR ICESat with its 65 m footprints, it's important to note that GEDI's 25m resolution is still relatively coarse compared to airborne LiDAR systems. Despite this limitation, GEDI provides valuable data for large-scale forest structure analysis, albeit with inherent uncertainties and potential errors (Duncanson et al., 2022; Potapov et al., 2021).

However, useful instruments do not necessarily produce accurate forest AGB maps. China encompasses diverse forest cover types and complex terrain, making it a challenging and highly variable environment for AGB estimation (Dai et al., 2011). China is among the countries with the highest degree of inaccuracy when it comes to estimating AGB using GEDI data. The AGB estimated by GEDI for China is more than double the estimates provided by the FAO (Dubayah et al., 2022) and other studies (Su et al., 2016; Huang et al., 2019; Chang et al., 2021). One potential reason for this is the undefined relationship between GEDI relative height (RH) metrics and forest AGB. Forest biomass estimation using remote sensing relies on the correlation between remotely sensed data and ground-based biomass measurements, which vary greatly depending on the forest type and terrain conditions such as elevation and slope. In this study, we collected new field plots co-located with GEDI footprints to develop relationships to convert GEDI RH metrics into forest AGB.

However, unlike SAR and optical data, GEDI data are not continuous. The discontinuous LiDAR data is often combined with optical and SAR data to generate wall-to-wall estimates (Guo et al., 2023; Chen, Ren, Bao, Zhang, Wang, Liu, Man and Liu, 2022; Silva et al., 2021). Their combined use can provide a more comprehensive and accurate estimation of forest carbon. Random forest (Breiman, 2001) is one of the most popular algorithm to extrapolate LiDAR footprints to wall-to-wall map using optical and SAR imagery (Baccini et al., 2012; Huang et al., 2019; Narine et al., 2020; Ngo et al., 2023; Yao et al., 2024).

In this study, we used random forest as well as a more advanced machine learning algorithm, LightGBM (Ke et al., 2017), to generate wall-to-wall AGB maps with a resolution of 25 m. LightGBM is a gradient boosting ensemble method that is known for its efficiency and ability to handle large-scale data (Yan et al., 2021). However, the universality of the two models across different different regions is not yet fully understood. We tested the performance of the two models in two contrasting regions, and validated the results using the field data. In addition, the trained models were applied to respective areas. The remote sensing data used in the models were from GEDI, Sentinel-1, Sentinel-2, and ALOS-2, all of which are open-source. We aim to propose a universally applicable method for estimating AGB in China, or if necessary demonstrate why different methods might be needed in different regions.

3.2 Data and methods

3.2.1 Study area

This study was conducted in two contrasting areas of China: one in northeastern China, and the other one in southwestern China (Fig. 3.1). Each of the study areas covers an area of 500 x 500 km, but the types of forests and terrain are completely different. The forest types differ in terms of their structure, species composition, and environmental conditions, providing an ideal opportunity to test the universality of remote sensing-based forest biomass mapping approaches which is of great significance for achieving large-scale forest AGB predictions in areas with complex forest compositions.

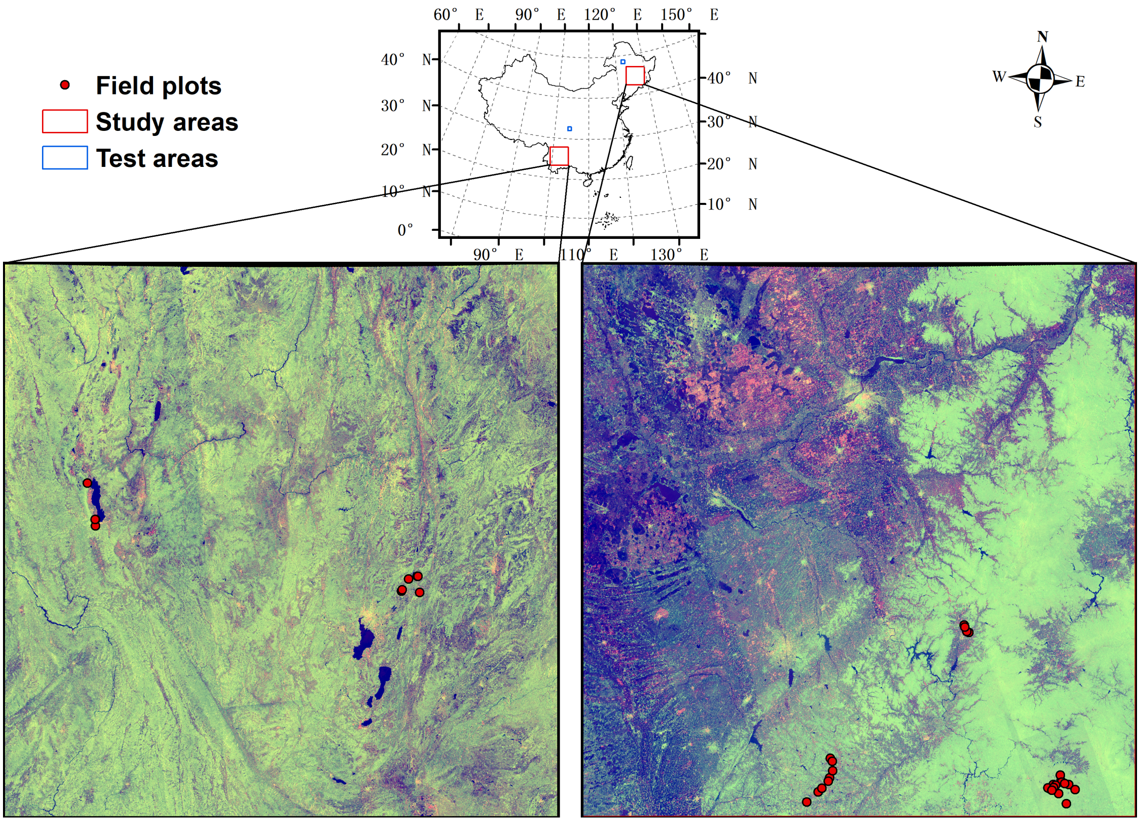


Figure 3.1: Location of the study area and field sites. The field plots are shown on the composited ALOS PALSAR-2 mosaic (HH in red, HV in green, and HV/HH in blue).

The northeastern study area, primarily located in Jilin Province as well as the southern part of Heilongjiang Province, is characterized by the temperate forest. The area experiences a continental monsoon climate, with long, cold winters and short, warm summers (Liu, Shan, Shu, Sun and Du, 2018). The average temperatures in January and July are -20 and 21 °C, respectively. The average annual precipitation ranges from 600 to 1100 mm (Dong et al., 2005). The forests are dominated by Korean pine (Pinus koraiensis) mixed with deciduous species (e.g. Betula platyphylla) (Wang, 2006) (Fig. 3.2). A portion of the primary forest has been destroyed by large-scale industrial logging, and replaced by secondary forests and plantations (Yu et al., 2011), which mainly consist of Betula platyphylla, Larix gmelinii, and Pinus koraiensis (Shi et al., 2015).

The southwestern study area is mainly situated in Yunnan Province. Forests in this region predominantly blanket the mountainous terrain, as this area is dominated by plateau and mountainous areas, with elevations ranging from 285 to 5350 m, and an average elevation of 2039 m (Digital Elevation Model). The average temperature fluctuates between roughly 9–11 °C in January and climbs to around 22 °C in July, while the average annual precipitation varies from 1100 to 1600 mm (Shi and Chen, 2018). The natural conditions described above contribute to the richness and diversity of Yunnan's forest resources, including needleleaf forest, subtropical evergreen broad-leaved forests and deciduous broadleaf forest (Zhang, Wang and Liu, 2019; Zhu et al., 2022; Frayer et al., 2014).

Figure 3.2: Field photos showing different forest types and terrains in the two study areas.

3.2.2 Field data of AGB

The field measurements were carried out during October and December 2021. We measured 26 plots in northeastern region with 24 plots were under GEDI footprints, and 16 plots in southwestern region with 12 plots were under GEDI footprints. In order to align with the GEDI data, all field plots were circular with a diameter of 25 m. We endeavored to maintain a substantial distance between field points (typically greater than 2 km) and ensured they were at least 50 meters away from non-forest areas. This strategy was adopted to avoid issues related to mixed pixels.

After identifying the GEDI footprints to be measured (Fig. 3.1), we navigated to the center of each footprint using a Garmin eTrex 20x handheld GPS device. To improve the accuracy of our measurements, we averaged the GPS readings over a period of 5 minutes. A tape was then used to delineate a circular plot, centered on the core of the footprint, with a diameter of 25 meters. Subsequently, we measured the height and DBH of trees within the plot that have a DBH $\geq$ 5 cm. This methodology ensures a comprehensive and accurate assessment of the field plot. DBH was measured at 1.3 m up the trunk from the ground using a diameter tape. For the majority of the field plots, tree height was estimated using a Vertex IV with a Transponder T3. However, for two plots in the northeastern region, the tree height was measured using a clinometer. The field AGB of individual tree was then obtained using tree height and DBH, calculated via the following allometric equation (Luo et al., 2018).

$$W = 0.1355 * (D^2 * H)^{0.817} \tag{3.1}$$

where, W is the AGB of each tree (kg), D and H are DBH (cm) and height (m) of the tree. We sum the AGB of all trees within the plot that have a DBH $\geq$ 5cm. This total AGB was then converted to a per hectare basis for standardization, expressed in Megagrams per hectare (Mg ha^{-1}).

3.2.3 LiDAR data

Canopy height metrics from GEDI

The GEDI instrument captures the waveform of the light that has bounced back from the Earth's surface to the GEDI sensor on the International Space Station, with a 25 m spatial footprint every 60 m along orbits. The GEDI instrument is equipped with three lasers. Out of these, two lasers operate at full power while the third one is divided into two coverage beams. These four beams are dithered across-track, creating a total of eight ground tracks of data, comprising four full-power beams and four coverage beams (Dubayah et al., 2020).

The dataset essentially presents the distribution of Earth's surface features, like vegetation and terrain, within the footprint of a laser shot. Specifically, it includes details on the vertical structure of the features, which is of great importance when assessing characteristics like canopy height or the vertical distribution of foliage in forests (Dubayah et al., 2020). In this study, L2A dataset acquired from 2019 to 2022 was used to produce continuous forest AGB map.

Airborne laser scanning data

Airborne laser scanning (ALS) are capable of collecting data on vegetation structure, which can potentially serve as a surrogate for traditional field measurements (Potapov et al., 2021). We employed ALS point clouds, collected in the summer of 2022, with a point density of 18 points/m^2 in our research. The ALS data consists two areas, the first area of 18 ha and the other area of 22 ha, co-located with 11 GEDI footprints respectively.

3.2.4 SAR data

To obtain as much information as possible on forest structure, we utilized both C-band and L-band SAR in this study.

Sentinel-1 C-band data

Sentinel-1 Interferometric Wide swath (IW) product acquired during 2020 and 2022 was used in the estimation. Sentinel-1 was accessed and processed in GEE, powered by Google

Cloud Platform. The Sentinel-1 Ground Range Detected (GRD) data had been calibrated and terrain-corrected, and the values were converted to decibels.

The backscatter coefficient of VV (vertical-vertical) and VH (vertical-horizontal) polarizations was exploited, as different polarizations carries information about the structure of the forest (Flores-Anderson et al., 2019). Their cross ratio (VV/VH) was also included, as the cross ratio of Sentinel-1 is likely to mitigate errors associated with the acquisition system (Veloso et al., 2017) and has been demonstrated to effectively differentiate between vegetation densities (Vreugdenhil et al., 2018). To reduce the impact of temporal variations and anomalies that might occur in individual images, we calculated temporal means of the data. To further reduce the speckle noise, we applied a focal mean speckle filter with a radius of 50 m to the data.

ALOS-2 PALSAR-2 L-band data

The annual mosaics of L-band SAR data from ALOS-2 PALSAR-2 was utilized in this study. The 25 m resolution yearly mosaic dataset was created by the JAXA, using data acquired in the target year. The ortho-rectified and slope corrected data was accessed in GEE, and we also applied a focal mean speckle filter with a radius of 50 m to the means of annual mosaic of 2020 to 2022 to reduce the impact of speckle noise. The DN of HV and HH were then converted to gamma naught values in decibel using the the following equation (Shimada et al., 2009).

$$\gamma_0 = 10\log_{10}(\mathrm{DN}^2) - 83.0\,\mathrm{dB} \tag{3.2}$$

where, γ_0 is the backscatter coefficient in dB. In addition to the both polarizations, we also prepared the ratio (HV/HH) data, which has been identified as effective strategies for enhancing the saturation point (Hayashi et al., 2019; Sarker et al., 2012).

3.2.5 Optical data

Optical data were also used to extrapolate LiDAR footprints to wall-to-wall image. The MultiSpectral Instrument (MSI) on board the Sentinel-2 satellite measures reflected radiance across 13 spectral bands. These bands span a range from the Visible and Near Infrared (VNIR) to the Short Wave Infrared (SWIR) spectra. This broad coverage enables detailed observations of various surface conditions and phenomena. The ortho-rectified and atmospherically corrected L2A surface reflectance (SR) dataset, excluding band 10, is publicly accessible in GEE. In addition to utilizing the mean values of each band of L2A imagery acquired between 2020 and 2022, we also computed the mean and maximum values of the NDVI, which provide valuable information regarding the average and peak level of vegetation growth.

3.2.6 Topographic data

The NASA Digital Elevation Model (NASADEM) is a global digital elevation model with a resolution of 30 m. NASADEM combines data from different satellite missions, including the SRTM, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and ICESat, to generate a seamless global elevation model (JPL, 2020). In this study, we utilized the terrain slope derived from NASADEM as an additional variable for analysis. The terrain slope was computed using a 3 × 3 moving window approach. The integration of terrain slope information provided valuable insights into the topographic characteristics of the study areas, enhancing the impact of topographic variations on the AGB predictions.

3.2.7 Forest AGB estimation

GEDI data filtering

Before proceeding with model fitting, we implemented multiple filters to identify and remove poor-quality footprints. L2A footprints collected from 2019 to 2022 were filtered using the following criteria (Dubayah et al., 2021).

(a) Only shots that have a quality_flag of 1 were used. This flag signifies that the shot meets a summation of energy level, sensitivity, amplitude, quality of real-time surface tracking, and the difference from DEM.

(b) Shots acquired during degrade period were excluded.

(c) Only shots with a sensitivity > 0.98 were used. A higher sensitivity indicates a greater likelihood of penetrating to the ground beneath a specific level of canopy cover.

(d) Shots were included if they were acquired at night, when the negative impact of solar illumination on the quality of GEDI waveform is eliminated, thus yielding better tree height prediction results (Liu, Cheng and Chen, 2021).

After we filtered the GEDI footprints, we then compared GEDI L1B data with ALS data to verify the quality of GEDI data, and to eliminate poor-quality footprints in GEDI. In order to compare with GEDI footprints, we used ALS data to simulate GEDI observations with three ground-finding algorithms: the Gaussian fitting, the lowest inflection point and the lowest maximum (Hancock et al., 2019).The simulated GEDI observations from ALS data was then compared with actual GEDI footprints (fig. 3.3).

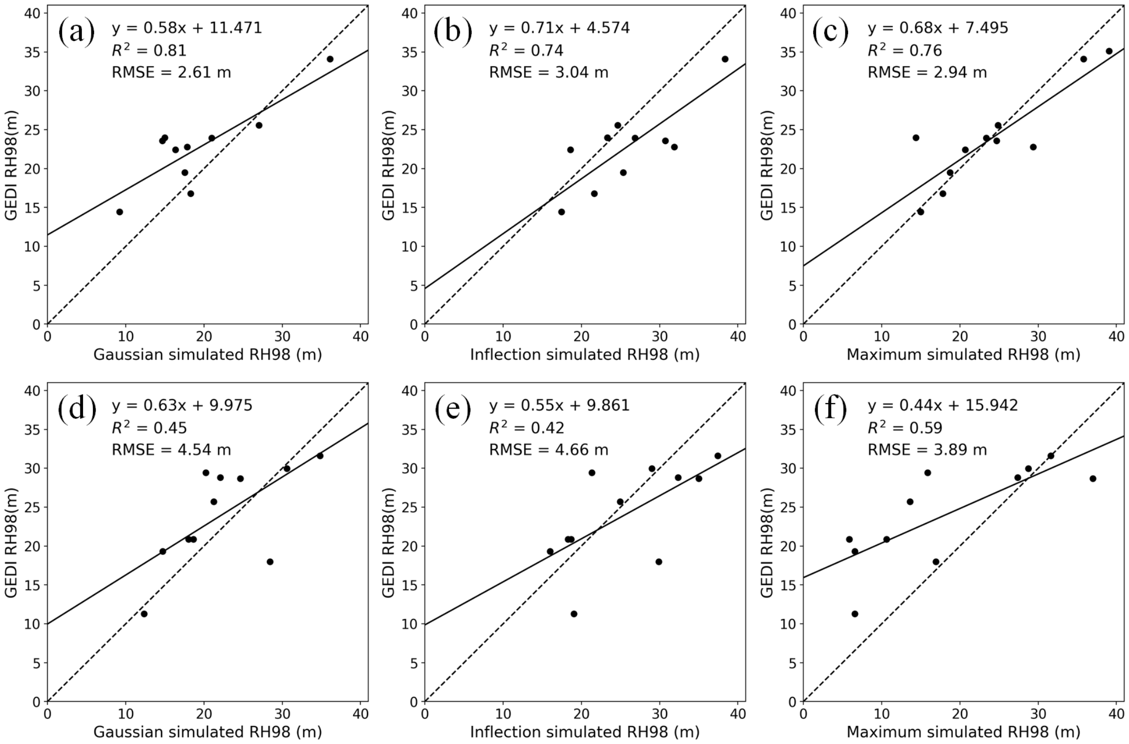


Figure 3.3: RH98 of actual GEDI against GEDI simulator. (a)-(c) is the scatter plot for the first area, all the footprints are comprised within full-power beam. (d)-(f) is the scatter plot for the second area, all the footprints are comprised within coverage power beam. The dotted line represents 1:1 line. The solid line represent the fitted line, and is extended to cover the entire range of the x-axis. This was done to provide a comprehensive visual representation of the relationships between GEDI RH metrics and ALS simulated RH metrics.

As expected, footprints of GEDI power beams have a higher relationship with ALS simulated footprints, with a higher R^2 and a smaller RMSE. Since this also had been proved in other study areas (Duncanson et al., 2020; Li et al., 2023; Liu, Cheng and Chen, 2021), we only included power beams in this study.

By comparing GEDI and ALS data, we also found that the geolocation errors (horizontal) of GEDI data in the China are substantial, with 17.9 m error for the first area and 9.8 m error for the second area. We noticed a significant difference in the geolocation error between the two areas, indicating that the geolocation error of some footprints is larger than others. This means that the RH metrics at the actual geolocation of some GEDI footprints may differ significantly from the RH metrics at the geolocation it shows, which could introduce large error into AGB predictions.

Because of this finding, we proceeded with an additional filtering process on the GEDI footprints using ALOS-2 PALSAR-2 HV, HH polarizations, and Sentinel-1 VH, VV polarizations, which have been proven to have a close relationship with forest height and AGB (Mitchard et al., 2012; Duncanson et al., 2020; Cartus et al., 2012; Ge et al., 2022; Santi et al., 2017). We fitted the relationship between GEDI RH98 and SAR imagery in the Northeast region. The reliability of the fitting relationships were assessed using R^2 and RMSE, calculated using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(SAR_i - f(RH_i))^2}{\sum_{i=1}^{n}(SAR_i - \overline{SAR})^2} \tag{3.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(SAR_i - f(RH_i))^2} \tag{3.4}$$

where SAR_i is the value of SAR imagery, $f(RH_i)$ is the fitted value, $overlineSAR$ is the average value of SAR_i. The R^2 and RMSE here can be influenced by the fitted function.

In northeastern region, cross-polarised SAR showed strong relationships with GEDI RH98 (Fig. 3.4a, c), while the relationships between co-polarised SAR and GEDI RH98 were not as expected (Fig. 3.4b, d). In contrast, in southwestern region, neither polarization showed a significant relationship with GEDI RH metrics, which might be substantially influenced by the steep slopes. There remains a substantial amount of GEDI data that does not meet the required quality standards. Generally, the backscatter of SAR from shorter trees is expected to be lower than that from taller trees. We therefore filtered the GEDI footprints in both regions using the relationships developed in northeastern regions, and only retained the GEDI data that fell within a range of plus and minus 3 from the fitted curve.

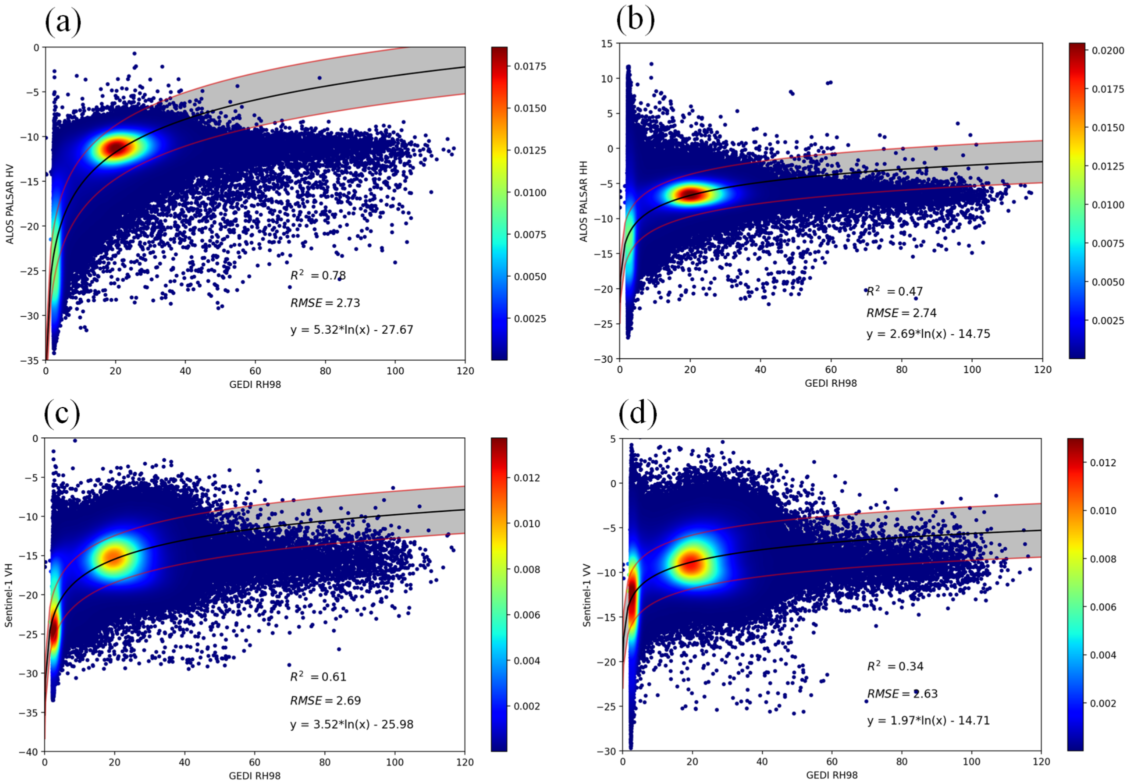


Figure 3.4: ALOS PALSAR-2 backscatter in (a) HV, (b) HH polarizations, and Sentinel-1 backscatter in (c) VH, (d) VV polarizations were plotted against GEDI RH98 for 1,405,451 GEDI footprints in the northeastern region. The black solid line represents the fitted line. The red solid line above the black solid line represents the fitted line plus 3, and the red solid line below the black solid line represents the fitted line minus 3. The color of each point in the scatter plot represents the estimated density of points at that location, calculated using a Gaussian kernel density estimate.

AGB of GEDI footprints

In order to obtain AGB values for the filtered GEDI footprints, we linked field AGB with GEDI RH metrics. We analysed the correlation between field AGB and RH values ranging from RH30 to RH98. In the northeastern region, the closest relationship was observed between field AGB and RH98, while in the mountainous southwestern region, field AGB exhibited a stronger correlation with RH80 (Fig. 3.5). We converted RH98 and RH80 to AGB using the fitted relationships for both regions, respectively.

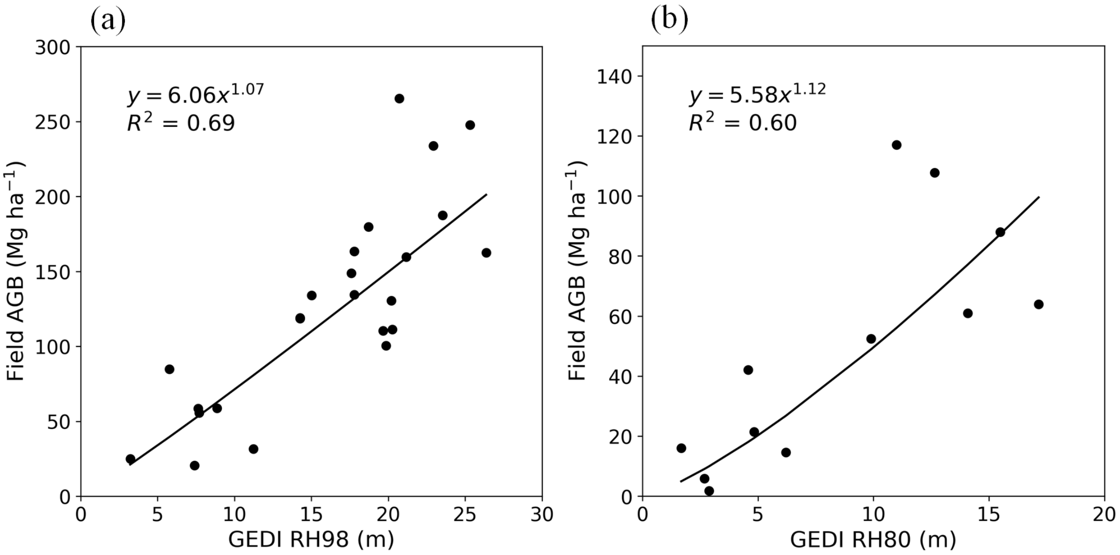


Figure 3.5: Field measured AGB are plotted against (a) RH98 in the northeastern region, and (b) RH80 in the southwestern region.

Accuracy assessment

We evaluated the model performance using a 5-fold cross-validation approach. The dataset was divided into five distinct subsets. Four subsets were employed for training, while the remaining one was used to test the model's accuracy. This procedure was repeated five times, ensuring each subset served as the test set once. Subsequent to these iterations, the results from all five test sets were aggregated to provide a comprehensive assessment of the model's accuracy.

The model performance and the accuracy of the results were evaluated using R^2 and RMSE, computed using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.6}$$

where y_i is the field measured AGB or the AGB derived from GEDI RH metrics, $\hat{y}_i$ is the model predicted value, $\bar{y}$ is the average value of y_i.

Wall-to-wall AGB map

In this study, we extrapolated the extensive GEDI footprints into continuous AGB maps using LightGBM and random forest in Python 3.8. A comprehensive set of 25 remote sensing layers were employed in the generation of wall-to-wall maps. These layers comprised a wide array of data, including bands and ratios from SAR, optical bands, the maximum and average NDVI values, data from the DEM and slope, along with gridded latitude and longitude information.

Using the 5-fold cross-validation, each trained model from the five iterations was employed to predict the AGB maps across the study area, resulting in five separate AGB maps for each of the two regions. To derive a robust and consolidated AGB map, we computed the mean of these five predictions for each pixel. Additionally, to assess the variability and reliability of our predictions, we calculated the standard deviation across the five results for each pixel, which was then utilized to generate an uncertainty map (Fig. 3.6).

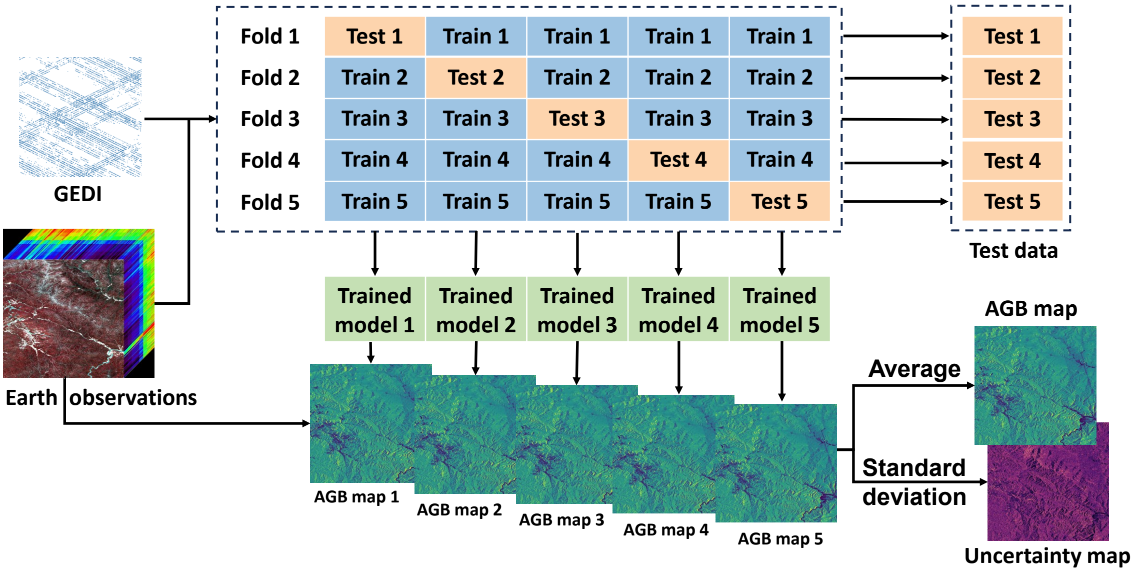


Figure 3.6: Framework for estimating AGB and its uncertainty.

To show the spatial distribution of forest AGB in both regions, the AGB maps were masked by a forest cover map in a 2021. The 2021 forest cover was derived from the Hansen et al forest map (Hansen et al., 2013), based on the 2000 forest cover map as a baseline. To generate the 2021 map, the degraded areas between 2000 and 2021 were excluded, and the forest areas that experienced an increase from 2000 to 2012 were added to the map (there is no information of forest gain after 2012).

3.3 Results

3.3.1 Model performance

Filtering GEDI data using parameters such as quality_flag, degrade, sensitivity, solar elevation (acquisition time), and beam intensity had been demonstrated to be an effective method for enhancing data quality. 1,405,451 and 2,042,330 GEDI footprints, filtered by the parameters above, were used to generate AGB maps for the northeastern and southwestern regions, respectively. The GEDI data was divided into training and test datasets at a ratio of 6:4. Scatter plots of the estimates compared to the test data are shown in Fig. 3.7, the variation of test data density is also illustrated. This initial analysis, which utilized GEDI footprints irrespective of GEDI RH metrics against SAR backscatter, indicated that both models exhibited strong performance in the northeastern region. However, neither model performed well in the southwestern region. There are many outliers in Fig. 4.6, which may be caused by geolocation uncertainties in the GEDI data (Fig. 3 and section 3.2.7), or steep slopes.

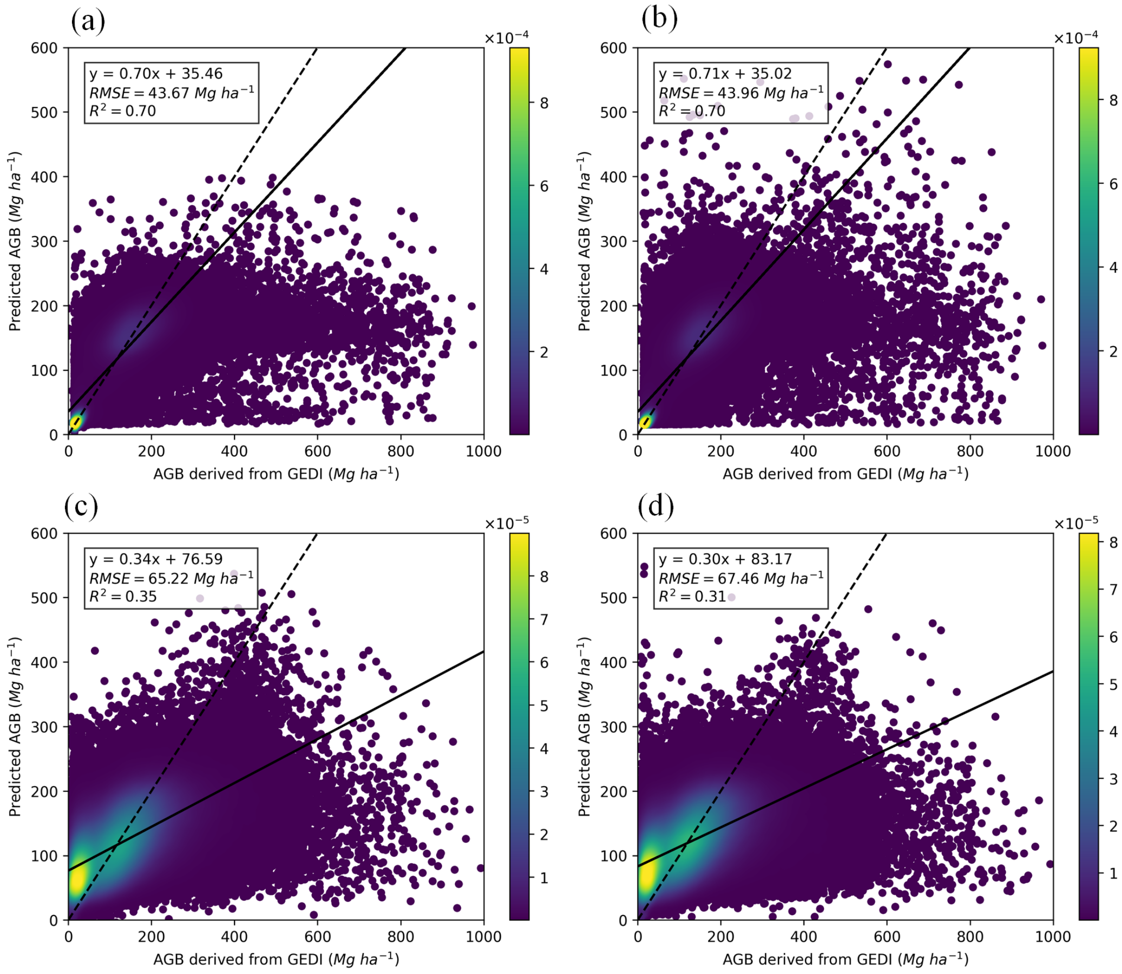


Figure 3.7: The predicted AGB values against AGB derived from GEDI RH metrics before filtering using earth observation data. (a) LightGBM in northeastern region, (b)random forest in northeastern region, (c) LightGBM in southwestern region, (d) random forest in southwestern region. The color of each point in the scatter plot represents the estimated density of points at that location.

We then utilized the GEDI footprints filtered using SAR data as described in section 3.2.7, with 689,013 and 673,744 footprints remaining for the northeastern and southwestern regions. We employed two different models to predict forest AGB. Predicted AGB from both models against test dataset of AGB derived from GEDI RH metrics are plotted in Fig. 3.8. The performance of both models in the northeastern region did not show significant improvement, except for a slight reduction in the RMSE and a notable reduction in the number of outliers. However, both models showed significant improvement in the southwestern region. The R^2 and RMSE of the LightGBM model improved from 0.35 and 65.22 Mg ha^{-1} to 0.62 and 37.71 Mg ha^{-1}, respectively. Similarly, the Random Forest model saw its R^2 and RMSE improve from 0.31 and 67.46 Mg ha^{-1} to 0.60 and 38.29 Mg ha^{-1}, respectively.

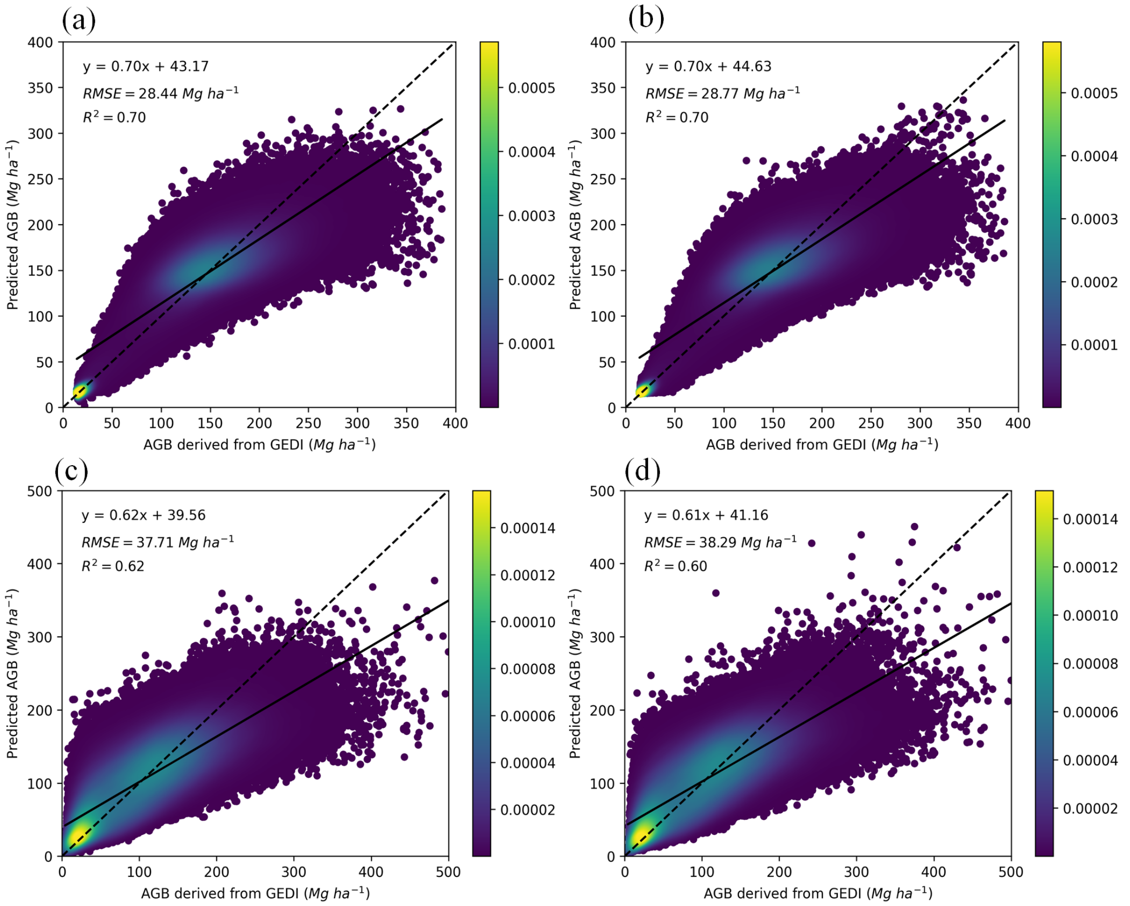


Figure 3.8: The predicted AGB values against AGB derived from GEDI RH metrics after filtering using earth observation data. (a) LightGBM in northeastern region, (b)random forest in northeastern region, (c) LightGBM in southwestern region, (d) random forest in southwestern region. The color of each point in the scatter plot represents the estimated density of points at that location.

Both models demonstrated similar accuracy in the two regions. However, LightGBM outperforms Random Forest in terms of computational speed. Table 3.1 presents the computational time required for training and prediction using the LightGBM and Random Forest models. The reduced computational time of LightGBM can be attributed to its optimized algorithm and efficient implementation. The notable difference in computational speed between LightGBM and Random Forest makes LightGBM a favorable choice for applications where efficient processing is crucial, especially when dealing with large-scale datasets.

Table 3.1. Computation time for LightGBM and random forest. NE represents the northeastern region, and SW represents the southwestern region.

Study area	Model	Number of GEDI footprints	Computation time (min)
NE	LightGBM	1,405,451	89
NE	random forest	1,405,451	233
NE	LightGBM	689,013	51
NE	random forest	689,013	145
SW	LightGBM	2,042,330	97
SW	random forest	2,042,330	329
SW	LightGBM	673,744	69
SW	random forest	673,744	187

3.3.2 Accuracy of the AGB maps

The accuracy of the AGB maps produced by both models was evaluated using 26 and 16 plots in the northeastern and southwest regions, respectively. Fig. 3.9 shows the scatter plot of field AGB and estimated AGB using filtered GEDI footprints. The estimated AGB from LightGBM and random forest correspond well with the field AGB in both regions. The R^2 values of estimated AGB from LightGBM are higher than those from random forest. In the northeastern region, LightGBM achieves an R^2 of 0.43, while in the southwestern region, it obtains an R^2 of 0.48. On the other hand, random forest yields an R^2 of 0.33 in the northeastern region and 0.55 in the southwestern region. The RMSE of both models in the two regions is very close, with a difference within 3 Mg ha^{-1}. Compared to the field data, both models have overestimated biomass in low biomass areas and underestimated biomass in high biomass areas, as is common in biomass mapping using machine learning (Shendryk, 2022; Huang et al., 2019). This trend is also evident when evaluating model performance using AGB derived from the GEDI footprints (Fig. 3.8).

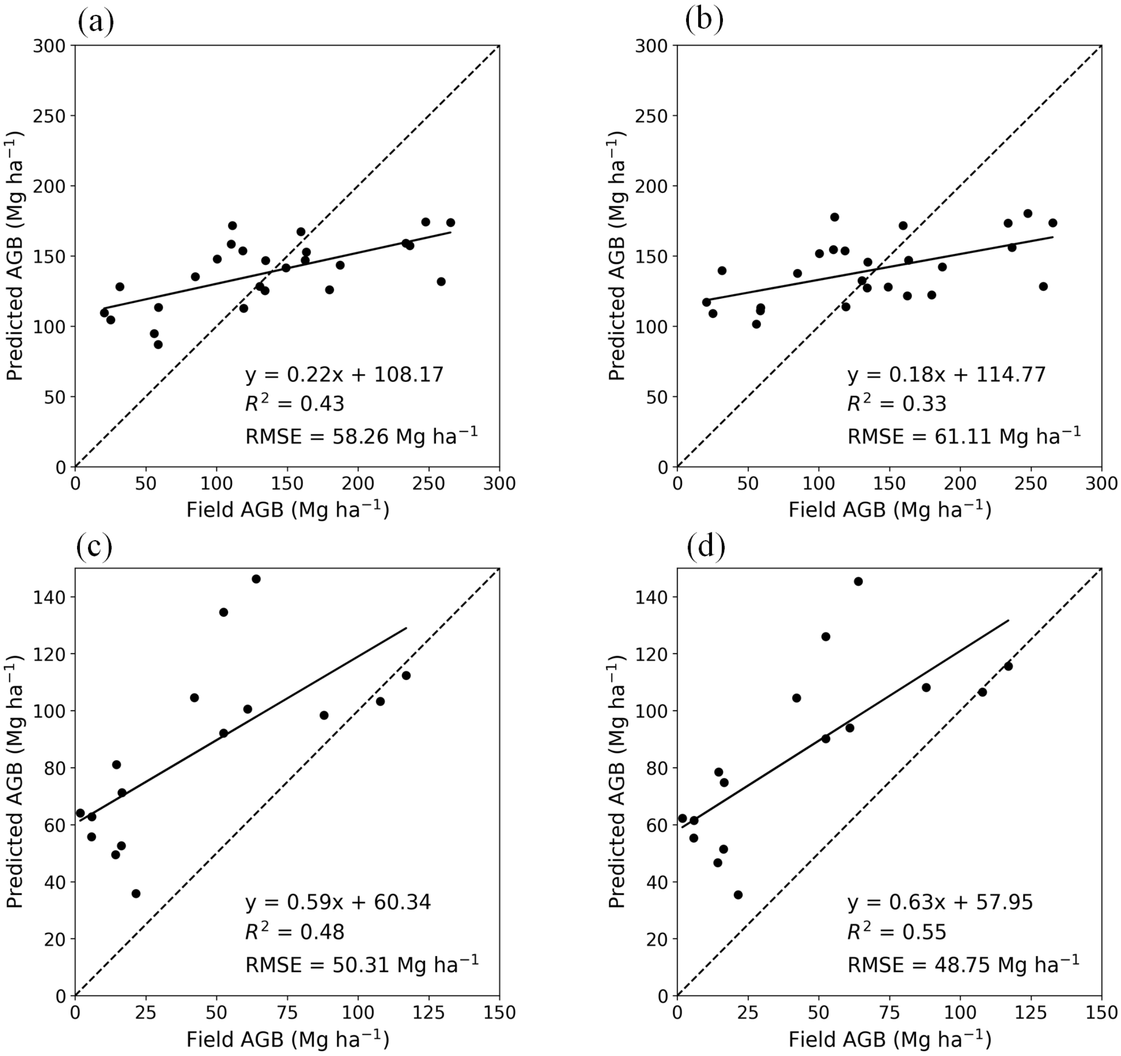


Figure 3.9: Scatter plot between predicted AGB and field measured AGB. (a) LightGBM in the northeastern region, (b)random forest in the northeastern region, (c) LightGBM in the southwestern region, (d) random forest in the southwestern region. The solid line represent the fitted line, and the dotted line represents 1:1 line.

3.3.3 Spatial distribution of AGB in both regions

The wall-to-wall forest AGB maps by LightGBM and random forest in both regions are shown in Fig. 3.10. In the northeastern region, high AGB values are predominantly concentrated in the eastern part of the study area. Conversely, in the southwestern region, the distribution of AGB exhibits a more dispersed pattern. In both the northeast and southwest regions, the spatial distribution patterns of AGB predicted by the LightGBM and Random Forest models show strong agreement, perhaps unsurprisingly given they are based on the same Earth Observation data inputs. This is further evidenced in the histograms, where both models generate similar overall AGB distributions 3.11. In the northeastern region, the mean difference in AGB estimates between LightGBM and Random Forest was -1.59 Mg ha^{-1}. Similarly, in the southwestern region, the mean AGB estimate difference was -1.44 Mg ha^{-1}. Although the overall spatial distribution patterns show strong agreement, difference maps in Fig. 3.10 reveal there are still discernible differences between the LightGBM and Random Forest AGB predictions. To better illustrate these spatial differences at a finer scale, zoomed-in 2500 m x 2500 m maps are presented in Fig. 3.12. Focusing on this local scale enables better visualization and analysis of the details and spatial heterogeneity in the AGB distribution. The AGB maps generated by the Random Forest model exhibit a deeper characterization of extreme values (Fig. 3.12b, e), both at the high and low ends of the AGB range. In contrast, the AGB maps generated by the LightGBM model display a relatively smoother transition between different AGB values (Fig. 3.12a, d), with more gradual change.

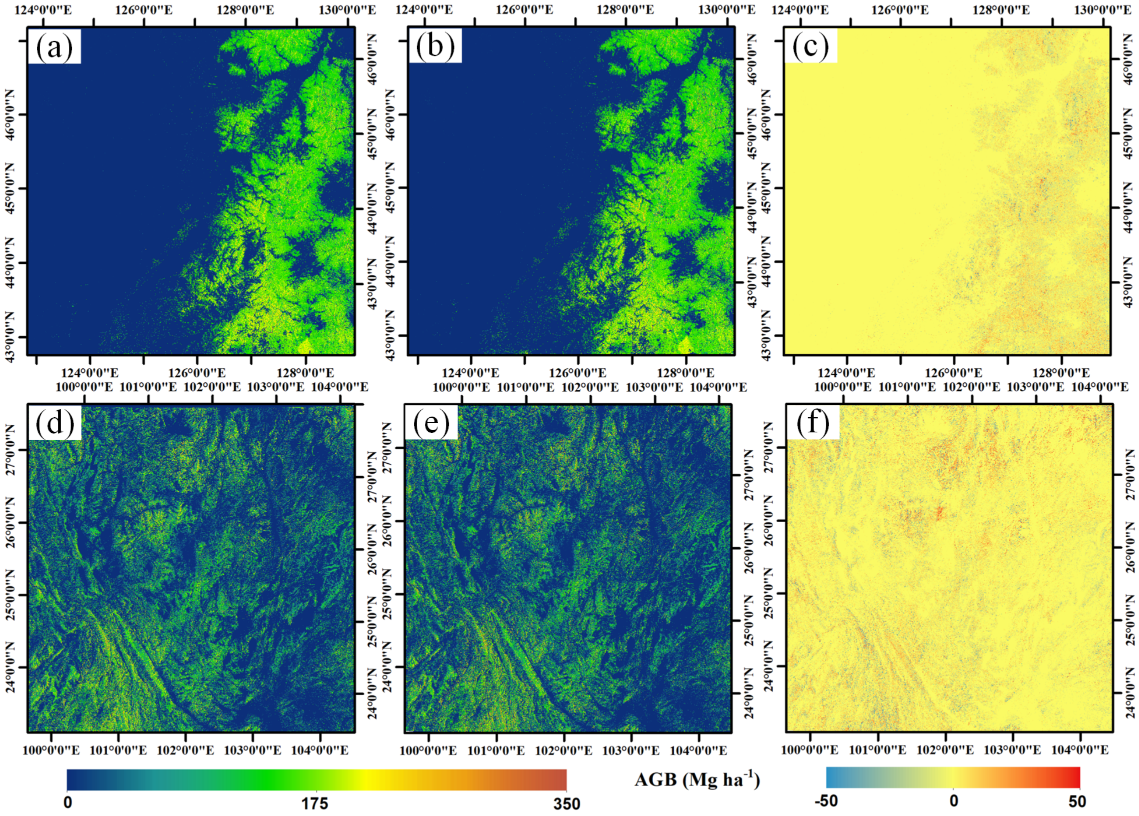


Figure 3.10: Spatial distribution of AGB and the differences between the AGB maps from the two models at 25 m resolution. (a) LightGBM in the northeastern region, (b) Random forest in the northeastern region, (c) Difference Map between LightGBM and RF Estimated AGB in the northeastern region (d) LightGBM in the southwestern region, (e) Random forest in the southwestern region, (f) Difference Map between LightGBM and RF Estimated AGB in the southwestern region.

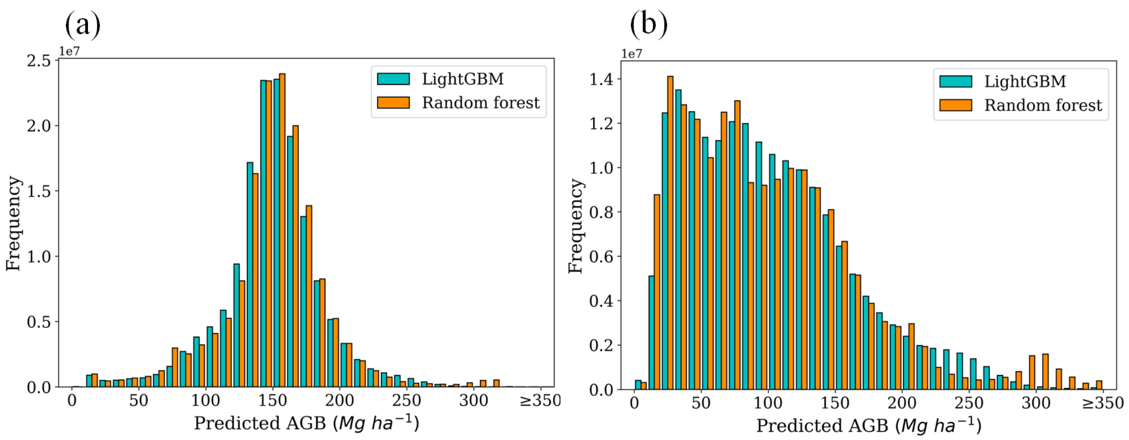


Figure 3.11: Frequency distribution histograms of AGB, in 10 Mg ha^{-1}, for (a)The northeastern region, (b)The southwestern region.

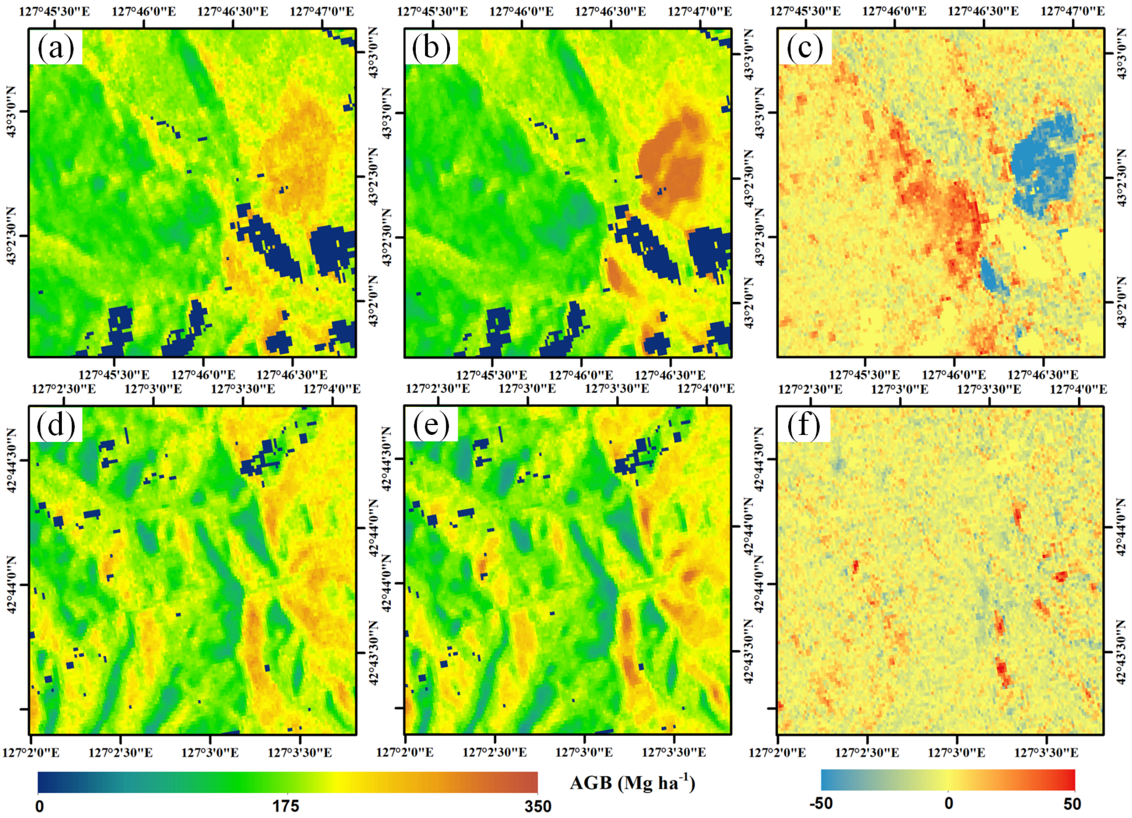


Figure 3.12: Spatial distribution of AGB for two 2500 m x 2500 m areas. (a) (d) LightGBM, (b) (e) Random forest, (c) (f) difference maps.

We also generated uncertainty maps based on the standard deviation of results from 5-fold cross-validation (Fig. 3.13). Both LightGBM and Random Forest showed relatively low uncertainty in the two study regions. For LightGBM, the uncertainty in the northeastern region was 3.28 Mg ha^{-1}, while in the southwestern region it was 4.22 Mg ha^{-1}. Meanwhile, Random Forest uncertainty was 2.96 Mg ha^{-1} in the northeast and 3.44 Mg ha^{-1} in the southwest. However, LightGBM exhibited slightly higher uncertainty, indicating it may be more sensitive to differences across folds in the training data. Further analysis is warranted to determine the factors contributing to the uncertainty patterns.

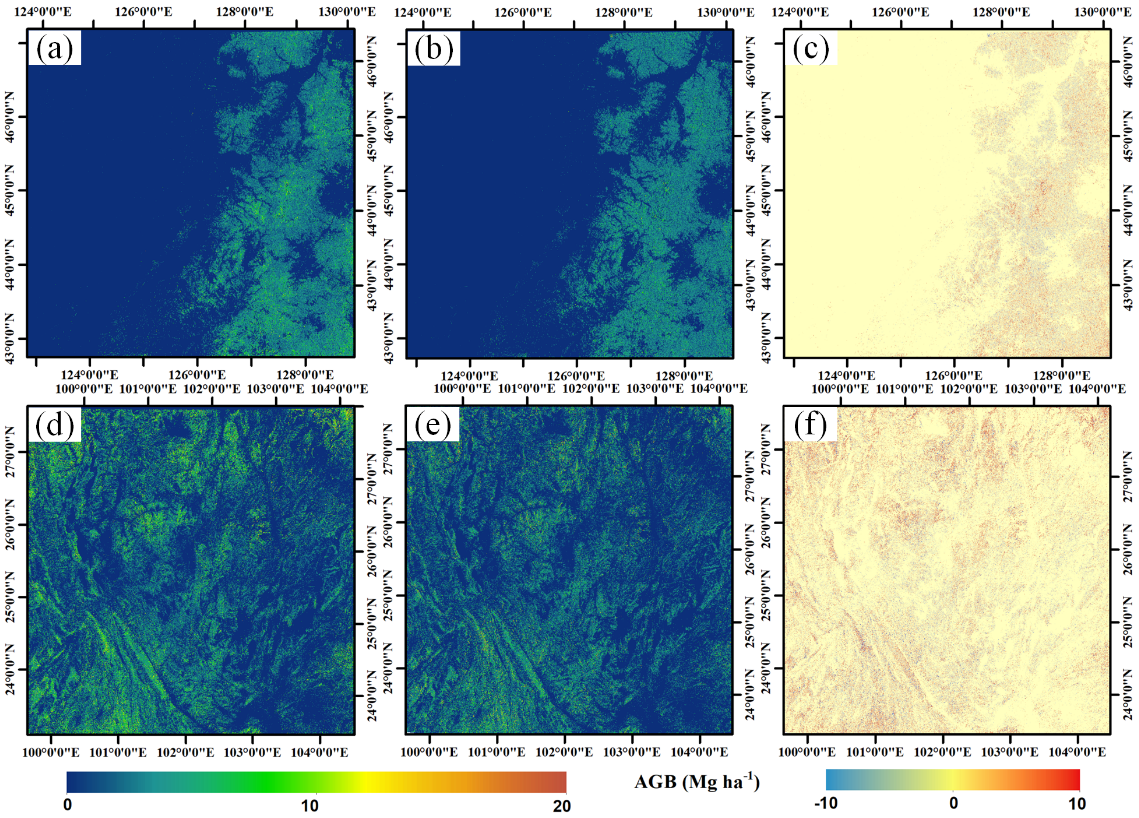


Figure 3.13: Spatial distribution of uncertainty and the differences between the uncertainty maps from the two models at 25 m resolution. (a) LightGBM in the northeastern region, (b) Random forest in the northeastern region, (c) Difference Map between LightGBM and RF uncertainty in the northeastern region, (d) LightGBM in the southwestern region, (e) Random forest in the southwestern region, (f) Difference Map between LightGBM and RF uncertainty in the southwestern region.

3.4 Discussion

3.4.1 Influence of steep slopes

Both regions are characterized by a significant amount of mountainous terrain, resulting in many GEDI footprints falling in areas with substantial slopes, especially in the southwestern region (Fig. 3.14). This could potentially impact the results, as steep slopes could lead to overestimation or underestimation of canopy height by GEDI data, depending on the steepness of the slope and spatial distribution of trees within footprints (Chen, 2010; Schneider et al., 2020). Slopes also impact the backscatter of the SAR data (Shimada et al., 2011). Ideally the SAR data should be corrected for increased or decreased backscatter caused by terrain slopes. However, the relationship between slope and backscatter is complex and depends on forest structure and properties, so slope corrections are not always feasible (Ramachandran et al., 2021). Additionally, some steep slopes can cause SAR layover and shadow effects, resulting in no useful SAR data for those areas (Samuele et al., 2021). These terrain factors introduce uncertainties in both the GEDI lidar and SAR data which should be considered, especially in regions with substantial topography.

As the slope of the GEDI footprints increases, there is a noticeable decrease in the R^2 values, and the RMSE gradually increases (Fig. 3.15) in both regions. Overall, the two models demonstrated similar accuracy in flat areas. However, in areas with steep slopes, LightGBM outperformed the random forest model, with a larger R^2 and smaller RMSE. We chose not to remove GEDI footprints from areas with steep slopes, because their removal could lead to unreliable AGB predictions in these steep areas, as much remaining forest is genuinely located on steep slopes in these regions. In the northeastern region, the average slope of GEDI is 10.23 degrees, and for the GEDI footprints within the Hansen forest mask, it is slightly higher at 11.97 degrees. The average slope of GEDI in the northeastern region measures 20.38 degrees, whereas the average slope of GEDI footprints within the Hansen forest mask stands at 21.18 degrees. More field data collected in areas with steep slopes are needed to evaluate the impact of removing GEDI footprints in steep areas on the accuracy of AGB estimation in these areas.

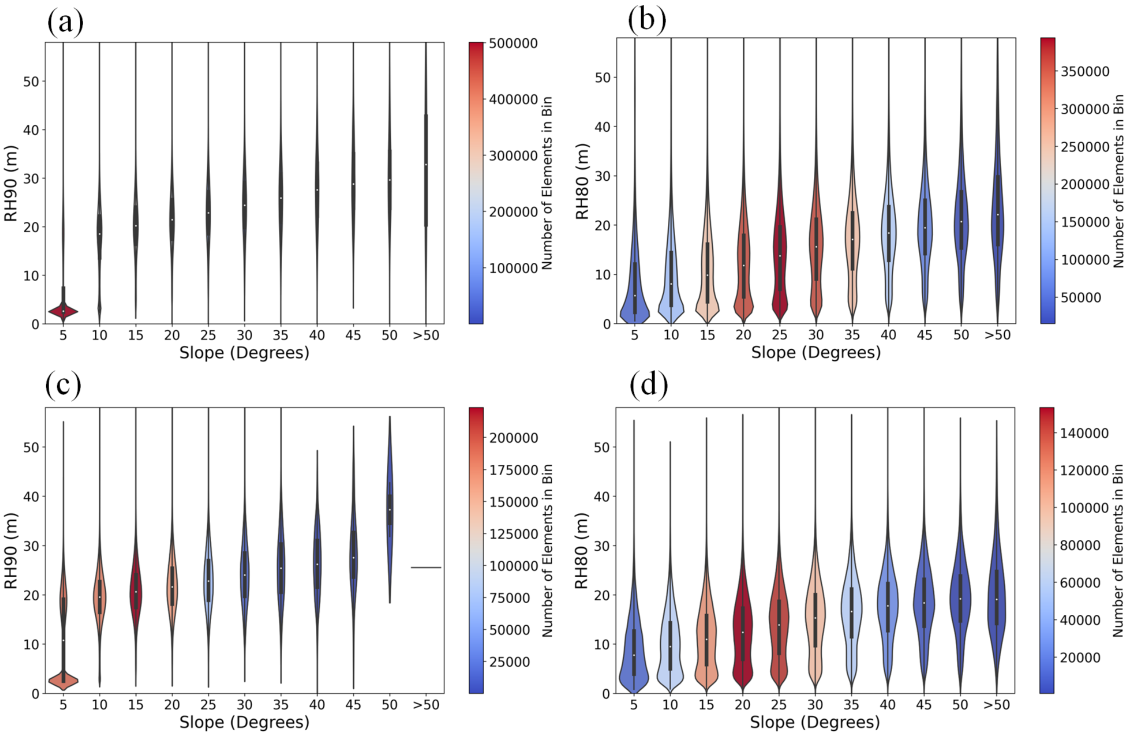


Figure 3.14: GEDI RH metrics are plotted against slope. The slope bin of 5 indicates slope from 0 to 5, and similar with all other bins. The colors of the boxes represent the quantity of GEDI footprints in different slope bins. (a) RH98 in northeastern region, (b) RH80 in southwestern region, (c) RH98 in northeastern region after filtering using SAR data, (d) RH80 in southwestern region after filtering using SAR data.

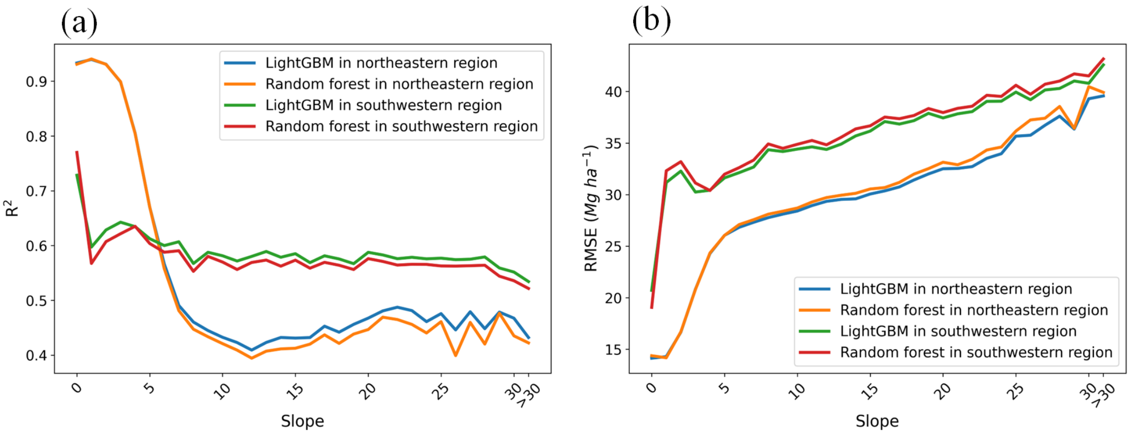


Figure 3.15: Model performance in both regions are plotted against slope.

3.4.2 Scalability of the approach

The models trained with GEDI and remote sensing imagery in the northeastern and southwestern regions were directly applied directly to two 100 x 100 km areas (Fig. 3.1). These test areas are located 90 km from the northeastern region and 450 km from the southwestern region, respectively. In the northern test area, the results of both models closely aligned with the AGB derived from GEDI data (Fig. 3.16). The LightGBM and random forest models achieved R^2 values of 0.76 and 0.77, with respective RMSE values of 31.01 and 29.44 Mg ha^{-1}. The fitted line and the 1:1 line between the model predictions and the AGB derived from GEDI were very close, indicating no significant saturation point within the AGB range below 300 Mg ha^{-1}. In the middle southern test area, the R^2 values were slightly lower, with LightGBM and random forest models achieving R^2 values of 0.55 and 0.56, respectively. However, the RMSE values remained relatively low, at 48.79 and 46.80 Mg ha^{-1}, respectively, indicating high accuracy in this area as well.

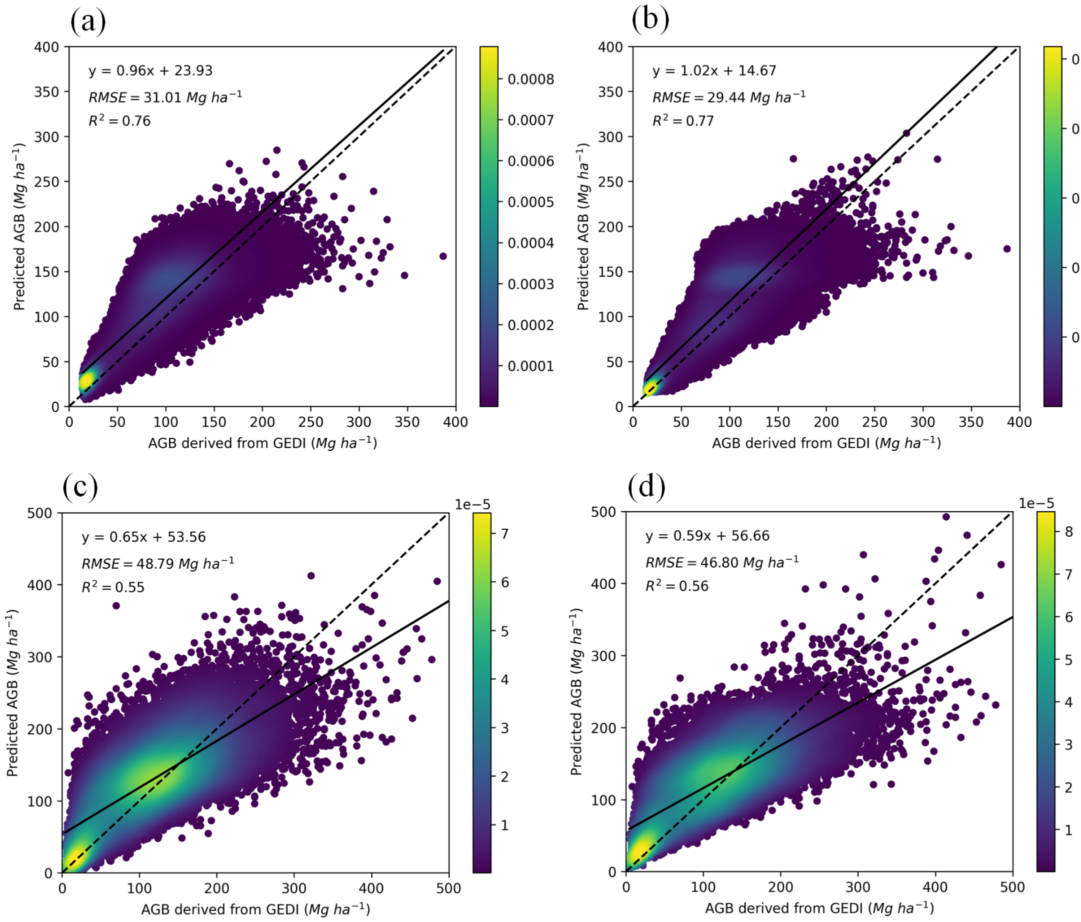


Figure 3.16: Performance of trained models in other area. (a) Trained LightGBM in the northern test region, (b) Trained random forest in the northern test region, (c) Trained LightGBM in the middle southern region, (d) Trained random forest in the middle southern region. The color of each point in the scatter plot represents the estimated density of points at that location.

3.4.3 Linking field AGB and GEDI RH metrics

In this study, we measured canopy height and DBH of trees in 42 field plots, with 36 of them co-located with GEDI footprints. There are various allometric equations to estimate forest AGB using height and DBH, depending on tree species and regions. The AGB estimates derived from different allometric equations can vary significantly. In this study, we used a general allometric equation developed for the trees in China (Luo et al., 2018). The selection of field plots significantly influence the conversion of GEDI RH metrics into AGB. It is crucial that the selection of field plots is representative. To this end, we have endeavored to ensure that our field points are evenly distributed spatially, thereby enhancing their representativeness.

However, it is important to recognize the uncertainties associated with this initial step. The allometric equations themselves are sources of uncertainty, as they are derived from field data that may not capture all variations in tree species and regional characteristics. Additionally, measurement errors in canopy height and DBH further contribute to the overall uncertainty in estimating AGB from RH metrics.

The primary implication of ignoring the uncertainty from the first step is an underestimation of the total uncertainty in the final biomass maps. The machine learning models rely on the AGB values derived from GEDI RH metrics as ground truth data. If these input values contain unquantified errors, the model's predictions will inherently reflect these inaccuracies, potentially leading to biased or overly confident estimates. This issue is compounded by the fact that machine learning models can obscure the propagation of uncertainty due to their complexity and non-linear nature.

To address this issue, future work should incorporate uncertainty quantification methods that propagate the uncertainties from the initial AGB estimation step through the entire upscaling process. Techniques such as Monte Carlo simulations, error propagation analysis, or Bayesian approaches can be employed to systematically account for and quantify the uncertainties at each stage.

3.4.4 Geolocation uncertainty of GEDI footprint

In this study, we used GEDI L2A version 2 dataset, which improved geolocation accuracy compared to version 1. However, upon comparison with ALS data, we found that the geolocation uncertainty in the GEDI data remains substantial. This may introduce errors into the AGB estimation of this study from two perspectives. Firstly, our field plots and GEDI footprints are at the same location and of the same size. However, there may be a discrepancy of about ten meters with the actual location of the GEDI footprints. This discrepancy could potentially result in the relationship we've established between field AGB and GEDI RH not being as accurate as intended. The effect had been somewhat mitigated by our strategy for selecting field points, which ensures that our field points were at least 50 meters away from non-forest areas. In addition, the geolocation uncertainty may introduce errors when matching GEDI footprints with continuous remote sensing images in the models. The geolocation uncertainty of GEDI footprints can be greatly improved through the use of ALS data. However, ALS data is scarce and expensive to acquire. Our research provides a reference for regions and countries where ALS data is lacking, especially for large regions. Our method of using known relationships between AGB and key remote sensing metrics to filter out GEDI footprints that are far from the expected values is shown here to be a reliable way to remove low reliability GEDI footprints, whatever the cause of error.

3.4.5 Implications of Data Usage in Model Calibration and Evaluation

In this study, we employed a 5-fold cross-validation approach to calibrate and validate our biomass estimation models using GEDI, Sentinel-1, ALOS-2 PALSAR-2, and Sentinel-2 data. While cross-validation is a widely accepted technique for assessing model performance, the use of the same dataset for both training and testing purposes may introduce certain challenges and potential overestimation of model accuracy (Ploton et al., 2020).

One primary concern is the risk of overfitting. Despite the segregation of data into training and testing sets through cross-validation, the entire GEDI dataset originates from the same spatial and temporal domain. This commonality may lead to an overly optimistic assessment of model performance, as the test data may not be truly independent from the training data (Meyer et al., 2018). The model might inadvertently learn patterns specific to the GEDI sampling strategy or the particular environmental conditions during the data collection period, potentially limiting its generalizability to other datasets or time periods.

Furthermore, spatial autocorrelation within the GEDI dataset could exacerbate this issue. Nearby GEDI footprints are likely to share similar environmental characteristics and biomass values, which may result in an artificially high agreement between predicted and observed values in the cross-validation process. This spatial dependency could lead to an underestimation of prediction errors when applying the model to new, spatially distinct areas.

While our current methodology provides valuable insights into AGB estimation across our study areas, acknowledging these limitations is crucial for accurate interpretation of our results and for guiding future research directions. The high performance observed in our cross-validation results should be interpreted with caution, considering the potential for optimistic bias due to data dependency.

3.4.6 Implications of data pre-filtering on accuracy estimation

The pre-filtering process applied to the GEDI footprints using SAR data from ALOS-2 PALSAR-2 and Sentinel-1, while beneficial in removing low-quality data, introduces the risk of overestimating the accuracy of biomass estimation models.

Our filtering process, which retained GEDI data within a ±3 dB range of the fitted curve between GEDI RH98 and SAR backscatter, is grounded in the well-established relationship between forest structure and SAR backscatter. Taller forests generally exhibit higher SAR backscatter (Mitchard et al., 2012; Woodhouse et al., 2012), providing a solid theoretical foundation for our approach. By leveraging this relationship, we aimed to identify and remove GEDI footprints with significant geolocation uncertainties, thereby enhancing the overall quality of our training data.

A critical consideration in interpreting our results is the issue of data independence. Although GEDI and SAR data provide complementary information about forest structure, they are not entirely independent. The use of SAR data to filter GEDI footprints may inadvertently remove some valid but anomalous forest structures that do not conform to the expected relationship. This could potentially limit the model's ability to capture the full range of forest structural diversity.

The stark contrast in improvement between the northeastern and southwestern regions highlights the significant impact of terrain on both GEDI and SAR data quality. In areas with complex topography, SAR data are subject to layover and shadow effects, which can compromise data quality. Our filtering approach, while primarily aimed at improving GEDI data quality, also inadvertently removed lower-quality SAR data in these challenging terrains. While this dual filtering effect contributes to the observed accuracy improvements, it underscores the need for cautious interpretation of our results, particularly in regions with varied topography.

Moreover, it is important to recognize that our approach may lead to an overestimation of accuracy, as we are effectively pre-filtering both GEDI and SAR data to retain instances where the products match well. This selective data retention could result in an overly optimistic assessment of model performance when applied to unfiltered data.

Despite these potential limitations, we argue that the accuracy improvements achieved through our filtering approach are beneficial for several reasons. Firstly, by removing GEDI footprints with high geolocation uncertainty, we enhance the spatial precision of our biomass estimates. Secondly, the elimination of low-quality SAR data in areas of complex terrain helps mitigate the confounding effects of topography on our models. Lastly, the resulting dataset, though potentially biased towards more "typical" forest structures, provides a more reliable foundation for biomass estimation in the majority of forest conditions.

3.5 Conclusions

In this study, we presented an experiment designed to determine the universality of remote sensing-based forest biomass mapping approaches at 25 m resolution, using new field plots co-located with GEDI footprints in two contrasting regions in northeastern and southwestern China. Both LightGBM and random forest performed well in the two regions compared to the GEDI validation data and field AGB. We also applied the models trained in these two contrasting regions to a northern area and a central-southern area, respectively, and achieved good results.Our preliminary test of the scalability of the models across different areas seems to demonstrate their robustness and ability to capture the relationships between the GEDI data and continuous remote sensing data in different areas with similar forest types. The promising outcomes can be attributed to several key factors. Firstly, the availability of a large volume of GEDI footprints and multiple sources of remote sensing imagery, including SAR and optical data, contributes to the accuracy and comprehensiveness of the prediction. In addition, the stringent removal of poor-quality data from the GEDI dataset enhances the reliability of the results. Our study demonstrates the significant potential of using GEDI data to produce high resolution AGB maps of China in the GEDI era (2019 and onwards).

LightGBM and random forest models exhibit similar performance in terms of accuracy. However, LightGBM outperforms random forest in terms of computational speed. With the growing number of space-borne sensors acquiring imagery and LiDAR footprints at high resolutions and frequencies, LightGBM's computational speed advantage empowers researchers to efficiently process large-scale and high-frequency AGB estimation (i.e. national scale annual estimates).

Chapter 4

Mapping forest AGB using deep learning method

Abstract

Accurate quantification of AGB is critical for understanding carbon stocks and changes in the context of climate change. In this study, we present a novel attention-based deep learning approach for forest AGB estimation, primarily utilizing openly accessible EO data, including: GEDI LiDAR data, C-band Sentinel-1 SAR data, ALOS-2PALSAR-2 data, and Sentinel-2 multispectral data. The attention UNet (AU) model achieved higher accuracy for biomass estimation compared to the conventional RF algorithm. Specifically, the AU model attained an R^2 of 0.66, RMSE of 43.66 Mg ha^{-1}, and bias of 0.14 Mg ha^{-1}, while RF resulted in an R^2 0.62, RMSE 45.87 Mg ha^{-1}, and bias 1.09 Mg ha^{-1}. However, the superiority of the deep learning approach was not uniformly observed across all tested models. ResNet101 only achieved an R^2 of 0.50, an RMSE of 52.93 Mg ha^{-1}, and a bias of 0.99 Mg ha^{-1}, while the UNet reported an R^2 of 0.65, an RMSE of 44.28 Mg ha^{-1}, and a substantial bias of 1.84 Mg ha^{-1}. Moreover, as spatial information is intrinsic to remote sensing data, it is crucial to investigate the dependence of deep learning models on spatial cues when estimating forest AGB. To explore the performance of AU in the absence of spatial information, fully connected (FC) layers were employed to eliminate spatial information from the remote sensing data. AU-FC achieved intermediate R^2 of 0.64, RMSE of 44.92 Mg ha^{-1}, and bias of -0.56 Mg ha^{-1}, outperforming RF but underperforming the AU model using spatial information. We

also generated 10m forest AGB maps across Guangdong for the year 2019 using AU, where a novel strategy was employed to mitigate boundary effects in patch-based predictions, a common challenge in deep learning models for image analysis. By overlapping and discarding edge pixels of patches, we significantly enhanced spatial coherence and predictive accuracy at boundary regions, thus ensuring more reliable AGB estimates. The AGB distributions from both AU and RF showed strong agreement with similar mean values; the mean forest AGB estimated by AU was 102.18 Mg ha^{-1} while that of RF was 104.84 Mg ha^{-1}. Additionally, it was observed through visual inspection that the AGB map generated by AU provided more realistic spatial information. Overall, this research substantiates the feasibility of employing deep learning for biomass estimation based on satellite data, highlighting its potential to enhance the accuracy and efficiency of forest carbon stock estimation.

4.1 Introduction

Forest ecosystems are fundamental to the global carbon cycle, holding 80% of the Earth's total vegetation biomass (Kindermann et al., 2008). Notably, of this forest AGB, 80% is stored above ground (Pan et al., 2013). Forest AGB have garnered significant attention given their profound implications for environmental and climate science (Mitchard, 2018). The accurate quantification and monitoring the dynamics of forest AGB are instrumental in enhancing our understanding of the global carbon cycle. Such insights provide a robust foundation for the formulation of evidence-based policies geared towards sustainable forest management and climate change mitigation (Herold et al., 2019).

Estimation methodologies for forest AGB are predominantly classified into two main categories: in-situ field measurements and remote sensing techniques. The most accurate, yet time-consuming small-scale approach for assessing AGB is based on in-situ destructive sampling (Kumar and Mutanga, 2017). Alternatively, field measurements typically involve selecting representative sites and measuring tree height and DBH using specialized instruments (Mitchard et al., 2009). In certain studies, wood density are also measured (Chave et al., 2005). Subsequently, these empirical measurements are fed into allometric equations to derive biomass estimates (Chave et al., 2014), otherwise they can be added based on databases of pre-measured values per species (Zanne et al., 2020). It's noteworthy, however, that such field-based methodologies, while precise, are labor-intensive and inherently limited in spatial coverage, often constrained to areas that are logistically feasible to access. In contrast, remote sensing techniques offer the advantages of large-scale monitoring combined with high spatiotemporal resolution (Saatchi et al., 2011; Santoro and Cartus, 2021*b*). Spaceborne remote sensing serves as an important approach for large-scale monitoring of AGB, providing explicit spatial detail and coverage, enabling high-resolution surveys of relatively extensive areas at a lower cost (Baccini et al., 2012; Brandt et al., 2018). Optical and SAR data serve as primary sources for estimating forest AGB. Optical data provides rich spectral information about vegetation properties based on reflection, while SAR penetrates into the canopy to derive canopy structure parameters (Gómez et al., 2014; Mitchard et al., 2011; Liao et al., 2020). However, they do not provide direct measurements of biomass (Woodhouse et al., 2012; Campbell et al., 2021). Conversely, LiDAR technology offers the capability for direct measurement of tree heights, which are closely related to forest AGB. However, spaceborne LiDAR data samples only a very small proportion of the landscape, and as such, it is often integrated with optical and SAR data for forest AGB estimation to provide wall to wall maps of estimated AGB.

Conventional machine learning techniques, such as RF, Maximum Entropy Modeling (MaxEnt), and Support Vector Machines (SVM), have been widely used for extrapolating discontinuous LiDAR data to generate wall-to-wall maps through integration with optical and SAR data (Saatchi et al., 2011; Liang et al., 2023; Dhanda et al., 2017).While these conventional machine learning models have shown promising results, they still have some inherent limitations that constrain their performance. Specifically, these models can incorporate some degree of spatial information through features like textures, their ability to capture complex spatial relationships is limited (Reichstein et al., 2019). Each pixel is often treated as an independent data instance, and while some pixel-based approaches can contain information on the neighborhood (Meng et al., 2016), they are typically much more limited in scope and information compared to spatially-aware deep learning models. However, forest AGB accumulation normally has an intrinsic spatial dependence that cannot be captured by treating pixels and samples independently. Neighboring pixels of forest exhibit spatial autocorrelation in AGB due to ecological factors and biophysical processes with spatial domains (Ploton et al., 2020). For instance, environmental variables like soil fertility and water availability that affect forest growth often have spatial patterns tied to topography and precipitation (Lewis et al., 2013). Therefore, accurate estimation of forest AGB requires characterization of these complex spatial patterns and relationships between neighboring pixels, rather than simplifying pixels as spatially independent data points. This suggests there is ample room for improvement by incorporating deep learning techniques, such as convolutional neural networks, graph neural networks or transformer models, which can automatically learn spatial features and patterns without relying on handcrafted features (Gu et al., 2018; Zhou et al., 2020; Dosovitskiy et al., 2020).

In the past few years, the field of remote sensing has begun to use deep learning to better exploit spatial and temporal information within datasets, achieving remarkable results in tasks such as classification and segmentation (Brandt et al., 2020; Tong et al., 2023; Mpakairi et al., 2023). The performance of deep learning largely depends on the availability and diversity of labeled datasets. An inadequate volume of data can result in model overfitting, while a lack of data diversity might compromise the model's generalization capacity (Long et al., 2021). Unfortunately, expansive and labeled datasets are not always readily available. Typically, for classification and detection tasks, manual annotations are made through visual interpretation (Brown et al., 2022). This approach, however, is not suitable for tasks like forest AGB estimation, as it is not possible to directly annotate the amount of biomass from remote sensing imagery. Forest AGB can be estimated using forest inventory plots, but the time involved in collecting individual plots, and the number needed to power deep learning algorithms, means that it is infeasible for sufficient forest inventory plots to be collected to directly train models.

The advent of GEDI has opened up possibilities for employing deep learning in biomass estimation. The GEDI instrument conducts high-resolution LiDAR observations of the Earth's three-dimensional structure, enabling precise inversion of canopy height (Liu, Cheng and Chen, 2021). The GEDI instrument integrates three lasers, with one split into two "coverage" beams and the others emitting full "power" beams. At any instance, four beams with a 25 m footprint diameter measure the ground, and these are alternately dithered across tracks, resulting in eight data tracks within a 4.2 km swath. Each data track has footprint centers distanced at 60 m along the track (Dubayah et al., 2020). With its unprecedented dense sampling as a LiDAR instrument in orbit, GEDI can serve as an ideal dataset to facilitate deep learning for forest attribute estimation.

In this study, we developed an approach to estimate forest AGB through the application of deep learning techniques to remote sensing data. We established a connection between field measured AGB and GEDI estimates of RH, thereby converting GEDI data into AGB to serve as labels. The framework we proposed relies on a UNet architecture incorporated with an attention mechanism, serving as the baseline model. Guangdong province, China, was selected as a case study region to demonstrate the applicability of the proposed models, owing to its diverse range of forest ecosystems and complex topographical features (Ren et al., 2013). The specific objectives of the study were as follows: i) to explore and validate the capabilities of deep learning methods for forest AGB estimation through the integration of multi-source remote sensing data; ii) to compare the results derived from deep learning methods with those obtained using the conventional RF method; iii) to analyze whether the improvements of AU over the RF method is due to AU's utilization of spatial information, or due to the increased flexibility and complexity of the AU algorithm compared to RF.

4.2 Data and methods

4.2.1 Study area

Guangdong province, situated in the southernmost region of mainland China, encompasses a total area of approximately 179,725 square kilometers. It boasts a diverse topography, ranging from coastal plains in the south to mountainous regions in the north, and is bordered by the South China Sea to its south. The province's climate is predominantly subtropical, characterized by mild winters and hot, humid summers (Cheng et al., 2021). Mean monthly temperatures range from approximately 16-19°C in January to 28-29°C in July (Tian et al., 2022). The province is subject to substantial precipitation, with average annual cumulative rainfall varying between 1,300-2,500 mm across regional gradients (Shen et al., 2019), fostering rich biodiversity and a variety of forest ecosystems. Guangdong is home to several forest types, including subtropical evergreen broad-leaved forests and monsoon rainforests (Ren et al., 2013). The area has undergone significant anthropogenic changes over the decades, marked by rapid urbanization and shifts in land use. From 1980 to 2020, the forested area in Guangdong province has experienced a noteworthy expansion, growing from 59,840 km^2

to 105,241 km^2 (Tian et al., 2022). Meanwhile, the province's forestry industry maintained its leading position in the nation in terms of total output value for several consecutive years (Zhang et al., 2018). Thus, Guangdong province has been selected as the case study for this research, aiming to delineate a high-resolution forest AGB map using remote sensing data to aid sustainable forest management and conservation efforts.

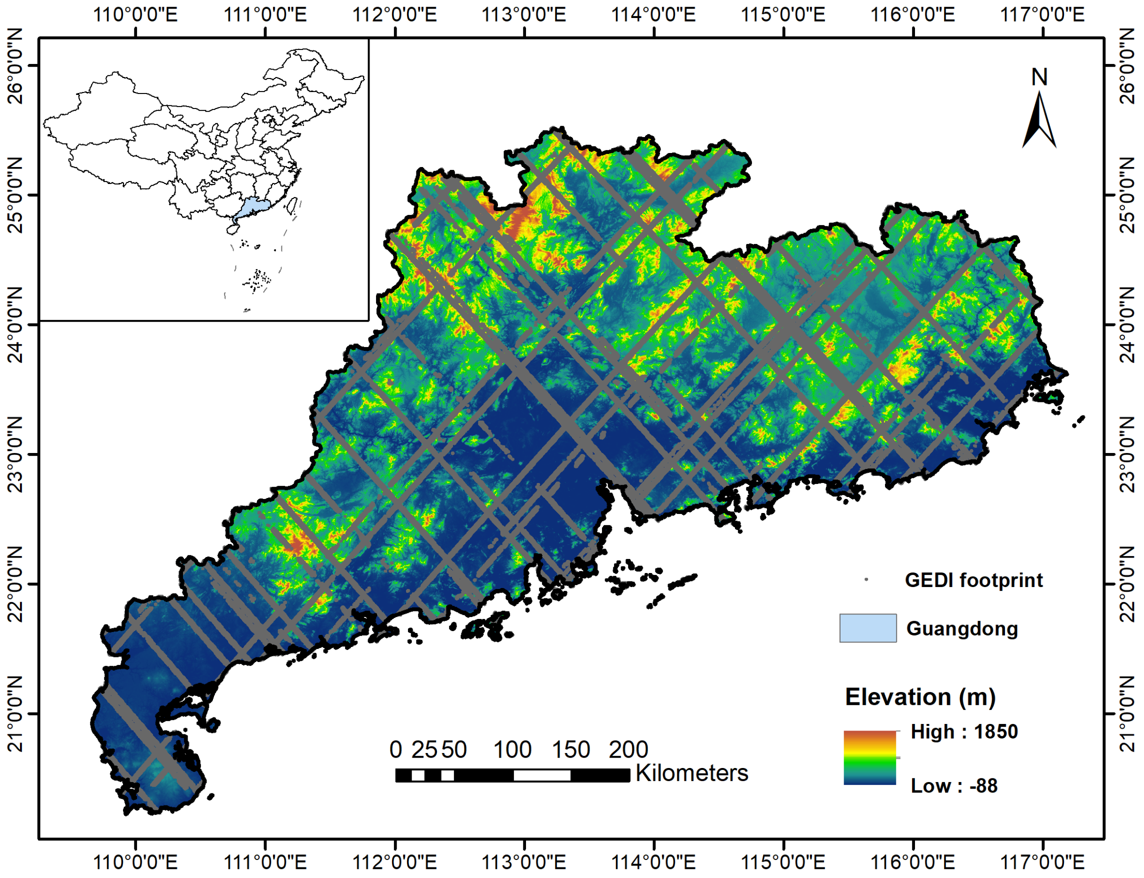


Figure 4.1: The location of Guangdong province and the spatial distribution of GEDI footprints shown over a Digital Elevation Model (DEM), the NASADEM (JPL, 2020).

4.2.2 Data

GEDI

GEDI Level 2A (L2A) and Level 2B (L2B) Version 2 data collected in 2019 were used in this study. GEDI L2A data provide footprint-level RH metrics and ground elevation derived from Level 1 waveform data. (Dubayah et al., 2021). RH metric refers to the height at which a specified quantile of the cumulative energy of the return signal is achieved. By capturing the full information of height distribution, the RH metrics quantify the vertical profile of forest within a GEDI footprint, and thus strongly correlate with AGB density and provide key inputs into AGB estimation models using GEDI data (Dubayah et al., 2020). GEDI L2B data includes canopy cover for each footprint. Canopy cover is a key metric in forest ecology, refers to the percentage of the ground area that is obscured by the vertical projection of the tree canopy.

The GEDI L2A and L2B data were accessed via GEE (Gorelick et al., 2017). Initially, we queried the GEDI L2 table index specific to our study area. Subsequent to this query, we utilized the retrieved GEDI file names to obtain the GEDI L2 vector data. Both the GEDI L2A and L2B data provide parameters that can be used to filter GEDI footprints with low quality. GEDI footprints that were acquired during the day (solar elevation ≥ 0), or those that were degraded (degrade $= 1$) or of low quality (quality flag $= 0$), were deemed unreliable and excluded from the analysis. For the GEDI footprints with GEDI L2B canopy cover of less than 0.8, the minimal sensitivity used in this study is 0.9. Conversely, for the footprints with a canopy cover of 0.8 or greater, the minimal sensitivity is 0.98. In this study, only power beams were used because their use results in higher accuracy when estimating tree height and biomass (Liu, Cheng and Chen, 2021; Duncanson et al., 2020). The Version 2 GEDI data, compared to Version 1, boast significant improvements in geolocation accuracy, with a mean geolocation uncertainty of 10.3 m. However, an uncertainty of 10.3 m is still substantial for our AGB map with a 10 m resolution. To reduce the geolocation uncertainty, footprints were filtered based on two criteria: firstly, where the absolute difference between the GEDI L2B canopy cover and the kNDVI derived from Sentinel-2 exceeded one standard deviation from the mean (Liang et al., 2023), and secondly, where footprints deviated from the fitted power curve of GEDI L2A RH98 and PALSAR HV by more than 2.5 dB.

Sentinel-1

We accessed and processed C-band (5.5 cm wavelength) Sentinel-1 data using GEE. Sentinel-1 GRD images in the IW mode, recorded during ascending orbits, were used in this study. These images were captured in dual-polarization, encompassing both the VV (Vertical Transmit, Vertical Receive) and VH (Vertical Transmit, Horizontal Receive) polarizations. Sentinel-1 GRD data have been subjected to essential preprocessing steps including thermal noise removal, radiometric calibration, and terrain correction, offering a ready-to-use dataset in decibels. We applied a focal mean filter with a radius of 50 m to the VV and VH polarizations to further mitigate noise and enhance data clarity. This filtering approach aids in suppressing extremes in SAR data (Bonafilia et al., 2020). We subsequently used the median values of the imagery acquired in 2019. Additionally, we incorporated the ratio of VV to VH polarizations into our analysis. The ratio serves as a more sensitive indicator for differentiating different vegetation characteristics, owing to its reduced sensitivity to ground characteristics (e.g. soil moisture) (Veloso et al., 2017; Vreugdenhil et al., 2020).

Sentinel-2

Sentinel-2 MSI L2A data, acquired in 2019, were used in this study. Sentinel-2 L2A products provide orthorectified surface reflectance images, which have been corrected for atmospheric effects. The MSI sensor collects high-resolution optical images at 13 spectral bands spanning from the visible and near-infrared to the shortwave infrared, with spatial resolutions of 10 m, 20 m, and 60 m (Drusch et al., 2012). We resampled all the bands to 10 m resolution using the bicubic method in GEE.To mitigate the effects of clouds, pre-filtering is initially applied to the Sentinel-2 images to retrieve granules with less than 10% cloud cover. Subsequently, the QA60 band is used to remove pixels affected by clouds and cirrus (Wang, Xiao, Liu, Wu, Qin, Steiner and Dong, 2020). The 5-day revisit time of the combined constellation ensures that sufficient data remain after cloud removal. We utilized the median values of all the bands, as well as the NDVI, kernel NDVI (kNDVI), and Normalized Difference Moisture Index (NDMI). The NDVI is sensitive to the greenness of vegetation (Rouse et al., 1974). kNDVI enhances sensitivity in dense canopy (Camps-Valls et al., 2021), and NDMI is responsive to vegetation water content (Wilson and Sader, 2002). In addition, we derived annual minimum NDVI, maximum NDVI, and the difference between them for each year. Minimum NDVI represents base greenness in low vegetation activity period, while maximum NDVI indicates peak greenness and helps to reduce the effects of the atmosphere (Martínez and Gilabert, 2009). Their difference quantifies annual vegetation dynamism, which helps account for vegetation seasonal variability when estimating forest AGB across time series. The multi-dimensional vegetation signals from NDVI, kNDVI and NDMI complement the optical bands, improving the representation of forest properties across different ecological conditions.

ALOS-2 PALSAR-2

As an L-band SAR (23.6 cm wavelength) system, ALOS-2 PALSAR-2 can penetrate vegetation canopies and interact with woody components (Ni et al., 2014), providing sensitivity to forest structure beneath the canopy layer. The L-band ALOS-2 PALSAR-2 mosaic data acquired in 2019 is directly accessible via GEE. The 25m PALSAR-2 yearly mosaic was created by merging strips of PALSAR-2 imagery, and has been ortho-rectificatied and slope corrected by JAXA (Shimada et al., 2014). For each annual mosaic, strip data were chosen based on visual inspection, prioritizing images with minimal surface moisture response, and exclusively using data from that specific year without resorting to prior years for gap-filling. We also applied a focal mean filter with a radius of 50 m to the HV and HH polarizations to reduce noise. This step was crucial as the presence of speckle can induce significant uncertainties in AGB change detection (Mermoz and Le Toan, 2016). The DN of HV and HH were converted into gamma naught values in decibel using the following equation:

$$\gamma_0 = 10\log_{10}(\mathrm{DN}^2) + CF \tag{4.1}$$

where, γ_0 is the backscatter coefficient in dB, CF is the calibration factor equals to -83.0dB (Shimada et al., 2009). We incorporated the PALSAR-2 HV/HH ratio, which helps improve saturation point of AGB (Sarker et al., 2012). In addition, the local incidence angle was integrated to mitigate the effects of terrain (Zhang, Wang and Liu, 2019; Das et al., 2015). Lastly, we resampled PALSAR-2 to a 10 m resolution using bicubic interpolation to ensure compatibility with other datasets.

Ancillary data

In addition to the aforementioned remote sensing data, we also utilized topographic data and latitude-longitude data in this study. The topographic data, including elevation and slope derived from NASADEM (JPL, 2020) using the 4-connected neighbors of each pixel, were incorporated to account for the variations in landscape and terrain, which significantly influence forest species and biomass accumulation (Stage and Salas, 2007). Latitude and longitude raster images were generated to represent the geographic coordinates of each pixel. Integrating longitude and latitude raster images can enrich models like RF by providing location information, which can potentially enhance the model's ability to capture and leverage spatial patterns and dependencies within the data, thereby improving prediction accuracy and model robustness. In addition, these latitude-longitude grids also provide additional geospatial contextual information beyond the patch scale to the deep learning models.

Remote sensing data processing for machine learning models

In order to estimate AGB using remote sensing data through deep learning and RF algorithms, it is imperative to transform both remote sensing imagery and footprint data into a format suitable for model input. Specifically, the initial sets of remote sensing imagery were segmented into image patches, each with dimensions of 64 pixels by 64 pixels and comprising 29 bands. Subsequently, GEDI footprints were employed to construct labels for the deep learning model. If the centroid of a pixel lies within a GEDI footprint, its label is assigned the value of AGB derived from GEDI; otherwise, it is set to -1 (Fig. 4.2). For the RF model, the mean value of pixels within the GEDI footprints was extracted using an area-weighted method. Partially intersected pixels along polygon boundaries contribute proportionally based on their overlapped area rather than fully counted. Then the mean value of each GEDI footprint was transformed to a patch used for a deep learning model without spatial information.

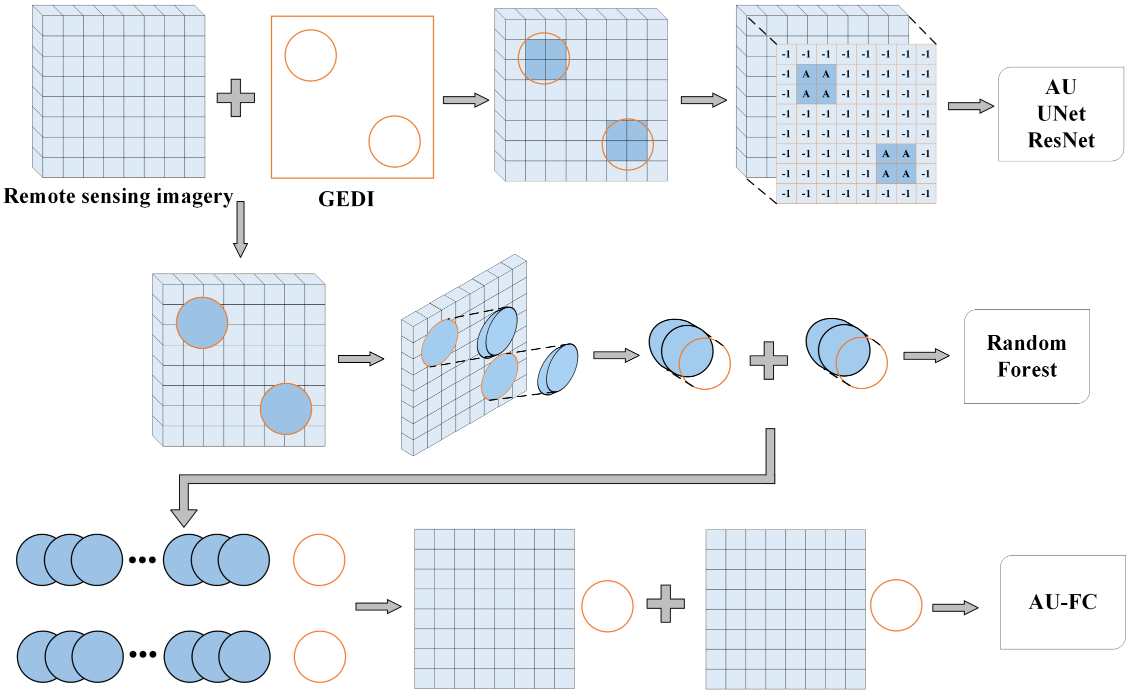


Figure 4.2: Illustration of data processing for deep learning models and RF. The squares represent pixels, and the orange circles represent GEDI footprints. A within a pixel in the top row mean a single value of AGB has been assigned to that pixel.

4.2.3 Forest AGB estimation

We developed a new approach to estimate forest AGB using a combination of satellite images and deep learning algorithms. The AGB derived from GEDI using the following equation were used as the label.

$$\text{AGB} = 5.58\text{RH80}^{1.12} \tag{4.2}$$

All remote sensing images were resampled to a 10m resolution using bicubic interpolation to extrapolate AGB derived from GEDI. In this study, we employed a 5-fold cross-validation approach (Yu, Wan, Dong, Zhang and Jia, 2022) to compare the accuracies of the deep learning models and RF, and utilized both AU and RF to generate forest AGB maps for Guangdong, China (Fig. 4.3). This involved conducting five distinct computations during the model accuracy comparison, utilizing the scatter plots with the highest precision for each model. Five AGB maps were generated using the trained models and subsequently, the mean of these five iterations was utilized as the final AGB map to obtain a more robust and reliable AGB estimation. In addition, an uncertainty map was derived by calculating the standard deviation of the 5 AGB predictions at each pixel location. Both our AGB and uncertainty maps were masked using a forest cover map. The forest cover map was extracted from the ESA WorldCover 10m v200 (Zanaga et al., 2022).

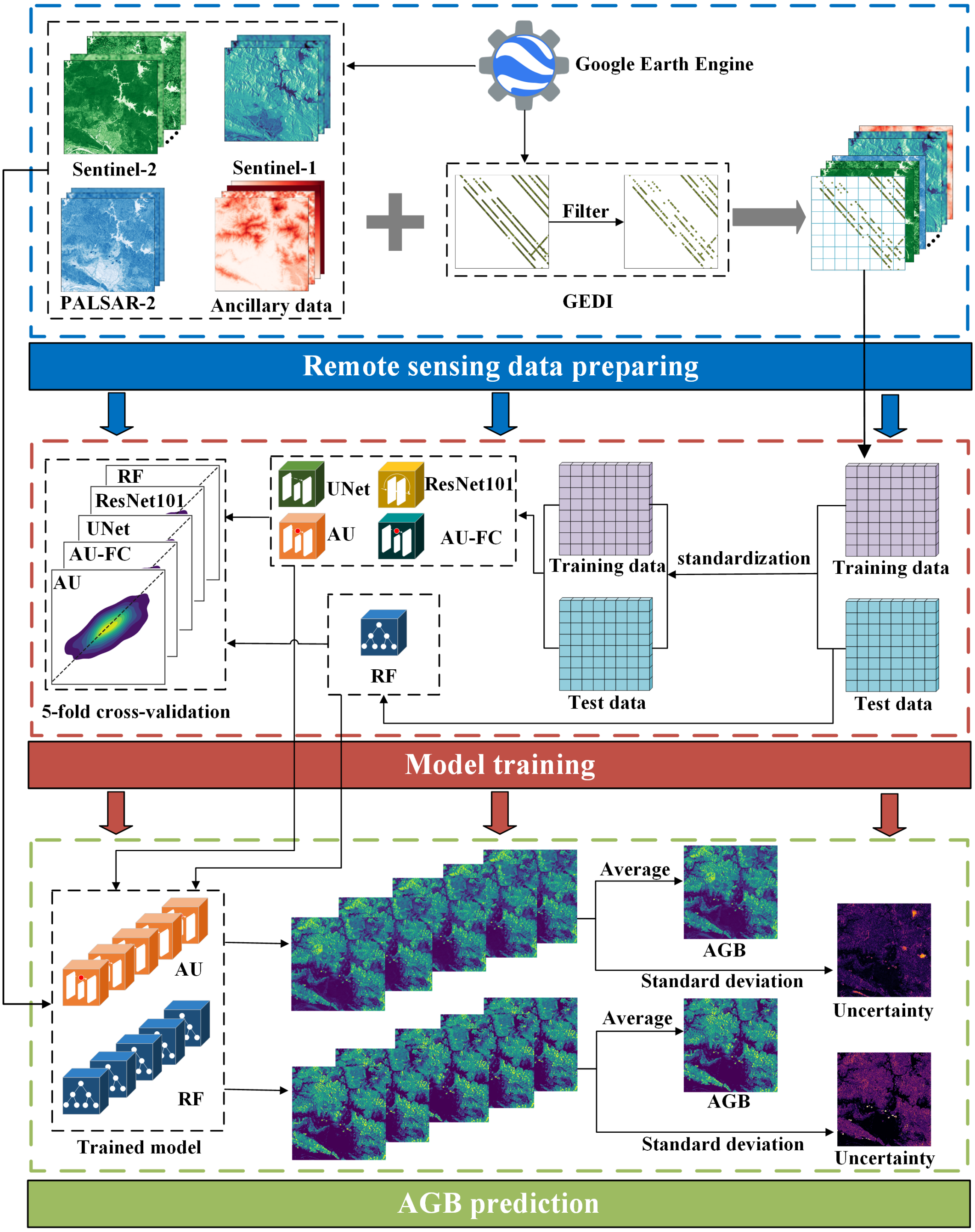


Figure 4.3: General framework.

Attention UNet

In this study, we aim to establish an end-to-end nonlinear mapping between remote sensing data and the corresponding AGB using the attention-based deep learning network. The mathematical formulation of this objective is as follows,

$$f = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \left\| f\left(X_i\right) - Y_i \right\|^2 \tag{4.3}$$

where f denotes the AU network model utilized in this study, $f(X_i)$ signifies the estimated AGB, and Y_i refers to the AGB values derived from GEDI data. Additionally, N represents the total number of samples present in the training dataset.

As illustrated in fig. 4.4, the architecture of the proposed AU model is fundamentally built upon the UNet framework. The model accepts remote sensing data with dimensions of 64×64×29 as input. The architecture is modular, consisting of both encoding and decoding stages, each containing three structurally identical blocks. In each encoding block, the input data undergo two successive 3×3 convolution layers, each accompanied by batch normalization (BN) and a Sigmoid activation function, before proceeding to a 2×2 max-pooling layer for spatial down-sampling. Following the encoding blocks, the bottom layer is subjected to up-sampling via a 2×2 kernel. The decoding blocks mirror their encoding counterparts, also featuring dual 3×3 convolution layers with batch normalization and Sigmoid activations. In addition, feature maps extracted at multiple scales are integrated through skip connections. This integration serves to combine coarse- and fine-level features from the remote sensing data. The architecture concludes with a 1×1 convolutional kernel to fine-tune the band dimensions of the final output. As a result, estimated AGB values with dimensions of 64×64 are obtained.

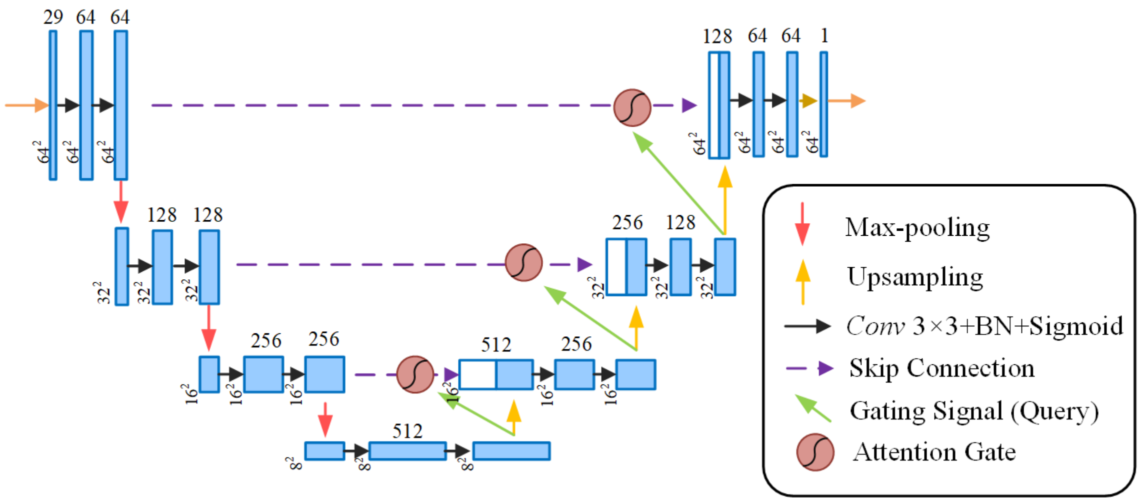


Figure 4.4: The architecture of AU used for estimating forest AGB from remote sensing data.

Extracting features that are strongly associated with AGB is a non-trivial task. To address this, our architecture incorporates an additive spatial self-attention mechanism in the skip connections. This modification enables the model to selectively emphasize salient features while attenuating less relevant features during the training process. By directing the focus of the model in this manner, we substantially enhance its predictive accuracy for AGB. Moreover, the integration of attention gates into the UNet framework introduces negligible computational overhead while delivering a notable improvement in both model sensitivity and accuracy. As depicted in Fig. 4.5, the implementation of the additive attention gate within the skip-connection segment is elaborated.

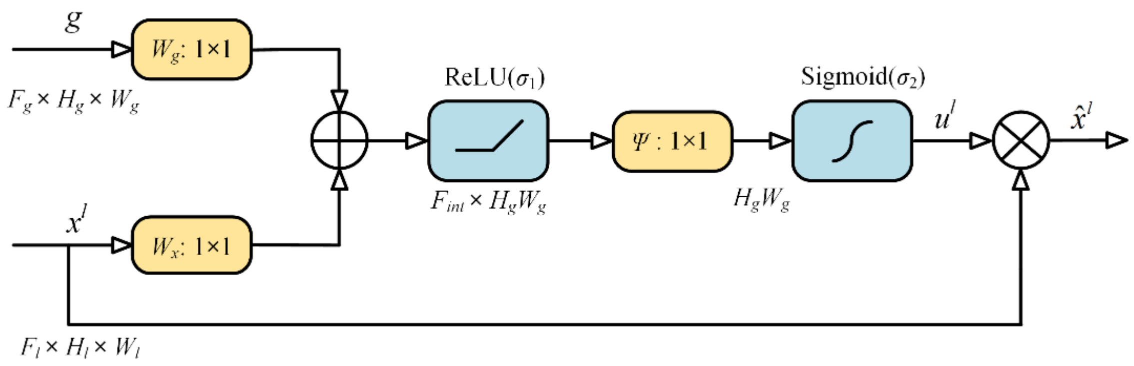


Figure 4.5: The structure of the attention gate. Rectified linear unit (ReLU) and Sigmoid are activation functions.

Here, F denotes the number of feature maps, while H and W represent the dimensions in terms of height and width of the feature maps, respectively. The mathematical formulation of the attention gate is provided below,

$$q_{att}^{l} = \psi^{T}\left(f_1\left(W_x^T x_i^l + W_g^T g_i + b_x + b_g\right)\right) + b_{\psi} \tag{4.4}$$

where W_x and W_g serve as the matrices for feature weighting, $x_i^l \in \mathbb{R}^{F_l}$ and $g_i \in \mathbb{R}^{F_g}$ correspond to the encoding matrix and the decoding matrix, respectively. The activation function f_1 is implemented using ReLU. Additionally, ψ represents a convolutional operator that employs a 1×1 kernel. The bias terms b_x, b_g and b_{ψ} are affiliated with their respective convolutional layers during the training process. Lastly, q_{att}^l functions as an intermediate representation in the computation. In the attention gate architecture, features extracted from coarser scales are used in gating to discern and eliminate irrelevant or noisy responses in the skip connections. This adaptive filtering mechanism is active during both the forward and backward propagation phases, thereby optimizing neuron activations and contributing to enhanced model performance.

The attention coefficient is formulated as given below,

$$u_i^l = f_2\left(q_{att}^l\left(x_i^l, g_i; \Theta_{att}\left(W_x, W_g, \psi, b_x, b_g, b_{\psi}\right)\right)\right) \tag{4.5}$$

where f_2 serves as the Sigmoid activation function, while Θ_{att} represents a set of parameters containing linear transformations $W_x \in \mathbb{R}^{F_l \times F_{\text{int}}}$, $W_g \in \mathbb{R}^{F_g \times F_{\text{int}}}$, $\psi \in \mathbb{R}^{F_{\text{int}} \times 1}$, and bias terms $b_x \in \mathbb{R}^{F_{\text{int}}}$, $b_g \in \mathbb{R}^{F_{\text{int}}}$, $b_{\psi} \in \mathbb{R}$.

The output of the attention gate, denoted as $\hat{x}_i^l$, is obtained through element-wise multiplication between the attention coefficient u_i^l and the encoding matrix x_i^l. As delineated in (4.6), this computational step amplifies the saliency of features that exhibit a robust correlation with AGB, while effectively attenuating the contributions from extraneous or less pertinent features.

$$\hat{x}_i^l = u_i^l * x_i^l \tag{4.6}$$

The update formulation for the convolutional parameters at layer $l-1$ is presented as follows:

$$\frac{\partial\left(\hat{x}_i^l\right)}{\partial\left(\Phi^{l-1}\right)} = \frac{\partial\left(u_i^l f\left(x_i^{l-1}; \Phi^{l-1}\right)\right)}{\partial\left(\Phi^{l-1}\right)} = u_i^l \frac{\partial\left(f\left(x_i^{l-1}; \Phi^{l-1}\right)\right)}{\partial\left(\Phi^{l-1}\right)} + \frac{\partial\left(u_i^l\right)}{\partial\left(\Phi^{l-1}\right)} x_i^l \tag{4.7}$$

Fully connected AU

The attention UNet fully connected (AU-FC) model was employed to validate the performance of AU in scenarios where spatial information is not utilized. This model assesses the capability of AU to effectively discern and learn pertinent features and patterns without leveraging the inherent spatial relationships within the remote sensing data. The input data for this model, based on the input data utilized for the RF, was transformed via a fully connected layer, feeding into 4096 nodes and then reshaping the nodes to 64*64 patches suitable for processing by the AU model (Fig. 4.2). We employed the fully connected layer to connect every input data to every output data (Liu, Kang, Zhang and Hou, 2018), and the fully connected layer was denoted as follows:

$$y = fc(x, w, b) \tag{4.8}$$

where x is the vector representing the input to a fully connected layer, w is the corresponding weight matrix, and b is the bias vector. The output of this layer is represented by the vector y. The input vector x encompasses M distinct elements, with each element denoted as x_i, where i represents the index of the element. It is crucial to note that the weight matrix w has dimension of $M \times K$, with K symbolizing the dimension of the output vector y. This layer operates by transforming the input vector x, subsequently generating each element, $y_{i'}$, in the output vector y, according to to the following function.

$$y_{i'} = \sum_{i} w_{ii'} x_i + b_{i'} \tag{4.9}$$

The data transformed to patches through the fully connected layer underwent the same processing steps as in the standard AU model. Subsequent to the predictions made by the AU model, the data were transformed back through two additional fully connected layers into discrete AGB predictions. The first fully connected layer condensed the spatial encoding into a 2048-vector embedding. The second layer mapped this to a single AGB value per sample. The fully connected layer essentially functions as a projection to align the inputs and outputs.

ResNet101

To benchmark the effectiveness of the AU model in estimating forest AGB, we compared it with the deep residual network (ResNet). This ResNets introduce residual learning, where the models learn residual functions with reference to the layer inputs, which enables the training of very deep networks by allowing the flow of gradients through the network during backpropagation (He et al., 2016). The core idea of ResNets is to add skip connections that bypass some layers, with their outputs being directly added to later layers. This mitigates the vanishing gradient problem associated with training deep neural networks especially as the depth increases. The extremely deep architecture of ResNet101, with over 100 layers, provides high modeling capacity to extract features at multiple levels of abstraction (Xu et al., 2020). ResNet101 potentially have the power to use more of the rich spatial information available in the remote sensing imagery, and thus could potentially outperform AU in their abilities to estimate AGB. We thus select ResNet101 to estimate forest AGB based on its powerful representational abilities.

Model setting

A. Loss function

The task of predicting AGB is cast as a regression problem, for which the loss function utilized in our model is articulated as follows:

$$\mathrm{L}_{\mathrm{loss}} = \mathrm{L}_{MSE} + \lambda\|\theta\|^2 \tag{4.10}$$

In this equation, the first term represents the Mean Squared Error (MSE) loss component, while the second term denotes the l_2 regularization term with a regularization parameter λ. $\theta = \{W, b\}$ denotes the training parameters: weights and bias of the network.

B. Data preprocessing

The image patches were partitioned at random into three distinct datasets: a training set, a validation set, and a test set, following a ratio of 7:2:1, respectively (Yu, Zhang, Gao and Jia, 2022). The training set comprises 39,116 samples, while the validation and test sets contain 1,1176 and 5,590 samples, respectively. Prior to being fed into the neural networks, both the remote sensing data and the associated AGB undergo standardization.

Standardization significantly enhances the ability of network to generalize across different datasets. The standardization process is mathematically represented by (4.11),

$$D_i' = \frac{D_i - D_{\text{mean\_train}}}{D_{\text{std\_train}}} \tag{4.11}$$

where D_i denotes either the remote sensing data or labels and D_i' represents the corresponding normalized data. $D_{\text{mean\_train}}$ and $D_{\text{std\_train}}$ signify the mean and standard deviation for each band in the training set, respectively. It should be noted that the input remote sensing data and labels in validation and test sets are also normalized using $D_{\text{mean\_train}}$ and $D_{\text{std\_train}}$, respectively. This strategy is employed to prevent data leakage and to ensure a robust evaluation of the generalization performance of the model.

C. Implementation details and computational configuration

The deep learning algorithms were executed on a dedicated server, configured with an NVIDIA RTX A5000 24GB GPU. The model was developed in Python 3.10 and utilized PyTorch 2.0.1 for backend computations. The Adam optimizer (Kingma and Ba, 2014) was employed in the deep learning methods. Furthermore, the initial learning rate was configured to 0.001 and was designed to decay by a factor of 0.1 at every 40th epoch. The training process was constrained to a maximum of 120 epochs. A batch size of 128 was employed, and an L_2 regularization term for weight decay was set to 1×10^{-5} as a countermeasure against overfitting (Tartaglione et al., 2018; Farebrother et al., 2018). In accordance with the deep supervision strategy, we employ a masking technique on regions devoid of labels during the loss computation, thereby addressing the inherent sparsity of the ground-truth labels.

Mitigating boundary effects in patch-based predictions

UNet and other deep learning models frequently demonstrate diminished accuracy and reliability at the boundary or edge regions of images (Innamorati et al., 2020), a phenomenon often referred to as the "boundary effect" or "edge effect". In addition, independent per-patch inference during prediction leads to spatial incoherence and discontinuities between adjacent patches predicted by the deep learning models. This is predominantly because each patch is processed and evaluated independently, without considering the spatial context and relational information from its neighboring patches. To mitigate the boundary effect, a strategy was deployed during the prediction phase where each patch has an overlap of 10 pixels with its neighboring patches. Subsequently, the outermost 3 pixels of each patch were discarded to alleviate the influence of potentially inaccurate predictions at the patch boundaries. The overlapping regions were then averaged to obtain a smooth transition between patches. This approach aims to enhance spatial coherence between neighboring patches and refine the model's predictive accuracy at the boundary regions, thus providing more reliable and consistent output in image analysis tasks.

Accuracy assessment

To quantitatively evaluate the predictive performance of various methods on AGB, three regression metrics are employed: the coefficient of determination, commonly known as R-squared (R^2), the RMSE, and the bias. In the regression problem, these three metrics convincingly reflect the predictive capability of methods, and the definitions are as follows,

$$\mathrm{R}^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_i^K \left(Y_i - \hat{Y}_i\right)^2}{\sum_i^K \left(Y_i - \bar{Y}\right)^2} \tag{4.12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\left(Y_i - \hat{Y}_i\right)^2} \tag{4.13}$$

$$\mathrm{Bias} = \frac{1}{K}\sum_{i=1}^{K}(\hat{Y}_i - Y_i) \tag{4.14}$$

where K denotes the cumulative number of observations, Y_i is the AGB derived from GEDI, $\hat{Y}_i$ denotes the predicted AGB, and $\bar{Y}$ is the mean of the AGB derived from GEDI.

4.3 Results

4.3.1 Model performance

To evaluate the accuracy and reliability of the deep learning models in estimating AGB, we conducted five separate AGB estimations for each of AU (Oktay et al., 2018), UNet (Ronneberger et al., 2015) without the attention mechanism and ResNet (He et al., 2016). For each time, we employed unique combinations of training and validation sets, ensuring a comprehensive understanding of the models' adaptability and consistency. The performance of the three deep learning methods across the five estimations remained consistent when evaluated using the test dataset, with the highest accuracy instance shown in Fig. 4.6. Based on the metrics of R^2, RMSE, and bias, AU clearly outperforms the other two models. Additionally, from the scatter plot, it is evident that AU has fewer outliers compared to UNet and ResNet101, further underscoring its reliability in AGB estimation. The R^2 values for AU and UNet are 0.65 and 0.64, respectively, and their RMSE values are closely matched at 44.02 Mg ha^{-1} and 44.81 Mg ha^{-1}. Notably, AU's bias is 0.19 Mg ha^{-1}, which is considerably lower than UNet's bias, which stands at 1.86 Mg ha^{-1}. In contrast, ResNet101's performance is poorer, with an R^2 of 0.48 and an RMSE of 53.89 Mg ha^{-1}. Its bias is 1.06 Mg ha^{-1}, placing it between AU and UNet.

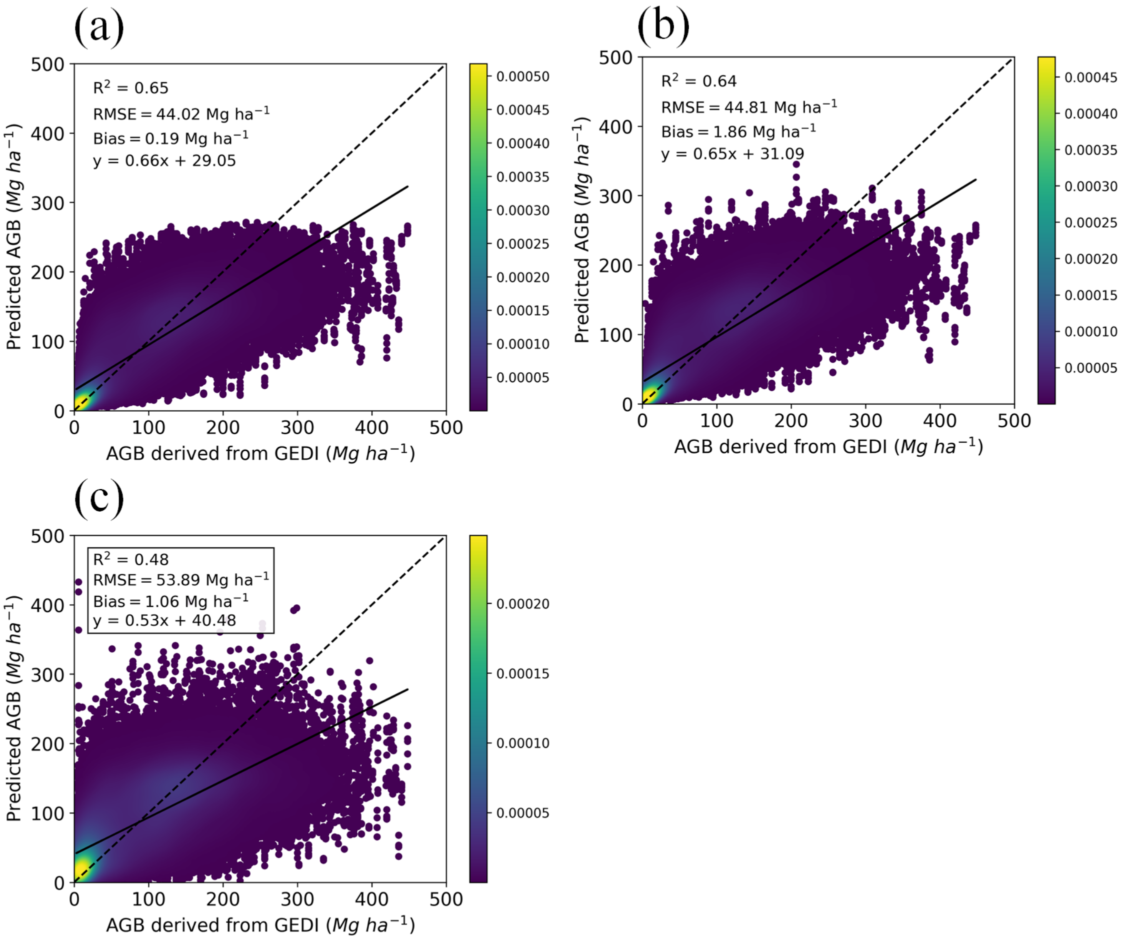


Figure 4.6: The predicted AGB values against AGB derived from GEDI RH metrics. (a) AU, (b) UNet, (c) ResNet101. The color represents the estimated density of points at that specific location.

To further assess the performance of deep learning models, we compared the deep learning models with a commonly utilized traditional machine learning method, the RF. To enable comparison with the RF approach, the deep learning AGB estimates for each individual GEDI footprint were aggregated by computing the mean prediction within each footprint. Scatter plots were generated between these mean AGB predictions and the corresponding AGB derived from GEDI (Fig. 4.7). Compared to the approach of scattering the predictions from individual pixels, employing footprint-level means yields an increased R^2 and a decreased RMSE, as evidenced by a comparison of Fig. 4.7 (a) (b) (c) with Fig. 4.6 (a) (b) (c). This enhancement can be attributed to the noise and variability attenuation achieved by averaging at the pixel level. By computing the mean, outliers and anomalies within each footprint are effectively smoothed out.

The performance of the AU model is superior to that of the RF, as demonstrated by its higher R^2 value, coupled with reduced RMSE and bias metrics. Notably, the minimal bias of the AU model indicates that, on average, its predictions are closely aligned with the true values, without consistently under- or over-estimating. This superiority underscores AU's ability to capture underlying patterns and trends more effectively than traditional machine learning methods. To explore whether the enhanced accuracy of AU over RF was solely attributed to its utilization of spatial information, an additional experiment was conducted comparing AU-FC (Fig. 4.7 (d)). The results revealed the prediction accuracy of AU-FC was intermediate between AU and RF models. This intermediate performance suggests that while the integration of spatial information by AU does contribute to its superior accuracy, other inherent characteristics of the model also play a significant role in optimizing prediction outcomes. It underscores the multifaceted nature of model accuracy, influenced not just by the incorporation of spatial context but also by the underlying architecture and learning mechanisms of the model. However, deep learning models are not necessarily superior to traditional machine learning algorithms. While the UNet boasts a higher R^2 and lower RMSE than RF, its bias is greater. Conversely, ResNet101 has a lower R^2 and higher RMSE compared to RF, with only its bias being marginally lower.

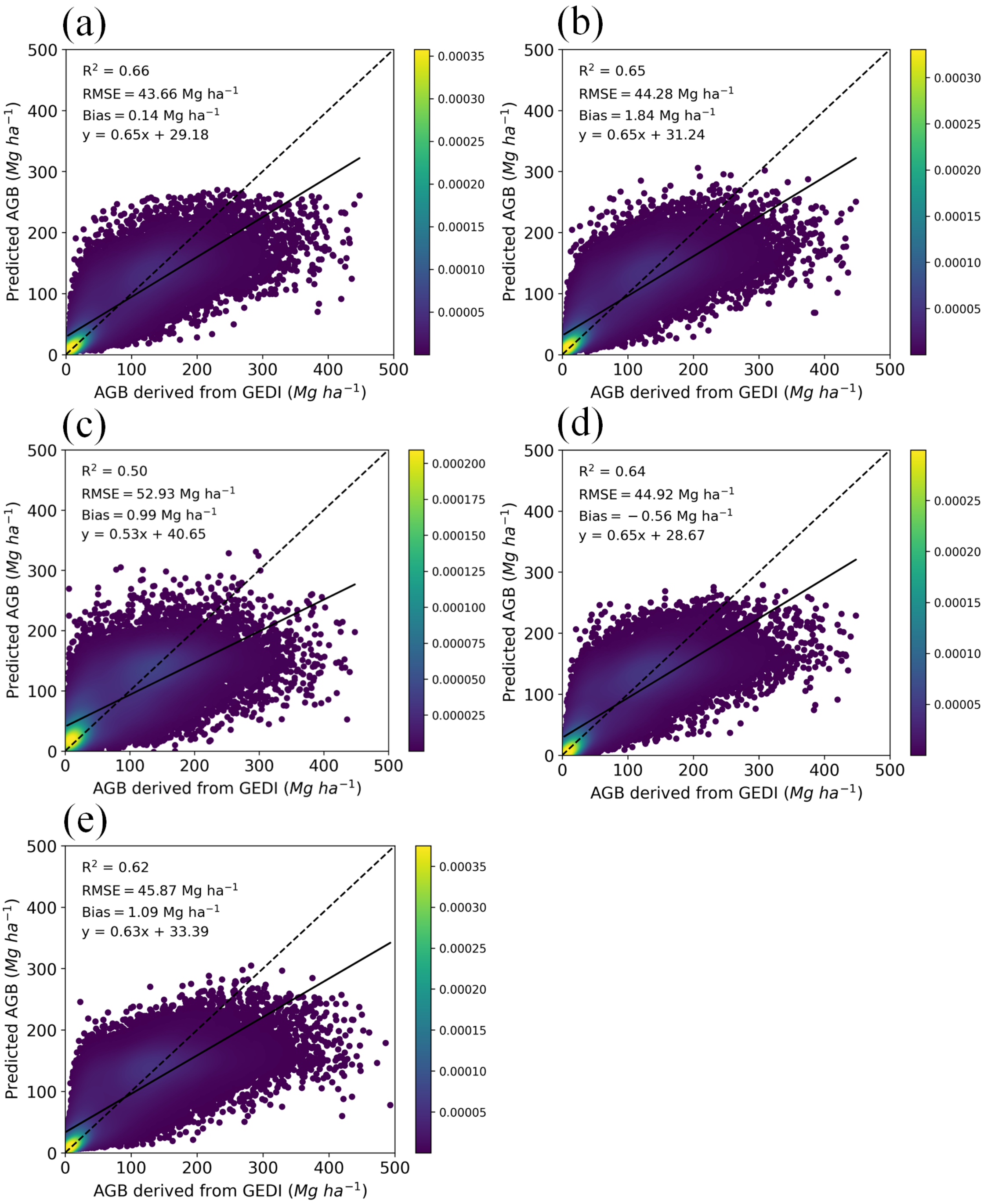


Figure 4.7: Mean predicted AGB values within GEDI footprints plotted against AGB derived from GEDI RH metrics. (a) AU, (b)UNet, (c) ResNet101, (d) AU-FC (e) RF.

4.3.2 Enhancement of spatial consistency

The AGB map initially generated by AU exhibited boundary effects as shown in (Fig. 4.8 (a)). Discontinuities and inconsistent values can be observed along the patch edges, likely resulting from the lack of contextual information exchange between neighboring image patches during the per-patch based training. To mitigate this deficiency, the boundary effect removal approach laid out in Section 4.2.3 of the methodology was applied. As illustrated subsequently in Fig. 4.8 (b), after undergoing the mitigation process, these boundary effects are no longer discernible. This indicates that the overlapping inference and averaging technique helped improve inter-patch coherence by reducing abrupt changes at patch boundaries caused by the boundary effect. The processed results after boundary effect removal clearly delineated the spatial variability and transitions in AGB levels across the landscape. The processed results demonstrated the road network, shown in the lower half of Fig. 4.8 (d), as a consistent low AGB region, accurately capturing its distinct linear footprint. In contrast, the representation by RF in Fig. 4.8 (c) did not elucidate the feature as distinctly, and is less smooth, with other high frequency angular artefacts visible.

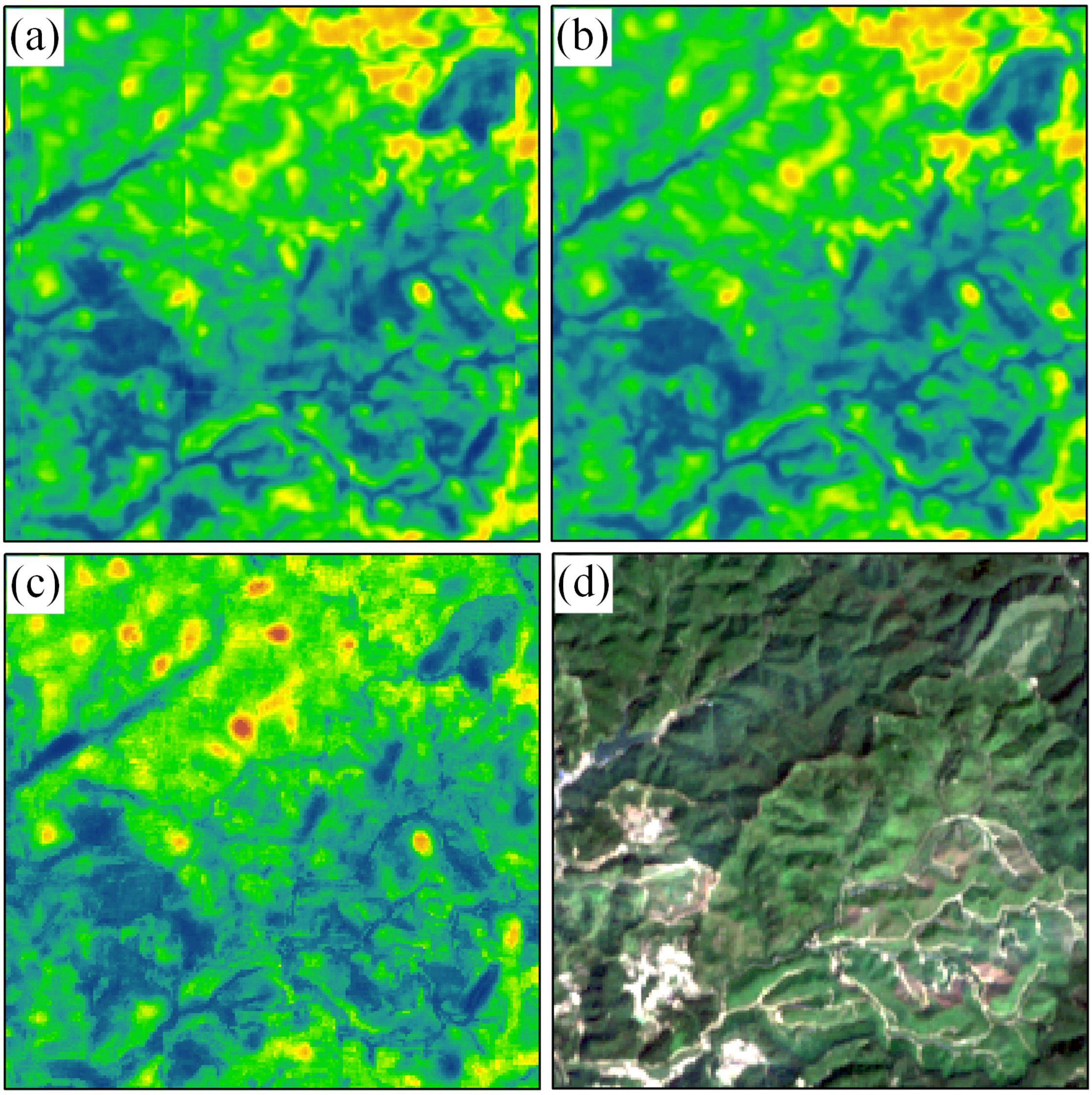


Figure 4.8: Local AGB mapping results for model comparison and boundary effect mitigation demonstration. (a) Example of a detail of the AGB map generated by AU, (b) The refined AGB map by AU after mitigating boundary effects, (c) AGB map produced by RF, (d) Sentinel-2 true-color composite image.

4.3.3 Spatial distribution of AGB and uncertainty map

The prediction and uncertainty maps by AU and RF at 10 m resolution are shown in Fig. 4.9, all the maps have been masked by the ESA forest cover product (Zanaga et al., 2022). It can be observed that the forest AGB maps from the two models exhibit similar trends, where the northern regions of Guangdong province have higher AGB, while the southern coastal areas manifest lower AGB values.

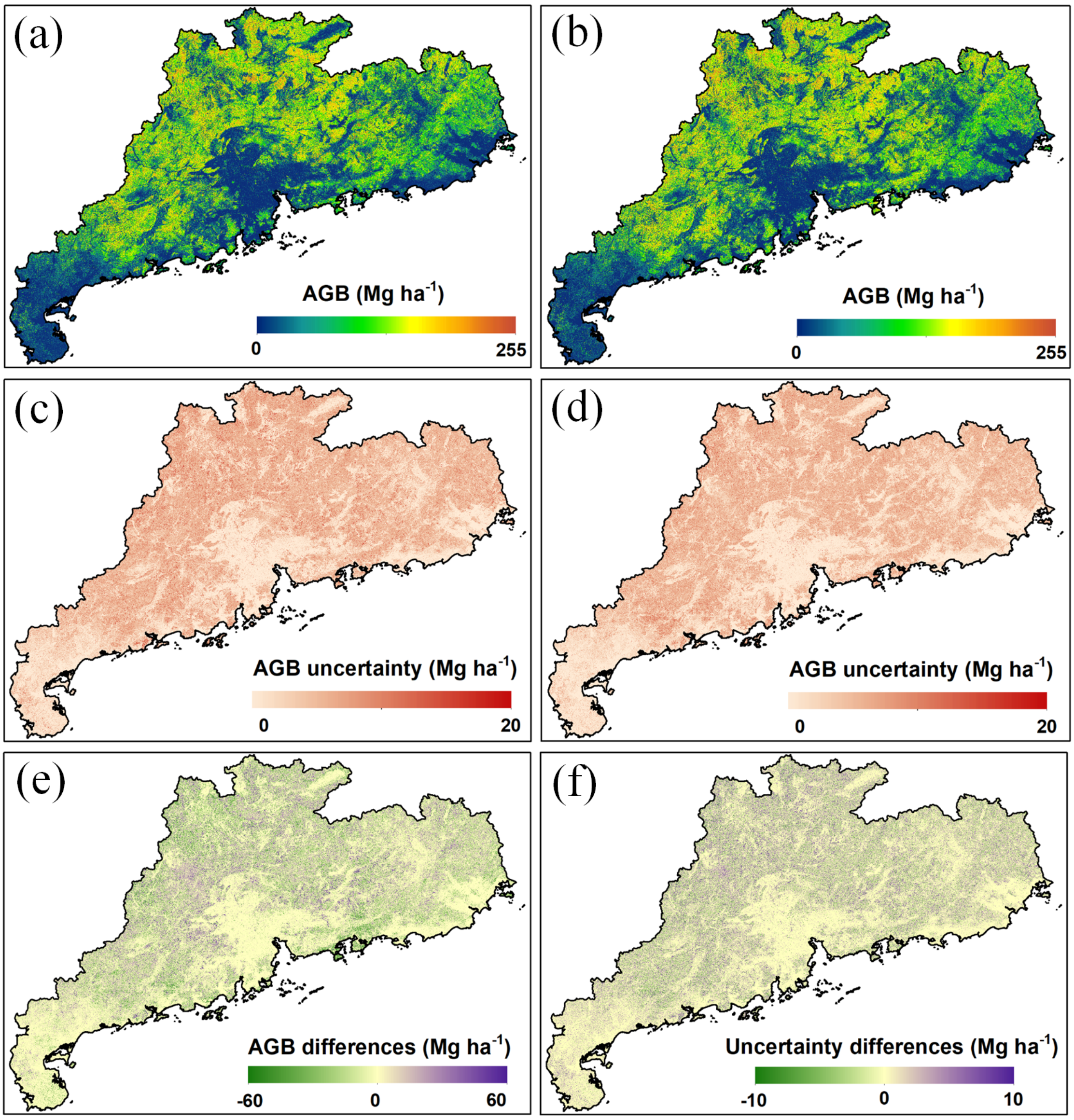


Figure 4.9: Comparative analysis of AGB maps and uncertainty maps at 10m resolution for AU and RF. (a) AU and (b) RF AGB maps, (c) AU uncertainty and (d) RF uncertainty maps, (e) differences in AGB maps: AU - RF, (f) differences in uncertainty maps: AU - RF.

The uncertainty associated with the forest AGB estimates exhibited a proportional relationship with the predicted AGB values; higher uncertainty was generally observed in areas with greater estimated AGB (Fig. 4.10). However, the RF model shows a more pronounced increase in uncertainty as AGB values rise, which is particularly evident from the elongation of the color gradient towards the upper right corner of the plot.

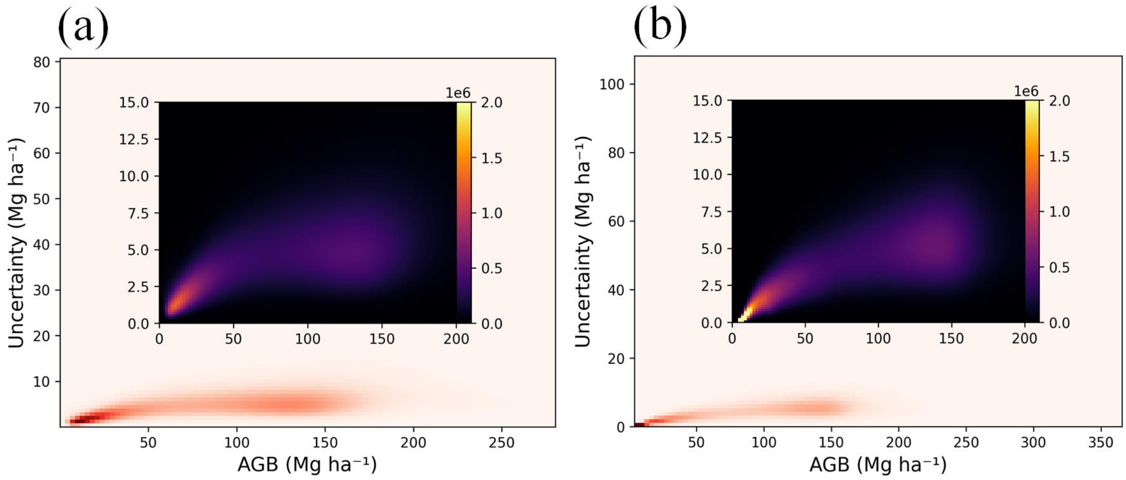


Figure 4.10: Relationship between predicted AGB and associated uncertainty for (a) the AU model and (b) the RF model. Insets in both panels show magnified views of the scatter plots, with identical scaling applied to both insets to facilitate direct comparison of data density and distribution patterns.

The AU model estimated the average forest AGB in Guangdong province to be 102.18 Mg ha^{-1} with an average uncertainty of 5.51 Mg ha^{-1}, while the RF model estimated it to be 104.84 Mg ha^{-1} with an average uncertainty of 5.12 Mg ha^{-1}. At the level of mean values, the estimations from both models appear closely aligned, yet the maps of AGB differences and uncertainty differences reveal distinct discrepancies in AGB and uncertainty between the two models (Fig. 4.9 (e) (f)).

4.3.4 Comparison with other AGB maps

To compare our AGB map against other AGB maps, we compared the AGB map generated by AU with the AGB map generated by RF, as well as the ESA CCI AGB map for 2019 (Santoro and Cartus, 2023) and the GEDI L4B version 2.1 (Dubayah et al., 2022), across multiple locations in Guangdong. The AU, RF, and CCI AGB maps are all for the year 2019, while the GEDI L4B 2.1 is based on GEDI data from 2019 to 2023. To better visualize the the spatial distribution differences between the AGB maps generated by AU and RF, we created enlarged images of specific areas (Fig. 4.11). The visual comparison of the four maps highlights that the 10-meter resolution AU and RF maps better reproduce the spatial patterns visible in the Sentinel-2 data. Fig. 4.11(a) shows a mountainous forest area where the AU and RF AGB maps exhibit similar spatial patterns. However, within circle A, the mean AGB derived from two GEDI footprints is 298.50 Mg ha^{-1}. The AU model estimated a mean AGB of 232.59 Mg ha^{-1} for this region, while the RF model estimated a value of 176.77 Mg ha^{-1}. Both models underestimated the AGB for circle A, as is evident in the scatter plot 4.7. However, the AU predictions are moderately closer to the GEDI reference, outperforming RF in this local region. Fig. 4.11(b) shows a mixed forest and non-forest area where the AU, RF, and CCI AGB maps all capture the low values within circle B. The AU, RF, and CCI AGB maps successfully characterized the zones in circle B by predicting low AGB values, as expected for non-forest land cover types. The AU model estimated a mean AGB of 16.16 Mg ha^{-1}, the RF model predicted an AGB of 19.54 Mg ha^{-1}, and the CCI map predicted a slightly higher mean AGB of 22.00 Mg ha^{-1} within circle B. The lower values from the AU network demonstrate its superior performance in representing low AGB conditions, at least in this portion of the study area. Fig. 4.11(c) shows a slightly larger area with a more complex landscape structure. The AU, RF, and CCI AGB maps all captured the low values within circle C. However, while the AU and RF AGB maps captured the low values within circle D, the CCI map did not. The GEDI L4B map did not accurately capture spatial detail in any of the three locations due to its lower resolution.

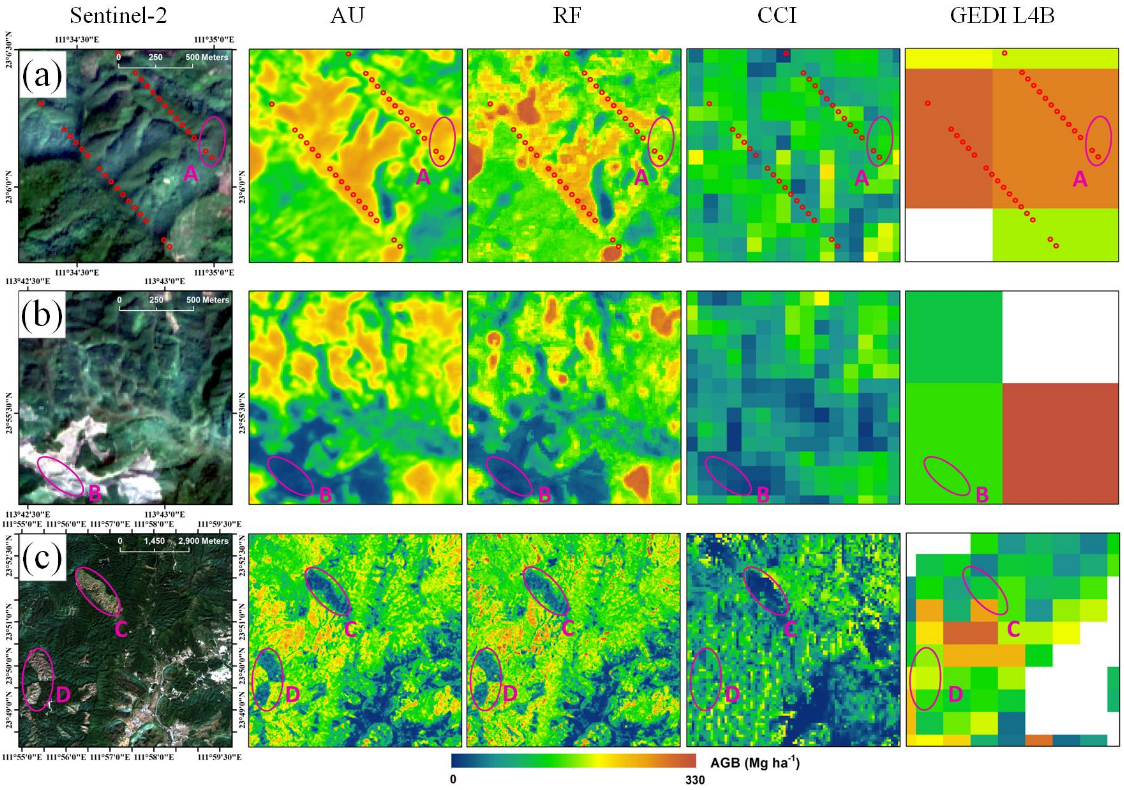


Figure 4.11: Comparison of AGB maps at three different locations: AU and RF (10m), CCI (100m), and GEDI L4B (1000m). The blank areas in the GEDI L4B map are due to data gaps. The red circles represent GEDI footprints. The pink circles represent four selected subregions used for comparing the AGB maps.

4.4 Discussion

4.4.1 Contributions of AU

AU was the best model we tested on the test dataset, exhibiting the highest R^2 value, alongside the lowest recorded RMSE and bias. For AGB mapping, high R^2 and low RMSE are desired to ensure precision, while low bias is required to avoid systematic under- or over-estimation when summing or averaging across values, as is critical when assessing carbon stocks and their changes through time. However, the use of deep learning methods for forest AGB estimation with GEDI data was not invariably superior to the RF approach. For instance, UNet exhibited a higher bias than RF, and ResNet101 showed inferior performance to RF in terms of R^2 and RMSE. The underlying reason for this disparity is that UNet lacks the capability to dynamically assign weight to the importance of different feature regions. Concurrently, ResNet is dependent solely on its sequentially stacked convolutional layers and lacks the skip connections inherent in UNet, which are crucial for amalgamating local and global contextual information.

Additionally, we employ a fully connected layer to transform the corresponding remote sensing imagery of each GEDI footprint into a patch suitable for deep learning applications. This methodology inherently precludes the utilization of spatial information within the data. Despite this limitation in exploiting spatial contextual cues inherent in the original data, the AU model on such transformed data still achieved superior accuracy over RF. This suggests that the superior performance of AU can be attributed not only to its effective utilization of spatial information between pixels but also to its exemplary integration of the advantages of attention mechanisms and the UNet architecture. The attention mechanism selectively emphasizes salient features and suppresses irrelevant ones, while the UNet encoder-decoder structure retains local and fine-grained spatial information through the skip connections. This allows AU to effectively learn multi-scale spatial features from the remote sensing inputs. Furthermore, both the Forest AGB map and the uncertainty map demonstrating high consistency with the widely-used RF attest to the reliability of the AU model. In summary, the AU model demonstrated good performance, proving to be a powerful tool for regression tasks, such as forest AGB estimation.

The apparent improvement in spatial detail observed in the AU-generated AGB maps (Fig. 4.11) is indeed promising. The model's ability to capture fine-scale variations in forest structure suggests a potential advancement in biomass mapping capabilities. However, the visual appeal and apparent detail of a map do not necessarily equate to improved accuracy or realism. Without comprehensive ground truthing, we cannot conclusively state that these finer details accurately represent real-world biomass distributions.

4.4.2 Efficiency analysis

Apart from accuracy, the efficiency of the models is also crucial, especially when deployed over large areas. We compared the training time of all the models, and the time taken to predict AGB across the entire province for the AU and RF. Both of the experiments were conducted on the server described in Section 4.2.3. The time consumption for training of all models, conducted through 5-fold cross-validation, along with the time spent by the AU and RF models in predicting AGB across the entire province using five different trained models, is shown in Table 4.1. During the training phase, it was observed that most deep learning models typically converged before or around 90 epochs; consequently, 120 epochs were selected as the maximum for training to ensure thorough learning without unnecessary computational expense. The deep convolutional networks entailed substantially longer training times than the conventional RF algorithm. Among the deep models, AU-FC took the most extensive computational time during training because each GEDI footprint was transformed into an individual patch. Our primary model, AU, utilized the second-longest duration for training, and notably, in comparison to UNet, the attention mechanism did not demand significantly additional time. When applying the 5 trained models to predict AGB across Guangdong province, the inference time of the AU and RF models was comparable. This is because the AU network utilized efficient GPU hardware acceleration, while the RF model leveraged multi-core CPU parallelism by distributing predictions across threads. Overall, the total training and inference time of AU was not substantially greater than the RF model. With future upgrades in hardware acceleration, such as utilization of high-performance computing clusters, the computational expenses of AU can potentially be further reduced. In particular, the inference speeds already matched those of RF, demonstrating the viability of deploying well-trained deep networks for expansive spatial and temporal predictions.

Table 4.1: Time consumption of the models (minutes)

Model	Training time	Prediction time
AU	587	1318
RF	201	1379
AU-FC	1570	-
UNet	560	-
ResNet101	350	-

4.4.3 Potential of deep learning for forest AGB estimation

The confluence of unprecedented data sources, augmented computational capabilities, and the latest advancements in deep learning presents exhilarating new opportunities to enhance our understanding of forest AGB derived from data. We have demonstrated through the use of FC layers that utilizing spatial information can improve the accuracy of AGB estimation. One of the advantages of deep learning is its capability to effectively utilize spatial information, presenting a promising future in the quantitative characterization of forest carbon dynamics. While advanced deep learning algorithms are continuously emerging in AI research (He et al., 2022; Liu, Lin, Cao, Hu, Wei, Zhang, Lin and Guo, 2021), they necessitate further refinement and adaptation for the field of remote sensing. This is particularly true for regression analysis tasks, such as estimating forest AGB using remote sensing data.

4.5 Conclusions

The deep learning approach we proposed offers an innovative computational methodology for estimating forest AGB, leveraging GEDI and remote sensing imagery. This study demonstrates the feasibility of mapping AGB over large areas by leveraging publicly accessible remote sensing data sources, including Sentinel-1, Sentinel-2, and ALOS-2 PALSAR-2, in conjunction with advanced deep learning methodologies and the open GEDI dataset as a reference. AU significantly reduces the overall bias in the estimation of forest AGB, concurrently enhancing R^2 and reducing RMSE. Moreover, it exhibits an enhanced capability to discern regions characterized by high and low AGB values with heightened precision. Given sufficient training data, such deep learning techniques may better capture complex ecological gradients and provide enhanced generalization across different forest types and disturbance conditions. Given the importance of AGB as an indicator of climate change impacts and mitigation, these methods for continuous mapping from freely available data represent a valuable tool for understanding and addressing pressing environmental challenges.

However, it is crucial to note that not all deep learning approaches can guarantee an enhancement in accuracy. Further research is imperative to devise suitable network architectures, optimize training strategies, including patch size, and to validate the methodologies across varied forest environments. Although deep learning algorithms typically demand more computational resources, it's worth noting that AGB maps do not necessarily require a high temporal resolution. In most cases, annual AGB maps are sufficient, and indeed a higher temporal frequency will often not be possible due to the need for 12 months of satellite data to achieve accurate results. Overall, the integration of emerging LiDAR datasets like GEDI with optical and SAR data through advanced deep learning has significant potential for improving large scale forest AGB estimation.

Chapter 5

Forest AGB change in China between 2007 and 2021

Abstract

Accurate estimation of forest AGB and its change is crucial for understanding the role of forests in the global carbon cycle and informing climate change mitigation strategies. However, large-scale mapping of forest AGB remains challenging due to the heterogeneity of forest ecosystems, the limited possible coverage of field data, and the limitations of remote sensing data to characterise AGB. This study presents a comprehensive assessment of forest AGB in China for the year 2021, and its changes since 2007, using a combination of the GEDI LiDAR data, multi-source satellite imagery, and machine learning algorithms. We developed a multi-step approach to estimate forest AGB at a national scale. First, we converted GEDI LiDAR metrics into AGB estimates using allometric equations specific to different regions of China, developed with the help of field plots. Next, we trained Random Forest models for each province using GEDI-derived AGB estimates and predictor variables from Sentinel-1, PALSAR-2, and Sentinel-2 data. The trained models were then applied to generate wall-to-wall AGB maps for each province and then merged into a national-scale AGB map. Our results show that the total forest AGB in China for 2021 was 13.06 Pg, with a mean AGB density of 108.75 Mg ha^{-1}. We observed an overall increase in carbon storage since 2007, with a net gain of 3.54 Pg C. The spatial patterns of AGB changes varied across the country, with the highest increases in the northeast and south central China. The validation of our

2021 AGB map using independent field data revealed the challenges associated with large-scale AGB estimation (R^2=0.56, RMSE = 49.8 Mg ha^{-1}), particularly in high-biomass forests where saturation limits the accuracy of remote sensing-based models. The generated high-resolution AGB map and the quantified biomass changes provide valuable insights into the spatial distribution and dynamics of forest biomass resources in China, supporting informed decision-making in forest management and climate change mitigation efforts.

5.1 Introduction

Forests are pivotal components of the global carbon cycle, acting as significant sinks and sources of atmospheric CO_2 (Mitchard, 2018; Pan et al., 2011). The dynamics of forest carbon stocks are influenced by a range of factors that can lead to either gains or losses in carbon storage (McNicol et al., 2018; Xu et al., 2021). Afforestation/reforestation efforts, which involve the establishment of new forests through planting or natural regeneration on currently non-forested lands, contribute to an increase in carbon stocks by creating new biomass pools (Fang et al., 2004; Bukoski et al., 2022). Similarly, the natural growth and accumulation of biomass in existing forests, driven by biological and ecological processes, can also result in an augmentation of carbon stocks over time (Yang et al., 2023). Conversely, deforestation and degradation collectively drive a significant diminution in forest carbon stocks (Harris et al., 2021; Sze et al., 2022). These processes occur through the removal of aboveground and belowground biomass, disturbance of soil carbon pools, and various anthropogenic and natural disturbances such as selective logging, fires, or insect outbreaks (Pearson et al., 2017; Qin et al., 2021).

Accurate quantification of forest AGB and its spatiotemporal dynamics is critical for understanding the terrestrial carbon budget, evaluating the potential of forests in mitigating climate change, and informing sustainable forest management strategies (Bustamante et al., 2016). This is particularly relevant in rapidly developing regions like China. China's forests are known to be changing rapidly, due to rapid landcover change, logging, large-scale afforestation/reforestation programs, commercial plantations, and climate change (Viña et al., 2016; Ke et al., 2019; Lü et al., 2012). While there is some certainty that the overall area of forest in China has increased in recent decades (Liu, 2014; Yu, Ciais, Piao, Houghton, Lu, Tian, Agathokleous, Kattel, Sitch, Goll et al., 2022), there is great uncertainty in the changes in carbon stocks of the forests of China, as opposed to changes in forest area. Similarly, there is uncertainty over which regions of China, that have diverse climate and land use histories, have higher or lower carbon stocks or rates of carbon stock change. It is likely that, due to this forest growth, China's land surface has recently become a carbon sink, contributing positively to the global carbon cycle (Piao et al., 2009; Xu, Yu, He, Wang, Gao, Wen, Li, Niu and Ge, 2018).

Remote sensing techniques have emerged as valuable tools for large-scale monitoring of forest structural attributes, including AGB, offering unparalleled spatial coverage and temporal resolution (Saatchi et al., 2011; Baccini et al., 2012; Santoro et al., 2022). Active remote sensing sensors, particularly light detection and ranging (LiDAR), have demonstrated remarkable potential for estimating AGB across various forest ecosystems (Dubayah et al., 2022). LiDAR systems can directly measure the three-dimensional structure of forest canopies, providing accurate estimates of canopy height and vertical distribution, which are strongly correlated with AGB (Dubayah et al., 2020). However, LiDAR data alone cannot provide wall-to-wall coverage of large regions due high acquisition costs from aircraft, and limited spatial coverage from space (Wulder et al., 2012). SAR data, particularly at longer wavelengths (e.g., L-band and P-band), can penetrate through the forest canopy and interact with the trunks and branches, enabling the retrieval of AGB information (Huang et al., 2018; Woodhouse et al., 2012; Mitchard et al., 2011). Multispectral passive remote sensors capture the spectral reflectance of solar radiation across various wavelengths, which is then used to infer the biophysical characteristics of vegetation. However, the ability of both multi-spectral and Synthetic Aperture Radar (SAR) technologies to estimate AGB is constrained by highly variable relationships by location and through time, and a diminishing sensitivity as biomass levels rise, a phenomenon often referred to as ”saturation” (Mitchard et al., 2012; Rodríguez-Veiga et al., 2016). A promising approach to overcome the limitations of individual sensors is to combine LiDAR, SAR, and passive optical data for estimating forest biomass. Numerous studies have demonstrated that the integration of these multi-sensor measurements can significantly improve the accuracy and reduce the uncertainties in AGB estimates (Zolkos et al., 2013; Dong, Mitchard, Yu, Hancock and Ryan, 2023).

In this study, we aim to quantify the changes in forest AGB across China between 2007 and 2021 by generating a new national-scale AGB map for the year 2021 and comparing it to an existing map, created using similar methods, for 2007 (Dong et al., 2024). We leverage GEDI data, in combination with Sentinel-1 C-band SAR, Sentinel-2 multispectral imagery, and PALSAR-2 L-band SAR data, to estimate forest AGB using machine learning algorithms. By comparing the 2021 AGB map with the previously published 2007 map (Dong et al., 2024), we assess the spatiotemporal patterns of AGB changes over the past 14 years. Our specific objectives are to: (1) develop a robust methodology for integrating GEDI LiDAR data with multi-sensor satellite imagery to estimate forest AGB across China in 2021; (2) evaluate the accuracy of the resulting AGB map using 20% of the GEDI data and independent field inventory data; (3) analyze the spatial patterns and magnitude of AGB changes between 2007 and 2021; and (4) explore the relationships between AGB changes and potential drivers, such as land-use change, climate variability, and forest management practices.

5.2 Methods

This study employed a multi-step approach to estimate AGB in China for 2021 and analyze its changes since 2007. First, GEDI L2A RH metrics were converted into AGB estimates using allometric equations we developed in Chapter 2 specific to different regions of China. Next, Random Forest models (Breiman, 2001) were trained for each first-level administrative unit in China. This included predictions for all 23 provinces and 5 autonomous regions individually. Additionally, municipalities were individually predicted, except for Chongqing, while other municipalities and Special Administrative Regions (SARs) were merged with adjacent units (see Appendix 1) to predict AGB using GEDI-derived AGB estimates and predictor variables from Sentinel-1, PALSAR-2, and Sentinel-2. The trained models were then applied to generate wall-to-wall AGB maps at 25 m resolution for each of the 29 level 1 units (hereafter provinces for simplicity). To ensure a smooth transition between neighboring provinces, a 2000-meter buffer was applied along the provincial boundaries during the AGB estimation process. When merging the provincial AGB maps, overlapping areas were handled by calculating the mean AGB values, resulting in a seamless national-scale AGB map. These provincial AGB maps were subsequently merged to create a national-scale map.

To facilitate the analysis of AGB changes in conjunction with the 2007 AGB map at 50 m resolution (Dong et al., 2024), the 2021 AGB map was resampled to 50 m using a cubic interpolation method. Forest extents were delineated using the tree cover layer from the European Space Agency's (ESA) worldcover 2021 product (Zanaga et al., 2022). Finally, the 2021 forest AGB map was compared with the 2007 map to quantify and analyze changes in forest AGB over the 14-year period.

5.2.1 Remote sensing data

LiDAR data

GEDI is a spaceborne LiDAR mission launched in December 2018 to provide high-resolution observations of forest vertical structure and canopy characteristics. GEDI is mounted on the International Space Station (ISS) and collects data globally between 51.6° N and 51.6° S latitudes (Dubayah et al., 2020). The GEDI instrument employs three lasers, each with a nominal footprint diameter of 25 m on the Earth's surface, separated by approximately 60 m along-track and 600 m across-track, resulting in a total of eight tracks of data per ISS orbit (Hancock et al., 2019). GEDI measures the waveform of the returned laser energy, which provides vertical information on forest structure, including canopy height, canopy cover, and the vertical distribution of leaf and woody material (Marselis et al., 2018). For the purpose of this study, the analysis was constrained to GEDI data footprints collected between the years 2020 and 2022. To accurately estimate AGB, we removed the poor-quality GEDI footprints (Chen, Dong, Yu, Woodhouse, Ryan, Liu, Georgiou and Mitchard, 2023), resulting in a total of 5,749,039 footprints used across China. We converted the GEDI L2A RH metrics into AGB using the allometric equations established for northern and southern China (Chapter 3).

SAR data

SAR data from two satellite missions, ALOS-2 PALSAR-2 and Sentinel-1, were used in this study to capture the structural properties of forest canopies. SAR sensors are active remote sensing instruments that emit microwave pulses and record the backscattered energy from the Earth's surface (Woodhouse, 2017). ALOS-2 PALSAR-2 is an L-band SAR sensor operated by the JAXA, which provides global observations with a spatial resolution of 25 m (Shimada et al., 2014). L-band SAR (wavelength 23 cm) has a higher penetration depth into forest canopies compared to shorter wavelengths, making it more sensitive to the woody components of vegetation (Mitchard et al., 2012). In this study, we used the median values of ALOS-2 PALSAR-2 yearly mosaic product for 2020 - 2022, which includes HH, HV polarizations and their ratio (HV divided by HH). Sentinel-1 is a C-band SAR mission operated by the ESA, consisting of two satellites (Sentinel-1A and Sentinel-1B) that provide global coverage with a revisit time of 6-12 days (Torres et al., 2012). The Sentinel-1 data used in this study were acquired in the Interferometric Wide (IW) mode, which has a spatial resolution of 10 m and includes VV, VH polarizations and their ratio (VV divided by VH). C-band SAR (wavelength 5.6 cm) is more sensitive to the upper canopy structure and leaves compared to L-band SAR (Sinha et al., 2015). The combination of L-band and C-band SAR data provides complementary information on forest structure, as they interact with different components of the forest canopy.

Passive optical data

Optical remote sensing data from the Sentinel-2 mission were used in this study to capture the spectral properties of forest canopies. Sentinel-2 is a multispectral imaging mission operated by the ESA, consisting of two satellites (Sentinel-2A and Sentinel-2B) that provide global coverage with a revisit time of 5 days at the equator (Gascon et al., 2017). In this study, we used the Sentinel-2 L2A product, which provides surface reflectance images that are radiometrically and geometrically corrected (Main-Knorn et al., 2017). To minimize the influence of cloud cover and seasonal variability on the AGB estimates, we generated median composite images using all cloud-free Sentinel-2 scenes from 2020 to 2022. For provinces with prolonged snow cover during the winter months, we excluded the winter scenes from the compositing process to avoid the confounding effects of snow on the spectral signatures of forests (Wang, Xiao, Bajgain, Starks, Steiner, Doughty and Chang, 2019).

5.2.2 Estimation of AGB changes and constraining

To estimate the changes in forest AGB between 2007 and 2021, we first resampled the 25 m resolution 2021 AGB map generated in this study to 50 m using a cubic interpolation method to match the resolution of the 2007 AGB map (Dong et al., 2024). We then extracted the forest AGB values from the resampled 2021 map using the forest extent derived from the European Space Agency's (ESA) worldcover 2021 product (Zanaga et al., 2022). Subsequently, the 2007 forest AGB map was subtracted from the resulting 50 m resolution forest AGB map for 2021 to calculate the AGB changes over the 14-year period. However, due to various sources of uncertainty, such as model limitations, saturation effects (likely impacting at a lower value in 2007 than 2021 due to improvements in satellite sensors available), and different allometric equations(Mitchard et al., 2013; Rodríguez-Veiga et al., 2017), some areas exhibited unrealistically large differences in AGB over the 14-year period.

To address this issue and ensure more realistic estimates of AGB changes, we applied a thresholding approach based on the maximum feasible growth rates of forests. In tropical regions, forest AGB accumulation can even exceed 10 Mg ha^{-1} yr^{-1} (Feldpausch et al., 2007; Requena Suarez et al., 2019), while boreal forest ecosystems generally exhibit lower rates of AGB accumulation (Gower et al., 2001; Jarvis and Linder, 2000). A study on forest biomass carbon pool changes in China, conducted at a 1 km resolution, showed that the maximum growth rate could reach 2 Mg C ha^{-1} yr^{-1} (Chen, Feng, Fu, Ma, Zohner, Crowther, Huang, Wu and Wei, 2023). Taking into account the high productivity of tropical forests, the existing estimates of forest carbon pool changes in China, and the high spatial resolution of our study, we set an upper limit of 5 Mg ha^{-1} yr^{-1} for AGB increase. This threshold allows for the reasonable constraint of overestimated AGB growth values while still capturing the dynamic changes in forest biomass at local scales. We applied this threshold to the calculated AGB changes, setting any value of ΔAGB exceeding 70 Mg ha^{-1} (i.e., 5 Mg ha^{-1} yr^{-1} × 14 years) to 70 Mg ha^{-1}. The resulting constrained ΔAGB map was then added to the 2007 forest AGB map to obtain the modified 2021 forest AGB map.

To address the impact of the chosen 70 Mg ha^{-1} threshold on our results, we conducted a sensitivity analysis. We applied a range of thresholds (50, 60, 70, 80, and 90 Mg ha^{-1}) to the calculated AGB changes and compared the resulting AGB change maps. Additionally, we created an AGB change map without applying any threshold.

5.3 Results

5.3.1 Model performance

The random forest models developed for each province in China to estimate forest AGB exhibited varying levels of performance when evaluated using the independent test dataset, which consisted of 20% of the GEDI footprints. At the national scale, the model achieved an R^2 of 0.52 and a RMSE of 54.8 Mg ha^{-1} (Fig.5.1). The national-scale scatter plot reveals a generally good agreement between the predicted and observed AGB values, with a majority of the points clustered around the 1:1 line. However, there is a noticeable tendency towards overestimation in the lower AGB ranges and underestimation in the higher AGB ranges. This pattern is likely due to the saturation effect of the remote sensing predictors in dense forest canopies with high biomass levels (Rodríguez-Veiga et al., 2016), difficulties of prediction at very low biomass values, and the general tendency of machine learning predictions to drag extreme values to the mean (Breiman, 2001).

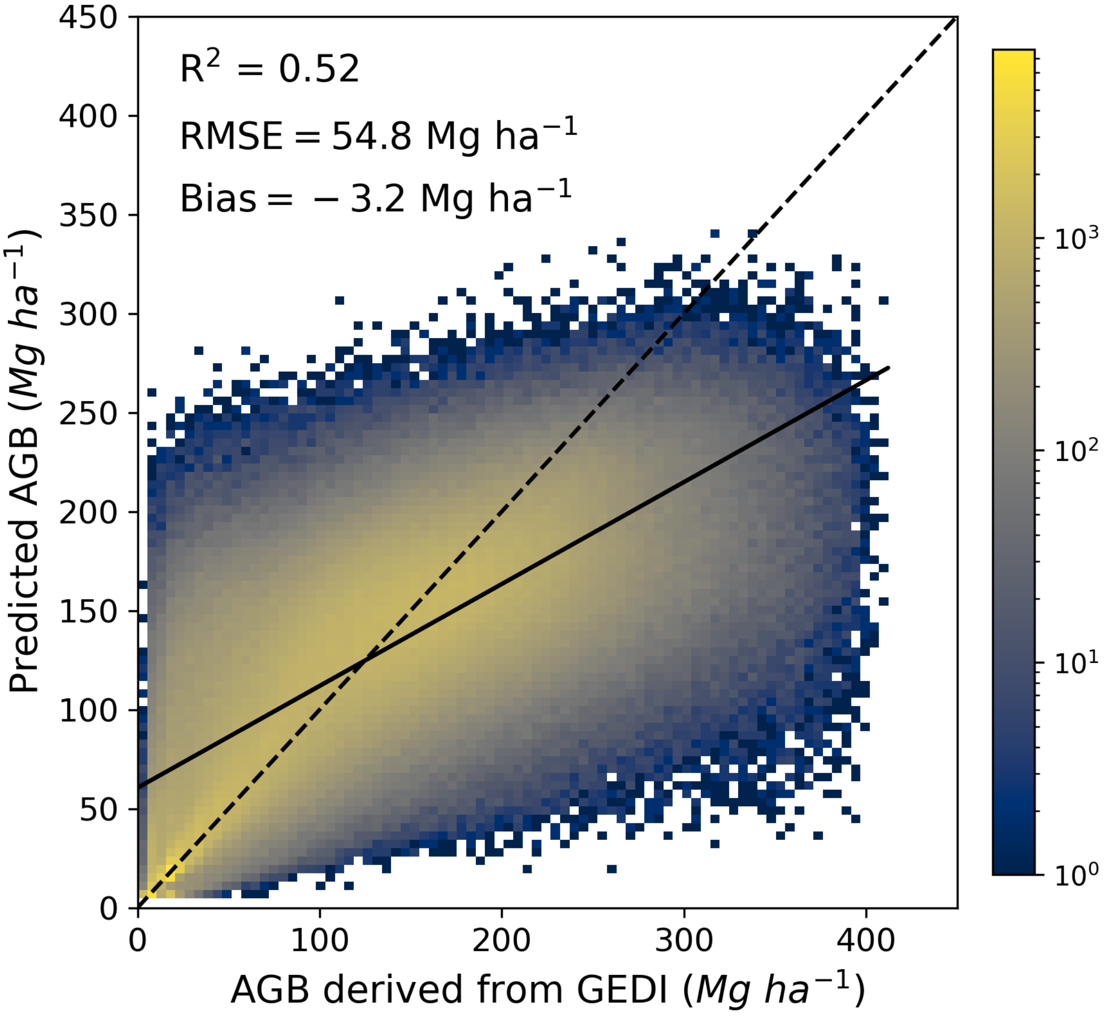


Figure 5.1: Comparison between predicted AGB using random forest and AGB derived from GEDI. Colours show the density of pixels in each bin. Dashed line shows 1:1 and solid line the line of best fit through the data.

At the provincial level, the model performance varied considerably (Appendix 1). Some provinces, such as the more arid Xinjiang and Neimenggu, exhibited strong relationships between the predicted and observed AGB values, with R^2 values of 0.89 and 0.88, and RMSE values of 18.6 Mg ha^{-1} and 18.4 Mg ha^{-1}, respectively. In contrast, wetter and more tropical provinces, like Zhejiang and Yunnan, had lower R^2 values of 0.32 and 0.50, and higher RMSE values of 59.4 Mg ha^{-1} and 69.9 Mg ha^{-1}. These variations in model performance might be related to differences in forest types, topography, average and maximum biomass, and the quality and quantity of the GEDI footprints available for each province.

5.3.2 Evaluation of AGB map

To further assess the accuracy of the 25 m resolution national AGB map, we conducted an independent validation using forest inventory data from two distinct regions in China: the Northeast and the Southwest. In the Northeastern region, 26 field plots were used for validation, while 16 field plots were used in the Southwestern region. These forest plots were not used in the model training process and served as an independent dataset for evaluating the performance of the national AGB map.

By combining the field inventory data from both regions, we obtained a total of 42 validation plots. The validation results for the 25 m resolution map yielded an R^2 value of 0.55 and an RMSE of 50.6 Mg ha^{-1} (Fig. 5.2 (a)). The R^2 value suggests that the national AGB map captures a significant portion of the spatial variability in forest biomass across the Northeastern and Southwestern regions. However, the RMSE of 50.6 Mg ha^{-1} indicates that there is still some discrepancy between the predicted and observed values, which may be attributed to the diverse forest types, topography, and environmental conditions encountered across these regions. In addition, the validation results reveal a noticeable saturation point around 150 Mg ha^{-1} (Fig. 5.2), indicating a limitation of the national AGB map in accurately estimating biomass values beyond this threshold.

Given the concerns about resolution discrepancies when comparing the 25 m 2021 AGB map with the 50 m 2007 AGB map, we undertook additional analyses. Firstly, the 25 m resolution AGB map was resampled to 50 m resolution, after which we conducted a new validation. The validation results for the resampled 50 m map yielded an R^2 value of 0.56 and an RMSE of 50.1 Mg ha^{-1} (Fig. 5.2 (b)), which were comparable to the original 25 m resolution map, indicating that the resampling process did not significantly impact the accuracy.

To further explore the impact of resolution, we directly trained and estimated AGB using data at a 50 m resolution, bypassing the 25 m step altogether. The validation results for this directly estimated 50 m resolution map, however, showed a lower R^2 value of 0.49 and a higher RMSE of 54.0 Mg ha^{-1} (Fig. 5.2 (c)), suggesting that estimating AGB directly at a coarser resolution resulted in a loss of spatial detail and overall accuracy compared to the approach of first estimating at 25 m and then resampling.

These findings underscore the importance of initially working at a finer resolution before resampling to match the reference dataset's resolution. While resampling from 25 m to 50 m retained much of the accuracy, directly modeling at the coarser resolution resulted in a loss of detail that impacted the model's predictive capability. This suggests that the approach of first generating high-resolution estimates and then adjusting to match historical datasets provides a better balance between maintaining spatial detail and achieving compatibility across different years.

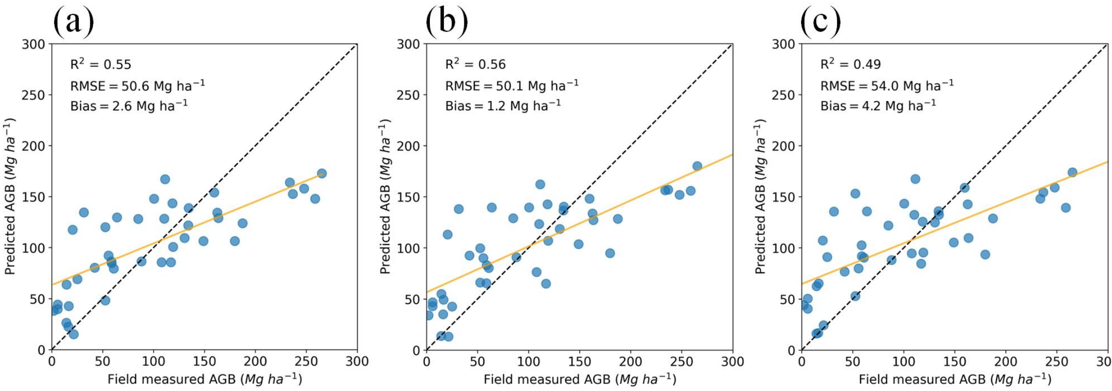


Figure 5.2: Comparison of predicted AGB against field-measured AGB across different spatial resolutions and modelling approaches. (a) Validation of the 25 m resolution AGB map, (b) Validation of the resampled 50 m resolution AGB map, and (c) Validation of the directly trained and estimated 50 m resolution AGB map. Each dot indicates one of the 42 validation plots. The dashed line shows 1:1 and yellow line shows the least squares regression fit to the data.

5.3.3 Forest AGB in China for the year 2021

The total forest AGB in China for the year 2021 was estimated to be 27.79 Pg, with a mean forest AGB density of 108.8 Mg ha^{-1}. To convert the AGB estimates to carbon storage, we applied a conversion factor of 0.47 (Eggleston et al., 2006). Based on this conversion, we estimated that the total carbon stock in China's forests in 2021 was 13.06 Pg C. The histogram of forest AGB distribution in China for the year 2021 (Fig. 5.3 a) reveals the majority of forest pixels having AGB values ranging from 50 to 200 Mg ha^{-1}. The histogram also shows a notable peak at the AGB value of 70 Mg ha^{-1}, which can be attributed to the application of the maximum AGB increase threshold of 70 Mg ha^{-1} over the 14-year period when estimating the AGB changes from 2007 to 2021, with values that were zero at the starting time point.

One of the reasons for the high frequency of pixels with an AGB value of 70 Mg ha^{-1} is the difference in forest extent between the ESA worldcover 2021 product (Zanaga et al., 2022) used for masking the 2021 AGB map and the forest cover dataset (Hansen et al., 2013) used for masking the 2007 AGB map. To a small degree this may be due to methodological and resolution differences between the two datasets, but to a larger degree this difference is likely driven by an increase in forest cover (particularly plantations and other planted or naturally regenerated forests) in China over the period, which has been discussed elsewhere (Yu, Ciais, Piao, Houghton, Lu, Tian, Agathokleous, Kattel, Sitch, Goll et al., 2022; WangYL et al., 2022; Cheng et al., 2023). Consequently, these areas will exhibit an AGB increase from 0 in 2007 to their actual AGB value in 2021, which could exceed the 70 Mg ha^{-1} threshold, and thus be constrained to 70 Mg ha^{-1} by our method.

To better understand the AGB changes within the consistent forest extent (areas that were forest in both 2007 and 2021), we reanalyzed AGB changes from 2007 to 2021 using only the pixels that were classified as forests in the 2007 Hansen forest extent (Fig. 5.3 b). The resulting histogram shows a more balanced distribution of AGB values, with an average AGB of 125.2 Mg ha^{-1}.

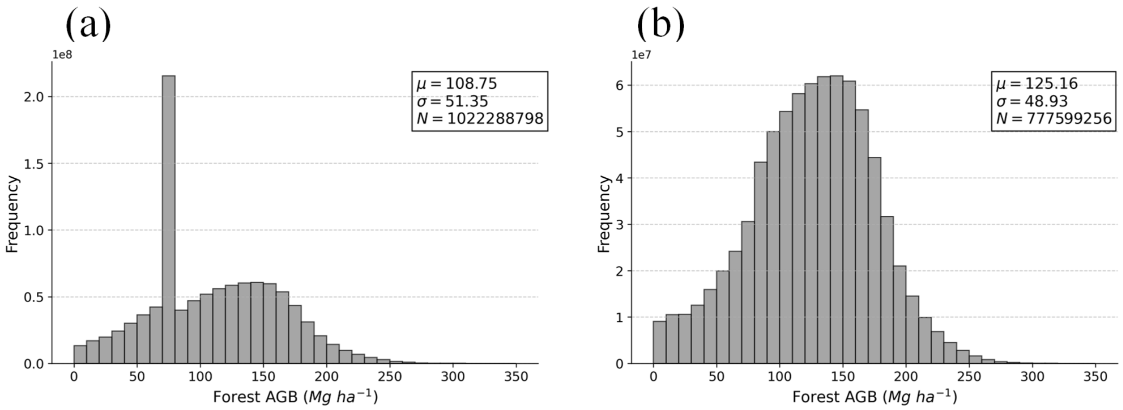


Figure 5.3: Frequency distribution of forest AGB in China for 2021 using different forest extent masks. Histogram of forest AGB distribution in 2021 using (a) the ESA worldcover 2021 forest mask (including new forest between 2007 and 2021, which was limited by our analysis to 70 Mg ha^{-1}), (b) the Hansen forest extent mask.

The modified forest AGB map for China in 2021 (Fig. 5.4) reveals the spatial distribution of aboveground biomass across the country's diverse forest ecosystems. The map highlights a distinct spatial pattern, with the highest AGB values mainly concentrated in the southern regions of China, particularly in the provinces of Yunnan, Sichuan, South East Xizang, and

Taiwan. These areas are characterized by a combination of favorable climatic conditions, complex topography, and diverse forest types, which contribute to the high accumulation of forest biomass (Xu, Shi, Fang, Zhou, Xu, Zhou, Tao, Ji, Xu, Li et al., 2018). The humid subtropical and tropical climates in these regions support the growth of dense, species-rich forests, such as subtropical evergreen broadleaf forests and tropical rainforests (Guo et al., 2013). In addition to southern China, high AGB densities were also observed in the eastern parts of Jilin Province. The high AGB values in eastern Jilin Province can be attributed to the presence of temperate mixed forests and old-growth forests, which have accumulated significant amounts of biomass over time (Zhang et al., 2017).

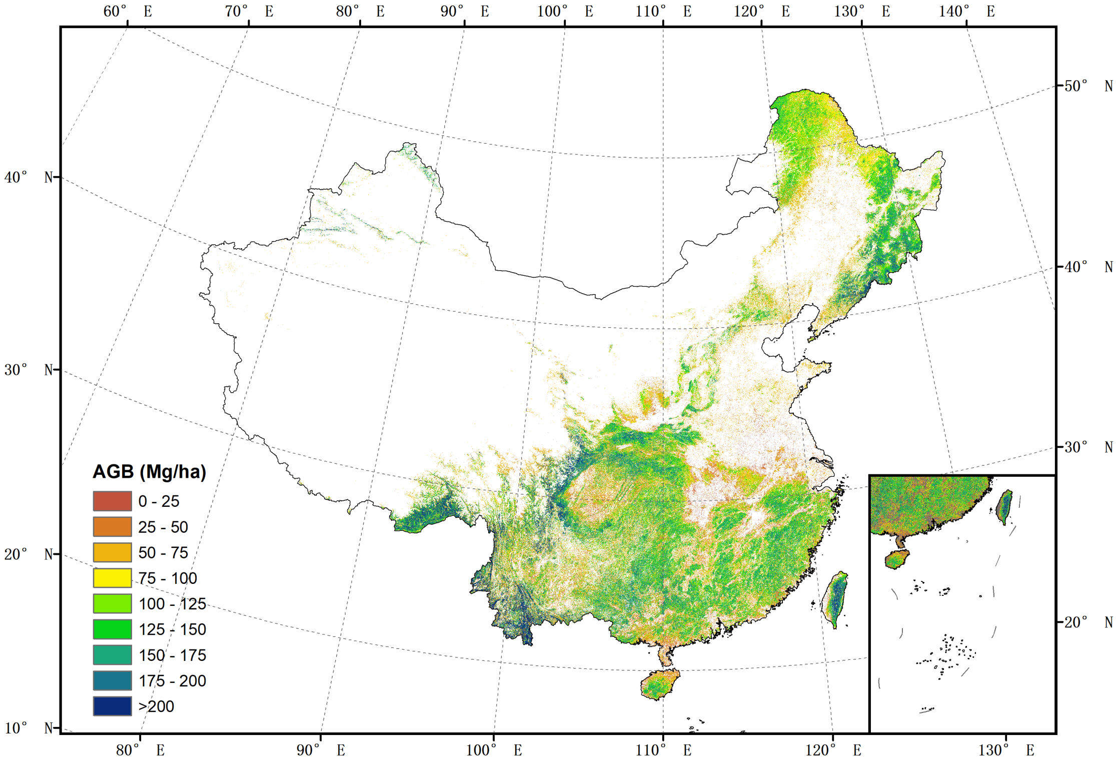


Figure 5.4: Distribution of forest AGB in China for 2021.

5.3.4 Forest AGB change in China from 2007 to 2021

To analyze the forest AGB change in China between 2007 and 2021, we subtract the 2007 forest AGB map from the modified 2021 forest AGB map (Fig. 5.5). Substantial AGB increases can be observed in the northeastern and central to southern China, while overall southwestern region show moderate losses.

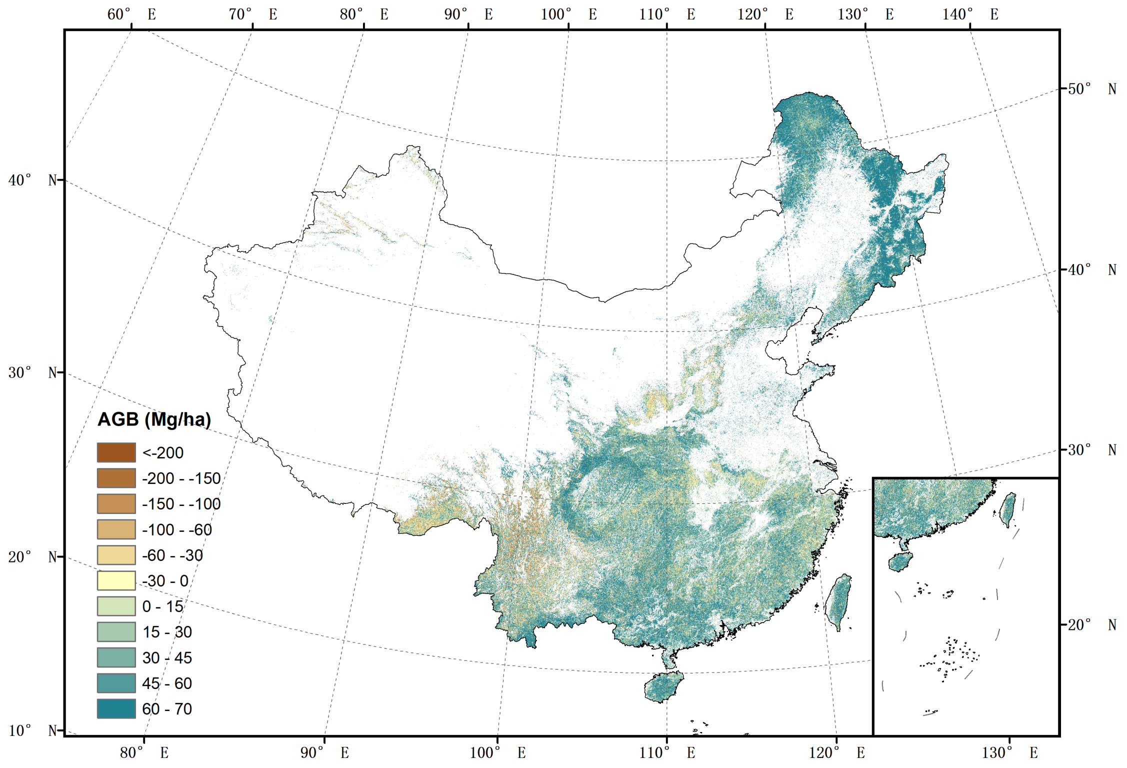


Figure 5.5: Forest AGB change between 2021 and 2007. The 2021 AGB map was masked using ESA worldcover 2021 map, and the 2007 AGB map was masked using Hansen 2007 map.

To quantify the changes in forest AGB and carbon storage in China between 2007 and 2021, we compared the AGB estimates derived from different forest extent maps 5.1. The 2007 AGB map was masked using the Hansen et al. (2013) forest cover map, while for the 2021 AGB map, we used both the Hansen et al. (2013) forest cover map and the ESA worldcover 2021 map to assess the influence of different forest extent scenarios on the estimated changes.

The results show that when considering only the area that was forest in 2007, the forest area remained relatively stable, decreasing slightly from 194.53 × 10^6 ha in 2007 to 194.40 × 10^6 ha in 2021. However, the average AGB density increased from 104.10 Mg ha^{-1} in 2007 to 125.16 Mg ha^{-1} in 2021, resulting in a total carbon stock increase in these stable forests from 9.52 Pg C to 11.44 Pg C. This increase of 1.92 Pg C can be attributed to the growth of existing forests within the consistent forest extent. When using the ESA worldcover 2021 map for the 2021 AGB estimates, thus including new forest, the forest area increased substantially to 255.57 × 10^6 ha, while the average AGB density (108.75 Mg ha^{-1})

was lower compared to the 2021 estimates based on the Hansen map. This difference can be explained by the inclusion of additional new forest areas in the ESA worldcover 2021 map, compared to the Hansen 2007 extent, which may encompass younger or less dense forests. Despite the lower average AGB density, the total carbon stock estimated using the ESA worldcover 2021 map was higher (13.06 Pg C) due to the larger forest area.

Table 5.1: Changes in forest AGB and carbon storage in China from 2007 to 2021 under different forest extent scenarios.

	2007 AGB map	2021 AGB map	2021 AGB map
Forest cover mask	Hansen forest cover 2007	Hansen forest cover 2007	ESA worldcover 2021
Forest area (10^6 ha)	194.5	194.4	255.6
Average AGB (Mg ha^{-1})	104.10	125.16	108.75
Total carbon (Pg C)	9.52	11.44	13.06

5.4 Discussion

5.4.1 Implications of Using the Same Data for Calibration and Validation

The methodology employed in this study, utilizing an 80:20 split of the dataset for model training and validation respectively, warrants critical examination due to its potential implications for result interpretation and model performance assessment. While this approach provides insights into model efficacy, it is imperative to consider the limitations and potential biases inherent in using subsets of the same dataset for both calibration and validation purposes.

A primary concern with this methodology is the potential lack of true independence between the training and validation datasets. Despite random allocation of data points to these subsets, both originate from the same spatial and temporal domain, potentially sharing common characteristics inherent to the data collection process and environmental conditions during the study period. This shared context may lead to an overly optimistic assessment of model performance, as the validation data may not fully represent the challenges the model would face when applied to truly independent datasets.

The issue is further compounded by spatial autocorrelation within the dataset. In remote sensing-based biomass estimation, proximate data points often exhibit similar environmental characteristics and biomass values. This spatial dependency can result in artificially high agreement between predicted and observed values in the validation process, potentially underestimating prediction errors when the model is applied to new, spatially distinct areas.

However, our results have been validated using independent field data, which strengthens the reliability of our findings. The results of this validation demonstrated the reliability of our model, showing its ability to accurately estimate biomass.

5.4.2 Saturation effect and limitations in high-density forest areas

The validation results of the 2021 national AGB map (Fig. 5.2) suggested a reduction in sensitivity in the map around 150 Mg ha^{-1}, indicating a limitation in accurately estimating biomass values in high-density forest areas. However, the validation against GEDI footprints (Fig. 4.6) suggested sensitivity to biomass continues until about 250 Mg ha^{-1}, without a great increase in uncertainty towards the top of the range (uncertainty if fairly high throughout). This saturation effect is a common challenge in remote sensing-based AGB estimation, particularly when using optical and radar data, as the sensitivity of the remote sensing signals to changes in AGB diminishes with increasing biomass (Rodríguez-Veiga et al., 2017), though the use of multiple EO datasets does reduce this to some degrees (Saatchi et al., 2011). Despite this limitation, the overall validation results demonstrate the effectiveness of the national AGB map in capturing the spatial patterns of forest biomass across a wide range of AGB values, providing valuable information for understanding the distribution and dynamics of forest biomass in China.

Comparing the 2021 AGB map with the previously published 2007 AGB map (Dong et al., 2024), we observed that the maximum AGB values in the 2007 map were generally lower than those in the 2021 map. This difference can be explained by either the growth of forest (Chen, Feng, Fu, Ma, Zohner, Crowther, Huang, Wu and Wei, 2023), but also potentially by the methodological approaches and data used in generating the two maps. For the 2007 AGB map, which was based on ICEsat GLAS, ENVISAT ASAR, ALOS-1 PALSAR-1 and Landsat-5, limitations in the quality of the inputs meant that forest height was limited to 25 m when averaged in 0.01 degrees x 0.01 degrees grid cells. While this approach effectively mitigated the potential overestimation of AGB in areas with tall trees, it also introduced a saturation effect that limited the map's ability to capture the full range of AGB values in high-density forests. In contrast, the 2021 AGB map was generated using

GEDI LiDAR data, which has a smaller footprint diameter of 25 m, along with Sentinel-1 C-band SAR, Sentinel-2 multispectral imagery, and PALSAR-2 L-band SAR data. The higher spatial resolution and direct measurement of forest vertical structure by GEDI allowed for a more accurate estimation of AGB, particularly in high-density forest areas (Dubayah et al., 2022; Duncanson et al., 2020). In conclusion, while the 2021 national AGB map represents an improvement over the 2007 map in terms of spatial resolution and accuracy, and sensitivity at high biomass, the spatial variability of biomass in high-density forest areas remains a challenge to any satellite-based estimation of forest biomass.

5.4.3 Challenges in large-scale forest AGB estimation

The variability in RMSE observed between the national scale and individual provinces underscores the complexities involved in estimating forest AGB across a vast and diverse landscape, where regional differences in forest characteristics and data quality can impact the model's performance at larger scales (Section 5.3.1). When the test sets from all provinces are aggregated, the resulting dataset encompasses a wider range of forest types, environmental conditions, and biomass values, leading to higher overall errors (Dai et al., 2011; Rodríguez-Veiga et al., 2019). In addition, the quality and representativeness of GEDI data and remote sensing data (Appendix 2) for model training and validation can vary between provinces, further contributing to the higher RMSE at the national scale (Duncanson et al., 2022). Despite the regional differences in model performance, the overall accuracy achieved at the national scale demonstrates the effectiveness of integrating GEDI LiDAR data with multi-source satellite imagery and machine learning algorithms for large-scale forest AGB estimation. The province-specific models developed in this study offer a more targeted approach that accounts for the unique characteristics of each region, ultimately contributing to a more accurate and spatially detailed assessment of China's forest biomass resources (Ståhl et al., 2016).

5.4.4 Forest growth rates

In this study, we imposed a maximum growth rate threshold of 5 Mg ha^{-1} yr^{-1} between 2007 and 2021, resulting in a cap of 70 Mg ha^{-1} over the 14-year period. This threshold was applied to limit unrealistic growth rates that were observed when comparing the two biomass maps, which indicated over 100 Mg ha^{-1} of biomass accumulation in some regions. These inflated values were partly due to scatter and differences in the saturation points between the two maps. Additionally, another significant factor contributing to these unrealistic growth rates was the use of different forest cover maps for 2007 and 2021. The differences in forest cover maps likely introduced discrepancies that led to exaggerated biomass changes in certain areas.

To empirically assess the impact of the 70 Mg ha^{-1} threshold, we conducted further analysis using alternative thresholds of 60 Mg ha^{-1} and 80 Mg ha^{-1}. When the 2021 AGB map was masked using the ESA WorldCover forest cover product, applying a 60 Mg ha^{-1} threshold resulted in a total carbon estimate that was 0.46 Pg C lower than that obtained with the 70 Mg ha^{-1} threshold, while the 80 Mg ha^{-1} threshold produced an estimate that was 0.39 Pg C higher than the 70 Mg ha^{-1} result (Table. 5.1, Table. 5.2). However, when the Hansen Global Forest Change dataset was used to mask the 2021 AGB map, the impact of the threshold change was less pronounced: the 60 Mg ha^{-1} threshold resulted in 0.24 Pg C less carbon than the 70 Mg ha^{-1} threshold, and the 80 Mg ha^{-1} threshold resulted in 0.17 Pg C more carbon (Table. 5.1, Table. 5.2).

Table 5.2: Comparison of Forest Area, Average AGB, and Total Carbon Estimates for 2021 Using Different Biomass Change Thresholds (60 Mg and 80 Mg) and Forest Cover Masks (ESA WorldCover and Hansen Global Forest Change)

	60 Mg threshold		80 Mg threshold	
	Hansen	ESA	Hansen	ESA
Forest area (10^6 ha)	194.4	255.6	194.4	255.6
Average AGB (Mg ha^{-1})	122.59	104.89	127.09	111.92
Total carbon (Pg C)	11.20	12.60	11.61	13.45

These findings illustrate that the choice of threshold has a significant impact on the estimated total carbon when using the ESA forest cover map, but this impact is somewhat mitigated when the Hansen forest cover map is applied. The differences observed can be attributed to the varying forest extents of the two forest cover products. When using the Hansen forest cover map, the different thresholds have little impact on the total carbon estimate, further confirming that the differences between our 2007 AGB map and the 2021 AGB map generally align with ecological patterns.

5.4.5 Implications of not propagating uncertainty in biomass estimation

The approach employed in this study, while providing valuable insights into biomass distribution, does not explicitly propagate uncertainty through the various stages of the estimation process. This limitation warrants careful consideration, as it has several important implications for the interpretation and application of our results.

In our biomass estimation approach, uncertainties are introduced at multiple stages: in field measurements, in the derivation of allometric equations, in the GEDI RH metrics, and in the machine learning models used for upscaling. We have made efforts to minimize the uncertainties at each step, such as using local allometric equations for China to calculate biomass and removing low-quality GEDI footprints. However, it is inevitable that some uncertainty remains. By not propagating these uncertainties through each step of the process, we risk underestimating the total uncertainty in our final biomass estimates.

It is important to note that a comprehensive quantification and propagation of uncertainties throughout the entire estimation process was beyond the scope of this study. Such an undertaking would require significant additional resources, including more extensive field data collection, development of complex uncertainty models for each step of the process, and potentially the use of computationally intensive methods

To address these limitations in future studies, several approaches could be considered:

1. Implementation of a Monte Carlo simulation framework to propagate uncertainties from field measurements through to the final biomass estimates.

2. Utilization of Bayesian hierarchical models that can naturally incorporate and propagate uncertainties at different levels of the estimation process.

3. Development of ensemble methods that capture model-related uncertainties in the upscaling process.

5.4.6 Ecological and historical context of biomass increase in Chinese forests

The observed ubiquitous increase in forest biomass across China, as indicated by our results, necessitates a deeper examination of the underlying ecological and historical factors. This trend is likely the result of a complex interplay of natural processes, anthropogenic influences, and environmental changes occurring over the past several decades.

One significant factor contributing to biomass increase is the implementation of large-scale afforestation and reforestation programs in China since the 1970s (Liu et al., 2008). Initiatives such as the Three-North Shelterbelt Program and the Grain for Green Program have led to substantial increases in forest cover and density across the country (Chen, Park, Wang, Piao, Xu, Chaturvedi, Fuchs, Brovkin, Ciais, Fensholt et al., 2019). These programs, aimed at combating desertification, soil erosion, and ecological degradation, have resulted in the establishment of new forests and the enhancement of existing ones, directly contributing to increased biomass.

CO_2 enrichment is another potential driver of biomass increase. Rising atmospheric CO_2 concentrations can enhance photosynthetic rates and water-use efficiency in plants, potentially leading to increased biomass accumulation (Zhu et al., 2016). However, the extent of this CO_2 fertilization effect may vary depending on other limiting factors such as nutrient availability and water stress.

Increasing temperatures associated with global climate change may also play a role in biomass accumulation, particularly in temperate and boreal regions of China. Warmer temperatures can extend growing seasons and enhance productivity in regions where cold temperatures were previously limiting (Piao et al., 2015). However, it's important to note that the effects of warming are not uniformly positive and can vary greatly depending on local conditions and species composition.

The observed biomass increase could also be partly attributed to forest succession following past disturbances. Many Chinese forests are in various stages of recovery from historical deforestation, agricultural abandonment, or natural disturbances. As these forests mature, they naturally accumulate biomass, contributing to the overall increasing trend.

Furthermore, changes in forest management practices over recent decades, including reduced timber harvesting in natural forests and improved silvicultural techniques, have likely contributed to biomass accumulation in many areas (Xiao et al., 2023).

It's crucial to recognize that these factors do not operate in isolation but interact in complex ways. For instance, the effects of CO2 enrichment and warming may be modulated by water availability and nutrient limitations. Similarly, the response of forests to these environmental changes can vary depending on their successional stage and management history.

5.4.7 Regional Variability and Model Performance

The analysis of variable importance across China's diverse regions and provinces revealed significant spatial heterogeneity in the factors influencing AGB estimation (Appendix 3). This variability underscores the complex interplay between environmental conditions, forest types, and anthropogenic factors across China's varied landscapes, highlighting the challenges in estimating forest AGB for the whole country.

In northern provinces like Jilin, characterized by temperate forests, our analysis showed that L-band SAR data (PALSAR HV) played a dominant role in AGB estimation. This prominence likely stems from the effectiveness of L-band SAR in capturing the structural characteristics of temperate forests, which often have less complex vertical structures compared to tropical forests. Topographic variables and spectral indices derived from optical sensors played secondary roles in these regions.

Conversely, southern provinces such as Yunnan, known for their subtropical and tropical forests, exhibited a different pattern of variable importance. Here, vegetation indices like NDVI and EVI emerged as crucial predictors, followed closely by topographic variables. This shift in importance likely reflects the region's more complex and diverse forest structures, where these indices can effectively capture variations in canopy density and vigor. The mountainous terrain of Yunnan also contributed to the increased significance of topographic variables in AGB estimation.

The general model for China, combining data from all provinces, presented a more balanced distribution of variable importance. While L-band SAR data remained a significant predictor, its dominance was less pronounced compared to region-specific models. Topographic variables and vegetation indices maintained important roles, reflecting China's diverse terrain and vegetation types. The increased importance of geographic coordinates in the national model highlights the substantial north-south and east-west gradients in forest characteristics across China.

To assess the efficacy of a generalized approach versus regionally tailored models, we developed a general 'all-China' model using 5% of the data from each province, ensuring representation across China's diverse landscapes while maintaining computational feasibility.

Results indicated that while the general model demonstrated reasonable performance across China with an R^2 of 0.36, RMSE of 62 Mg ha^{-1}, and bias of 13.8 Mg ha^{-1} (Fig. 5.6), it exhibited lower accuracy when applied to specific provinces compared to the regionally calibrated models (Fig. 5.2). This discrepancy in performance can be attributed to several factors, including the enhanced local adaptability of province-specific models, the varying importance of predictor variables across regions, and potential bias in the general model towards more extensively sampled or dominant forest types.

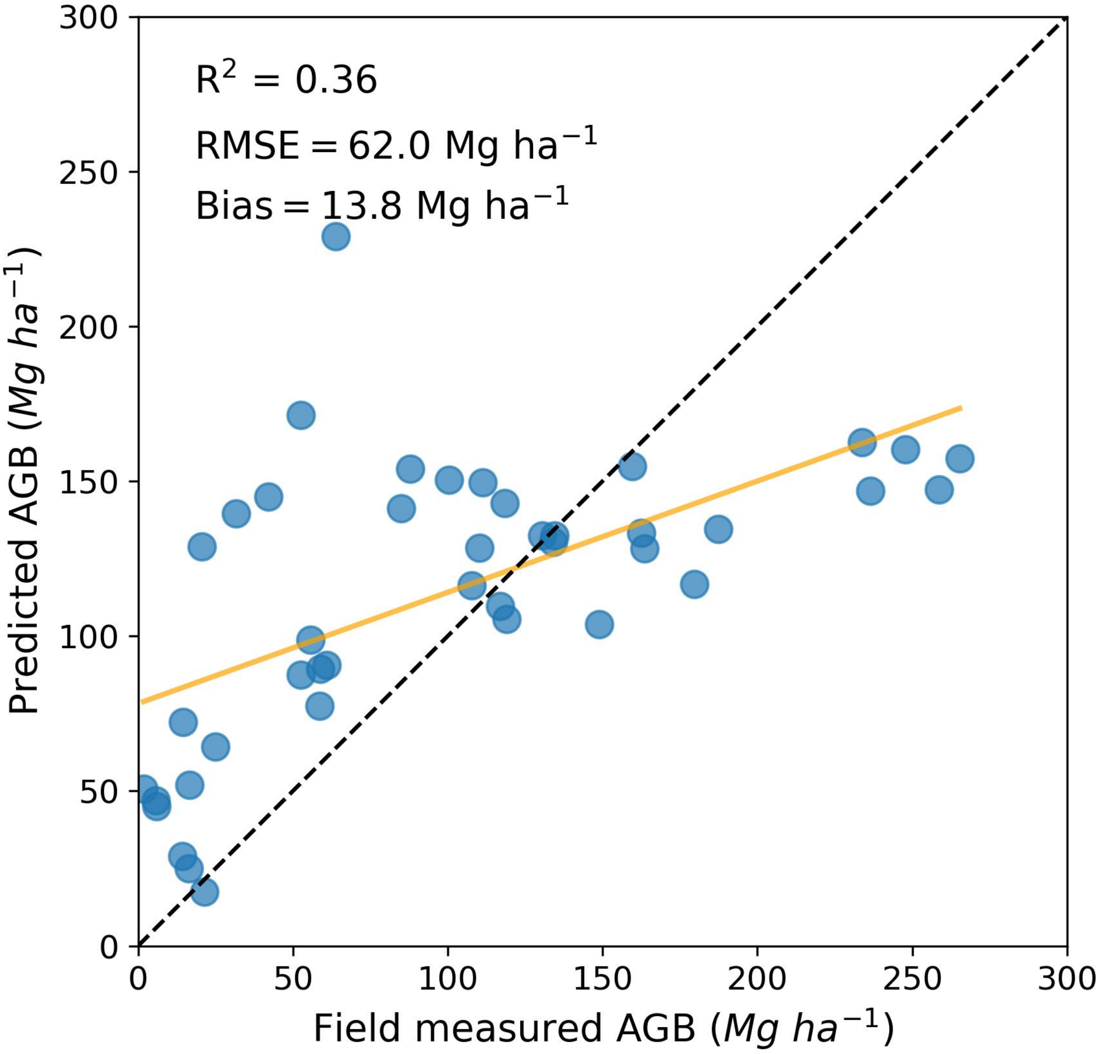


Figure 5.6: Comparison of predicted AGB from the 'all-China' model versus field-measured AGB. The dashed line indicates a 1:1 relationship, while the solid yellow line shows the least squares regression fit to the data.

These findings highlight a critical trade-off between model generalizability and local accuracy in large-scale AGB estimation. While a general model offers the advantage of broader applicability and reduced computational demands, it may sacrifice precision in specific regions. Conversely, regionally calibrated models can provide more accurate local estimates but require more extensive data collection and computational resources.

The observed variations in model performance and variable importance across China's diverse landscapes underscore the complexity of developing robust AGB estimation models at large scales. They emphasize the need for careful consideration of regional ecological and environmental factors in model development and application. Future research could explore hierarchical modeling approaches that combine the strengths of both general and regional models, potentially capturing broad-scale patterns while still accounting for regional specificities. Such approaches could lead to more robust and accurate AGB estimates across diverse landscapes, advancing our understanding of forest carbon dynamics and their role in mitigating climate change.

5.5 Conclusions

This study presents a comprehensive assessment of forest aboveground biomass (AGB) in China for the year 2021 and its changes since 2007 using a combination of GEDI LiDAR data, multi-source satellite imagery, and machine learning algorithms. By integrating province-specific models into a national-scale mapping framework, we generated a high-resolution (50 m) forest AGB map for China that captures the spatial variability of biomass across diverse forest ecosystems. The total forest AGB in China for 2021 was estimated to be 27.79 Pg, with a mean AGB density of 108.75 Mg ha^{-1}.

The comparison of the 2021 AGB map with the previously published 2007 map revealed an overall increase in forest carbon storage over the past 14 years, with a net gain of 3.54 Pg C when considering both forest area expansion (61.04 M ha) and biomass densification. This increase can be attributed to China's ambitious afforestation and reforestation programs, as well as improved forest management practices. However, the observed spatial patterns of AGB changes also highlighted regional variations and the influence of different forest extent datasets on the estimated biomass dynamics.

In conclusion, this study demonstrates the potential of integrating GEDI LiDAR data with multi-source satellite imagery and machine learning algorithms for large-scale forest

AGB mapping in China. The generated high-resolution AGB map and the quantified biomass changes provide valuable insights into the spatial distribution and dynamics of forest biomass resources, supporting informed decision-making in forest management and climate change mitigation efforts.

Chapter 6

Discussion

6.1 AGB change in China

Our assessment of forest AGB changes in China between 2007 and 2021 highlights the importance of forests as a significant carbon sink and their potential role in mitigating climate change. The observed net gain of 3.54 Pg C (0.25 Pg C yr^{-1}) over the 14-year period underscores the effectiveness of sustainable forest management and conservation practices in enhancing carbon storage. If variations in forest area are disregarded, the forest AGB carbon sink is quantified at 1.92 Pg C (0.14 Pg C yr^{-1}). This observation aligns with prior research outcomes, which have documented an augmentation in forest cover and biomass across China, attributable to extensive afforestation and reforestation initiatives (WangYL et al., 2022; Jiang et al., 2016).

It is crucial to consider the limitations and uncertainties associated with large-scale AGB estimation using remote sensing data. The application of a growth rate threshold of 5 Mg ha^{-1} yr^{-1} in our study aimed to constrain overestimated AGB growth values and provide more realistic estimates of AGB changes. While this threshold was informed by the existing literature on forest productivity and carbon pool changes in China (Chen, Feng, Fu, Ma, Zohner, Crowther, Huang, Wu and Wei, 2023), it is important to acknowledge that the actual growth rates of forests can vary substantially depending on factors such as forest type, age, and environmental conditions.

The assessment of forest AGB changes between 2007 and 2021 in this study relied on the use of different land cover maps for masking the AGB maps in each respective year. The

2007 AGB map was masked using the Hansen map, while the 2021 AGB map was masked using the ESA WorldCover map. Although both land cover maps are considered to be of high quality, differences in the methodologies, data sources, and classification schemes used in their production can lead to significant discrepancies in the estimated forest area. These differences in forest area estimates can have a substantial impact on the calculation of AGB changes over time. The availability of long time series of land cover data would greatly benefit future research on forest AGB dynamics.

6.2 Deep learning approaches for forest AGB estimation

Deep learning has emerged as a promising approach for improving the accuracy of forest aboveground biomass (AGB) estimation, as demonstrated in Chapter 4 of this thesis. The AU model, which incorporates an attention mechanism, showcased superior performance compared to traditional machine learning methods, such as RF. This highlights the potential of deep learning to capture complex patterns and relationships in remote sensing data, leading to more accurate AGB estimates.

However, it is essential to acknowledge the limitations and challenges associated with deep learning in the context of forest AGB estimation. One key concern is the computational resource requirements, particularly the need for powerful GPUs, which can hinder the scalability of deep learning approaches to large-scale applications (Lang et al., 2023). This was evident in Chapter 5, where the RF algorithm was employed for generating the 2021 national-scale AGB map due to its lower computational demands. Future research should focus on developing computationally efficient deep learning architectures and strategies for scaling up these approaches to support national and global-scale forest AGB estimation efforts.

Another aspect to consider is the interpretability of deep learning models. While deep learning approaches have demonstrated impressive performance in various domains, including forest AGB estimation, their complex architectures can make it challenging to interpret the underlying decision-making processes. This lack of interpretability can be a barrier to the adoption of deep learning in operational settings, where transparency and explainability are crucial. Future research should explore methods for enhancing the interpretability of deep learning models, such as attention mechanisms and visualization techniques, to facilitate their integration into forest management and policy-making processes.

The heterogeneous performance across different deep learning architectures, as observed in Chapter 4, underscores the importance of model selection and optimization. Future research should focus on developing guidelines and best practices for selecting and fine-tuning deep learning architectures for specific applications, such as forest AGB estimation. This may involve exploring novel architectures, such as transformer-based models, that can better capture the spatial and contextual information inherent in remote sensing data (Fayad et al., 2024).

Moreover, the integration of multi-modal data sources presents an exciting opportunity for improving the accuracy and robustness of deep learning-based forest AGB estimation. Future research should investigate the fusion of various remote sensing data types, such as LiDAR, SAR, and multispectral imagery, within multi-modal deep learning frameworks to leverage their complementary information.

In conclusion, deep learning has shown significant potential for advancing the accuracy of forest AGB estimation, as demonstrated in this thesis. However, challenges related to computational resources, interpretability, and model selection need to be addressed to fully harness the power of deep learning in this domain. Future research should focus on developing computationally efficient architectures, enhancing model interpretability, establishing best practices for model selection and optimization, and exploring the integration of multi-modal data sources. By addressing these challenges and opportunities, deep learning can play a pivotal role in improving our understanding of forest carbon dynamics and informing climate change mitigation strategies.

6.3 AGB map validation

Forest AGB map validation is a critical aspect of ensuring the reliability and usability of the generated maps for various applications, such as carbon stock assessment and forest management. In this thesis, field measurements of tree height and DBH were used to calculate plot-level AGB through allometric equations, which served as the primary validation data for the AGB maps. While this approach provides a direct link between ground truth and estimated AGB values, it is important to discuss alternative validation methods and the challenges associated with the scale mismatch between field plots and AGB maps.

Unmanned Aerial Vehicle (UAV) data and Airborne Laser Scanning (ALS) data offer promising opportunities for validating forest AGB maps. UAVs equipped with high-resolution cameras and LiDAR sensors can capture detailed structural information of forest canopies at a fine spatial scale. This enables the generation of high-density point clouds and 3D models of forest stands, from which plot-level AGB can be estimated using allometric equations or volume-based approaches. Similarly, ALS data, which provides highly accurate canopy height and structure information, can be used to derive plot-level AGB estimates for validation purposes. These remote sensing-based validation methods have the advantage of covering larger areas compared to field measurements, thus providing a more representative sample of the forest landscape.

Terrestrial Laser Scanning (TLS) is another emerging technology that can be used for forest AGB map validation. TLS instruments capture high-resolution 3D point clouds of individual trees or forest plots, enabling the accurate estimation of tree-level attributes such as DBH, height, and volume. These measurements can be used to calculate plot-level AGB using allometric equations, similar to the approach used with field measurements. TLS data offers the advantage of non-destructive sampling and provides highly detailed structural information that can improve the accuracy of AGB estimates. However, the limited spatial coverage and high cost of TLS data acquisition may restrict its applicability for large-scale validation efforts.

One of the major challenges in validating forest AGB maps is the scale mismatch between field plots and the spatial resolution of the maps. This scale mismatch can occur when field plots are either smaller or larger than the pixel size of AGB maps. In both cases, discrepancies in AGB estimates can arise due to spatial heterogeneity within pixels or the lack of representativeness of field plots. To address this issue, a multi-scale validation approach can be employed, where field plots are aggregated to match the spatial resolution of the AGB maps, or high-resolution remote sensing data, such as UAV or ALS data, can be used to provide plot-level AGB estimates at a resolution comparable to the field measurements. By considering the scale mismatch and selecting appropriate validation methods, the accuracy and reliability of forest AGB maps can be effectively assessed.

Another consideration in AGB map validation is the temporal mismatch between field measurements and the acquisition of remote sensing data used for map generation. Forest ecosystems are dynamic, and changes in AGB can occur due to natural disturbances, such as fires or storms, or anthropogenic activities, such as logging or land-use change. To ensure the accuracy of AGB maps, it is important to minimize the temporal gap between field measurements and remote sensing data acquisition. This can be achieved by coordinating field campaigns with satellite or airborne data collection efforts and by incorporating multi-temporal remote sensing data to capture forest dynamics.

While field measurements of tree height and DBH provide a direct link to ground truth, alternative validation methods, such as UAV, ALS, and TLS data, offer complementary approaches that can improve the spatial coverage and representativeness of validation samples. Addressing the scale mismatch between field plots and AGB maps through multi-scale validation and high-resolution remote sensing data is crucial for accurate map validation. Additionally, minimizing the temporal gap between field measurements and remote sensing data acquisition is important for capturing forest dynamics. Future research should focus on integrating multiple validation approaches and developing robust protocols for AGB map validation to support the development of reliable and actionable forest carbon monitoring systems.

6.4 Uncertainty and scale considerations in AGB estimation

The estimation of AGB using remote sensing data and machine learning methods inherently involves a complex interplay of uncertainties and scale-dependent factors.

The foundation of our AGB estimates lies in field measurements and allometric models, both of which introduce initial uncertainties into our analysis. Field measurements of tree height and diameter are subject to both random and systematic errors, stemming from instrument limitations and observer biases. The subsequent application of allometric equations to convert these measurements into biomass estimates compounds these uncertainties, particularly when generalized models are applied across diverse forest types (Chave et al., 2014).

As we transition from point-based field measurements to spatially continuous remote sensing data, we encounter additional layers of uncertainty. Each remote sensing platform used in this study – ICESat, GEDI, Sentinel-1/-2, Landsat, and ALOS PALSAR – carries its own error characteristics related to sensor calibration, atmospheric effects, and spatial resolution. The fusion of these multi-sensor data, while enhancing our ability to capture various aspects of forest structure, also introduces complexities in error propagation. For instance, the integration of optical and radar data requires careful consideration of their respective sensitivities to different forest properties and how these interact in the biomass estimation process (Réjou-Méchain et al., 2017).

The application of machine learning models for AGB estimation further complicates the uncertainty landscape. Model uncertainties arise from the selection of input features, architecture choices, and the quality and representativeness of training data. The potential for overfitting or underfitting, and the challenges in model transferability across diverse forest ecosystems, add to the complexity of error characterization.

In addressing these uncertainties, we employed a 5-fold cross-validation approach in our regional studies (Chapters 3 and 4). While this method provides valuable insights into model stability and performance, it may not fully capture all sources of uncertainty, particularly those arising from systematic biases or limitations in spatial representativeness of our training samples. As we scaled up to national-level AGB estimation, the complexity of error propagation increased.

The implications of changing resolutions across our analysis chain present another critical consideration. The aggregation of high-resolution data (e.g., 25m Sentinel-2) to coarser resolutions (e.g., 50m or 1km) inevitably leads to information loss and potential biases in AGB estimates. This scale dependency affects not only the relationship between AGB and remote sensing variables but also the integration of reference datasets with varying spatial extents. The fusion of data from multiple sensors with different native resolutions further complicates this issue, potentially introducing artifacts or misrepresentations in our final AGB maps.

6.4.1 Uncertainty in Carbon Stock and Flux Estimates

While our study has focused on the estimation of above-ground biomass (AGB) across China using remote sensing and machine learning techniques, it is crucial to acknowledge the importance of quantifying uncertainties in derived carbon stock and flux measures. Although our current research does not provide specific numerical uncertainties for these estimates, addressing this aspect is vital for the robust application of our findings in carbon accounting and climate change mitigation strategies.

The conversion of AGB estimates to carbon stocks introduces additional layers of uncertainty. This conversion typically relies on assumptions about carbon content in biomass, which can vary across species, forest types, and environmental conditions (Thomas and Martin, 2012). Moreover, the spatial and temporal variability in carbon allocation patterns within forest ecosystems adds complexity to these calculations. Future research should aim to incorporate region-specific biomass-to-carbon conversion factors and their associated uncertainties to provide more accurate and reliable carbon stock estimates.

Quantifying uncertainties in carbon flux measurements derived from multi-temporal AGB estimates presents even greater challenges. These fluxes, which represent changes in carbon stocks over time, are subject to compounded uncertainties from both the initial AGB estimates and the temporal change detection process. Factors such as inter-annual variability in forest growth rates, disturbance events, and land-use changes contribute to the complexity of accurately measuring carbon fluxes at large scales.

To address these limitations in future studies, several approaches could be considered:

1. Probabilistic modeling: Implementing Monte Carlo simulations or Bayesian approaches to propagate uncertainties from AGB estimates through to carbon stock and flux calculations. This would allow for the generation of confidence intervals around mean estimates.

2. Multi-scale validation: Integrating ground-based carbon flux measurements (e.g., from eddy covariance towers) with remote sensing-based estimates to assess and calibrate large-scale carbon flux models.

3. Uncertainty budgets: Developing comprehensive uncertainty budgets that account for all sources of error in the carbon estimation process, from field measurements and allometric equations to remote sensing data and modeling approaches.

4. Ensemble methods: Utilizing ensemble approaches that combine multiple models or data sources to provide more robust estimates of carbon stocks and fluxes, along with associated uncertainty measures.

5. Time series analysis: Incorporating advanced time series analysis techniques to better

capture and quantify temporal dynamics in carbon stocks, accounting for seasonal variations and long-term trends.

By addressing these aspects in future work, we can enhance the utility of remote sensing-based biomass and carbon estimates for policy makers, forest managers, and climate scientists. Improved uncertainty quantification will not only increase confidence in our estimates but also provide valuable information for risk assessment and decision-making in the context of climate change mitigation and adaptation strategies.

Chapter 7

Conclusion

This thesis presents a comprehensive assessment of forest AGB and its changes in China using a combination of field data, remote sensing observations, and machine learning algorithms. The research findings contribute to our understanding of the spatial distribution and dynamics of forest biomass resources in China, providing valuable insights for forest management and climate change mitigation efforts.

In Chapter 2, we generated the first high-resolution forest AGB map of China for the year 2007 by integrating ICESat GLAS Lorey's height data, L-band SAR data from ALOS PALSAR, C-band SAR data from Envisat ASAR, and optical data from Landsat-5, along with field measurements. The resulting map provided a benchmark for estimating the total forest carbon stock in China at 9.52 Pg C, with an average AGB density of 104 Mg ha^{-1}.

Chapter 3 focused on two study regions in northeastern and southwestern China, where we conducted field measurements to assess the accuracy of GEDI LiDAR data in estimating tree height and AGB. We developed region-specific allometric equations to convert GEDI relative height metrics to AGB and employed random forest and LightGBM models to generate 25 m resolution AGB maps for the year 2021. The results highlighted the importance of filtering GEDI data to remove low-quality footprints and the influence of terrain slope on the accuracy of AGB estimates.

In Chapter 4, we explored the potential of deep learning techniques for forest AGB estimation in Guangdong Province. The AU model, which leverages LiDAR and multi-source satellite data, demonstrated superior performance compared to the traditional random forest approach. We also proposed a novel strategy to mitigate boundary effects in patch-based predictions, improving the spatial consistency and accuracy of the resulting AGB map.

Chapter 5 presented a national-scale assessment of forest AGB in China for the year 2021 and its changes since 2007. By employing region-specific allometric equations and random forest models, we estimated the total forest carbon stock in China for 2021 at 13.06 Pg C, with a mean AGB density of 108.8 Mg ha^{-1}. Comparing the 2021 and 2007 AGB maps revealed a net gain of 3.54 Pg C, with significant spatial variations in AGB changes across the country.

The research presented in this thesis highlights the importance of integrating field measurements, remote sensing data, and advanced machine learning techniques for accurate and large-scale forest AGB estimation. The high-resolution AGB maps generated in this study provide a valuable resource for understanding the spatial distribution and dynamics of forest biomass in China, supporting informed decision-making in forest management and climate change mitigation strategies.

However, some challenges and limitations remain. The validation of large-scale AGB maps using field data is critical for assessing their accuracy, but the scale mismatch between field plots and satellite observations can introduce uncertainties. Future research should focus on developing improved validation strategies, such as the integration of high-resolution remote sensing data (e.g., UAV or airborne LiDAR) to bridge the scale gap.

Moreover, the use of different land cover maps for masking the AGB maps in 2007 and 2021 highlights the potential impact of forest area discrepancies on the assessment of AGB changes over time. The availability of consistent, long time series of high-resolution land cover data would greatly benefit future research on forest AGB dynamics, enabling more accurate and reliable assessments of changes in forest carbon storage.

Another important aspect to consider is the computational resources required for large-scale AGB mapping using advanced machine learning techniques, particularly deep learning models. While these methods have shown promising results in improving AGB estimation accuracy, their application at national or global scales may be limited by the availability of computational resources. Future research should explore strategies for optimizing deep learning models and developing computationally efficient frameworks for large-scale AGB mapping.

In conclusion, this thesis contributes to the growing body of knowledge on forest AGB estimation and mapping using remote sensing and machine learning techniques. However, further research is needed to address the challenges and limitations identified in this study, such as improving validation strategies, developing consistent long-term land cover datasets, and optimizing computational resources for large-scale AGB mapping. By addressing these challenges and building upon the findings of this thesis, we can continue to advance our understanding of forest biomass dynamics and their role in the global carbon cycle, ultimately contributing to the development of effective strategies for sustainable forest management and climate change mitigation.

References

Araza, A., De Bruin, S., Herold, M., Quegan, S., Labriere, N., Rodriguez-Veiga, P., Avitabile, V., Santoro, M., Mitchard, E. T., Ryan, C. M. et al. (2022), 'A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps', *Remote Sensing of Environment* **272**, 112917.

Avitabile, V., Herold, M., Lewis, S., Phillips, O., Aguilar-Amuchastegui, N., Asner, G., Brienen, R., DeVries, B., Gatti, R. G., Feldpausch, T. et al. (2014), Comparative analysis and fusion for improved global biomass mapping, *in* 'International Conference GV2M, Avignon, France', pp. 251–252.

Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., Friedl, M. et al. (2012), 'Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps', *Nature climate change* **2**(3), 182–185.

Baghdadi, N., Le Maire, G., Fayad, I., Bailly, J. S., Nouvellon, Y., Lemos, C. and Hakamada, R. (2013), 'Testing different methods of forest height and aboveground biomass estimations from icesat/glas data in eucalyptus plantations in brazil', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **7**(1), 290–299.

Balzter, H., Rowland, C. and Saich, P. (2007), 'Forest canopy height and carbon estimation at monks wood national nature reserve, uk, using dual-wavelength sar interferometry', *Remote Sensing of Environment* **108**(3), 224–239.

Bartholome, E. and Belward, A. S. (2005), 'Glc2000: a new approach to global land cover mapping from earth observation data', *International Journal of Remote Sensing* **26**(9), 1959–1977.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B. et al. (2010), 'Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate', *Science* **329**(5993), 834–838.

Besnard, S., Santoro, M., Cartus, O., Fan, N., Linscheid, N., Nair, R., Weber, U., Koirala, S. and Carvalhais, N. (2021), 'Global sensitivities of forest carbon changes to environmental conditions', *Global Change Biology* **27**(24), 6467–6483.

Blackard, J., Finco, M., Helmer, E., Holden, G., Hoppus, M., Jacobs, D., Lister, A., Moisen, G., Nelson, M., Riemann, R. et al. (2008), 'Mapping us forest biomass using nationwide forest inventory data and moderate resolution information', *Remote sensing of Environment* **112**(4), 1658–1677.

Bonafilia, D., Tellman, B., Anderson, T. and Issenberg, E. (2020), Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1, *in* 'Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops', pp. 210–211.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L. and Kimes, D. S. (2008), 'Regional aboveground forest biomass using airborne and spaceborne lidar in québec', *Remote Sensing of Environment* **112**(10), 3876–3890.

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L. and Asner, G. P. (2018), 'An above-ground biomass map of african savannahs and woodlands at 25 m resolution derived from alos palsar', *Remote sensing of environment* **206**, 156–173.

Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., Diouf, A. A. et al. (2020), 'An unexpectedly large count of trees in the west african sahara and sahel', *Nature* **587**(7832), 78–82.

Brandt, M., Wigneron, J.-P., Chave, J., Tagesson, T., Penuelas, J., Ciais, P., Rasmussen, K., Tian, F., Mbow, C., Al-Yaari, A. et al. (2018), 'Satellite passive microwaves reveal recent climate-induced carbon losses in african drylands', *Nature ecology & evolution* **2**(5), 827–835.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**, 5–32.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S. et al. (2022), 'Dynamic world, near real-time global 10 m land use land cover mapping', *Scientific Data* **9**(1), 251.

Bukoski, J. J., Cook-Patton, S. C., Melikov, C., Ban, H., Chen, J. L., Goldman, E. D., Harris, N. L. and Potts, M. D. (2022), 'Rates and drivers of aboveground carbon accumulation in global monoculture plantation forests', *Nature Communications* **13**(1), 4206.

Bustamante, M. M., Roitman, I., Aide, T. M., Alencar, A., Anderson, L. O., Aragão, L., Asner, G. P., Barlow, J., Berenguer, E., Chambers, J. et al. (2016), 'Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity', *Global change biology* **22**(1), 92–109.

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., Culvenor, D., Avitabile, V., Disney, M., Armston, J. et al. (2015), 'Nondestructive estimates of above-ground biomass using terrestrial laser scanning', *Methods in Ecology and Evolution* **6**(2), 198–208.

Campbell, M. J., Dennison, P. E., Kerr, K. L., Brewer, S. C. and Anderegg, W. R. (2021), 'Scaled biomass estimation in woodland ecosystems: Testing the individual and combined capacities of satellite multispectral and lidar data', *Remote Sensing of Environment* **262**, 112511.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L. et al. (2021), 'A unified vegetation index for quantifying the terrestrial biosphere', *Science Advances* **7**(9), eabc7447.

Cao, L., Coops, N. C., Innes, J. L., Sheppard, S. R., Fu, L., Ruan, H. and She, G. (2016), 'Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne lidar data', *Remote Sensing of Environment* **178**, 158–171.

Carreiras, J. M., Vasconcelos, M. J. and Lucas, R. M. (2012), 'Understanding the relationship between aboveground biomass and alos palsar data in the forests of guinea-bissau (west africa)', *Remote Sensing of Environment* **121**, 426–442.

Cartus, O., Santoro, M. and Kellndorfer, J. (2012), 'Mapping forest aboveground biomass in the northeastern united states with alos palsar dual-polarization l-band', *Remote Sensing of Environment* **124**, 466–478.

Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T. and Durand, P. (2001), 'Sensitivity of space-borne sar data to forest parameters over sloping terrain. theory and experiment', *International journal of remote sensing* **22**(12), 2351–2376.

Chang, Z., Hobeichi, S., Wang, Y.-P., Tang, X., Abramowitz, G., Chen, Y., Cao, N., Yu, M., Huang, H., Zhou, G. et al. (2021), 'New forest aboveground biomass maps of china integrating multiple datasets', *Remote Sensing* **13**(15), 2892.

Change, I. P. O. C. (2006), 'Guidelines for national greenhouse gas inventories', *CD ou no site: www. ipcc. ch. Guia de Boas Práticas* .

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T. et al. (2005), 'Tree allometry and improved estimation of carbon stocks and balance in tropical forests', *Oecologia* **145**, 87–99.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S. and Perez, R. (2004), 'Error propagation and scaling for tropical forest biomass estimates', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**(1443), 409–420.

Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G. and Zanne, A. E. (2009), 'Towards a worldwide wood economics spectrum', *Ecology letters* **12**(4), 351–366.

Chave, J., Muller-Landau, H. C., Baker, T. R., Easdale, T. A., Steege, H. t. and Webb, C. O. (2006), 'Regional and phylogenetic variation of wood density across 2456 neotropical tree species', *Ecological applications* **16**(6), 2356–2367.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C. et al. (2014), 'Improved allometric models to estimate the aboveground biomass of tropical trees', *Global change biology* **20**(10), 3177–3190.

Chen, B., Chen, F., Ciais, P., Zhang, H., Lü, H., Wang, T., Chevallier, F., Liu, Z., Yuan, W.

and Peters, W. (2022), ‘Challenges to achieve carbon neutrality of china by 2060: status and perspectives’, *Science Bulletin* **67**(20), 2030–2035.

Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R. et al. (2019), ‘China and india lead in greening of the world through land-use management’, *Nature sustainability* **2**(2), 122–129.

Chen, L., Ren, C., Bao, G., Zhang, B., Wang, Z., Liu, M., Man, W. and Liu, J. (2022), ‘Improved object-based estimation of forest aboveground biomass by integrating lidar data from gedi and icesat-2 with multi-sensor images in a heterogeneous mountainous region’, *Remote Sensing* **14**(12), 2743.

Chen, L., Wang, Y., Ren, C., Zhang, B. and Wang, Z. (2019), ‘Assessment of multi-wavelength sar and multispectral instrument data for forest aboveground biomass mapping using random forest kriging’, *Forest ecology and management* **447**, 12–25.

Chen, M., Dong, W., Yu, H., Woodhouse, I., Ryan, C. M., Liu, H., Georgiou, S. and Mitchard, E. T. (2023), ‘Multimodal deep learning for mapping forest dominant height by fusing gedi with earth observation data’, *arXiv preprint arXiv:2311.11777* .

Chen, Q. (2010), ‘Retrieving vegetation height of forests and woodlands over mountainous areas in the pacific coast region using satellite laser altimetry’, *Remote Sensing of Environment* **114**(7), 1610–1627.

Chen, Y., Feng, X., Fu, B., Ma, H., Zohner, C., Crowther, T., Huang, Y., Wu, X. and Wei, F. (2023), ‘Maps with 1 km resolution reveal increases in above-and belowground forest biomass carbon pools in china over the past 20 years’, *Earth System Science Data* **15**(2), 897–910.

Cheng, J., Bambrick, H., Yakob, L., Devine, G., Frentiu, F. D., Williams, G., Li, Z., Yang, W. and Hu, W. (2021), ‘Extreme weather conditions and dengue outbreak in guangdong, china: Spatial heterogeneity based on climate variability’, *Environmental research* **196**, 110900.

Cheng, K., Su, Y., Guan, H., Tao, S., Ren, Y., Hu, T., Ma, K., Tang, Y. and Guo, Q. (2023), ‘Mapping china’s planted forests using high resolution imagery and massive amounts of crowdsourced samples', *ISPRS Journal of Photogrammetry and Remote Sensing* **196**, 356–371.

Clark, D. B. and Kellner, J. R. (2012), 'Tropical forest biomass estimation and the fallacy of misplaced concreteness', *Journal of Vegetation Science* **23**(6), 1191–1196.

Crippen, R., Buckley, S., Agram, P., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalle, M., Martin, J., Neumann, M. et al. (2016), 'Nasadem global elevation model: Methods and progress', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **41**, 125–128.

Cushman, K., Muller-Landau, H. C., Condit, R. S. and Hubbell, S. P. (2014), 'Improving estimates of biomass change in buttressed trees using tree taper models', *Methods in Ecology and Evolution* **5**(6), 573–582.

Dai, L., Wang, Y., Su, D., Zhou, L., Yu, D., Lewis, B. J. and Qi, L. (2011), 'Major forest types and the evolution of sustainable forestry in china', *Environmental Management* **48**, 1066–1078.

Das, A., Agrawal, R. and Mohan, S. (2015), 'Topographic correction of alos-palsar images using insar-derived dem', *Geocarto International* **30**(2), 145–153.

Delang, C. O. and Yuan, Z. (2015), 'China's grain for green program', *A review of the largest ecological restoration and rural development program in the world. Switzerland: Springer International Publishing* .

Dhanda, P., Nandy, S., Kushwaha, S., Ghosh, S., Murthy, Y. K. and Dadhwal, V. (2017), 'Optimizing spaceborne lidar and very high resolution optical sensor parameters for biomass estimation at icesat/glas footprint level using regression algorithms', *Progress in Physical Geography* **41**(3), 247–267.

Dong, J., Kaufmann, R. K., Myneni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., Buermann, W., Alexeyev, V. and Hughes, M. K. (2003), 'Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks', *Remote sensing of Environment* **84**(3), 393–410.

Dong, W., Mitchard, E. T. A., Santoro, M., Chen, M. and Wheeler, C. E. (2023), '2007 forest aboveground biomass map for china'.
URL: *https://doi.org/10.7488/ds/7480*

Dong, W., Mitchard, E. T., Santoro, M., Chen, M. and Wheeler, C. E. (2024), 'A new circa 2007 biomass map for china differs significantly from existing maps', *Scientific Data* **11**(1), 287.

Dong, W., Mitchard, E. T., Yu, H., Hancock, S. and Ryan, C. M. (2023), 'Forest aboveground biomass estimation using gedi and earth observation data through attention-based deep learning', *arXiv preprint arXiv:2311.03067* .

Dong, X., Li-min, D., Guo-fan, S., Lei, T. and Hui, W. (2005), 'Forest fire risk zone mapping from satellite images and gis for baihe forestry bureau, jilin, china', *Journal of forestry research* **16**(3), 169–174.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al. (2020), 'An image is worth 16x16 words: Transformers for image recognition at scale', *arXiv preprint arXiv:2010.11929* .

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., Chazdon, R. L., Weishampel, J. F. and Prince, S. (2002), 'Estimation of tropical forest structural characteristics using large-footprint lidar', *Remote sensing of environment* **79**(2-3), 305–319.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. et al. (2012), 'Sentinel-2: Esa's optical high-resolution mission for gmes operational services', *Remote sensing of Environment* **120**, 25–36.

Dubayah, R., Armston, J., Healey, S. P., Bruening, J. M., Patterson, P. L., Kellner, J. R., Duncanson, L., Saarela, S., Ståhl, G., Yang, Z. et al. (2022), 'Gedi launches a new era of biomass inference from space', *Environmental Research Letters* **17**(9), 095001.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S. et al. (2020), 'The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography', *Science of remote sensing* **1**, 100002.

Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H. and Luthcke, S. (2021), 'Gedi l2a elevation and height metrics data global footprint level v002. nasa eosdis land processes daac'.

Duncanson, L., Kellner, J. R., Armston, J., Dubayah, R., Minor, D. M., Hancock, S., Healey, S. P., Patterson, P. L., Saarela, S., Marselis, S. et al. (2022), 'Aboveground biomass density models for nasa's global ecosystem dynamics investigation (gedi) lidar mission', *Remote Sensing of Environment* **270**, 112845.

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C. A., Armston, J., Luthcke, S. B., Hofton, M. et al. (2020), 'Biomass estimation from simulated gedi, icesat-2 and nisar across environmental gradients in sonoma county, california', *Remote Sensing of Environment* **242**, 111779.

Duncanson, L., Rourke, O. and Dubayah, R. (2015), 'Small sample sizes yield biased allometric equations in temperate forests', *Scientific reports* **5**(1), 17153.

Eggleston, S., Leandro, B., Kyoko, M., Todd, N. and Tanabe, K. (2006), 'Inter-governmental panel on climate change (ipcc) guidelines for national green house gas inventories', *Published by the Institute for Global Environmental Strategies (IGES), Hayama, Japan* .

Englhart, S., Keuck, V. and Siegert, F. (2011), 'Aboveground biomass retrieval in tropical forests—the potential of combined x-and l-band sar data use', *Remote sensing of environment* **115**(5), 1260–1271.

Fang, J., Chen, A., Peng, C., Zhao, S. and Ci, L. (2001), 'Changes in forest biomass carbon storage in china between 1949 and 1998', *Science* **292**(5525), 2320–2322.

Fang, J., Guo, Z., Hu, H., Kato, T., Muraoka, H. and Son, Y. (2014), 'Forest biomass carbon sinks in e ast a sia, with special reference to the relative contributions of forest expansion and forest growth', *Global change biology* **20**(6), 2019–2030.

Fang, J., Piao, S., He, J. and Ma, W. (2004), 'Increasing terrestrial vegetation activity in china, 1982–1999', *Science in China Series C: Life Sciences* **47**, 229–240.

FAO (2020), *Global Forest Resources Assessment 2020: Main report*, Food and Agriculture Organization of the United Nations, Rome.

Farebrother, J., Machado, M. C. and Bowling, M. (2018), ‘Generalization and regularization in dqn’, *arXiv preprint arXiv:1810.00123* .

Fassnacht, F., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P. and Koch, B. (2014), ‘Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass’, *Remote sensing of environment* **154**, 102–114.

Fayad, I., Ciais, P., Schwartz, M., Wigneron, J.-P., Baghdadi, N., de Truchis, A., d’Aspremont, A., Frappart, F., Saatchi, S., Sean, E. et al. (2024), ‘Hy-tec: a hybrid vision transformer model for high-resolution and large-scale mapping of canopy height’, *Remote Sensing of Environment* **302**, 113945.

Feldpausch, T. R., Banin, L., Phillips, O. L., Baker, T. R., Lewis, S. L., Quesada, C. A., Affum-Baffoe, K., Arets, E. J., Berry, N. J., Bird, M. et al. (2011), ‘Height-diameter allometry of tropical forest trees’, *Biogeosciences* **8**(5), 1081–1106.

Feldpausch, T. R., PRATES-CLARK, C. d. C., FERNANDES, E. C. and Riha, S. J. (2007), ‘Secondary forest growth deviation from chronosequence predictions in central amazonia’, *Global Change Biology* **13**(5), 967–979.

Flores-Anderson, A. I., Herndon, K. E., Thapa, R. B. and Cherrington, E. (2019), ‘Sar handbook: Comprehensive methodologies for forest monitoring and biomass estimation. nasa’.

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley Jr, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Hughes, M. J. and Laue, B. (2017), ‘Cloud detection algorithm comparison and validation for operational landsat data products’, *Remote sensing of environment* **194**, 379–390.

Food and Agriculture Organization of the United Nations (FAO) (2010), Global forest resources assessment 2010 - country report: China, Fra 2010/042, Food and Agriculture Organization of the United Nations, Rome.

Forkuor, G., Zoungrana, J.-B. B., Dimobe, K., Ouattara, B., Vadrevu, K. P. and Tondoh, J. E. (2020), ‘Above-ground biomass mapping in west african dryland forest using sentinel-1 and 2 datasets-a case study’, *Remote Sensing of Environment* **236**, 111496.

Frayer, J., Müller, D., Sun, Z., Munroe, D. K. and Xu, J. (2014), ‘Processes underlying 50 years of local forest-cover change in yunnan, china’, *Forests* **5**(12), 3257–3273.

Gao, Y., Lu, D., Li, G., Wang, G., Chen, Q., Liu, L. and Li, D. (2018), ‘Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region’, *Remote Sensing* **10**(4), 627.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A. et al. (2017), ‘Copernicus sentinel-2a calibration and products validation status’, *Remote Sensing* **9**(6), 584.

Ge, S., Su, W., Gu, H., Rauste, Y., Praks, J. and Antropov, O. (2022), ‘Improved lstm model for boreal forest height mapping using sentinel-1 time series’, *Remote Sensing* **14**(21), 5560.

Gómez, C., White, J. C., Wulder, M. A. and Alejandro, P. (2014), ‘Historical forest biomass dynamics modelled with landsat spectral trajectories’, *ISPRS Journal of Photogrammetry and Remote Sensing* **93**, 14–28.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R. (2017), ‘Google earth engine: Planetary-scale geospatial analysis for everyone’, *Remote sensing of Environment* **202**, 18–27.

Gower, S., Krankina, O., Olson, R., Apps, M., Linder, S. and Wang, C. (2001), ‘Net primary production and carbon allocation patterns of boreal forest ecosystems’, *Ecological applications* **11**(5), 1395–1411.

Grace, J., Mitchard, E. and Gloor, E. (2014), ‘Perturbations in the carbon budget of the tropics’, *Global Change Biology* **20**(10), 3238–3255.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. et al. (2018), ‘Recent advances in convolutional neural networks’, *Pattern recognition* **77**, 354–377.

Guo, Q., Du, S., Jiang, J., Guo, W., Zhao, H., Yan, X., Zhao, Y. and Xiao, W. (2023), 'Combining gedi and sentinel data to estimate forest canopy mean height and aboveground biomass', *Ecological Informatics* **78**, 102348.

Guo, Z., Hu, H., Li, P., Li, N. and Fang, J. (2013), 'Spatio-temporal changes in biomass carbon sinks in china's forests from 1977 to 2008', *Science China Life Sciences* **56**, 661–671.

H. Nguyen, T., Jones, S., Soto-Berelov, M., Haywood, A. and Hislop, S. (2019), 'Landsat time-series for estimating forest aboveground biomass and its dynamics across space and time: A review', *Remote Sensing* **12**(1), 98.

Hamdan, O., Aziz, H. K. and Hasmadi, I. M. (2014), 'L-band alos palsar for biomass estimation of matang mangroves, malaysia', *Remote Sensing of Environment* **155**, 69–78.

Han, H., Guo, X. and Yu, H. (2016), Variable selection using mean decrease accuracy and mean decrease gini based on random forest, *in* '2016 7th ieee international conference on software engineering and service science (icsess)', IEEE, pp. 219–224.

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L. I., Kellner, J. R. and Dubayah, R. (2019), 'The gedi simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions', *Earth and Space Science* **6**(2), 294–310.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R. et al. (2013), 'High-resolution global maps of 21st-century forest cover change', *science* **342**(6160), 850–853.

Harding, D. J. and Carabajal, C. C. (2005), 'Icesat waveform measurements of within-footprint topographic relief and vegetation vertical structure', *Geophysical research letters* **32**(21).

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M. C., Herold, M., Houghton, R. A. et al. (2021), 'Global maps of twenty-first century forest carbon fluxes', *Nature Climate Change* **11**(3), 234–240.

Hayashi, M., Motohka, T. and Sawada, Y. (2019), 'Aboveground biomass mapping using alos-2/palsar-2 time-series images for borneo's forest', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **12**(12), 5167–5177.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P. and Girshick, R. (2022), Masked autoencoders are scalable vision learners, *in* 'Proceedings of the IEEE/CVF conference on computer vision and pattern recognition', pp. 16000–16009.

He, K., Zhang, X., Ren, S. and Sun, J. (2016), Deep residual learning for image recognition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 770–778.

Henry, M., Picard, N., Trotta, C., Manlay, R., Valentini, R., Bernoux, M. and Saint André, L. (2011), 'Estimating tree biomass of sub-saharan african forests: a review of available allometric equations'.

Herold, M., Carter, S., Avitabile, V., Espejo, A. B., Jonckheere, I., Lucas, R., McRoberts, R. E., Næsset, E., Nightingale, J., Petersen, R. et al. (2019), 'The role and need for space-based forest biomass-related measurements in environmental management and policy', *Surveys in Geophysics* **40**, 757–778.

Hou, X. (2001), 'Vegetation atlas of china', *Chinese Academy of Science, the editorial board of vegetation map of China* pp. 113–124.

Houghton, R., Hall, F. and Goetz, S. J. (2009), 'Importance of biomass in the global carbon cycle', *Journal of Geophysical Research: Biogeosciences* **114**(G2).

Huang, H., Liu, C., Wang, X., Zhou, X. and Gong, P. (2019), 'Integration of multi-resource remotely sensed data and allometric models for forest aboveground biomass estimation in china', *Remote Sensing of Environment* **221**, 225–234.

Huang, X., Ziniti, B., Torbick, N. and Ducey, M. J. (2018), 'Assessment of forest above ground biomass estimation using multi-temporal c-band sentinel-1 and polarimetric l-band palsar-2 data', *Remote Sensing* **10**(9), 1424.

Hunter, M., Keller, M., Victoria, D. and Morton, D. C. (2013), 'Tree height and tropical forest biomass estimation', *Biogeosciences* **10**(12), 8385–8399.

Innamorati, C., Ritschel, T., Weyrich, T. and Mitra, N. J. (2020), 'Learning on the edge: Investigating boundary filters in cnns', *International Journal of Computer Vision* **128**, 773–782.

Jarvis, P. and Linder, S. (2000), 'Constraints to growth of boreal forests', *Nature* **405**(6789), 904–905.

Jiang, F., Chen, J. M., Zhou, L., Ju, W., Zhang, H., Machida, T., Ciais, P., Peters, W., Wang, H., Chen, B. et al. (2016), 'A comprehensive estimate of recent carbon sinks in china using both top-down and bottom-up approaches', *Scientific Reports* **6**(1), 22130.

Jiyuan, L., Mingliang, L., Xiangzheng, D., Dafang, Z., Zengxiang, Z. and Di, L. (2002), 'The land use and land cover change database and its relative studies in china', *Journal of Geographical Sciences* **12**, 275–282.

Joshi, N., Mitchard, E. T., Brolly, M., Schumacher, J., Fernández-Landa, A., Johannsen, V. K., Marchamalo, M. and Fensholt, R. (2017), 'Understanding 'saturation'of radar signals over forests', *Scientific reports* **7**(1), 3505.

JPL, N. (2020), 'Nasadem merged dem global 1 arc second v001 [data set]. nasa eosdis land processes daac', *2020* .

Jucker, T., Bongalov, B., Burslem, D. F., Nilus, R., Dalponte, M., Lewis, S. L., Phillips, O. L., Qie, L. and Coomes, D. A. (2018), 'Topography shapes the structure, composition and function of tropical forest landscapes', *Ecology letters* **21**(7), 989–1000.

Kasischke, E. S., Tanase, M. A., Bourgeau-Chavez, L. L. and Borr, M. (2011), 'Soil moisture limitations on monitoring boreal forest regrowth using spaceborne l-band sar data', *Remote Sensing of Environment* **115**(1), 227–232.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017), 'Lightgbm: A highly efficient gradient boosting decision tree', *Advances in neural information processing systems* **30**.

Ke, S., Qiao, D., Zhang, X. and Feng, Q. (2019), 'Changes of china's forestry and forest products industry over the past 40 years and challenges lying ahead', *Forest policy and economics* **106**, 101949.

Kindermann, G., McCallum, I., Fritz, S. and Obersteiner, M. (2008), 'A global forest growing stock, biomass and carbon map based on fao statistics', *Silva Fennica* **42**(3), 387–396.

Kingma, D. P. and Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Kumar, L. and Mutanga, O. (2017), 'Remote sensing of above-ground biomass'.

Lang, N., Jetz, W., Schindler, K. and Wegner, J. D. (2023), 'A high-resolution canopy height model of the earth', *Nature Ecology & Evolution* **7**(11), 1778–1789.

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R. and Wegner, J. D. (2022), 'Global canopy height regression and uncertainty estimation from gedi lidar waveforms with deep ensembles', *Remote sensing of environment* **268**, 112760.

Larjavaara, M. and Muller-Landau, H. C. (2013), 'Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest', *Methods in Ecology and Evolution* **4**(9), 793–801.

Lary, D. J., Alavi, A. H., Gandomi, A. H. and Walker, A. L. (2016), 'Machine learning in geosciences and remote sensing', *Geoscience Frontiers* **7**(1), 3–10.

Lefsky, M. A. (2010), 'A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system', *Geophysical Research Letters* **37**(15).

Lefsky, M. A., Cohen, W. B., Parker, G. G. and 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 interest to forest, landscape, and global ecologists', *BioScience* **52**(1), 19–30.

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Del Bom Espirito-Santo, F., Hunter, M. O. and de Oliveira Jr, R. (2005), 'Estimates of forest canopy height and aboveground biomass using icesat', *Geophysical research letters* **32**(22).

Lewis, S. L., Sonké, B., Sunderland, T., Begne, S. K., Lopez-Gonzalez, G., Van Der Heijden, G. M., Phillips, O. L., Affum-Baffoe, K., Baker, T. R., Banin, L. et al. (2013), 'Aboveground biomass and structure of 260 african tropical forests', *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**(1625), 20120295.

Li, N., Xie, G., Zhang, C., Xiao, Y., Zhang, B., Chen, W., Sun, Y. and Wang, S. (2015), 'Biomass resources distribution in the terrestrial ecosystem of china', *Sustainability* **7**(7), 8548–8564.

Li, X., Wessels, K., Armston, J., Hancock, S., Mathieu, R., Main, R., Naidoo, L., Erasmus, B. and Scholes, R. (2023), 'First validation of gedi canopy heights in african savannas', *Remote Sensing of Environment* **285**, 113402.

Liang, M., Duncanson, L., Silva, J. A. and Sedano, F. (2023), 'Quantifying aboveground biomass dynamics from charcoal degradation in mozambique using gedi lidar and landsat', *Remote Sensing of Environment* **284**, 113367.

Liao, Z., He, B. and Quan, X. (2020), 'Potential of texture from sar tomographic images for forest aboveground biomass estimation', *International journal of applied earth observation and geoinformation* **88**, 102049.

Ling, J., Zhang, H. and Lin, Y. (2021), 'Improving urban land cover classification in cloud-prone areas with polarimetric sar images', *Remote Sensing* **13**(22), 4708.

Liu, A., Cheng, X. and Chen, Z. (2021), 'Performance evaluation of gedi and icesat-2 laser altimeter data for terrain and canopy height retrievals', *Remote Sensing of Environment* **264**, 112571.

Liu, J. (2014), 'Forest sustainability in c hina and implications for a telecoupled world', *Asia & the Pacific Policy Studies* **1**(1), 230–250.

Liu, J., Li, S., Ouyang, Z., Tam, C. and Chen, X. (2008), 'Ecological and socioeconomic effects of china's policies for ecosystem services', *Proceedings of the National academy of Sciences* **105**(28), 9477–9482.

Liu, J., Liang, M., Li, L., Long, H. and De Jong, W. (2017), 'Comparative study of the forest transition pathways of nine asia-pacific countries', *Forest Policy and Economics* **76**, 25–34.

Liu, K., Kang, G., Zhang, N. and Hou, B. (2018), 'Breast cancer classification based on fully-connected layer first convolutional neural networks', *IEEE Access* **6**, 23722–23732.

Liu, K., Shen, X., Cao, L., Wang, G. and Cao, F. (2018), 'Estimating forest structural attributes using uav-lidar data in ginkgo plantations', *ISPRS journal of photogrammetry and remote sensing* **146**, 465–482.

Liu, Q., Shan, Y., Shu, L., Sun, P. and Du, S. (2018), 'Spatial and temporal distribution of forest fire frequency and forest area burnt in jilin province, northeast china', *Journal of Forestry Research* **29**(5), 1233–1239.

Liu, X. (2008), 'Airborne lidar for dem generation: some critical issues', *Progress in physical geography* **32**(1), 31–49.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021), Swin transformer: Hierarchical vision transformer using shifted windows, *in* 'Proceedings of the IEEE/CVF international conference on computer vision', pp. 10012–10022.

Long, Y., Xia, G.-S., Li, S., Yang, W., Yang, M. Y., Zhu, X. X., Zhang, L. and Li, D. (2021), 'On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-aid', *IEEE Journal of selected topics in applied earth observations and remote sensing* **14**, 4205–4230.

Louet, J. and Bruzzi, S. (1999), Envisat mission and system, *in* 'IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293)', Vol. 3, IEEE, pp. 1680–1682.

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G. and Moran, E. (2016), 'A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems', *International Journal of Digital Earth* **9**(1), 63–105.

Lu, D., Mausel, P., Brondızio, E. and Moran, E. (2004), 'Relationships between forest stand parameters and landsat tm spectral responses in the brazilian amazon basin', *Forest ecology and management* **198**(1-3), 149–167.

Lü, Y., Fu, B., Feng, X., Zeng, Y., Liu, Y., Chang, R., Sun, G. and Wu, B. (2012), 'A policy-driven large scale ecological restoration: quantifying ecosystem services changes in the loess plateau of china', *PloS one* **7**(2), e31782.

Luckman, A., Baker, J., Kuplich, T. M., Yanasse, C. d. C. F. and Frery, A. C. (1997), 'A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne sar instruments', *Remote Sensing of Environment* **60**(1), 1–13.

Lumbierres, M., Méndez, P. F., Bustamante, J., Soriguer, R. and Santamaría, L. (2017), 'Modeling biomass production in seasonal wetlands using modis ndvi land surface phenology', *Remote Sensing* **9**(4), 392.

Luo, Y., Wang, X. and Ouyang, Z. (2018), 'A china's normalized tree biomass equation dataset'.

Luo, Y., Wang, X., Ouyang, Z., Lu, F., Feng, L. and Tao, J. (2019), 'Chinallometree 1.0: China's normalized tree biomass equation dataset', *Earth Syst. Sci. Data Discuss.* **1**, 1–75.

Luo, Y., Zhang, X., Wang, X. and Lu, F. (2014), 'Biomass and its allocation of chinese forest ecosystems: Ecological archives e095-177', *Ecology* **95**(7), 2026–2026.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. and Gascon, F. (2017), Sen2cor for sentinel-2, *in* 'Image and signal processing for remote sensing XXIII', Vol. 10427, SPIE, pp. 37–48.

Mallet, C. and Bretar, F. (2009), 'Full-waveform topographic lidar: State-of-the-art', *ISPRS Journal of photogrammetry and remote sensing* **64**(1), 1–16.

Margolis, H. A., Nelson, R. F., Montesano, P. M., Beaudoin, A., Sun, G., Andersen, H.-E. and Wulder, M. A. (2015), 'Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of north america', *Canadian Journal of Forest Research* **45**(7), 838–855.

Marselis, S. M., Tang, H., Armston, J. D., Calders, K., Labrière, N. and Dubayah, R. (2018), 'Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in lopé national park, gabon', *Remote sensing of environment* **216**, 626–634.

Martínez, B. and Gilabert, M. A. (2009), 'Vegetation dynamics from ndvi time series analysis using the wavelet transform', *Remote sensing of environment* **113**(9), 1823–1842.

McDowell, N. G., Allen, C. D., Anderson-Teixeira, K., Aukema, B. H., Bond-Lamberty, B., Chini, L., Clark, J. S., Dietze, M., Grossiord, C., Hanbury-Brown, A. et al. (2020), 'Pervasive shifts in forest dynamics in a changing world', *Science* **368**(6494), eaaz9463.

McNicol, I. M., Ryan, C. M. and Mitchard, E. T. (2018), 'Carbon losses from deforestation and widespread degradation offset by extensive growth in african woodlands', *Nature communications* **9**(1), 3045.

Meng, S., Pang, Y., Zhang, Z., Jia, W. and Li, Z. (2016), 'Mapping aboveground biomass using texture indices from aerial photos in a temperate forest of northeastern china', *Remote Sensing* **8**(3), 230.

Mermoz, S. and Le Toan, T. (2016), 'Forest disturbances and regrowth assessment using alos palsar data from 2007 to 2010 in vietnam, cambodia and lao pdr', *Remote Sensing* **8**(3), 217.

Meyer, F. (2019), 'Spaceborne synthetic aperture radar: Principles, data access, and basic processing techniques', *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation* pp. 21–64.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. and Nauss, T. (2018), 'Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation', *Environmental Modelling & Software* **101**, 1–9.

Mitchard, E. T. (2018), 'The tropical forest carbon cycle and climate change', *Nature* **559**(7715), 527–534.

Mitchard, E. T., Feldpausch, T. R., Brienen, R. J., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., Lewis, S. L., Lloyd, J., Quesada, C. A., Gloor, M. et al. (2014), 'Markedly divergent estimates of a mazon forest carbon density from ground plots and satellites', *Global ecology and biogeography* **23**(8), 935–946.

Mitchard, E. T., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. L. and

Brown, S. (2013), ‘Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps’, *Carbon balance and management* **8**, 1–13.

Mitchard, E. T., Saatchi, S. S., Lewis, S., Feldpausch, T., Woodhouse, I. H., Sonké, B., Rowland, C. and 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 Sensing of Environment* **115**(11), 2861–2873.

Mitchard, E. T., Saatchi, S. S., White, L. J., Abernethy, K. A., Jeffery, K. J., Lewis, S. L., Collins, M., Lefsky, M. A., Leal, M. E., Woodhouse, I. H. et al. (2012), ‘Mapping tropical forest biomass with radar and spaceborne lidar in lopé national park, gabon: overcoming problems of high biomass and persistent cloud’, *Biogeosciences* **9**(1), 179–191.

Mitchard, E. T., Saatchi, S. S., Woodhouse, I. H., Nangendo, G., Ribeiro, N., Williams, M., Ryan, C. M., Lewis, S. L., Feldpausch, T. and Meir, P. (2009), ‘Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different african landscapes’, *Geophysical Research Letters* **36**(23).

Montesano, P., Nelson, R., Sun, G., Margolis, H., Kerber, A. and Ranson, K. (2009), ‘Modis tree cover validation for the circumpolar taiga–tundra transition zone’, *Remote Sensing of Environment* **113**(10), 2130–2141.

Motohka, T., Shimada, M., Uryu, Y. and Setiabudi, B. (2014), ‘Using time series palsar gamma nought mosaics for automatic detection of tropical deforestation: A test study in riau, indonesia’, *Remote Sensing of Environment* **155**, 79–88.

Mpakairi, K. S., Dube, T., Sibanda, M. and Mutanga, O. (2023), ‘Fine-scale characterization of irrigated and rainfed croplands at national scale using multi-source data, random forest, and deep learning algorithms’, *ISPRS Journal of Photogrammetry and Remote Sensing* **204**, 117–130.

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G. and Leblon, B. (2015), ‘Savannah woody structure modelling and mapping using multi-frequency (x-, c- and l-band) synthetic aperture radar data’, *ISPRS Journal of Photogrammetry and Remote Sensing* **105**, 234–250.

Narine, L. L., Popescu, S. C. and Malambo, L. (2020), 'Using icesat-2 to estimate and map forest aboveground biomass: A first example', *Remote Sensing* **12**(11), 1824.

Ngo, Y.-N., Ho Tong Minh, D., Baghdadi, N. and Fayad, I. (2023), 'Tropical forest top height by gedi: From sparse coverage to continuous data', *Remote Sensing* **15**(4), 975.

Ngomanda, A., Obiang, N. L. E., Lebamba, J., Mavouroulou, Q. M., Gomat, H., Mankou, G. S., Loumeto, J., Iponga, D. M., Ditsouga, F. K., Koumba, R. Z. et al. (2014), 'Site-specific versus pantropical allometric equations: which option to estimate the biomass of a moist central african forest?', *Forest Ecology and Management* **312**, 1–9.

Ni, W., Zhang, Z., Sun, G., Guo, Z. and He, Y. (2014), 'The penetration depth derived from the synthesis of alos/palsar insar data and aster gdem for the mapping of forest biomass', *Remote Sensing* **6**(8), 7303–7319.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B. et al. (2018), 'Attention u-net: Learning where to look for the pancreas', *arXiv preprint arXiv:1804.03999* .

Overman, J. P. M., Witte, H. J. L. and Saldarriaga, J. G. (1994), 'Evaluation of regression models for above-ground biomass determination in amazon rainforest', *Journal of tropical Ecology* **10**(2), 207–218.

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. et al. (2011), 'A large and persistent carbon sink in the world's forests', *science* **333**(6045), 988–993.

Pan, Y., Birdsey, R. A., Phillips, O. L. and Jackson, R. B. (2013), 'The structure, distribution, and biomass of the world's forests', *Annual Review of Ecology, Evolution, and Systematics* **44**, 593–622.

Pandit, S., Tsuyuki, S. and Dube, T. (2018), 'Estimating above-ground biomass in sub-tropical buffer zone community forests, nepal, using sentinel 2 data', *Remote Sensing* **10**(4), 601.

Pearson, T. R., Brown, S., Murray, L. and Sidman, G. (2017), 'Greenhouse gas emissions from tropical forest degradation: an underestimated source', *Carbon balance and management* **12**, 1–11.

Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S. and Wang, T. (2009), 'The carbon balance of terrestrial ecosystems in china', *Nature* **458**(7241), 1009–1013.

Piao, S., Fang, J., Zhu, B. and Tan, K. (2005), 'Forest biomass carbon stocks in china over the past 2 decades: Estimation based on integrated inventory and satellite data', *Journal of Geophysical Research: Biogeosciences* **110**(G1).

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., Liu, R., Mao, J., Myneni, R. B., Peng, S. et al. (2015), 'Detection and attribution of vegetation greening trend in china over the last 30 years', *Global change biology* **21**(4), 1601–1609.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N. et al. (2020), 'Spatial validation reveals poor predictive performance of large-scale ecological mapping models', *Nature communications* **11**(1), 4540.

Portner, H. O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B., Biesbroek, R. et al. (2022), Climate change 2022: Impacts, adaptation and vulnerability, Technical report, IPCC.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E. et al. (2021), 'Mapping global forest canopy height through integration of gedi and landsat data', *Remote Sensing of Environment* **253**, 112165.

Qin, Y., Xiao, X., Wigneron, J.-P., Ciais, P., Brandt, M., Fan, L., Li, X., Crowell, S., Wu, X., Doughty, R. et al. (2021), 'Carbon loss from forest degradation exceeds that from deforestation in the brazilian amazon', *Nature Climate Change* **11**(5), 442–448.

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. et al. (2019), 'The european space agency biomass mission: Measuring forest above-ground biomass from space', *Remote Sensing of Environment* **227**, 44–60.

Ramachandran, N., Saatchi, S., Tebaldini, S., d'Alessandro, M. M. and Dikshit, O. (2021), 'Evaluation of p-band sar tomography for mapping tropical forest vertical backscatter and tree height', *Remote Sensing* **13**(8), 1485.

Ranson, K., Sun, G., Kharuk, V. and Kovacs, K. (2001), 'Characterization of forests in western sayani mountains, siberia from sir-c sar data', *Remote Sensing of Environment* **75**(2), 188–200.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, f. (2019), 'Deep learning and process understanding for data-driven earth system science', *Nature* **566**(7743), 195–204.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J. and Hérault, B. (2017), 'biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests', *Methods in Ecology and Evolution* **8**(9), 1163–1167.

Ren, H., Chen, H., Li, L., Li, P., Hou, C., Wan, H., Zhang, Q. and Zhang, P. (2013), 'Spatial and temporal patterns of carbon storage from 1992 to 2002 in forest ecosystems in guangdong, southern china', *Plant and soil* **363**, 123–138.

Requena Suarez, D., Rozendaal, D. M., De Sy, V., Phillips, O. L., Alvarez-Dávila, E., Anderson-Teixeira, K., Araujo-Murakami, A., Arroyo, L., Baker, T. R., Bongers, F. et al. (2019), 'Estimating aboveground net biomass change for tropical and subtropical forests: Refinement of ipcc default rates using forest plot data', *Global Change Biology* **25**(11), 3609–3624.

Rodríguez-Veiga, P., Quegan, S., Carreiras, J., Persson, H. J., Fransson, J. E., Hoscilo, A., Ziółkowski, D., Stereńczak, K., Lohberger, S., Stängel, M. et al. (2019), 'Forest biomass retrieval approaches from earth observation in different biomes', *International Journal of Applied Earth Observation and Geoinformation* **77**, 53–68.

Rodríguez-Veiga, P., Saatchi, S., Tansey, K. and Balzter, H. (2016), 'Magnitude, spatial distribution and uncertainty of forest biomass stocks in mexico', *Remote Sensing of Environment* **183**, 265–281.

Rodríguez-Veiga, P., Wheeler, J., Louis, V., Tansey, K. and Balzter, H. (2017), 'Quantifying forest biomass carbon stocks from space', *Current Forestry Reports* **3**, 1–18.

Ronneberger, O., Fischer, P. and Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18', Springer, pp. 234–241.

Rosen, P., Hensley, S., Shaffer, S., Edelstein, W., Kim, Y., Kumar, R., Misra, T., Bhan, R. and Sagi, R. (2017), The nasa-isro sar (nisar) mission dual-band radar instrument preliminary design, *in* '2017 IEEE international geoscience and remote sensing symposium (IGARSS)', IEEE, pp. 3832–3835.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. et al. (1974), 'Monitoring vegetation systems in the great plains with erts', *NASA Spec. Publ* **351**(1), 309.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S. et al. (2011), 'Benchmark map of forest carbon stocks in tropical regions across three continents', *Proceedings of the national academy of sciences* **108**(24), 9899–9904.

Saatchi, S. S., HOUGHTON, R. A., Dos Santos Alvala, R., Soares, J. V. and Yu, Y. (2007), 'Distribution of aboveground live biomass in the amazon basin', *Global change biology* **13**(4), 816–837.

Samuele, D. P., Filippo, S., Orusa, T. and Enrico, B.-M. (2021), 'Mapping sar geometric distortions and their stability along time: A new tool in google earth engine based on sentinel-1 image time series', *International Journal of Remote Sensing* **42**(23), 9135–9154.

Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., Mura, M., Zolli, C., Maselli, F., Chiesi, M., Bottai, L. and Chirici, G. (2017), 'The potential of multifrequency sar images for estimating forest biomass in mediterranean areas', *Remote Sensing of Environment* **200**, 63–73.

Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J. E., Hall, R. J., Pathe, C., Schmullius, C., Schepaschenko, D., Shvidenko, A. et al. (2015), 'Forest growing stock volume of the northern hemisphere: Spatially explicit estimates for 2010 derived from envisat asar', *Remote Sensing of Environment* **168**, 316–334.

Santoro, M. and Cartus, O. (2021*a*), 'ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2010, 2017 and 2018, v3'.

Santoro, M. and Cartus, O. (2021*b*), 'Global datasets of forest above-ground biomass for the years 2010, 2017 and 2018, v2, centre for environmental data analysis'.

Santoro, M. and 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, v4', `https://dx.doi.org/10.5285/af60720c1e404a9e9d2c145d2b2ead4e`. Accessed: 2023-04-21.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabilie, V., Araza, A., De Bruin, S., Herold, M., Quegan, S., Rodríguez Veiga, P. et al. (2020), 'The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations', *Earth System Science Data Discussions* **2020**, 1–38.

Santoro, M., Cartus, O., Wegmüller, U., Besnard, S., Carvalhais, N., Araza, A., Herold, M., Liang, J., Cavlovic, J. and Engdahl, M. E. (2022), 'Global estimation of above-ground biomass from spaceborne c-band scatterometer observations aided by lidar metrics of vegetation structure', *Remote Sensing of Environment* **279**, 113114.

Santoro, M., Wegmüller, U., Lamarche, C., Bontemps, S., Defourny, P. and Arino, O. (2015), 'Strengths and weaknesses of multi-year envisat asar backscatter measurements to map permanent open water bodies at global scale', *Remote Sensing of Environment* **171**, 185–201.

Santoro, M. et al. (2021), 'Supplement of the global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations'.

Sarker, M. L. R., Nichol, J., Ahmad, B., Busu, I. and Rahman, A. A. (2012), 'Potential of texture measurements of two-date dual polarization palsar data for the improvement of forest biomass estimation', *ISPRS Journal of Photogrammetry and Remote Sensing* **69**, 146–166.

Schneider, F. D., Ferraz, A., Hancock, S., Duncanson, L. I., Dubayah, R. O., Pavlick, R. P. and Schimel, D. S. (2020), 'Towards mapping the diversity of canopy structure from space with gedi', *Environmental Research Letters* **15**(11), 115006.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D. and DiMarzio, J. P. (2005), 'Overview of the icesat mission', *Geophysical research letters* **32**(21).

Schwartz, M., Ciais, P., Ottlé, C., De Truchis, A., Vega, C., Fayad, I., Brandt, M., Fensholt, R., Baghdadi, N., Morneau, F. et al. (2024), 'High-resolution canopy height map in the landes forest (france) based on gedi, sentinel-1, and sentinel-2 data with a deep learning approach', *International Journal of Applied Earth Observation and Geoinformation* **128**, 103711.

Seidel, D., Fleck, S., Leuschner, C. and Hammett, T. (2011), 'Review of ground-based methods to measure the distribution of biomass in forest canopies', *Annals of Forest Science* **68**, 225–244.

Shen, W., Li, M., Huang, C., He, T., Tao, X. and Wei, A. (2019), 'Local land surface temperature change induced by afforestation based on satellite observations in guangdong plantation forests in china', *Agricultural and Forest Meteorology* **276**, 107641.

Shendryk, Y. (2022), 'Fusing gedi with earth observation data for large area aboveground biomass mapping', *International Journal of Applied Earth Observation and Geoinformation* **115**, 103108.

Shi, B., Gao, W. and Jin, G. (2015), 'Effects on rhizospheric and heterotrophic respiration of conversion from primary forest to secondary forest and plantations in northeast china', *European journal of soil biology* **66**, 11–18.

Shi, H. and Chen, J. (2018), 'Characteristics of climate change and its relationship with land use/cover change in yunnan province, china', *International journal of climatology* **38**(5), 2520–2537.

Shimada, M., Isoguchi, O., Motooka, T., Shiraishi, T., Mukaida, A., Okumura, H., Otaki, T. and Itoh, T. (2011), Generation of 10m resolution palsar and jers-sar mosaic and forest/non-forest maps for forest carbon tracking, *in* '2011 IEEE International Geoscience and Remote Sensing Symposium', IEEE, pp. 3510–3513.

Shimada, M., Isoguchi, O., Tadono, T. and Isono, K. (2009), 'Palsar radiometric and geometric calibration', *IEEE Transactions on Geoscience and Remote Sensing* **47**(12), 3915–3932.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R. and Lucas, R. (2014), 'New global forest/non-forest maps from alos palsar data (2007–2010)', *Remote Sensing of environment* **155**, 13–31.

Shugart, H., Saatchi, S. and Hall, F. (2010), 'Importance of structure and its measurement in quantifying function of forest ecosystems', *Journal of Geophysical Research: Biogeosciences* **115**(G2).

Silva, C. A., Duncanson, L., Hancock, S., Neuenschwander, A., Thomas, N., Hofton, M., Fatoyinbo, L., Simard, M., Marshak, C. Z., Armston, J. et al. (2021), 'Fusing simulated gedi, icesat-2 and nisar data for regional aboveground biomass mapping', *Remote Sensing of Environment* **253**, 112234.

Simard, M., Pinto, N., Fisher, J. B. and Baccini, A. (2011), 'Mapping forest canopy height globally with spaceborne lidar', *Journal of Geophysical Research: Biogeosciences* **116**(G4).

Sinha, S., Jeganathan, C., Sharma, L. K. and Nathawat, M. S. (2015), 'A review of radar remote sensing for biomass estimation', *International Journal of Environmental Science and Technology* **12**, 1779–1792.

Soenen, S. A., Peddle, D. R., Hall, R. J., Coburn, C. A. and Hall, F. G. (2010), 'Estimating aboveground forest biomass from canopy reflectance model inversion in mountainous terrain', *Remote Sensing of Environment* **114**(7), 1325–1337.

Song, J. (2015), 'Bias corrections for random forest in regression using residual rotation', *Journal of the Korean Statistical Society* **44**(2), 321–326.

Spawn, S. A., Sullivan, C. C., Lark, T. J. and Gibbs, H. K. (2020), 'Harmonized global maps of above and belowground biomass carbon density in the year 2010', *Scientific Data* **7**(1), 112.

Stage, A. R. and Salas, C. (2007), ‘Interactions of elevation, aspect, and slope in models of forest species composition and productivity’, *Forest Science* **53**(4), 486–492.

Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S. P., Patterson, P. L., Magnussen, S., Næsset, E., McRoberts, R. E. et al. (2016), ‘Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation’, *Forest Ecosystems* **3**, 1–11.

Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S. and Fang, J. (2016), ‘Spatial distribution of forest aboveground biomass in china: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data’, *Remote Sensing of Environment* **173**, 187–199.

Sze, J. S., Carrasco, L. R., Childs, D. and Edwards, D. P. (2022), ‘Reduced deforestation and degradation in indigenous lands pan-tropically’, *Nature Sustainability* **5**(2), 123–130.

Tartaglione, E., Lepsøy, S., Fiandrotti, A. and Francini, G. (2018), ‘Learning sparse neural networks via sensitivity-driven regularization’, *Advances in neural information processing systems* **31**.

Thomas, S. C. and Martin, A. R. (2012), ‘Carbon content of tree tissues: a synthesis’, *Forests* **3**(2), 332–352.

Tian, L., Tao, Y., Fu, W., Li, T., Ren, F. and Li, M. (2022), ‘Dynamic simulation of land use/cover change and assessment of forest ecosystem carbon storage under climate change scenarios in guangdong province, china’, *Remote Sensing* **14**(10), 2330.

Tong, X.-Y., Xia, G.-S. and Zhu, X. X. (2023), ‘Enabling country-scale land cover mapping with meter-resolution satellite imagery’, *ISPRS Journal of Photogrammetry and Remote Sensing* **196**, 178–196.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M. et al. (2012), ‘Gmes sentinel-1 mission’, *Remote sensing of environment* **120**, 9–24.

Tucker, C., Brandt, M., Hiernaux, P., Kariryaa, A., Rasmussen, K., Small, J., Igel, C., Reiner, F., Melocik, K., Meyer, J. et al. (2023), ‘Sub-continental-scale carbon stocks of individual trees in african drylands’, *Nature* **615**(7950), 80–86.

Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M. and Schmullius, C. (2018), ‘Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne lidar, and sar and optical satellite data in mexico’, *Carbon balance and management* **13**, 1–20.

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M. and Kmoch, A. (2020), ‘Vertical accuracy of freely available global digital elevation models (aster, aw3d30, merit, tandem-x, srtm, and nasadem)’, *Remote Sensing* **12**(21), 3482.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F. and Ceschia, E. (2017), ‘Understanding the temporal behavior of crops using sentinel-1 and sentinel-2-like data for agricultural applications’, *Remote sensing of environment* **199**, 415–426.

Viña, A., McConnell, W. J., Yang, H., Xu, Z. and Liu, J. (2016), ‘Effects of conservation policy on china’s forest recovery’, *Science advances* **2**(3), e1500965.

Vreugdenhil, M., Navacchi, C., Bauer-Marschallinger, B., Hahn, S., Steele-Dunne, S., Pfeil, I., Dorigo, W. and Wagner, W. (2020), ‘Sentinel-1 cross ratio and vegetation optical depth: A comparison over europe’, *Remote Sensing* **12**(20), 3404.

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C. and Strauss, P. (2018), ‘Sensitivity of sentinel-1 backscatter to vegetation dynamics: An austrian case study’, *Remote Sensing* **10**(9), 1396.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T. and Studnicka, N. (2006), ‘Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner’, *ISPRS journal of Photogrammetry and Remote Sensing* **60**(2), 100–112.

Wai, P., Su, H. and Li, M. (2022), ‘Estimating aboveground biomass of two different forest types in myanmar from sentinel-2 data with machine learning and geostatistical algorithms’, *Remote Sensing* **14**(9), 2146.

Wang, C. (2006), ‘Biomass allometric equations for 10 co-occurring tree species in chinese temperate forests’, *Forest Ecology and Management* **222**(1-3), 9–16.

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R. B. and Chang, Q. (2019), 'Estimating leaf area index and aboveground biomass of grazing pastures using sentinel-1, sentinel-2 and landsat images', *ISPRS Journal of Photogrammetry and Remote Sensing* **154**, 189–201.

Wang, J., Xiao, X., Liu, L., Wu, X., Qin, Y., Steiner, J. L. and Dong, J. (2020), 'Mapping sugarcane plantation dynamics in guangxi, china, by time series sentinel-1, sentinel-2 and landsat images', *Remote sensing of environment* **247**, 111951.

Wang, S., Zhang, Y., Ju, W., Chen, J. M., Ciais, P., Cescatti, A., Sardans, J., Janssens, I. A., Wu, M., Berry, J. A. et al. (2020), 'Recent global decline of co2 fertilization effects on vegetation photosynthesis', *Science* **370**(6522), 1295–1300.

Wang, Y., Song, Q., Du, Y., Wang, J., Zhou, J., Du, Z. and Li, T. (2019), 'A random forest model to predict heatstroke occurrence for heatwave in china', *Science of the total environment* **650**, 3048–3053.

WangYL, W. H. et al. (2022), 'Thesizeoftheland carbonsinkinchina', *nature* **603**, E7rE9.

Watson, J. E., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., Thompson, I., Ray, J. C., Murray, K., Salazar, A. et al. (2018), 'The exceptional value of intact forest ecosystems', *Nature ecology & evolution* **2**(4), 599–610.

Wenhua, L. (2004), 'Degradation and restoration of forest ecosystems in china', *Forest Ecology and Management* **201**(1), 33–41.

White, J. C., Wulder, M. A., Varhola, A., Vastaranta, M., Coops, N. C., Cook, B. D., Pitt, D. and Woods, M. (2013), 'A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach', *The Forestry Chronicle* **89**(6), 722–723.

Williamson, G. B. and Wiemann, M. C. (2010), 'Measuring wood specific gravity... correctly', *American Journal of Botany* **97**(3), 519–524.

Wilson, E. H. and Sader, S. A. (2002), 'Detection of forest harvest type using multiple dates of landsat tm imagery', *Remote Sensing of Environment* **80**(3), 385–396.

Woodhouse, I. H. (2017), *Introduction to microwave remote sensing*, CRC press.

Woodhouse, I. H., Mitchard, E. T. A., Brolly, M., Maniatis, D. and Ryan, C. M. (2012), 'Radar backscatter is not a'direct measure'of forest biomass', *Nature climate change* **2**(8), 556–557.

Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., Hilker, T., Bater, C. W. and Gobakken, T. (2012), 'Lidar sampling for large-area forest characterization: A review', *Remote sensing of environment* **121**, 196–209.

Xiao, Y., Wang, Q., Tong, X. and Atkinson, P. M. (2023), 'Thirty-meter map of young forest age in china', *Earth System Science Data* **15**(8), 3365–3386.

Xu, L., Saatchi, S. S., Shapiro, A., Meyer, V., Ferraz, A., Yang, Y., Bastin, J.-F., Banks, N., Boeckx, P., Verbeeck, H. et al. (2017), 'Spatial distribution of carbon stored in forests of the democratic republic of congo', *Scientific Reports* **7**(1), 15030.

Xu, L., Saatchi, S. S., Yang, Y., Yu, Y., Pongratz, J., Bloom, A. A., Bowman, K., Worden, J., Liu, J., Yin, Y. et al. (2021), 'Changes in global terrestrial live biomass over the 21st century', *Science Advances* **7**(27), eabe9829.

Xu, L., Shi, Y., Fang, H., Zhou, G., Xu, X., Zhou, Y., Tao, J., Ji, B., Xu, J., Li, C. et al. (2018), 'Vegetation carbon stocks driven by canopy density and forest age in subtropical forest ecosystems', *Science of the Total Environment* **631**, 619–626.

Xu, L., Yu, G., He, N., Wang, Q., Gao, Y., Wen, D., Li, S., Niu, S. and Ge, J. (2018), 'Carbon storage in china's terrestrial ecosystems: A synthesis', *Scientific Reports* **8**(1), 2806.

Xu, W., Xiao, Y., Zhang, J., Yang, W., Zhang, L., Hull, V., Wang, Z., Zheng, H., Liu, J., Polasky, S. et al. (2017), 'Strengthening protected areas for biodiversity and ecosystem services in china', *Proceedings of the National Academy of Sciences* **114**(7), 1601–1606.

Xu, Z., Sun, K. and Mao, J. (2020), Research on resnet101 network chemical reagent label image classification based on transfer learning, *in* '2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT', IEEE, pp. 354–358.

Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., Ma, C., Yan, J. and Wang, X. (2021), 'Lightgbm: accelerated genomically designed crop breeding through ensemble learning', *Genome Biology* **22**, 1–24.

Yang, H., Ciais, P., Frappart, F., Li, X., Brandt, M., Fensholt, R., Fan, L., Saatchi, S., Besnard, S., Deng, Z. et al. (2023), 'Global increase in biomass carbon stock dominated by growth of northern young forests over past decade', *Nature Geoscience* **16**(10), 886–892.

Yang, L., Liang, S. and Zhang, Y. (2020), 'A new method for generating a global forest aboveground biomass map from multiple high-level satellite products and ancillary information', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**, 2587–2597.

Yao, S., Tan, K., Wang, Y., Zhang, W., Liu, S. and Yang, J. (2024), 'Estimating terrain elevations at 10 m resolution by integrating random forest machine learning model and icesat-2, sentinel-1, and sentinel-2 satellite remotely sensed data', *International Journal of Applied Earth Observation and Geoinformation* **132**, 104010.

Yin, G., Zhang, Y., Sun, Y., Wang, T., Zeng, Z. and Piao, S. (2015), 'Modis based estimation of forest aboveground biomass in china', *PloS one* **10**(6), e0130143.

Yu, D., Zhou, L., Zhou, W., Ding, H., Wang, Q., Wang, Y., Wu, X. and Dai, L. (2011), 'Forest management in northeast china: history, problems, and challenges', *Environmental Management* **48**, 1122–1135.

Yu, H., Wan, X., Dong, Z., Zhang, Z. and Jia, J. (2022), 'Estimation of reference voltages for time-difference electrical impedance tomography', *IEEE Transactions on Instrumentation and Measurement* **71**, 1–10.

Yu, H., Zhang, Z., Gao, Y. and Jia, J. (2022), 'Multiscale voltage reconstruction with attention-based network for volume fraction prediction of industrial oil–water two-phase flow by eit', *IEEE Transactions on Instrumentation and Measurement* **71**, 1–9.

Yu, Z., Ciais, P., Piao, S., Houghton, R. A., Lu, C., Tian, H., Agathokleous, E., Kattel,

G. R., Sitch, S., Goll, D. et al. (2022), 'Forest expansion dominates china's land carbon sink since 1980', *Nature Communications* **13**(1), 5374.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F. and Arino, O. (2022), 'ESA Worldcover 10 m 2021 v200'.

Zanne, A. E., Lopez-Gonzalez, G., Coomes, D. A., Ilic, J., Jansen, S., Lewis, S. L., Miller, R. B., Swenson, N. G., Wiemann, M. C. and Chave, J. (2020), 'Data from: Towards a worldwide wood economics spectrum'.

Zhang, G., Wang, M. and Liu, K. (2019), 'Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china', *International Journal of Disaster Risk Science* **10**, 386–403.

Zhang, X., Huang, G., Liu, L., Zhai, M. and Li, J. (2018), 'Ecological and economic analyses of the forest metabolism system: A case study of guangdong province, china', *Ecological Indicators* **95**, 131–140.

Zhang, Y., Song, C., Band, L. E. and Sun, G. (2019), 'No proportional increase of terrestrial gross carbon sequestration from the greening earth', *Journal of Geophysical Research: Biogeosciences* **124**(8), 2540–2553.

Zhang, Y., Yao, Y., Wang, X., Liu, Y. and Piao, S. (2017), 'Mapping spatial distribution of forest age in china', *Earth and Space Science* **4**(3), 108–116.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J. and Ryu, S.-R. (2004), 'Estimating aboveground biomass using landsat 7 etm+ data across a managed landscape in northern wisconsin, usa', *Remote sensing of environment* **93**(3), 402–411.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C. and Sun, M. (2020), 'Graph neural networks: A review of methods and applications', *AI open* **1**, 57–81.

Zhu, J., Hu, H., Tao, S., Chi, X., Li, P., Jiang, L., Ji, C., Zhu, J., Tang, Z., Pan, Y. et al. (2017), 'Carbon stocks and changes of dead organic matter in china's forests', *Nature Communications* **8**(1), 151.

Zhu, X. and Liu, D. (2015), ‘Improving forest aboveground biomass estimation using seasonal landsat ndvi time-series’, *ISPRS Journal of Photogrammetry and Remote Sensing* **102**, 222–231.

Zhu, Z., Deng, X., Zhao, F., Li, S. and Wang, L. (2022), ‘How environmental factors affect forest fire occurrence in yunnan forest region’, *Forests* **13**(9), 1392.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A. et al. (2016), ‘Greening of the earth and its drivers’, *Nature climate change* **6**(8), 791–795.

Zolkos, S. G., Goetz, S. J. and Dubayah, R. (2013), ‘A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing’, *Remote sensing of environment* **128**, 289–298.

Appendix A

Appendix 1

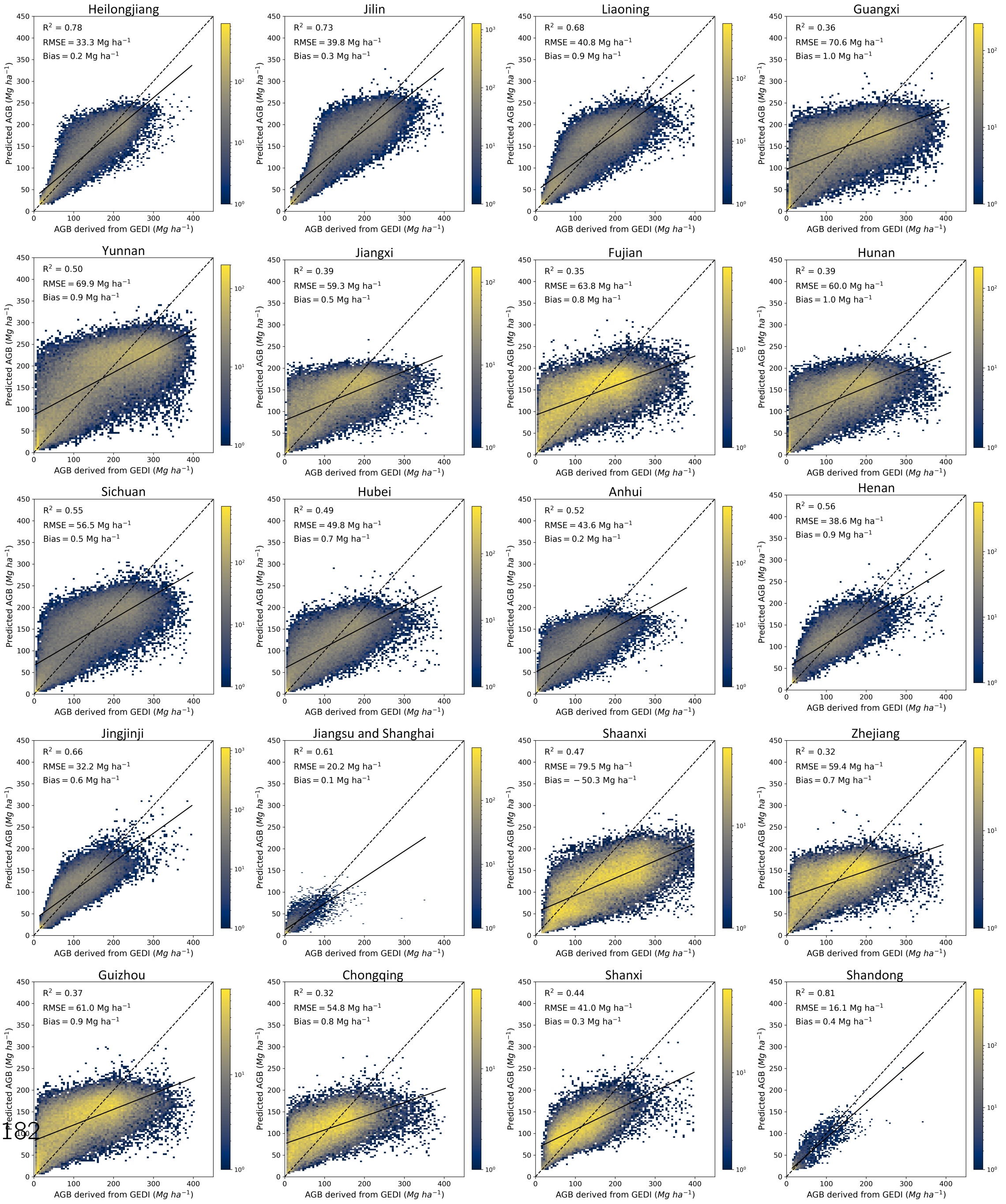


Figure A.1: Model performance for each province (Part 1).

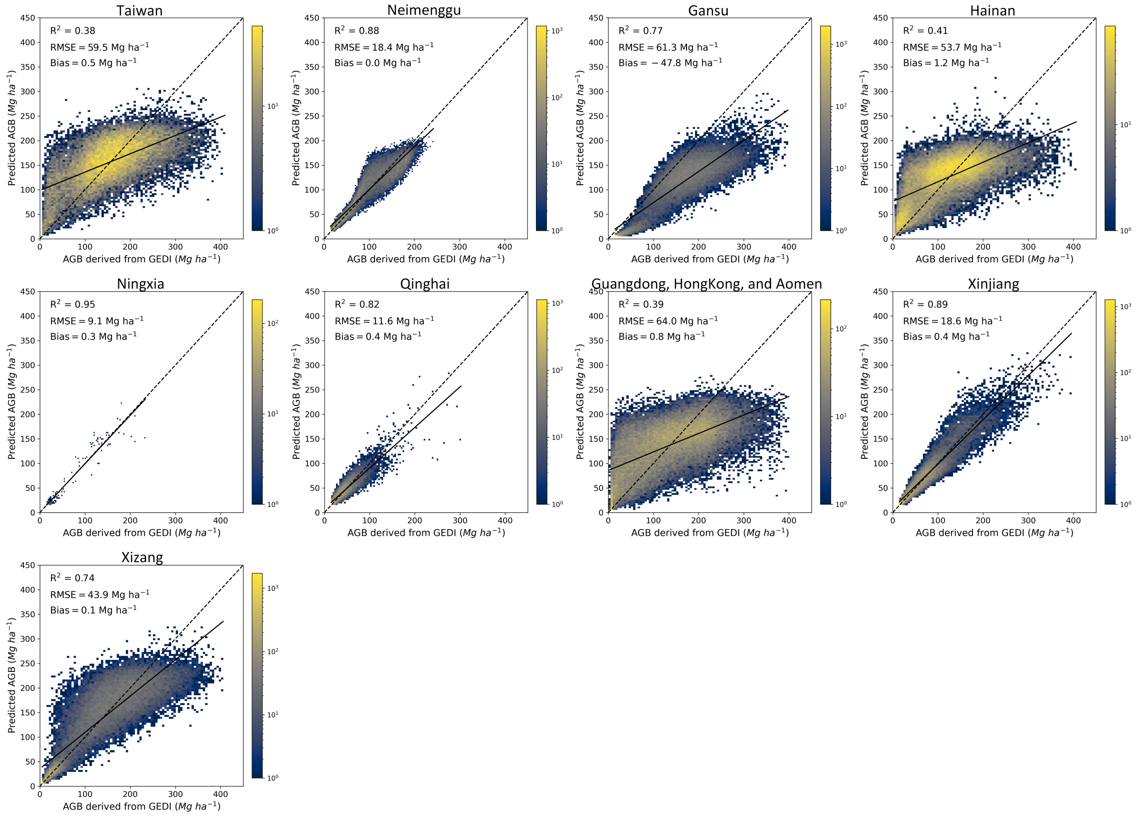


Figure A.2: Model performance for each province (Part 2).

Appendix B

Appendix 2

Table B.1: A Summary of remote sensing data used for each province

Province	Sentinel-1	Sentinel-2	PALSAR-2	Cloud (%)
Heilongjiang	DESCENDING	Apr-Sep, 2020-2022	2020-2022	20
Jilin	DESCENDING	Apr-Sep, 2020-2022	2020-2022	20
Liaoning	DESCENDING	Apr-Sep, 2020-2022	2020-2022	10
Guangxi	ASCENDING	whole year	2020-2022	10
Yunnan	ASCENDING	whole year	2020-2022	10
Jiangxi	ASCENDING	whole year	2020-2022	10
Fujian	ASCENDING	whole year	2020-2022	10
Hunan	ASCENDING	whole year	2020-2022	10
Sichuan	ASCENDING	whole year	2020-2022	20
Hubei	ASCENDING	whole year	2020-2022	10
Anhui	ASCENDING	whole year	2020-2022	10
Henan	ASCENDING	whole year	2020-2022	10
Jingjinji	ASCENDING	whole year	2020-2022	10
Jiangsu and Shanghai	ASCENDING	whole year	2020-2022	15
Shaanxi	ASCENDING	whole year	2020-2022	10
Zhejiang	ASCENDING	whole year	2020-2022	10
Guizhou	ASCENDING	whole year	2020-2022	15
Chongqing	ASCENDING	whole year	2020-2022	10
Shanxi	ASCENDING	whole year	2020-2022	10
Shandong	ASCENDING	whole year	2020-2022	10
Taiwan	ASCENDING	whole year	2020-2022	20
Neimenggu	ASCENDING/DESCENDING	Apr-Sep, 2020-2022	2020-2022	20
Gansu	ASCENDING	Apr-Sep, 2020-2022	2020-2022	20
Hainan	ASCENDING	whole year	2020-2022	20
Ningxia	ASCENDING	whole year	2020-2022	10
Qinghai	ASCENDING	Mar-Dec, 2020-2022	2021 majority, 2022 fill gaps	20
Guangdong, HongKong, and Aomen	ASCENDING	whole year	2020-2022	10
Xinjiang	DESCENDING	Apr-Sep, 2020-2022	2020-2022	20
Xizang	DESCENDING	Mar-Dec, 2020-2022	2020-2022	20

Appendix C

Appendix 3

Layers	Chongqing	Sichuan	Guizhou	Jilin	Heilongjiang	Hubei	Jiangsu	Xizang	Gansu	Hunan	Shandong
PALSAR_HV	0.0767	0.3723	0.0989	0.6142	0.5929	0.1301	0.5023	0.6424	0.8400	0.0540	0.7918
PALSAR_HH	0.0277	0.0196	0.0238	0.0104	0.0079	0.0263	0.0156	0.0109	0.0094	0.0224	0.0115
PALSAR_angle	0.0250	0.0192	0.0228	0.0117	0.0088	0.0212	0.0132	0.0099	0.0055	0.0256	0.0063
P_ratio	0.0262	0.0181	0.0238	0.0096	0.0073	0.0211	0.0135	0.0098	0.0052	0.0241	0.0087
Elevation	0.0378	0.0358	0.0403	0.0253	0.0135	0.0275	0.0262	0.0437	0.0077	0.0363	0.0120
slope	0.1405	0.0572	0.0957	0.0560	0.0539	0.2549	0.0481	0.0157	0.0131	0.2042	0.0103
Longitude	0.0379	0.0346	0.0299	0.0405	0.0172	0.0437	0.0232	0.0131	0.0083	0.0434	0.0124
Latitude	0.0356	0.0300	0.0332	0.0272	0.1160	0.0340	0.0254	0.0122	0.0170	0.0364	0.0126
SB1	0.0275	0.0200	0.0258	0.0102	0.0081	0.0211	0.0143	0.0102	0.0053	0.0239	0.0082
SB2	0.0210	0.0164	0.0204	0.0091	0.0073	0.0170	0.0117	0.0076	0.0044	0.0194	0.0057
SB3	0.0221	0.0129	0.0225	0.0095	0.0088	0.0173	0.0140	0.0069	0.0047	0.0207	0.0052
SB4	0.0220	0.0153	0.0213	0.0100	0.0075	0.0214	0.0214	0.0071	0.0048	0.0708	0.0053
SB5	0.0373	0.0174	0.0314	0.0091	0.0094	0.0193	0.0296	0.0088	0.0048	0.0247	0.0073
SB6	0.0200	0.0103	0.0177	0.0076	0.0062	0.0140	0.0156	0.0075	0.0035	0.0177	0.0066
SB7	0.0154	0.0086	0.0143	0.0061	0.0051	0.0106	0.0090	0.0063	0.0026	0.0135	0.0051
SB8	0.0178	0.0103	0.0161	0.0067	0.0054	0.0121	0.0109	0.0076	0.0029	0.0154	0.0044
SB8A	0.0165	0.0093	0.0148	0.0059	0.0049	0.0110	0.0088	0.0067	0.0027	0.0141	0.0044
SB9	0.0291	0.0200	0.0264	0.0107	0.0080	0.0213	0.0173	0.0139	0.0051	0.0267	0.0073
SB11	0.0263	0.0264	0.0260	0.0105	0.0092	0.0180	0.0159	0.0118	0.0050	0.0234	0.0062
SB12	0.0816	0.0226	0.0359	0.0106	0.0082	0.0199	0.0146	0.0110	0.0055	0.0313	0.0058
NDVI	0.0308	0.0850	0.1574	0.0114	0.0092	0.0553	0.0197	0.0393	0.0057	0.0456	0.0075
EVI	0.0700	0.0338	0.0594	0.0207	0.0286	0.0624	0.0453	0.0454	0.0058	0.0608	0.0082
Max_EVI	0.0260	0.0190	0.0246	0.0108	0.0081	0.0204	0.0153	0.0111	0.0052	0.0242	0.0078
Min_EVI	0.0294	0.0198	0.0254	0.0116	0.0099	0.0218	0.0161	0.0112	0.0059	0.0240	0.0087
Dif_EVI	0.0240	0.0176	0.0215	0.0096	0.0074	0.0193	0.0143	0.0099	0.0053	0.0221	0.0073
Sentinel1_VV	0.0211	0.0118	0.0200	0.0088	0.0098	0.0162	0.0109	0.0054	0.0039	0.0197	0.0059
Sentinel1_VH	0.0289	0.0204	0.0254	0.0161	0.0117	0.0215	0.0144	0.0074	0.0051	0.0276	0.0102
S_ratio	0.0259	0.0164	0.0254	0.0102	0.0098	0.0217	0.0134	0.0071	0.0056	0.0281	0.0072

Table C.1: Feature importances for different layers (Part 1)

Layers	Gansu	Taiwan	Anhui	Yunnan	Henan	Shanxi	Guangxi	Shaanxi	Fujian	Ningxia
PALSAR_HV	0.7397	0.0725	0.3005	0.0772	0.4162	0.2457	0.0864	0.0877	0.0493	0.9177
PALSAR_HH	0.0145	0.0341	0.0193	0.0195	0.0191	0.0270	0.0237	0.0240	0.0238	0.0280
PALSAR_angle	0.0089	0.0278	0.0178	0.0190	0.0172	0.0284	0.0222	0.0246	0.0231	0.0023
P_ratio	0.0084	0.0266	0.0189	0.0195	0.0161	0.0217	0.0270	0.0215	0.0278	0.0020
Elevation	0.0126	0.1225	0.0434	0.0309	0.0316	0.0362	0.0297	0.0591	0.0389	0.0023
slope	0.0199	0.1701	0.0720	0.0593	0.1117	0.1234	0.0701	0.2052	0.0891	0.0033
Longitude	0.0135	0.0350	0.0273	0.0469	0.0435	0.0337	0.0307	0.0356	0.0488	0.0027
Latitude	0.0276	0.0437	0.0507	0.0636	0.0265	0.0377	0.0313	0.0751	0.0458	0.0023
SB1	0.0087	0.0264	0.0197	0.0217	0.0176	0.0263	0.0256	0.0289	0.0262	0.0013
SB2	0.0074	0.0206	0.0155	0.0180	0.0140	0.0186	0.0232	0.0173	0.0244	0.0011
SB3	0.0078	0.0189	0.0148	0.0176	0.0140	0.0171	0.0393	0.0179	0.0295	0.0013
SB4	0.0078	0.0322	0.0165	0.0168	0.0137	0.0191	0.0294	0.0182	0.0242	0.0012
SB5	0.0080	0.0229	0.0145	0.0261	0.0145	0.0166	0.0336	0.0207	0.0452	0.0007
SB6	0.0059	0.0152	0.0150	0.0127	0.0119	0.0130	0.0193	0.0126	0.0203	0.0009
SB7	0.0043	0.0122	0.0103	0.0115	0.0093	0.0114	0.0161	0.0105	0.0152	0.0008
SB8	0.0048	0.0140	0.0118	0.0131	0.0101	0.0126	0.0194	0.0118	0.0194	0.0010
SB8A	0.0044	0.0124	0.0105	0.0117	0.0094	0.0112	0.0171	0.0108	0.0173	0.0008
SB9	0.0085	0.0250	0.0206	0.0218	0.0173	0.0232	0.0279	0.0212	0.0278	0.0085
SB11	0.0081	0.0230	0.0176	0.0189	0.0151	0.0193	0.0259	0.0187	0.0342	0.0008
SB12	0.0089	0.0281	0.0180	0.0235	0.0160	0.0266	0.0336	0.0212	0.0271	0.0013
NDVI	0.0096	0.0382	0.1233	0.2487	0.0210	0.0305	0.0715	0.1007	0.1370	0.0037
EVI	0.0101	0.0352	0.0404	0.0944	0.0291	0.0667	0.1616	0.0258	0.0674	0.0062
Max_EVI	0.0084	0.0259	0.0193	0.0203	0.0205	0.0246	0.0254	0.0275	0.0256	0.0011
Min_EVI	0.0096	0.0255	0.0200	0.0201	0.0187	0.0245	0.0247	0.0230	0.0239	0.0018
Dif_EVI	0.0085	0.0233	0.0175	0.0180	0.0162	0.0206	0.0222	0.0215	0.0213	0.0019
Sentinel1_VV	0.0062	0.0188	0.0124	0.0127	0.0136	0.0181	0.0175	0.0164	0.0180	0.0016
Sentinel1_VH	0.0085	0.0248	0.0160	0.0174	0.0196	0.0244	0.0265	0.0221	0.0267	0.0015
S_ratio	0.0093	0.0251	0.0161	0.0188	0.0166	0.0217	0.0229	0.0204	0.0226	0.0017

Table C.2: Feature importances for different layers (Part 2)

Layers	Neimenggu	Zhejiang	Hainan	Liaoning	Guangdong	Xinjiang	Jiangxi	Jingjinji	China
PALSAR_HV	0.8588	0.0569	0.0767	0.5764	0.0493	0.8607	0.0510	0.6035	0.30631
PALSAR_HH	0.0072	0.0262	0.0223	0.0130	0.0220	0.0101	0.0213	0.0276	0.01780
PALSAR_angle	0.0044	0.0279	0.0204	0.0136	0.0215	0.0057	0.0247	0.0156	0.01888
P_ratio	0.0045	0.0273	0.0230	0.0112	0.0251	0.0054	0.0260	0.0129	0.01765
Elevation	0.0063	0.0403	0.0372	0.0262	0.0314	0.0056	0.0442	0.0170	0.02847
slope	0.0087	0.1678	0.0798	0.0519	0.0559	0.0088	0.1146	0.0428	0.08649
Longitude	0.0128	0.0434	0.0465	0.0320	0.0321	0.0102	0.0313	0.0160	0.05512
Latitude	0.0116	0.0366	0.0446	0.0164	0.0363	0.0058	0.0327	0.0201	0.05840
SB1	0.0045	0.0278	0.0243	0.0129	0.0245	0.0038	0.0252	0.0147	0.01916
SB2	0.0038	0.0222	0.0213	0.0109	0.0260	0.0028	0.0218	0.0108	0.01550
SB3	0.0041	0.0207	0.0331	0.0120	0.0412	0.0027	0.0287	0.0098	0.02134
SB4	0.0042	0.0247	0.0241	0.0114	0.0283	0.0098	0.0230	0.0098	0.02201
SB5	0.0042	0.0233	0.0258	0.0132	0.0360	0.0043	0.0352	0.0090	0.02209
SB6	0.0036	0.0188	0.0181	0.0089	0.0229	0.0023	0.0170	0.0076	0.01336
SB7	0.0027	0.0152	0.0143	0.0075	0.0169	0.0019	0.0134	0.0068	0.01051
SB8	0.0030	0.0175	0.0163	0.0078	0.0201	0.0026	0.0164	0.0077	0.01242
SB8A	0.0028	0.0155	0.0140	0.0072	0.0190	0.0022	0.0146	0.0065	0.01093
SB9	0.0044	0.0302	0.0243	0.0126	0.0324	0.0044	0.0253	0.0144	0.01975
SB11	0.0058	0.0280	0.0274	0.0122	0.0219	0.0041	0.0447	0.0108	0.01825
SB12	0.0049	0.0253	0.0278	0.0114	0.0290	0.0064	0.0249	0.0117	0.01947
NDVI	0.0053	0.0815	0.1634	0.0133	0.1701	0.0113	0.1858	0.0156	0.02982
EVI	0.0056	0.0634	0.0899	0.0400	0.1042	0.0047	0.0434	0.0246	0.07142
Max_EVI	0.0044	0.0273	0.0242	0.0115	0.0235	0.0041	0.0238	0.0131	0.01908
Min_EVI	0.0054	0.0274	0.0241	0.0146	0.0225	0.0051	0.0232	0.0155	0.02006
Dif_EVI	0.0041	0.0241	0.0208	0.0111	0.0203	0.0039	0.0206	0.0125	0.01721
Sentinel1_VV	0.0034	0.0238	0.0151	0.0095	0.0152	0.0032	0.0182	0.0121	0.01292
Sentinel1_VH	0.0045	0.0292	0.0207	0.0189	0.0308	0.0040	0.0233	0.0176	0.01833
S_ratio	0.0050	0.0278	0.0205	0.0134	0.0216	0.0039	0.0257	0.0139	0.01726

Table C.3: Feature importances for different layers (Part 3)