分类号 密级

U D C 编号 10486

武漢大學

硕 士 学 位 论 文

（学术学位）

利用多传感器遥感的方法绘制森林地上生物量地图

研究生姓名 ： Israr Ahmad

学号 ： 2022276190015

校内导师姓名、职称 ： Prof. Dr. Zhenfeng SHAO

学 科 、 专 业 名 称 ： 摄 影 测 量 与 遥 感

研 究 方 向 ： 地上生物量

二〇二四年五月

MSc Dissertation of Wuhan University

Forest Aboveground Biomass Mapping Using a Multi-Sensor Remote Sensing Approach

By

Israr Ahmad

Supervised by

Prof. Dr. Zhenfeng SHAO

Wuhan University

May 2024

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者（签名）：Israr Ahmad

年 2024 月 05 日 24

ABSTRACT

Climate change poses a significant challenge globally, however, developing countries are particularly vulnerable to its impacts. Forests emerge as a vital component in this context due to their function in the Earth's carbon exchange system and as reservoirs of terrestrial carbon. Forests not only contribute to the livelihood of a substantial portion of the global population but also serve as sanctuaries for rich biodiversity. However, these critical ecosystems are threatened by accelerating deforestation rates, primarily driven by agricultural expansion in developing regions, including the three continents: Asia, Africa, and Latin America. This deforestation disrupts forest carbon storage, threatens biodiversity, increases urban pollution, and climate change issues.

China's 23.63% forest cover is rapidly diminishing due to agricultural expansion and population growth, intensifying the country's vulnerability to local climate variations. Rapid urban development in cities like Wuhan has caused a surge in pollution levels, adversely affecting the local climate and ecosystem. Wuhan's air pollution surpasses safe limits, leading to respiratory issues and environmental deterioration. Accurate mapping of aboveground biomass in small urban forest parks is important to address the impacts of urbanization on local climate, recreational spaces, and pollution from population growth. Mapping accurate AGB met some challenges. This study identifies two key scientific issues: firstly, the lack of accurate forest ground sample collection, and secondly, addresses the limitation of ground sample collection from restricted and remote areas.

The main content of this thesis includes:

1) Importance of forest biomass and state-of-the-art measurement methods, emphasizing the role of forests in carbon sequestration and local-level climate change mitigation.
2) Focus on sampling strategies for ground samples, data acquisition, and analysis methods, including the use of Sentinel-2 and Sentinel-1 data, and introduces the GEDI L4B dataset for validation.
3) Results and comparison of the biomass assessment and modeling analysis for the Maanshan Forest Park. It includes stand structure analysis, vegetation index analysis, and model evaluations.

The research is centered on the Maanshan Forest Park in the Wuchang District of Wuhan, China. The study utilizes Sentinel-1 and Sentinel-2 data to estimate forest aboveground biomass using two different validation datasets i.e., (Ground Samples and GEDI L4B mean AGB) and their estimation results comparison, despite encountering some challenges in establishing robust statistical relationships between vegetation indices and AGB. Machine learning models, specifically the Random Forest (RF) and Support Vector Machine (SVM) algorithms, showed potential contributions in enhancing AGB estimation, with varying degrees of success. RF notably outperformed the SVM and Stepwise Regression models in predicting AGB.

RF model predicted AGB within the range of 47.37 Mg ha^{-1} to 86.62 Mg ha^{-1}, exhibiting a significant statistical association (R^2=0.87, RMSE=28.19 mg/ha), SVM performed (R^2=0.51, RMSE=34.3 mg/ha) and SR performed (R^2=0.34, RMSE=26.47 mg/ha) with field-measured AGB. Additionally, RF model predicted AGB within the range of 50.37 Mg ha^{-1} to 90.62 Mg ha^{-1}, with a significant statistical association (R^2=0.83, RMSE=25.75 mg/ha), SVM showed (R^2=0.48, RMSE=21.51 mg/ha) and SR performed (R^2=0.31, RMSE=30.16 mg/ha) with GEDI L4B validation dataset. The RF algorithm proved to be effective in AGB estimation.

The main initiative of this study lies in the ground sampling method devised to accurately map aboveground biomass (AGB). This method incorporates real-time ground observation and takes into account the shape of the plots and precise locations from each corner of the sampling plot for accurate sampling of forest trees to map AGB with improved accuracy.

In conclusion, the study highlights the importance of forests in mitigating local-level climate change and emphasizes the need to preserve and enhance forest carbon stocks. It demonstrates the challenges and opportunities in using advanced technologies for forest conservation. In the future, longer wavelength radar data can be explored for improved biomass estimation through deeper canopy penetration. And to develop specialized allometric equations and integrate GEDI data with field measurements to enhance the accuracy of above-ground biomass mapping.

Keywords: AGB Mapping; SAR; Sentinel-2; Global Ecosystem Dynamics Investigation; Machine Learning

摘 要

在气候变化的背景下，全球各个国家尤其是发展中国家面临着重大挑战。森林作为陆地生态系统中的碳库，是生态系统的重要组成部分，在全球碳循环中发挥着重要作用。森林生态系统不仅为人类的生存和发展做出了重大贡献，还在维护生物多样性方面发挥着重要作用。发展中国家受到农业扩张的影响，尤其在是亚洲、非洲和拉丁美洲三个大洲的生态系统正面临着森林过度砍伐的威胁。同时，这也会对碳存储和生物多样造成影响。

由于农业扩张和人口增长，中国的森林覆盖面积减少了 23.63%，这加剧了气候变化背景下生态系统的脆弱性。因此，亟需采取行动减少森林砍伐，保护森林生物多样性。武汉等城市的快速发展导致环境污染水平激增，对当地气候和生态系统产生了不利的影响。武汉的空气污染程度超过了安全阈值，可能会引起了呼吸道疾病和环境恶化等问题。要解决城市化对当地气候、休闲空间和人口增长造成的问题，准确绘制小型城市森林公园的地上生物量图非常重要。绘制准确的 AGB 图遇到了一些挑战。 本研究发现了两个关键的科学问题：一个是缺乏准确的森林地面样本采集，另一个是从限制区和偏远地区采集地面样本具有局限性。

本论文的主要内容包括:

1）全面概述了森林生物量的重要性和测量方法，强调了森林在固碳和缓解区域尺度气候变化的重要作用。

2）聚焦采样策略、数据获取和分析方法 ，包括哨兵-2 号和哨兵-1 号数据的使用以及用于验证的 GEDI L4B 数据集。

3）针对马鞍山森林公园的生物量评估和建模分析，分别阐述了其结果和有效性, 包括林分结构分析、植被指数分析和模型评估。

本研究以中国武汉市武昌区的马鞍山森林公园为研究对象。研究结果表明，尽管在建立植被指数与地上生物量（Aboveground Biomass，AGB）之间的稳健关系方面存在挑战，但使用哨兵-1 号和哨兵-2 号数据在估算森林生物量方面具有良好的效果。机器学习模型，尤

其是随机森林（Random Forest，RF）和支持向量机（Support Vector Machines，SVM）算法，在估算地上生物量方面表现出不同的优势。随机森林在 AGB 估算方面明显优于支持向量机算法和逐步回归模型。

实验结果表明，使用优化随机森林模型预测的马鞍山森林公园 AGB 范围在 47.37 ~ 86.62 Mg ha^{-1} 之间，与实测 AGB 存在显著的相关性（R^2 =0.87，RMSE=28.19 mg/ha）。支持向量机算法预测的结果与实测 AGB 的 R^2 为 0.51，均方根误差为 34.3 mg/ha；逐步回归模型预测结果与实测 AGB 的 R^2 为 0.34，均方根误差为 26.47 mg/ha。此外，使用随机森林模型预测的 AGB 范围在 50.37 ~ 90.62 Mg ha^{-1} 之间，与 GEDI L4B 验证数据集存在显著的相关性（R^2 =0.83，RMSE=25.75 mg/ha）。支持向量机算法预测结果与 GEDI L4B 验证数据集的 R^2 为 0.48，均方根误差为 21.51 mg/ha，逐步回归模型预测结果与 GEDI L4B 验证数据集的 R^2 为 0.31，均方根误差为 30.16 mg/ha。总的来说，随机森林算法在估算 AGB 方面效果更佳。

本研究的主要创新点在于采用地面取样法精确绘制地上生物量（AGB）图。此方法结合了实时地面观测，并考虑到地块的形状和采样地块每个角落的精确位置，对林木进行精确取样，从而提高了绘制 AGB 地图的准确性。

总之，本研究强调了森林在减缓地方一级气候变化方面的重要性，并强调了保护和提高森林碳储量的必要性。它表明了在利用先进技术保护森林方面存在的挑战和机遇。未来，可以探索更长波长的雷达数据，通过更深的冠层穿透来改进生物量估算。并开发专门的异速方程，将 GEDI 数据与实地测量相结合，以提高地上生物量绘图的准确性。

关键词: AGB 测绘；合成孔径雷达；哨兵-2；全球生态系统动态调查；机器学习

Table of contents

List of figures

List of tables

List of abbreviations

AFOLU	Agriculture, Forestry, and Other Land Use
AGB	Aboveground Biomass
ASM	Angular Second Moment
BEF	Biomass Expansion Factors
BOA	Bottom of Atmosphere
CO_2	Carbon Dioxide
DBH	Diameter at Breast Height
EC	European Commission
ECV	Essential Climate Variable
ESA	European Space Agency
FAO	Food and Agriculture Organization
GEDI	Global Ecosystem Dynamics Investigation
GHG	Greenhouse Gas
GIS	Geographic Information System
GLCM	Grey Level Co-Occurrence Matrix
GMES	Global Monitoring for Environment and Security
GRASS	Geographic Resources Analysis Support System
GRD	Ground Range Detection
Gt	Gigatons
ICESat	Ice Cloud and Land Elevation Satellite
LiDAR	Light Detection and Ranging
LULC	Land Use Land Cover
MFP	Maanshan Forest Park
MLR	Multiple linear regression
NASA	National Aeronautics and Space Administration
PPM	Parts Per Million
RBF	Radial Basis Function
REDD	Reducing Emissions from Deforestation and Forest Degradation
RF	Random Forest
RMSE	Root Mean Square Error
SCP	Semi-Automatic Classification Plugin
SNAP	Sentinel Application Platform
SOC	Soil Organic Carbon
SR	Stepwise Regression
SVM	Support Vector Machine
TOA	Top of Atmosphere
UNFCCC	United Nations Framework Convention on Climate Change
WD	Wood Density

CHAPTER 1 INTRODUCTION

1.1 Background

Climate change, propelled by escalating greenhouse gas (GHG) emissions, stands as a major environmental challenge in today's world [1]. A primary greenhouse gas (GHG), the carbon dioxide (CO_2) concentration has risen from 277 parts per million (ppm) during the pre-industrial era to 409.85 ppm in 2019. This increase has been linked to temperature rise, changed patterns of precipitation, and erratic extreme weather occurrences [2]. The role of human activities in this climatic shift is undeniable, with the period 2006-2015 witnessing a land temperature increase of 1.53°C over the 1850-1900 baseline. Deforestation and forest degradation have been significant contributors to greenhouse gas (GHG) emissions, with estimates indicating a notable increase over the years. In 1970, the estimated GHG emissions from these activities were around 490 ± 180 gigatons (Gt) of carbon dioxide (CO_2). By 2010, this figure had risen to approximately 680 ± 300 Gt CO_2 [3]. Agriculture, Forestry, and Other Land Use (AFOLU) sectors are responsible for nearly 23% of anthropogenic GHG emissions.

Spanning a substantial portion (31%) of the planet's land area, forests, are vital for their role in carbon storage, sequestering an estimated 2-4 gigatons of carbon annually [4]. They are crucial for climate change mitigation, supporting biodiversity, and providing livelihoods for 7.888 billion people globally. The global forest biomass, particularly AGB is essential for determining the carbon status of forests, representing a significant portion of the forest's total carbon pool [5]. Efforts to reduce GHG emissions through initiatives like reduced deforestation and forest degradation (REDD+) underline the need of precisely measuring the carbon and biomass of forests.

Despite forests pivotal role in climate regulation, there is still little reliable data on the biomass and carbon status of forests, especially across national and subnational scales. Many believe that significant afforestation and reforestation efforts are essential to limit global warming to 1.5°C or 2°C. Initiatives like REDD+ offer cost-effective strategies for climate mitigation [3]. However, the success of these initiatives hinges on the precise measurement of forest biomass, including all living biomass above and below the ground, and other carbon pools like deadwood, litter, and soil

organic carbon (SOC) [6]. China underscores the urgency in addressing deforestation and its impact on climate change. The local community's dependency on forestry resources, combined with socio-economic challenges and the absence of alternative fuels, further exacerbates deforestation pressures. This situation underscores the crucial importance of sustainable forest management and conservation strategies. These efforts are vital for mitigating the impacts of climate change and safeguarding global forest resources.

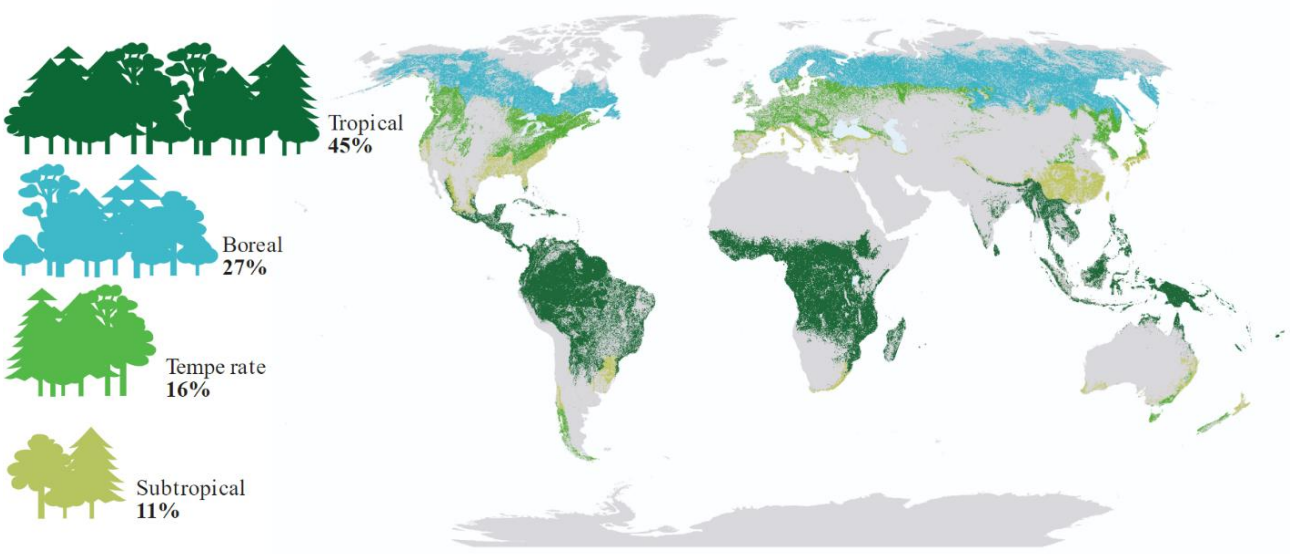

Figure 1 Global forest area proportion and distribution across climatic zones (FAO, 2020b)

The purpose of this research is to assess forest biomass in MFP by integrating field inventorying techniques and the latest GEDI L4B dataset along with advanced remote sensing and machine learning approaches. This research seeks to inform effective forest management and conservation strategies that can mitigate climate change impacts local scale, preserve biodiversity, and support the livelihoods of local communities.

1.2 Study area

Maanshan Forest Park is located in Wuchang district, Wuhan city. Wuhan is in the middle of Hubei province, central China, and the capital of the province. The city is located at the confluence of the Yangtze River and Hanshui River. Maanshan Forest Part is located with Easte Lake and Yujia Mountain covering a total area of 8,117,343 m^2. Wuhan is the economic, political,

and financial center and the most populous city of central China, with a population of over 11 million. The city of Wuhan is considered a transportation center for travel to nine provinces, which means that many seaways, expressways and railways surround it.

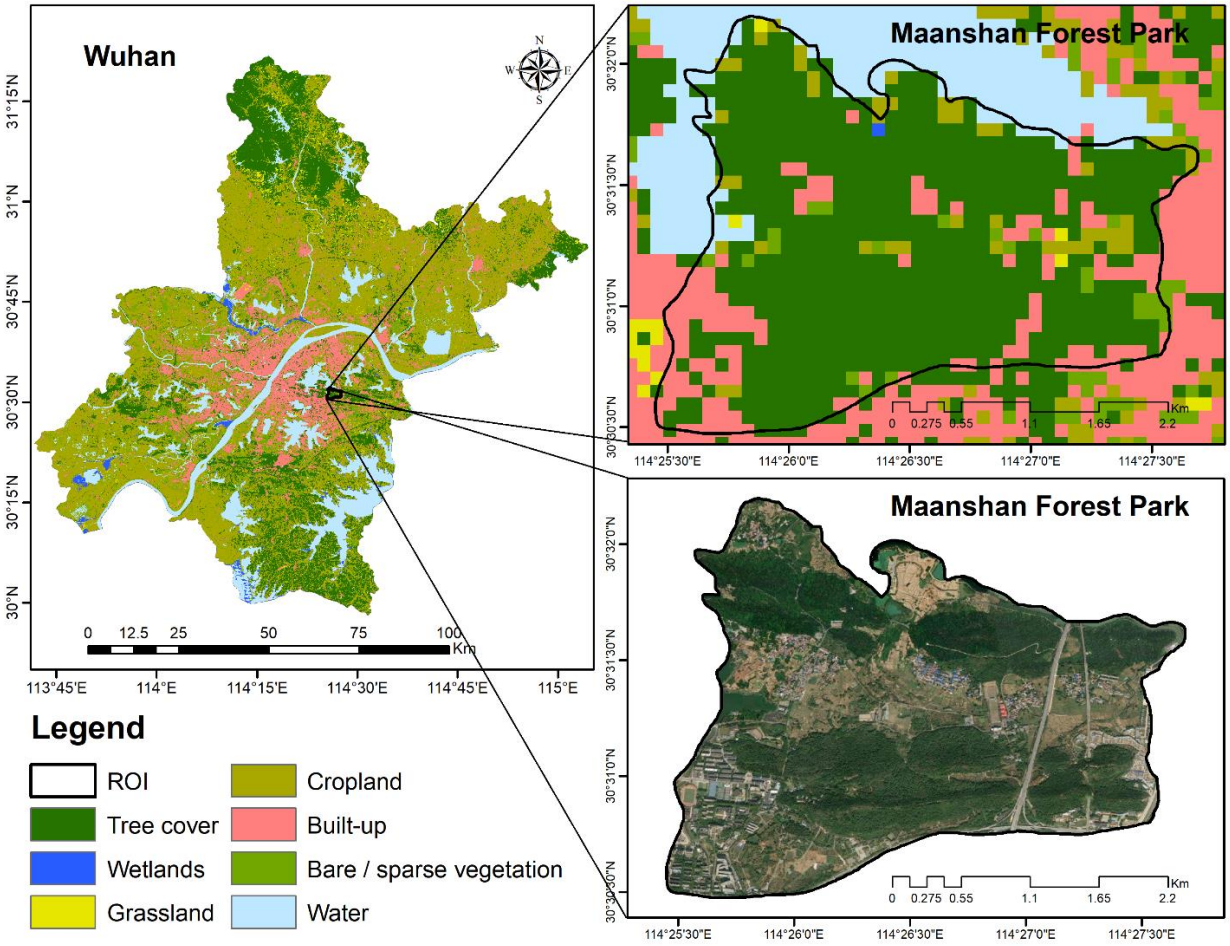

Figure 2 Study area

There are seven urban districts in Wuhan and six districts in suburban or rural areas. The urban area consists of three major towns divisions, i.e., Wuchang, Hankou, and Hanyang, by Yangtze River and Han River. The terrain consists of Hilly and Plain areas with an average altitude below 50m. During the hottest month mean temperature recorded as 29.8°C, where maximum temperature recorded as 39.4°C, and the mean temperature during the coldest month as 3°C, where max-minimum temperature recorded as -18.1°C [7]. There are four LULC types, i.e., water, urban, forest, and cropland. In 2017 the percentage of LULC in each type was analyzed as; 8.1% water,

20.2% urban, 14.8% forest, and 56.9% cropland [8]. The study area is dominant with the lakes and the Yangtze River also passes through the middle.

1.3 Scientific questions

Urban climate change poses a significant threat globally, with developing countries bearing a disproportionate impact. Forests are crucial for their role in carbon sequestration and as sources of livelihood and biodiversity, are under threat from deforestation, exacerbated by the use of trees as a fuel source by urban societies contributing to environmental degradation and loss of green spaces, and population growth, especially in developing countries like China. Effective forest management and conservation strategies are imperative to mitigate climate change impacts, preserve biodiversity, and support local communities. However, achieving these goals are affected by the limitation of enough ground samples and accurate assessment of AGB.

1) What are the current biomass levels in MFP, and how can they be accurately assessed using field inventorying techniques, GEDI L4B, and remote sensing data?
2) How effective are machine learning algorithms in estimating AGB in MFP?

1.4 Research objectives

The primary objective of this research is to comprehensively assess the biomass in MFP. Specific objectives include:

1) Quantifying aboveground biomass (AGB) using field inventorying and GEDI L4B mean AGB along with remote sensing data.
2) Exploring the relationship between satellite data and forest AGB to develop accurate estimation models.
3) Exploring the relationship and AGB estimation trend between two different validation datasets i.e. field inventory data and GEDI L4B.

1.5 Structure of a thesis

Chapter 1 elaborates on the general introduction and background information, setting the context for the research topic.

Chapter 2 provides the literature review and related work, covering theoretical studies on the significance of forest biomass, forest measurements, the importance of vegetation indices, GEDI L4B data, remote sensing data, and machine learning models.

Chapter 3 provides descriptive studies that outline the materials and methods employed, including field data collection methods, satellite remote sensing techniques, variable extractions, data analysis, and the models used.

Chapter 4 presents the results and validation of the research, encompassing biomass assessment, model evaluations, correlation analysis, above-ground biomass (AGB) mapping, and comparisons.

Chapter 5 concludes the thesis by summarizing the main conclusions and providing insights into future work. (Fig. 3) effectively visualizes the logical flow.

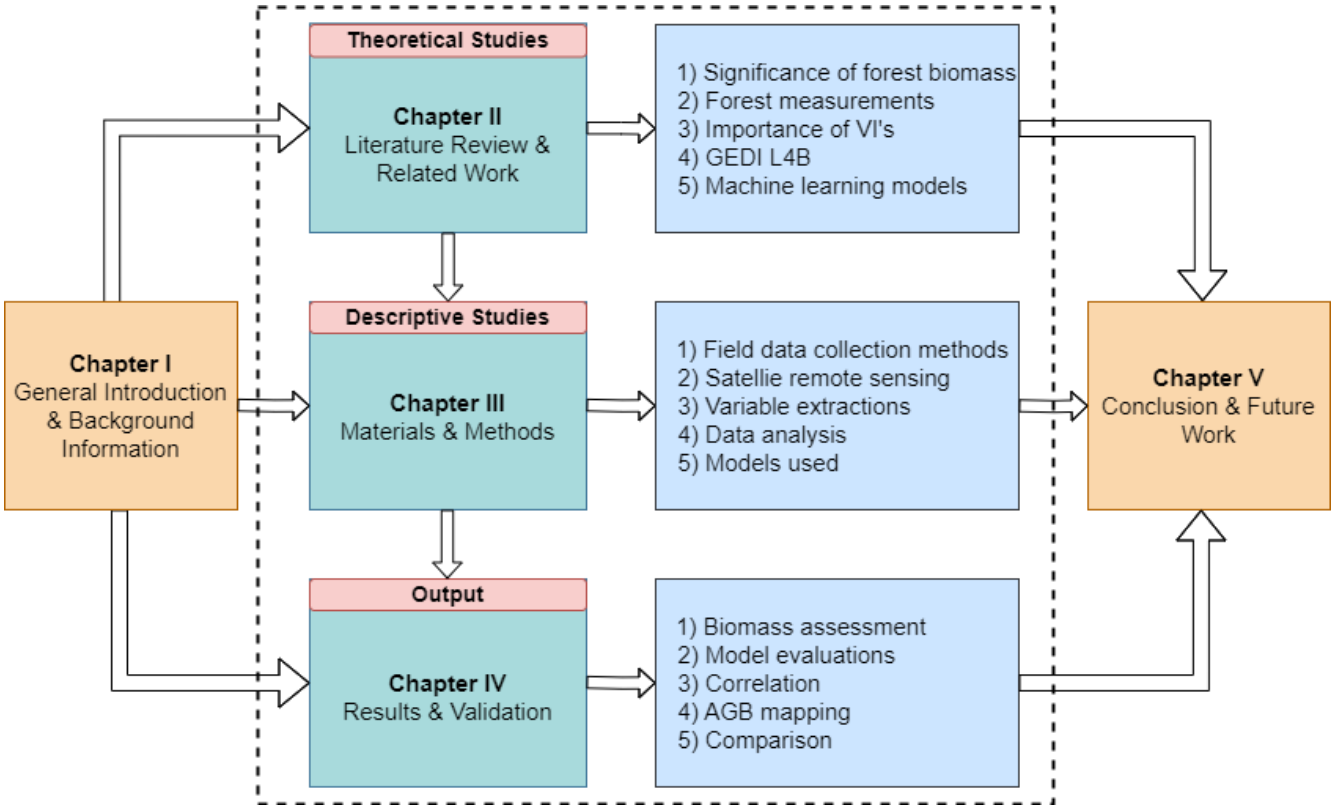

Figure 3 Thesis structure diagram

CHAPTER 2 LITERATURE REVIEW

2.1 Significance of forest biomass

Urban climate change, recognized as one of the most pressing global threats, is primarily caused by greenhouse gas emissions, especially rising carbon dioxide (CO_2) levels. These increasing CO_2 levels impact temperatures, rainfall patterns, and extreme weather events in urban areas [1, 9]. Between 1750, when the Industrial Era began, and 2019, CO_2 concentrations increased dramatically from about 277 parts per million (ppm) to 409.85 ± 0.1 ppm, with anthropogenic activities being the main culprits behind this dramatic increase [2, 10]. This escalation has notably warmed land temperatures by 1.53°C from 2006-2015 compared to the period 1850-1900, highlighting the severity of climate change impacts [3].

Forests, covering 4.06 billion hectares or 31% of the Earth's surface, play a crucial role in carbon sequestration, accounting for 80% of terrestrial carbon and significantly mitigating climate change effects [4, 11]. Even though forests play a vital role in absorbing greenhouse gases, cutting down trees and damaging existing forests are releasing a significant amount of these gases into the atmosphere. In fact, between 2007 and 2016, human activities related to land use, including agriculture, forestry, and deforestation, were responsible for nearly 23% of all greenhouse gas emissions [3]. The global mean forest carbon is estimated at 75 Mg C ha^{-1}, with forests annually storing 2-4 gigatons of carbon, highlighting their significance for the global carbon cycle and are critical for 1.6 billion people worldwide [12].

AGB, or aboveground biomass, makes up between 70% and 90% of the total biomass in forests, and emerges as a critical carbon pool within forest ecosystems. Changes in land usage and disturbances like fires and logging have an immediate effect on it, thereby playing a significant role in carbon fluxes [13]. Initiatives such as REDD+ provide financial incentives for developing countries to combat deforestation. Consequently, accurate measurement of carbon and biomass in forests becomes imperative for effective climate mitigation strategies [14].

Protected areas serve as crucial biodiversity conservation tools and play a key role in preventing deforestation and preserving forest carbon stocks. However, they face anthropogenic pressures that threaten their integrity, particularly in developing regions [15]. This comprehensive analysis

underscores the intricate relationship between forests, climate change, and anthropogenic activities. It highlights the critical role of forest ecosystems in carbon sequestration, the alarming rates of deforestation, and the significance of protected areas in mitigating climate change impacts. Initiatives like REDD+ and the designation of protected areas are vital for preserving forest carbon stocks, yet the effectiveness of these measures depends on addressing improving forest management techniques and identifying the root causes of deforestation.

2.2 Measurements of forest biomass using remotely sensed characteristics and forest inventories

Accurate aboveground biomass (AGB) measurement is essential for carbon stock assessment but challenging due to the limitations of direct, destructive sampling methods, which are labor-intensive, costly, and potentially harmful to biodiversity [16]. Non-destructive alternatives involve using allometric equations derived from forest inventory data, including tree diameter, height, and wood density, which are reliable but require precise selection to avoid significant estimation errors [17]. Wood Density (WD) and Biomass Expansion Factors (BEF) offer another non-destructive inventory-based method, though their scalability is limited [18, 19].

Remote sensing, combined with field data, is favored for its ability to cover extensive areas efficiently, making it a popular AGB estimation method endorsed by the UN-REDD program. It encompasses passive sensors (optical remote sensing), and active sensors like Radar and LiDAR, each with unique advantages and challenges [13]. Passive sensors are widely used for their correlation with vegetation indices and AGB but are limited by sunlight dependency and cloud cover interference [20]. Radar, unaffected by these conditions, offers potential for AGB retrieval through backscatter analysis, despite issues like data saturation and complexity [21]. LiDAR technology, though expensive, provides precise 3D forest structure data, greatly benefiting AGB estimation. Despite their drawbacks, these remote sensing methods remain critical for advancing AGB estimation efforts, especially in resource-constrained settings.

Satellite-mounted LiDAR systems, such as those used in the Global Ecosystem Dynamics Investigation (GEDI) and the Ice Cloud and Land Elevation Satellite (ICESat-2) missions by the National Aeronautics and Space Administration (NASA), are essential for assessing vertical forest

structures. The GEDI mission, in particular, is noteworthy for being the first spaceborne LiDAR mission specifically designed for this purpose [22, 23]. While LiDAR offers exceptional capabilities for AGB estimation, its high cost can be a barrier. This limits access to LiDAR data and sensors, particularly for developing nations. Researchers are actively exploring alternative methods that can address this gap. Three main categories of remote sensing techniques are being investigated for AGB estimation:

1) Passive/Optical methods: These techniques rely on capturing reflected sunlight, offering a cost-effective approach but with limitations in penetrating dense canopies.
2) Active/SAR/Microwave methods: These methods utilize radar waves that can penetrate clouds and foliage, making them valuable in all weather conditions.
3) LiDAR: As mentioned previously, LiDAR provides highly detailed 3D data but can be cost-prohibitive.

(Fig. 4) visually compares the effectiveness of these three common methods in penetrating forest canopies [24, 25].

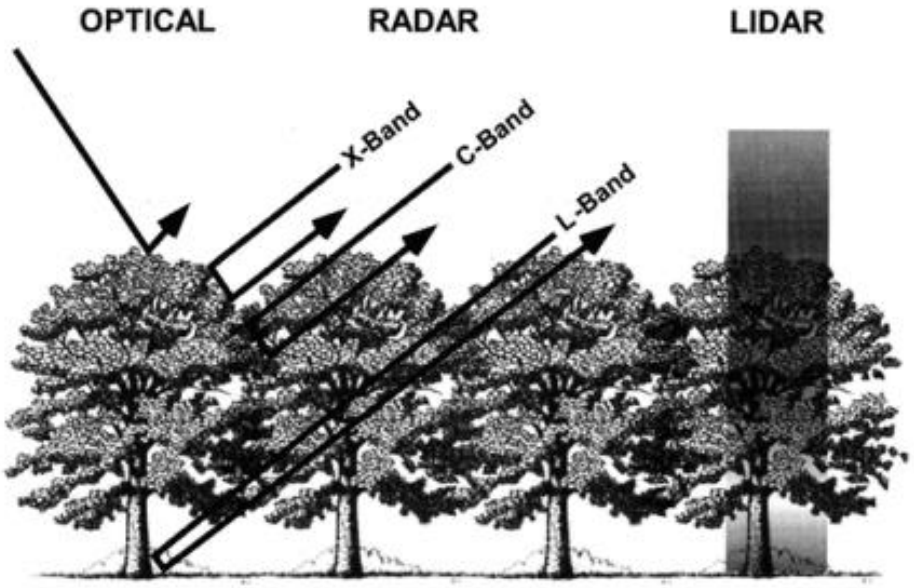

Figure 4 Forest canopy penetrating comparison of three remote sensing techniques

2.3 Importance of VI's in AGB mapping

Vegetation indices play a crucial role in enhancing the mapping of aboveground biomass (AGB) in diverse forest landscapes. Sensors with higher spatial and spectral resolution, such as Sentinel-2 benefit from the inclusion of indices like the red-edge band, improving their sensitivity to large AGB values and overall mapping accuracy compared to sensors like Landsat-8. Despite variations in sensor performance, the ability to predict lower AGB values remains consistent, particularly important for early forest successional stages [26]. Calibrating AGB models within specific AGB domains helps mitigate bias, enhancing the accuracy of AGB estimation. This improved mapping accuracy is vital for addressing sustainable development goals, aiding in natural resource management, and the sustainable use of biomass [27]. Vegetation indices are key in improving AGB mapping accuracy, especially in data-scarce regions, contributing to the understanding of carbon cycling and ecosystem functions.

2.4 AGB mapping using GEDI L4B

The GEDI L4B dataset plays a crucial role in advancing the estimation and monitoring of aboveground biomass (AGB) in forests worldwide [23]. Its three-dimensional measurements of forest canopy structure provide valuable insights into biomass distribution and density, enabling more accurate and comprehensive AGB assessments. This dataset is particularly valuable for areas where ground-based measurements are limited or unavailable, offering a cost-effective and efficient alternative for AGB estimation.

One of the key advantages of the GEDI L4B dataset is its ability to capture vertical forest structure, including canopy height and vertical distribution of biomass, which are critical factors in AGB estimation. By providing detailed information on forest structure, GEDI data enhances understanding of carbon dynamics in forests and helps identify areas with high carbon stocks and potential for carbon sequestration. This information is essential for developing effective forest management strategies and policies aimed at mitigating climate change.

Furthermore, the global coverage of the GEDI dataset is essential for assessing the effectiveness of conservation and restoration efforts and for guiding sustainable land use practices. Overall, the

GEDI L4B dataset represents a significant advancement in our ability to accurately assess and monitor AGB in forests, contributing to global efforts to combat climate change and preserve biodiversity.

2.5 Machine learning and statistical algorithms for AGB estimation

Algorithms like linear, quadratic, exponential, and logarithmic regression equations are the foundation for estimating Above-Ground Biomass (AGB) using field inventory characteristics. The foundation of these algorithms is the establishment of an empirical link between AGB and other remote-sensing metrics [28]. The direct application and ease of use of parametric techniques like multiple linear regression (MLR), non-linear regression, and linear regression, make them attractive for researchers aiming to quantify biomass using remotely sensed data [29]. Creating an empirical link between AGB and remote sensing parameters is the first step in these strategies, which is then applied to image pixels to estimate biomass [30].

Despite their widespread use, parametric models come with a significant caveat: they require the fulfillment of certain statistical assumptions. These consist of the data's normal distribution, the explanatory factor's lack of multicollinearity, and the presumption of a linear relationship between the variables and the results [24, 30]. Such assumptions can be limiting, as they do not always hold in complex environmental data, leading to potential errors and inaccuracies in biomass estimation. Moreover, parametric models often struggle to capture the intricate, non-linear relationships that exist between forest variables and remote sensing parameters, further complicating their use in diverse forest ecosystems [31].

In contrast to the limitations of parametric methods, non-parametric algorithms, particularly machine learning models, have emerged as powerful tools for AGB estimation. These models, which consist of a Support Vector Machine (SVM) and Random Forest (RF), do not rely on the traditional assumptions required by parametric approaches. Rather, they can manage intricate and non-linear correlations among variables, which makes them ideal for examining the complicated interplay between forest biomass and remote sensing information [32].

Because machine learning algorithms can handle big datasets with resilience and flexibility and can incorporate a variety of input factors without requiring data transformation or assumption

compliance, they are being used more and more. These models have been shown to outperform traditional parametric approaches in terms of accuracy and reliability, particularly when applied to heterogeneous forest landscapes [33]. The Random Forest algorithm, for example, has been widely praised for its efficiency in handling overfitting, its ability to manage missing values, and its provision of variable importance measures, which offer insights into the factors most strongly influencing biomass estimation [34].

2.6 Summary of this chapter

This chapter provides a detailed review of forest biomass importance and measurement methods. It highlights the forest's role in carbon sequestration and climate change mitigation, accounting for 80% of terrestrial carbon storage and annually sequestering 2-4 gigatons of carbon. Deforestation and degradation contribute significantly to greenhouse gas emissions, emphasizing the need for accurate biomass monitoring.

Remote sensing, combined with field data, offers a promising solution for efficient biomass estimation. Passive optical sensors, radar, and LiDAR systems are examined, each with unique advantages and challenges. Optical sensors are hindered by sunlight and cloud cover, while radar offers AGB retrieval potential through backscatter analysis, despite data saturation issues. LiDAR provides precise 3D forest structure data, enhancing AGB estimation accuracy. The significance of vegetation indices (VIs), particularly from high-resolution sensors like Sentinel-2, in AGB mapping improvement is discussed. The chapter also explores the GEDI L4B dataset, a global LiDAR dataset providing high-resolution canopy structure measurements, beneficial in areas with limited ground-based data. Additionally, the review discusses the shift from parametric to non-parametric machine learning algorithms, like Random Forest and Support Vector Machine, for AGB estimation. These models handle complex relationships between forest variables and remote sensing parameters, often outperforming traditional regression approaches in diverse forest landscapes.

CHAPTER 3 MATERIALS AND METHODS

3.1 Methods for field data collection

Forestry relies heavily on sampling techniques for several reasons. Compared to measuring every single tree, sampling offers: reduced cost, faster data collection, and improved accuracy. However, smaller sample sizes can result in fewer errors, including those made by field workers during measurements and data entry, inaccuracies related to the equipment used, and errors that occur during field inventory and data processing [35].

Subjective and probabilistic samplings are the two main types of sampling techniques. Personal experience and judgment are used in subjective sampling to choose sample units that are representative of the total population. The population that needs to be sampled may not be correctly characterized by the subjective sampling approach. In contrast, probability sampling relies on the theory of probability to choose sampling units that accurately represent the total population. For this research a specific type of sampling method was used to ensure all areas of the forest had a fair chance of being included in the data collection. This method, called stratified random sampling, involves dividing the forest into smaller sections (strata) based on relevant characteristics. Then, random samples are chosen from each section. This approach is preferred over simply selecting any tree (simple random sampling) because it guarantees better representation of the forest's diversity. Other common probability sampling techniques include systematic sampling and stratified random sampling, but stratified random sampling was most appropriate for this particular study collecting field inventory data [36].

3.1.1 Shape of sampling plot

There are a variety of shapes to choose from when setting up sample areas in a forest. These shapes can be squares, rectangles, circles, or even other geometric designs. Each shape has its own advantages and considerations. Square plots, for example, offer several advantages. They provide a uniform area for sampling and are relatively easy to lay out and measure [36]. Additionally, square plots can be more efficient in terms of space utilization compared to circular plots, as they can be arranged in a grid pattern with minimal gaps between plots.

In this field inventory collection, square sampling plots are used to sample tree ground data. These plots size was 10 meters square, which is equivalent to 0.1 hectares each. The choice of square plots was influenced by their ease of setup and maintenance, making them practical for the terrain and conditions of the study area. This approach aligns with common practices in forestry sampling, where the goal is to simplify the sampling process while ensuring representative and reliable data collection [31].

3.1.2 Forest feature assessment

To estimate the size and amount of trees in the sample plots, the diameter of all trees wider than 5 centimeters at a specific height was measured. This height, called diameter at breast height (DBH), is a standard measure in forestry and is typically around 1.3 meters above the ground [37]. The measuring tape was used to measure the diameter of each tree, recording the value in centimeters on a data collection form. DBH is commonly measured in forest inventories due to its strong correlation with other tree variables [39, 40]. After excluding error measurements, a total of 435 trees were included from the 46 sampling plots in the Maanshan Forest Park (MFP). The sampling was conducted in October 2022 and the GIS coordinates were recorded at the four corners of each plot. Careful attention was paid to the precise location where the DBH measurements were taken. (Fig. 5) illustrates the method used for measuring tree DBHs.

Figure 5 DBH measurement location for trees

In each sampling plot, the height of every tree was meticulously measured and recorded in the data collection form. Tree height is a crucial characteristic that can be directly measured in forestry projects. It is commonly measured to estimate the potential timber yield of a forest stand and assess the success of silvicultural treatments. Additionally, tree height plays a significant role in ecosystem studies, as it influences light interception, carbon sequestration, and habitat structure for various wildlife species. Therefore, accurate measurement and recording of tree height are essential for obtaining reliable data for forest management and conservation purposes. It is essential to carefully choose the appropriate terminology, particularly in the context of forest inventory tasks, as several types of heights can be measured, including tree height, bole height, and merchantable height. For this study, tree height, which refers to the vertical distance from a tree's base to its tip, was the chosen metric [40] (Fig. 6). In general, measuring the height of trees in forests is a challenging, expensive, and time-consuming task [41].

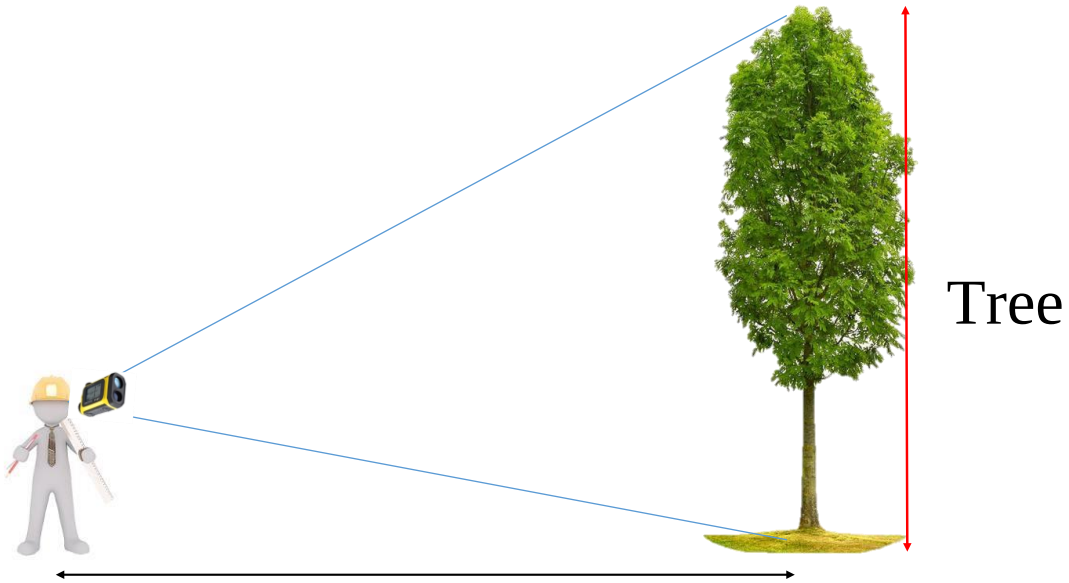

Figure 6 Height of a tree

A handheld laser rangefinder was used to measure the height of the trees. It is simple to use and accurate, estimating height using ultrasonic technology [42]. (Fig. 7) shows the locations of the sampling plots that were set up in the MFP for forest inventorying.

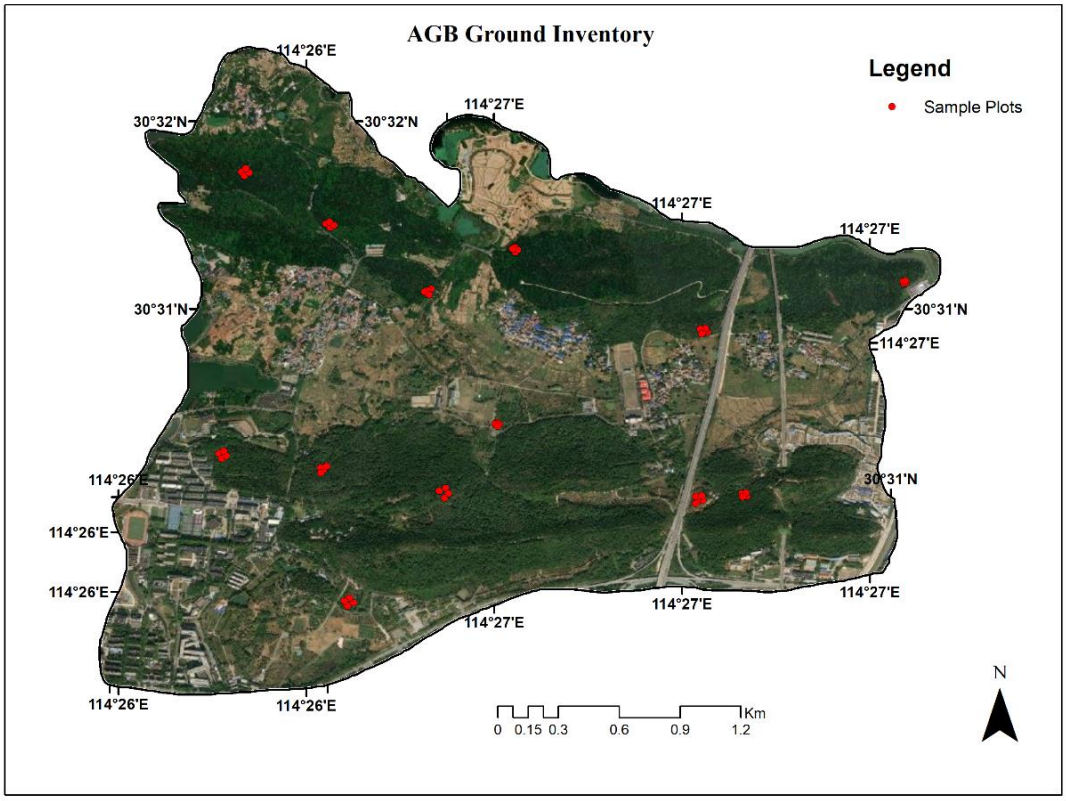

Figure 7 Field samples in MFP

3.2 Satellite remote sensing

3.2.1 Global ecosystem dynamics investigation L4B

GEDI, pioneering the first orbiting full-waveform LiDAR on the ISS, offers unparalleled global insights into the vertical profiles of forests since its 2018 launch. GEDI's LiDAR system meticulously captures the 3D forest structure with 25-meter footprints, staggered by 60 meters along its path and 600 meters across, utilizing trio lasers for eight observation tracks [23]. The GEDI L4B data product extrapolates 1 km² AGBD averages from extensive mission observations, translating waveforms into biomass predictions.

In this research, GEDI's Level 4B mean aboveground biomass density (AGBD) for AGB validation. Although 10 data files are included in the cloud-optimized GeoTIFF dataset of GEDI L 4 B. All of the ten data files provide mean aboveground biomass density (AGBD) estimates for

the period 2019-04-18 to 2021-08-04. The data file selected for this research is shown in (Table 1):

Table 1 GEDI L4B data file

File Name	Description	Unit	Value	Type
Dataset				
GEDI04_B MU.tif	Mean aboveground biomass density	Mg ha^{-1}	-9999	Float32

To generate the validation dataset, the samples from MU data file were collected between April and August of 2021 for MFP study area. (Fig. 8) shows GEDI mean aboveground biomass density (AGBD) samples of MFP.

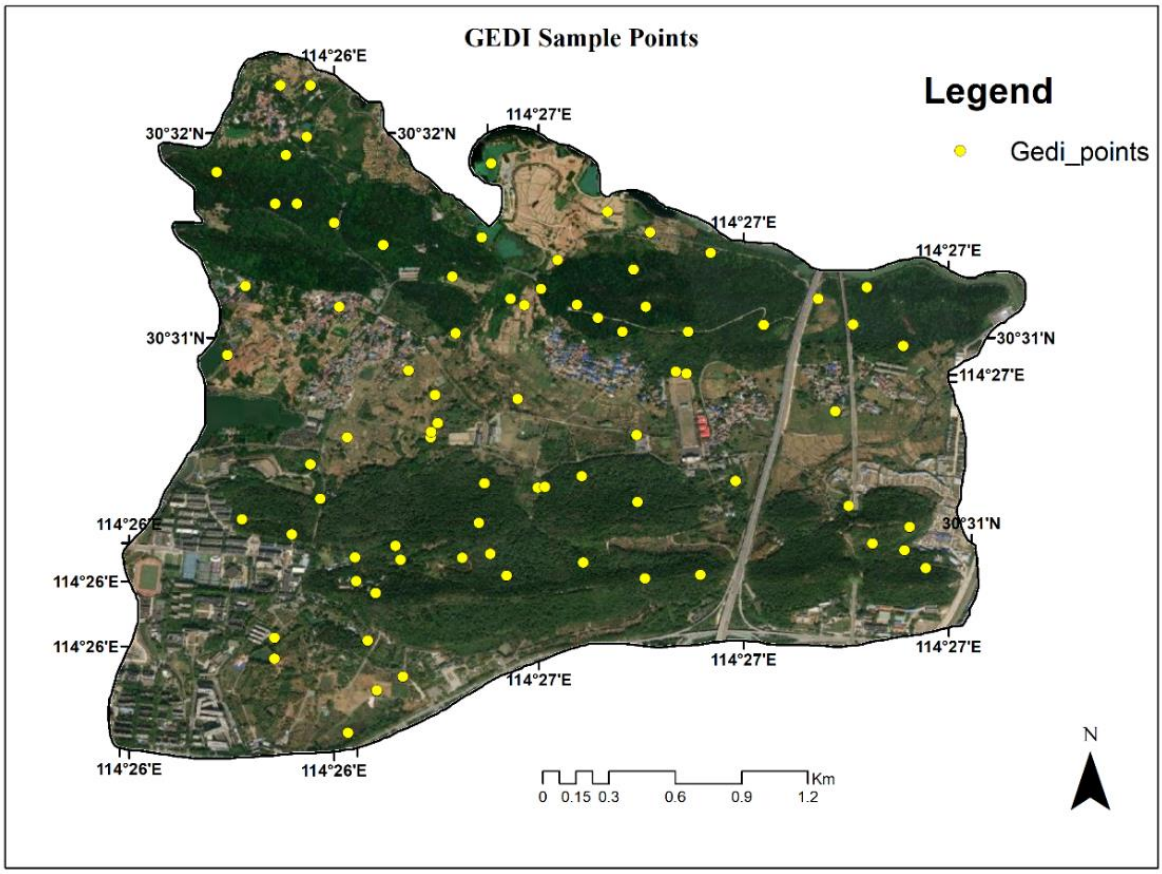

Figure 8 GEDI samples points

3.2.2 Sentinel-2 MSI optical data acquisition and processing

The Global Monitoring for Environment and Security (GMES), launched in 1998, is the European Commission's (EC) Copernicus program. The European Space Agency (ESA) oversees the program's space component, including the Copernicus Sentinel satellite constellations [43]. As of October 2017, the Copernicus program's S2 mission is operational, and satellite data is accessible for free [44].

The two satellites of the Sentinel-2 constellation mission are Sentinel-2A (launched on June 23, 2015) and Sentinel-2B (launched on March 7, 2017) [45]. The high revisit time of the S2 constellation is five days. Their Multispectral Imager (MSI) uses a 290 km field of view and gathers data in 13 spectral bands utilizing pushbroom technology. There are three spatial resolutions available for the data generated by the S2 constellation: 10 m, 20 m, and 60 m. At a height of 786 kilometers, the mission runs in a sun-synchronous orbit. S2 data have been used in a number of research pertaining to vegetation [46]. (Table 2) contains information about each of S2's spectral bands [47].

Table 2 Sentinel-2 spectral bands

Bands	Description	Central Wavelength(nm)	Resolution(m)
1	Coastal aerosol	443	60
2	Blue	490	10
3	Green	560	10
4	Red	665	10
5	Vegetation Red Edge 1	705	20
6	Vegetation Red Edge 2	740	20
7	Vegetation Red Edge 3	783	20
8	NIR	842	10
8A	Narrow NIR	865	20
9	Water Vapor	945	60
10	Cirrus	1375	60
11	SWIR 1	1610	20
12	SWIR 2	2190	20

For this research I chose a specific satellite image from Sentinel-2 to analyze the forest. This image was acquired on August 29, 2022, and importantly, had very minimal cloud cover, only 0.4

percent, from which my part of study area was having zero clouds. The selected date is crucial because it is very close to the date when the forest inventory data was collected on the ground.

To ensure consistency and leverage the higher resolution of 10 meters, bands 2, 3, 4, and 8 are utilize [49, 50]. The image used is a Level-1C product, which already includes geometric and radiometric corrections. The area of interest within the Sentinel-2 image is relatively small. To save time processing the entire image, only the relevant portion is achieved by using a shapefile of the study area to "clip" out the specific region from the larger image. This clipping is done using QGIS software, a program commonly used for geographic data analysis. Once clipped, the resulting image containing only the study area, saved and loaded back into QGIS for further analysis, such as calculating vegetation indices.

3.2.3 Acquisition and processing of sentinel-1 data

The Sentinel-1 mission is a joint project between the European Union and the European Space Agency. As mentioned earlier, Copernicus was formerly referred to as GMES. This is an attempt by Europe to offer security and environmental information services [43].

The European Space Agency's Sentinel-1 mission plays a crucial role in Earth observation, utilizing two satellites (S1A and S1B) to capture radar data. These satellites carry a powerful C-Band Synthetic Aperture Radar (SAR) instrument, capable of generating images with exceptional detail. Resolutions can reach as fine as 5 meters, and the swath width, the area covered in a single image, can span up to 400 kilometers. The mission offers various imaging modes, each suited for specific applications.

In this study, the Interferometric Wide (IW) mode is used due to its effectiveness in land surface research [50]. The IW mode offers several advantages for this purpose. It captures data with a spatial resolution of 5 by 20 meters, allowing to zoom in on specific features. Another key benefit is the dual polarization capability, meaning the data can be analyzed from different perspectives (VV and VH). This provides a more comprehensive understanding of the land surface characteristics [57, 58].

The Sentinel-1 image of 20 September 2022 for the study area is downloaded using the Google Earth Engine platform, which has already undergone pre-processing steps that are essential for

preparing the Sentinel-1 data for subsequent analysis, to derive meaningful insights into above ground biomass [53]. Following are the preprocessing steps that are performed by Sentinel-1 Toolbox before uploading to GEE:

1) Thermal noise removal: This reduces unwanted noise in the data.
2) Radiometric calibration: This converts the raw data into units of backscatter coefficient (sigma- naught, σ°) in decibels (dB) using the following equation (developed by ESA, 2021) [55];

$$\sigma^0(dB) = 10 \log \sigma^0 \quad (1)$$

where, 'σ^0' is the backscatter for a certain polarization, and '$\sigma^0(dB)$' is the normalized radar cross-section.

This allows for comparison between images and easier interpretation of radar backscatter strength [61].

3) Terrain correction (orthorectification): This accounts for distortions caused by the terrain's topography, projecting the image onto a flat reference plane.
4) Orbit file application: This ensures the image is accurately positioned based on the satellite's orbit information.

In last the image is scaled to 10- meter resolution to match the resolution of Sentinel-2 data and exported from Google Earth Engine, and imported to SNAP software to improve the image visual appeal, the Non-linear Median Filter (3 x 3) is used [56]. Subsequently, the image is clipped by the study area shape file to match the dimensions of the study area. The resulting image, containing VH and VV polarizations, is then saved as GeoTIFF raster file. This step ensured that the data is properly aligned with the study area, setting the stage for further analysis and interpretation of the Sentinel-1 imagery for AGB estimation.

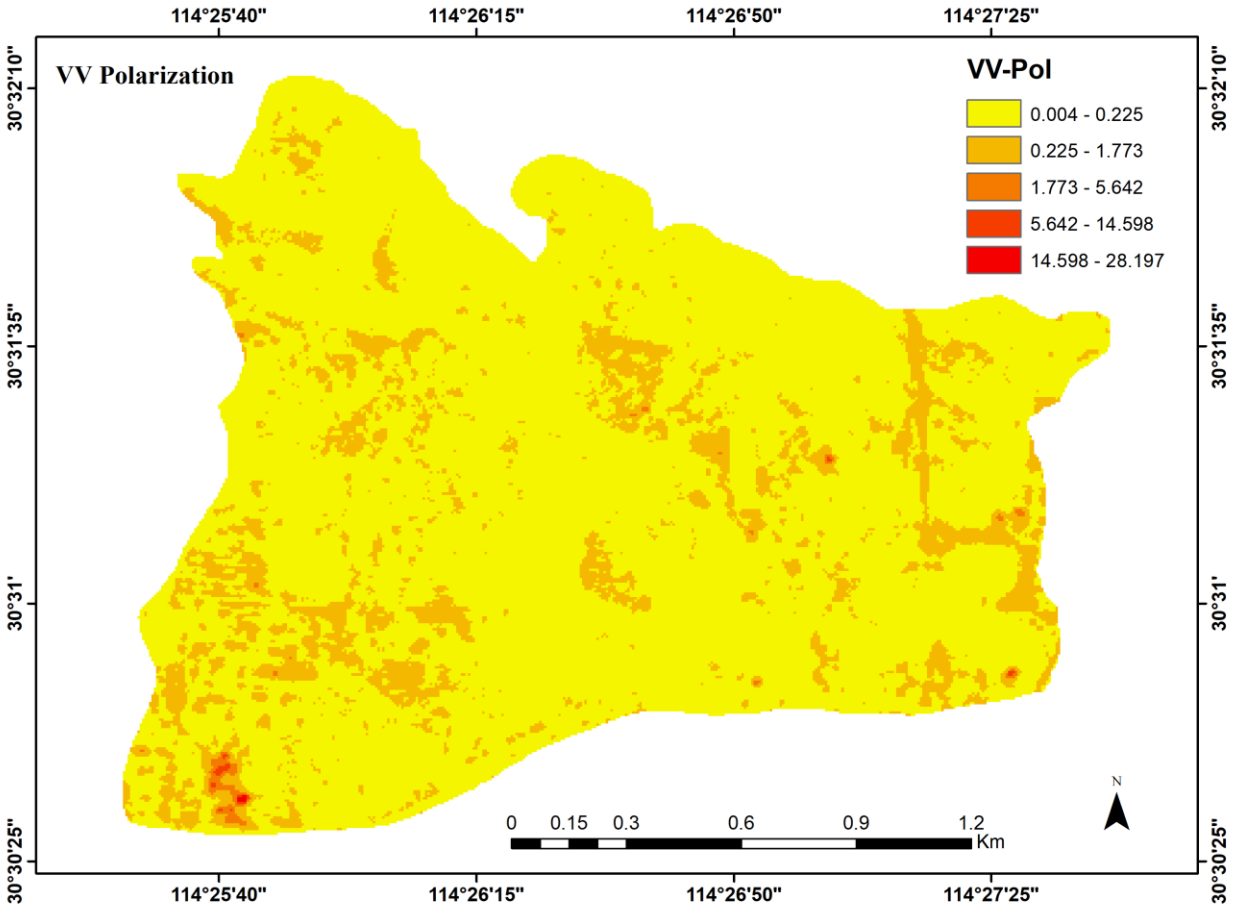

Figure 9 VV polarized data from the study area

The VV and VH polarization maps of the study area provide valuable insights into the vegetation characteristics essential for AGB estimation. The VV map (Fig. 9), with a range from 0.004 to 28.197, indicates the backscatter intensity sensitive to vertical vegetation structures. Areas with higher VV values suggest denser vegetation or rougher surfaces, aiding in identifying areas with potentially higher biomass. In the VV polarization, dense vegetation areas typically appear bright due to the strong backscatter from the vegetation canopy. This brightness indicates high radar reflectivity, which is associated with dense vegetation cover and potentially higher AGB. In contrast, non-vegetated or less vegetated areas reflect lower radar reflectivity. By visually inspecting the VV and VH visualizations of the study area, patterns and variations in radar backscatter are identified that are indicative of different vegetation types, densities, and biomass levels.

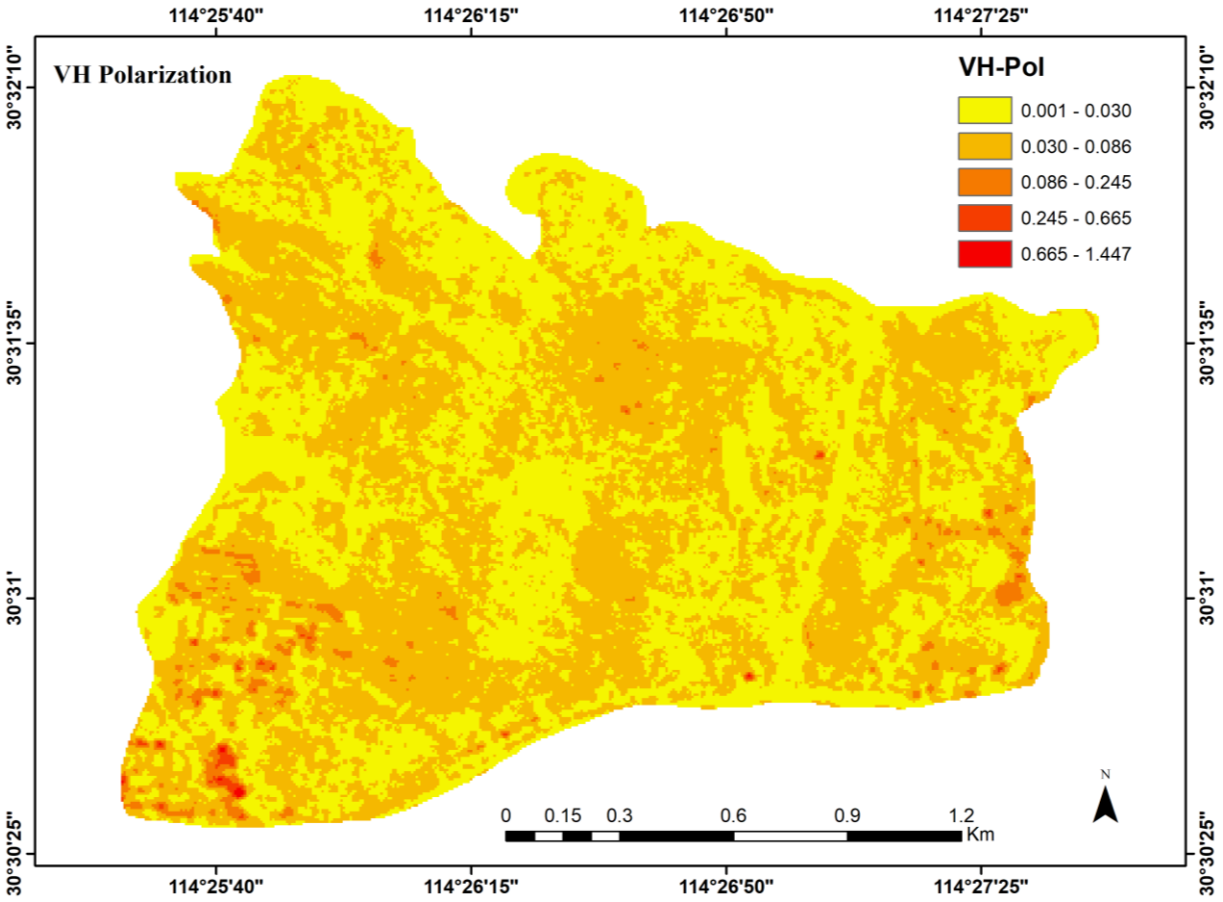

Figure 10 VH polarized data from the study area

On the other hand, the VH map (Fig. 10), with values ranging from 0.001 to 1.447, provides information on both vertical and horizontal vegetation structures. Higher VH values may indicate more scattering from volume structures like buildings. Comparing the spatial patterns and ratios of VH to VV backscatter can further elucidate the vegetation structure and density variability within the area. VH polarization can provide complementary information to VV. In the VH polarization, areas with different surface properties exhibit varying backscatter characteristics. For example, water bodies and smooth surfaces may appear bright due to specular reflection, while rough surfaces such as bare soil or urban areas may appear darker due to increased scattering. Vegetated areas may also exhibit moderate to high backscatter depending on vegetation structure and density.

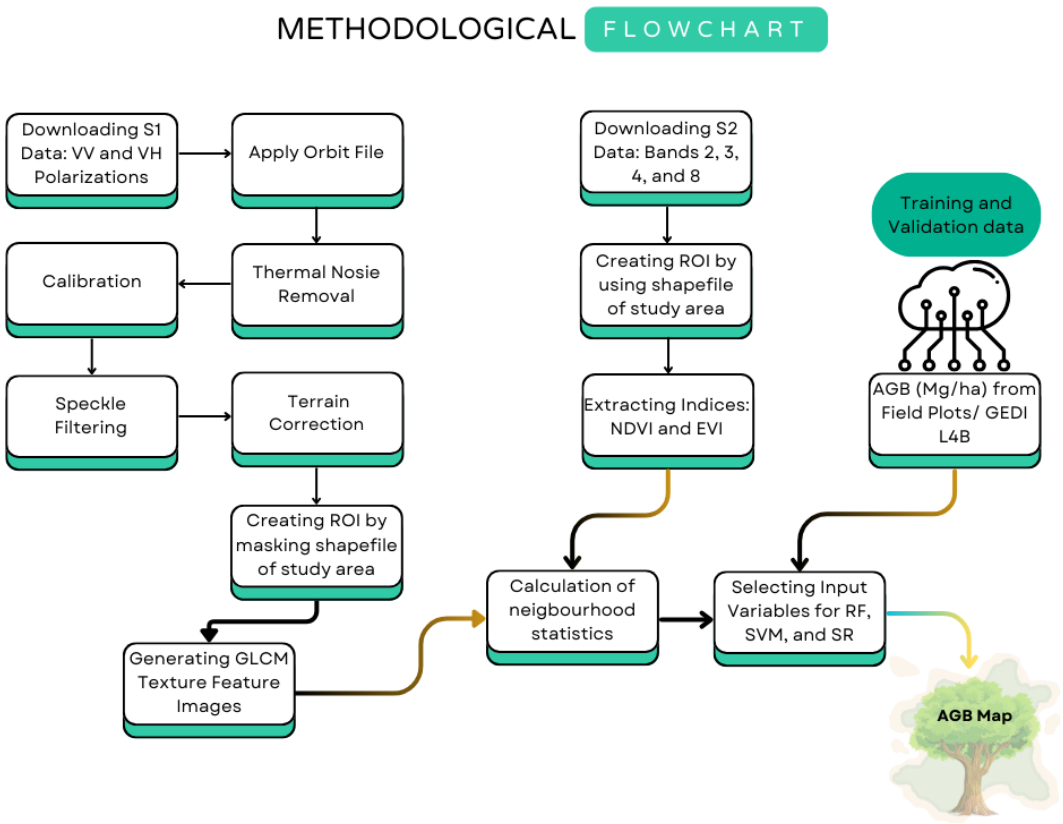

Figure 11 Workflow methodology for estimating and mapping AGB

3.3 Variables extraction from satellite data

3.3.1 VIs

Vegetation Indices (VIs) are combinations of spectral bands in the visible and near-infrared regions of the electromagnetic spectrum [57]. These VIs offer a straightforward method for gathering details about vegetation properties from a big collection of remote sensing data. They are an effective way to gauge the productivity and vigor of the plants [58]. VIs have also shown statistically significant associations and have been employed for AGB calculation [59]. Chlorophyll, water content, leaf angle, density, canopy structure, and other factors are some of the causes of this association [60]. Various kinds of VIs fall into the categories of broadband and narrowband, for example, reviewed about 35 VIs for their research. In a similar vein, [61] reported distinct VIs in their research.

3.3.2 EVI

The Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) are employed in this study due to their widespread use and significant correlations with Aboveground Biomass (AGB) [62]. Due to its availability, ease of use, and suitability for a variety of ecosystem types, EVI is also one of the most used indices. The Enhanced Vegetation Index (EVI) ratio includes the red and near-infrared (NIR) bands of the electromagnetic spectrum, which are related to the "greenness" of the vegetation canopy [63]. The EVI is more sensitive to canopy structure among the two vegetation indices, but the NDVI is more sensitive to chlorophyll [64]. In investigations about vegetation, these two indicators work well together. Moreover, the EVI's value spans from -1 to 1. The formula used to calculate EVI is as follows [65];

$$EVI = 2.5 \times \frac{(\mathrm{NIR}-\mathrm{Red})}{(\mathrm{NIR}+6\times\mathrm{Red}-7.5\times\mathrm{Blue}+1)} \quad (2)$$

EVI for the research area is determined using QGIS software. The above formula for the EVI is written using a raster calculator after the S2 image for the research region was input. The raster is stored for later use, and an EVI map is created using the QGIS software's layout manager (Fig. 19).

3.3.3 NDVI

The Normalized Difference Vegetation Index (NDVI) is a widely used tool to identify areas with vegetation [66]. It works because healthy, green plants reflect more near-infrared light (NIR) while absorbing the most visible light, compared to unhealthy or sparse vegetation. This difference is captured by the NDVI, which uses specific wavelengths within the electromagnetic spectrum: red light (0.62-0.75 micrometers) and near-infrared (0.75-1.3 micrometers) [67]. Leaves, rich in chlorophyll, act like tiny solar panels. They absorb the red light for energy and reflect the non-visible near-infrared light. This reflection is also influenced by the structure of the leaves and how they are arranged on the plant [68]. NDVI values range from -1 to +1. Negative values typically indicate water bodies, while values close to zero suggest bare ground. Values between 0.2 and 0.8 represent areas with some vegetation, but it might be sparse or unhealthy. As the NDVI value increases above 0.8, the vegetation becomes denser and greener. Values near 1 indicate areas with the most lush and healthy vegetation, like rainforests. The NDVI formula is as follows [69];

$$NDVI = \frac{(\text{NIR}-\text{Red})}{(\text{NIR}+\text{Red})} \quad (3)$$

In this study, QGIS is used to calculate the Normalized Difference Vegetation Index (NDVI) for the research area. The NDVI equation is applied using a raster calculator after importing the Sentinel-2 image into the QGIS software. The resulting raster is saved for further analysis, and an NDVI map is generated using QGIS's layout manager (Fig. 17).

3.3.4 Grey level co-occurrence matrix (GLCM)

The spatial link between neighboring pixels is taken into consideration by the two-dimensional GLCM (Gray-Level Co-occurrence Matrix) dependence matrix [70]. For this investigation, three GLCM metrics were computed for the S1 data. They were Contrast, Entropy, and Angular Second Moment (ASM). The GLCM metrics were computed using the 'r.texture' geo-algorithm from the Geographic Resources Analysis Support System (GRASS) plugin in the QGIS software, utilizing a 3 x 3 pixel moving window size [71]. The degree of disorder in the image is measured by entropy. ASM, sometimes known as "Energy" gauges an image's textural homogeneity [72]. The quantity of local differences in the image is measured by contrast [73]. Strong connections have also been found between the AGB and the Entropy, ASM, and Contrast [81, 82]. The three metrices formulas for entropy, ASM, and contrast are mentioned in (Table 3).

Table 3 GLCM texture for sentinel 1

GLCM Texture	Equations	Description
Contrast	$\sum_{i,j=0}^{N-1} p_{i,j} \times (i-j)^2$	Contrast is a measure of how much the brightness of pixels in an image differ from their neighbors, High contrast in an image indicates sharp edges or surface irregularities.
ASM	$\sum_{i,j=0}^{N-1} \left(p_{i,j}\right)^2$	ASM is a statistical measure that analyzes the spatial distribution of pixel values in an image, revealing the degree of order or randomness.
Entropy	$\sum_{i,j=0}^{N-1} p_{i,j} \times (-ln^{p_{i,j}})$	Entropy reflects the unpredictability of pixel values. Higher entropy indicates a more complex and detailed image, with less uniformity in pixel intensity.

Where, '$P_{i,j}$' is calculated as $P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{n-1} V_{i,j}}$, and it is the probability of finding the GLCM relationship at cell (i,j); is such that $\Sigma_{i,j=0}^{N-1}(P_{i,j}) = 1$, 'N' is the number of grey levels in the image as specified by the number of levels in the quantization, and '$V_{i,j}$' is the grey level value in a cell (i,j) of the image window.

3.4 Data analysis

3.4.1 Statistical evaluation of field measurements

The statistical evaluation of field measurements included calculating mean diameter, tree volume, density, and aboveground biomass (AGB) using established allometric equation. The data is further analyzed using regression analysis to understand the relationships between variables and to assess the accuracy of the measurements.

3.4.2 Mean diameter

Using following equation [76], the quadratic mean diameter for every plot is calculated;

$$Mean\ DBH = \sqrt{\frac{d_i^2}{n}} \qquad (4)$$

Because it has a closer relationship to volume than the average of DBHs, the quadratic mean diameter is thought to be more beneficial [77]. The mean DBH for this study is defined as the quadratic mean diameter.

3.4.3 Tree volume

Tree volume was determined using the formula below [78];

$$Volume\ of\ tree(m^3) = \frac{\pi}{4} \times DBH^2 \times H \times FF \qquad (5)$$

Where 'DBH' is the diameter at breast height, 'H' is the height of a tree, and 'FF' stands for Form Factor.

For commercial reasons, the volume of a tree is one of the most important variables in forestry, and it is also easily related to other variables in forestry [79].

3.4.4 Density

The number of trees growing in a specific area is called tree density [80]. This measurement is important because it helps us understand the different stages of growth within a forest [81].

$$\text{Tree Density} = (\text{Number of Trees}) / (\text{Area}) \quad (6)$$

3.4.5 Field AGB

Initially, the field inventory data collected is used to calculate AGB using following allometric [82];

$$AGB = 0.0673 \times (\rho D^2 H) \times 0.976 \quad (7)$$

Where the symbol 'ρ' is assigned a value of 0.375, 'D' represents the Diameter at Breast Height, and 'H' represents the tree height.

3.4.6 Converting to hectare units

To analyze the data at a landscape level and facilitate comparisons across larger areas, the forest variables volume (in cubic meters m^3), density, and aboveground biomass (in megagrams, Mg) were scaled up to hectares (ha). This upscaling process involved dividing each variable by 0.1 [83], effectively converting the measurements from a smaller unit area (likely square meters) to a larger unit area (hectares). This allows for a more comprehensive understanding of forest resources and ecosystem services at the hectare level.

3.4.7 Statistical analysis

Regression analysis is used to understand how the variables are related. This technique helps identify linear relationships (straight line) between variables. In cases where the relationship isn't linear, a different method called power regression is used [84]. This is common in studies of Aboveground Biomass (AGB) because the link between variables isn't always perfectly straight [85]. A key metric used to assess the strength of this relationship is the Coefficient of Determination (R-squared). Higher R-squared values indicate a stronger connection between the

variables [86]. In AGB research, R-squared is commonly used to gauge how well the variables can be used to predict biomass. Here's a general guide to interpreting R-squared values [87]:

1) Below 0.2: Weak relationship
2) Between 0.2 and 0.5: Moderate relationship
3) Above 0.5: Strong relationship

The presence of a negative sign simply indicates that the variables change in opposite directions. For example, as one variable increases, the other decreases. A significance level of 0.05 is used for all statistical tests conducted in RStudio. This software is also employed to generate the relevant data visualizations.

3.5 Models for analyzing remotely sensed parameters from satellites

3.5.1 Random forest

Regression and classification are two applications for the non-parametric machine learning technique known as RF [88]. It has been effectively applied in research combining remote sensing methods with AGB estimates. It was created by [89] and employs regression using the ensemble learning approach [90]. When compared to the outcomes of the individual models, the ensemble model results will typically be superior.

Among RF's benefits are: 1) increased precision 2) Resistance to noise and outliers 3) Processing speed 4) the relevance of predictor factors is estimated [91]. In the case of RF, numerous predictor factors can also be employed [89]. When employing the RF approach, decision trees are built as much as possible without pruning. In this instance, bootstrapping is used to randomly choose two-thirds of the training data through repeated resampling of the data with replacement.

By combining the predictions made by several individual trees, this procedure increases the final prediction's flexibility. The model's errors are estimated using validation samples drawn from the remaining one-third of the data, which is referred to as out-of-bag (OOB) data. This term is used because the model is not trained on this portion of the data and therefore does not "see" it during the training process [92]. The basic equation of RF model is:

$$\hat{Y} = \frac{1}{B}\sum_{b=1}^{B} f_b(X) \quad (8)$$

Where, '$\hat{Y}$' is the predicted output (e.g., AGB), 'B' is the number of trees in the forest, and 'f_b' is the prediction of the b[th] tree for input X.

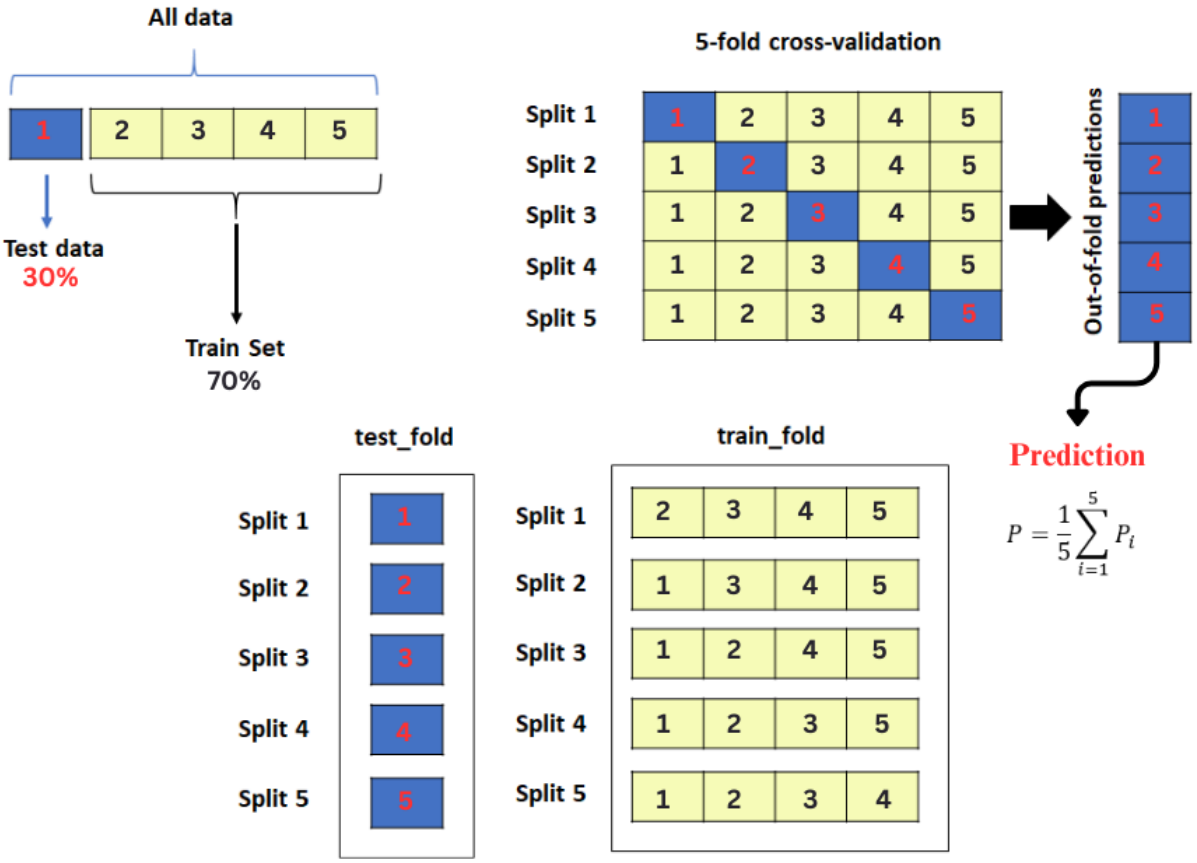

Figure 12 Random forest 5-fold cross-validation

The RF algorithm requires two inputs. First is "mtry," which denotes the number of predictor variables considered at each decision tree node to partition the data. The second is "ntree," which represents the total number of trees the model utilizes. The statistical program RStudio made use of the "raster," "caret," and "randomForest" packages. In the RF models, the default values for ntree and mtry were 500 and 1/3, respectively, for the regression [92]. The R 'caret' package was utilized to determine which variables in a model were most crucial [93]. The ideal amount of variables is crucial for the model's increased accuracy [85]. Additionally, it makes interpretations simpler and boosts the variable's predictive potential. To assess the model's predictive ability, a

validation dataset is typically separated apart from the main dataset. However, I employed a 5-fold cross-validation strategy in this study due to the short dataset [94], (Fig. 12).

3.5.2 Support vector machine

Support Vector Machines (SVMs) are a powerful classification tool widely used for tasks like Aboveground Biomass (AGB) estimation. This technique relies on statistical learning theory to find the best separation line (or boundary) between different categories in data, even in complex situations [106]. SVM focuses on training data points closest to this boundary, called support vectors. By maximizing the distance between these support vectors, SVM aims to create the most accurate separation between classes. Originally designed for simple, straight-line separations, SVMs can handle more complex data by using a technique called the kernel trick [97]. This trick essentially projects the data into a higher-dimensional space, allowing for more intricate boundaries. There are different options (kernels) for this projection, with the radial basis function (RBF) being a popular choice due to its good accuracy and stability in AGB estimation research [95, 110]. See also (Fig. 13).

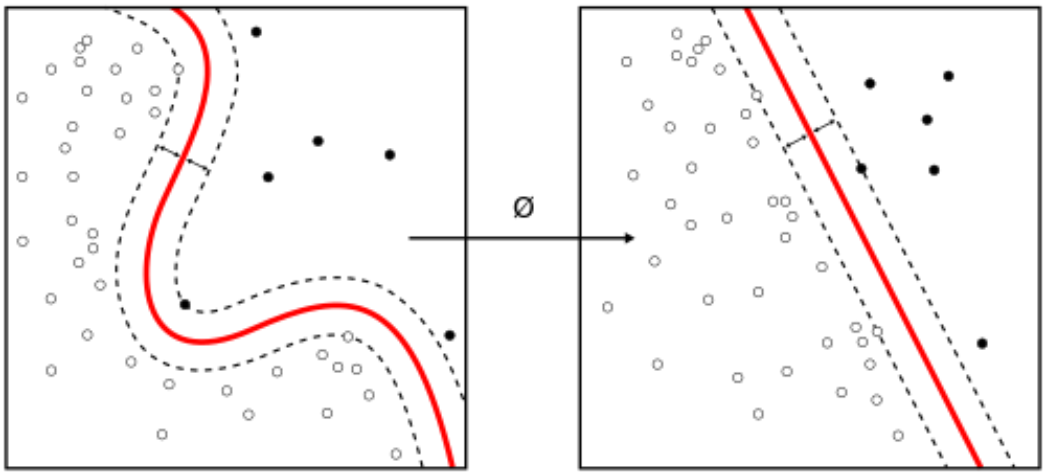

Figure 13 Illustration of the kernel function in SVM

Support Vector Machines (SVMs) rely on two main settings: kernel width (γ) and regularization parameter (C). Kernel width (γ) controls how much influence individual data points have on the classification boundary. Regularization parameter (C) balances the trade-off between fitting the training data perfectly and avoiding overfitting. A higher C prioritizes perfectly fitting the training

data, while a lower C allows for more flexibility to avoid overfitting. Utilizing the 'e1071' package, the SVM is implemented in RStudio. Due to the small dataset size, a 5-fold cross-validation method, similar to that described for the RF model, was used. Additionally, the 'caret' package in RStudio was utilized to select variables in a manner similar to the RF model. The basic equation of the SVM model is:

$$f(X) = sign(w \cdot X + b) \quad (9)$$

Where, '$f(X)$' is the predicted AGB value, 'w' represents the importance of each feature, is the input vector, 'X' containing the values of the features, and 'b' is the bias term, which allows the SVM to shift the decision boundary to better fit the data.

3.5.3 Stepwise regression

The stepwise regression technique acts like a detective, it selects the most influential variables, one by one, to create a simpler and more effective model. There are three main approaches to stepwise regression: 1) Forward Selection, 2) Backward Elimination, and 3) Mixed Selection [45]. Forward selection starts with an empty model and then carefully adds the most important variables (building blocks) one at a time. This continues until all the truly influential variables are included. In contrast, backward elimination starts with all the variables in the model (all the furniture) and then removes the least important ones (the unnecessary items) until only the most relevant variables (essential furniture) remain. Variables are then removed one by one, with each removal being guided by the same criterion. Subsequently, the variables that hold no importance for the model are eliminated individually. This method similarly relies on a predetermined standard. In contrast, "Mixed Selection or Bidirectional Elimination" combines the two previously described methods of "Forward Selection" and "Backward Elimination."

In this study, I applied SR's "Backward Elimination" method. The method has been frequently used for estimating AGB utilizing parameters from remote sensing. The data's normality distribution was examined through the application of the Shapiro-Wilk Test. Using the RStudio program, the variable's Spearman correlation coefficient is determined [99]. Where appropriate, the data is additionally logarithmically transformed. The variance inflation factor is used to evaluate the collinearity (VIF). Variables with a VIF greater than 10 were eliminated, additionally,

it confirms that the models satisfied the fundamentals of linear regression. Once more, the 5-cross validation approach is implemented using the "caret" package. Following I have shown the backend process of stepwise regression:

$$Y_i = \beta_0 + \beta_i X_i + \varepsilon_i \quad (10)$$

Where, 'Y_i' is ground reference/GEDI L4B AGB (Mg ha^{-1}), 'β_0' is intercept, can be calculated as: $\beta_0 = \bar{y} - \beta_i \bar{x}_i$, '$\beta_i$' is the coefficient of the explanatory variable: $\beta_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, '$X_i$' is explanatory variable from sentinel 1 and 2, and 'ε_i' is the error terms of the model: $\varepsilon_i = y_i - \hat{y}_i$.

3.5.4 AGB mapping and model validation

To evaluate how well the model performs, two metrics were used: R-squared (R^2) and root mean square error (RMSE). R-squared indicates how well the model explains the data's variation. A higher R-squared value suggests a better fit. RMSE measures the average difference between the predicted values and the actual values. A lower RMSE indicates higher model accuracy [31]. The RMSE and R^2 are given as;

$$R^2 = 1 - \frac{\Sigma_i (y_i - \hat{y}_i)^2}{\Sigma_i (y_i - \bar{y})^2} \quad (11)$$

Where 'R^2' is the coefficient of determination, $\Sigma_i (y_i - \hat{y}_i)^2$ is the sum of squares of residuals (RSS), and $\Sigma_i (y_i - \bar{y}_i)^2$ is the total sum of squares (TSS).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (12)$$

Where measured AGB in the field is represented by 'y_i', and '$\hat{y}_i$' is the predicted AGB and n is the total number of the plots.

While both R-squared (R^2) and root mean square error (RMSE) were used to evaluate the models, RMSE was given more weight when choosing the final model [31]. This means that a lower RMSE, indicating a smaller average difference between predicted and actual values, was considered more important than a higher R^2. The final models, chosen based on their RMSE

performance, were then used to create a spatial map of Aboveground Biomass (AGB) across the study area.

The R "raster" package is used to perform spatial prediction of the AGB. This is accomplished by using the 'predict()' method, which takes as inputs the finished models and the raster object. For additional processing, the AGB raster for the MFP is exported.

3.6 Summary of this chapter

This chapter sums up the materials and methods used in the study, focusing on field sampling techniques, satellite remote sensing data acquisition and processing, and the various modeling approaches employed for AGB estimation. The field sampling strategy involved a stratified random sampling approach, where the study area (Maanshan Forest Park) was divided into smaller strata, and samples were randomly selected from each stratum. Square sampling plots of 10 meters by 10 meters were used, a common practice in forestry due to their ease of setup and maintenance.

The chapter then interduce the acquisition and processing of satellite remote sensing data, including Sentinel-2 optical imagery and Sentinel-1 radar data, and vegetation indices (NDVI and EVI) are calculated. In addition to the satellite data, the chapter also introduces the GEDI L4B dataset, a global LiDAR product that provides three-dimensional measurements of forest canopy structure. This dataset was used for AGB validation purposes.

The core of the chapter focuses on the data analysis and modeling approaches employed in the study. These include statistical evaluations of the field measurements, such as mean diameter, tree volume, density, and AGB calculations. Additionally, the chapter explains the use of three modeling techniques for AGB estimation: Random Forest (RF), Support Vector Machine (SVM), and Stepwise Regression (SR). The chapter outlines the specific steps involved in each modeling approach, including the use of 5-fold cross-validation, variable importance assessments, and model performance evaluations using metrics like coefficient of determination (R^2) and root mean square error (RMSE). The final models are then used to generate AGB maps for the study area.

CHAPTER 4 RESULTS AND VALIDATION

4.1 Biomass assessment of maanshan forest park

4.1.1 Stand structure analysis of the MFP

(Table 4) summarizes the key structural characteristics and aboveground biomass (AGB) stock of MFP. The mean AGB is observed as 29.67 Mg ha^{-1}, with a standard deviation of 41.36 Mg ha^{-1}. This value indicates a moderate level of biomass accumulation within MFP, though the high standard deviation suggests some variability across the sampled area.

The mean tree diameter at breast height (DBH) is 59.64 cm, with a standard deviation of 6.03 cm. This suggests a mature forest stand, with trees having a substantial diameter and likely contributing significantly to the overall AGB.

The mean density of the MFP stand is 142.90 trees ha^{-1}, with a standard deviation of 32.37 trees ha^{-1}. This indicates a moderately dense forest, with a potential range from open canopy conditions to more crowded stands. Investigating the relationship between tree density and AGB could be insightful for understanding forest management practices and their impact on biomass accumulation. Mean tree volume is recorded as 37.82 m^3 ha^{-1} (cubic meters per hectare) with a standard deviation of 50.38 m^3 ha^{-1}. This aligns with the observed AGB stock, suggesting that most of the biomass is stored within the trees themselves.

The mean tree height is 10.62 meters, with a standard deviation of 1.99 meters. This implies a moderately tall canopy, although there may be some variation in tree heights within the stand. Understanding the distribution of tree heights is important for characterizing forest structure and light availability within the MFP system.

Table 4 Forest structural characteristics

Forest	Mean AGB (Mg ha^{-1})	Mean DBH (cm)	Mean Density/Ha	Mean Tree Volume (m^3ha^{-1})	Mean Tree Height (m)
MFP	29.67 ± 41.36	59.64 ± 6.03	142.90 ± 32.37	37.82 ± 50.38	10.62 ± 1.99

Analysis (Table 5) using Spearman's rank correlation revealed strong, positive associations between AGB and structural features. Stands with higher AGB tend to have larger trees (DBH & height) and greater total volume.

Table 5 Spearman correlation coefficients

Variables	Volume (m^3ha^{-1})	Mean DBH (cm)	Mean Tree Height (m)
AGB ($Mgha^{-1}$)	0.95	0.90	0.74
Volume (m^3ha^{-1})		0.89	0.73
Mean DBH (cm)			0.72

An analysis of tree diameters at breast height (DBH) in the MFP revealed a clear trend as shown in (Figure 14). The most common DBH range is between 60 and 74 centimeters, with 127 trees falling into this category. This represents roughly 36.8% of the total number of trees measured. The DBH measurements also showed variation across tree species. The largest diameter recorded belonged to a plane tree, reaching 130 centimeters. In contrast, the smallest diameter belonged to a pine tree, measuring only 16 centimeters.

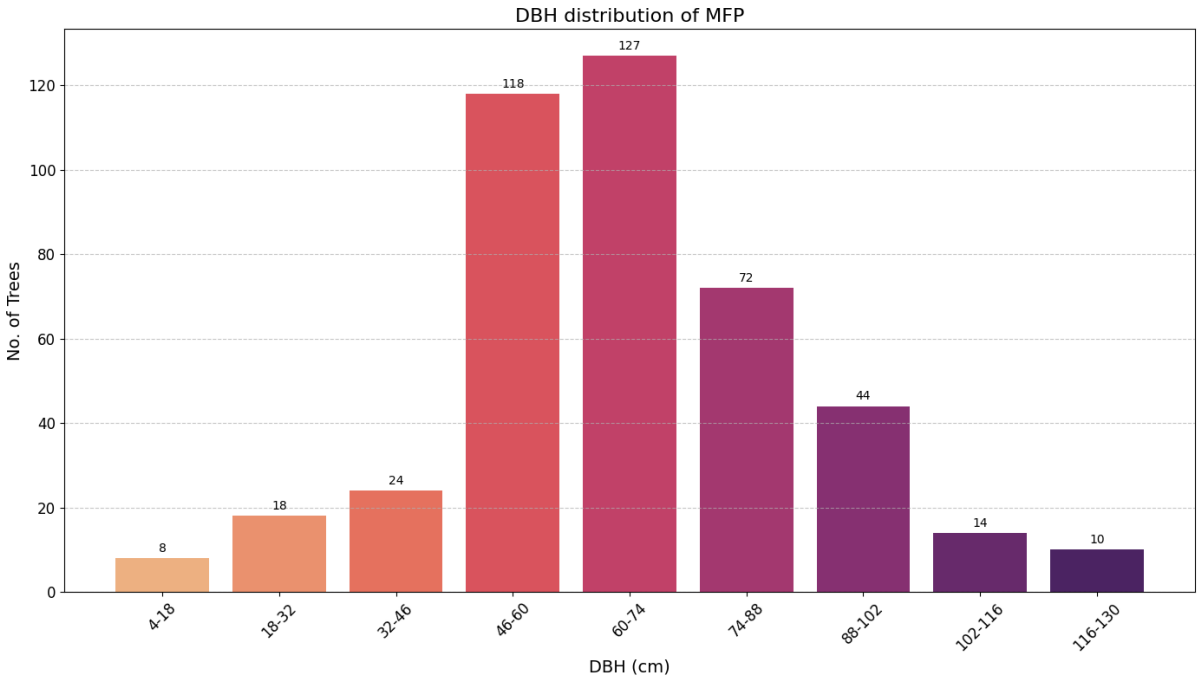

Figure 14 DBH distribution

The height interval class of 5-7 meters stood out with the highest number of trees at 120, accounting for 27.6% of the total trees in the forest plots of (MFP), as indicated in (Fig. 15). Among the trees in the MFP, the Chinese tallow tree attained the greatest height at 18 meters, while the Osmanthus fragrans is the shortest at 1.7 meters. This distribution underscores the diverse height ranges and species composition within the forest plot, reflecting a dynamic ecosystem.

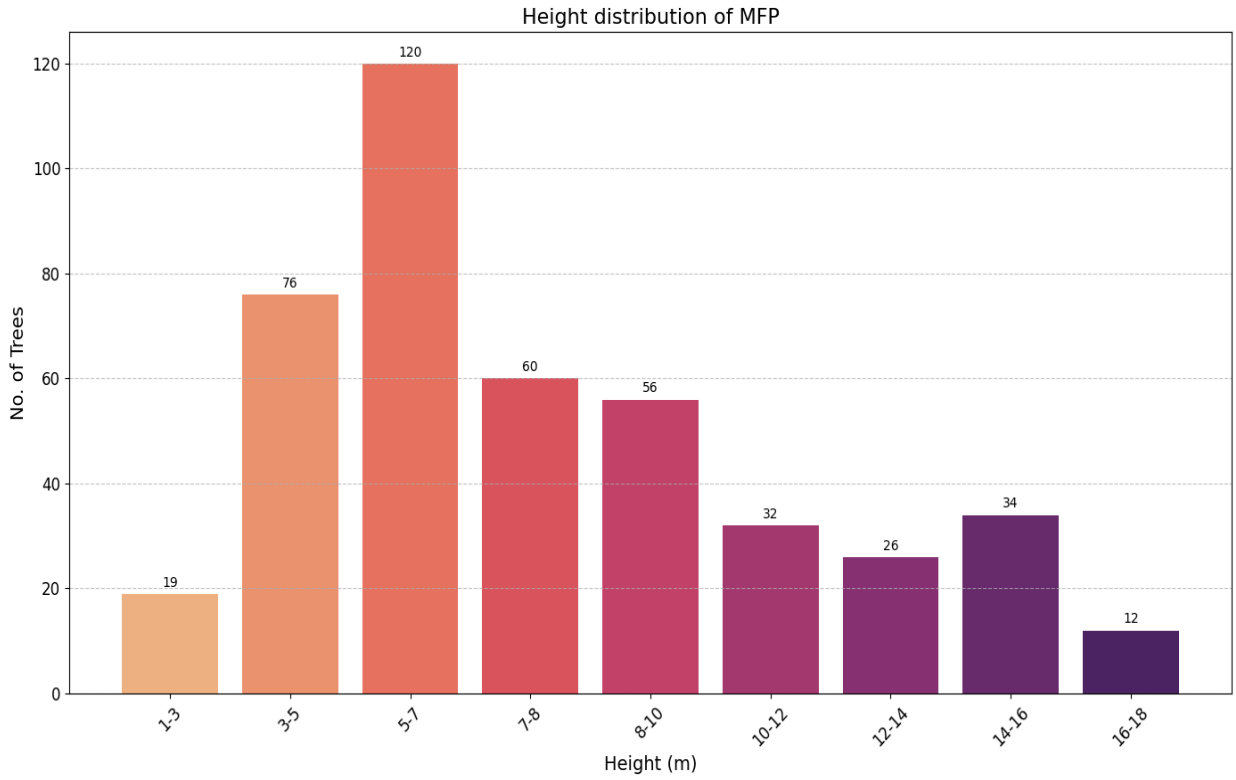

Figure 15 Height distribution

(Fig. 16) depicts a gradient from lighter to darker colors, indicating an increase in DBH (cm), which can also be observed on the x-axis and the y-axis represents AGB (kg). The visualization illustrates a positive relationship between AGB (kg) and tree height (m), with higher AGB values corresponding to taller trees. This color gradient provides a visual representation of how AGB increases with both height and diameter at breast height DBH (cm), highlighting the importance of these factors in estimating biomass. The gradual change in color intensities suggests a strong correlation between AGB and the measured tree characteristics, hinting at the role of tree maturity in biomass accumulation.

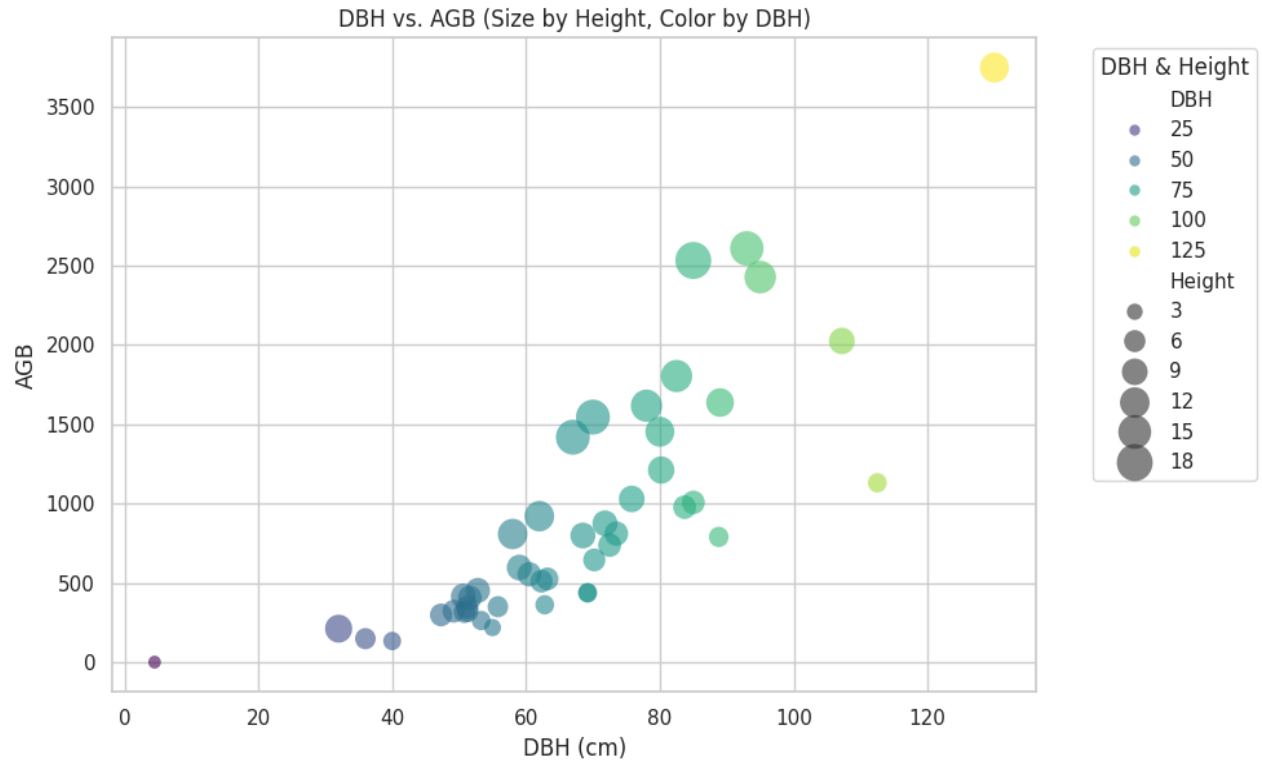

Figure 16 AGB against the mean DBH and the mean height

4.1.2 Indices estimation of the MFP

The Normalized Difference Vegetation Index (NDVI) exhibited a wide range from -0.31 to 0.83 across the MFP, reflecting significant variability in vegetation cover. The average NDVI of 0.60 ± 0.19 indicated a moderate level of vegetation health within the area. (Fig. 18) illustrates a left-skewed distribution of NDVI values, with the majority (72.87%) falling within the 0.6-0.8 class, representing healthy vegetation. This distribution pattern suggests that the MFP is predominantly covered by moderately dense to dense vegetation, highlighting the overall health and density of the vegetation within the plot. The observed NDVI values also indicate potential areas of interest for further investigation, such as regions with lower NDVI values that may indicate stress or sparse vegetation, which could be linked to specific land use or environmental conditions within the MFP. Understanding these variations in NDVI can provide valuable insights for land management and conservation efforts in the area.

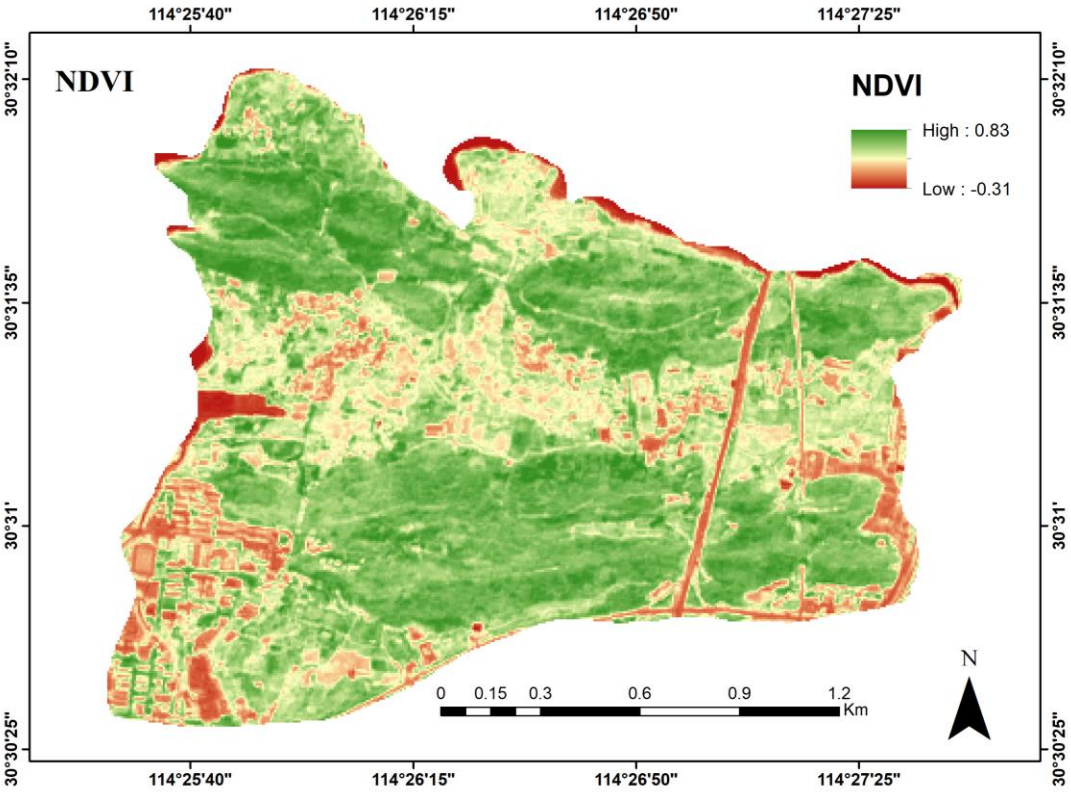

Figure 17 NDVI visualization map of MFP

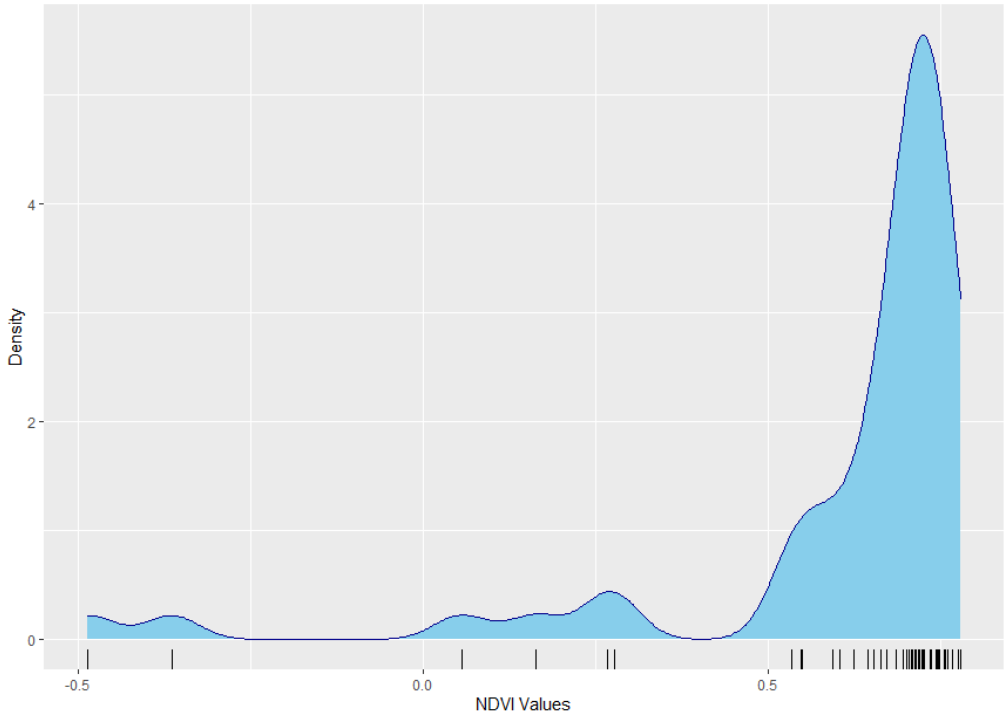

Figure 18 NDVI density Plot

Table 6 Percentage of NDVI within each class interval

Sr. No.	Interval of NDVI classes	Percentage of total pixels
1	> 0	1.56
2	0.2 – 0.4	4.35
3	0.4 – 0.6	21.22
4	0.6 – 0.8	72.87

The EVI values vary widely across the MFP, ranging from -0.15 to 0.91, (Fig. 19) shows the visualization of EVI estimation of MFP. Within this range, EVI values span from -0.15, indicating the lowest values, to 0.91, representing the highest values observed. The mean EVI value for the MFP is calculated to be 0.45 ± 0.16, indicating a moderate overall level of vegetation health. When examining the distribution of EVI values within the MFP, a bimodal pattern is evident (Fig. 20). The interval between 0.5 and 0.7 accounts for the majority of EVI values, comprising 61.19% of the total values or 54.19% of the total observations. Following this, the interval between 0.3 and 0.5 records 20.036% of the total values, indicating a secondary peak in EVI distribution (Table 7). This bimodal distribution suggests a complex pattern of vegetation health and density within the MFP, highlighting areas of both high and moderate vegetation cover.

Table 7 Percentage of EVI within each class interval

Sr. No.	Interval of EVI classes	Percentage of total pixels
1	> 0	2.77
2	0.3-0.5	20.036
3	0.5-0.7	61.19
4	0.7-0.9	16.004

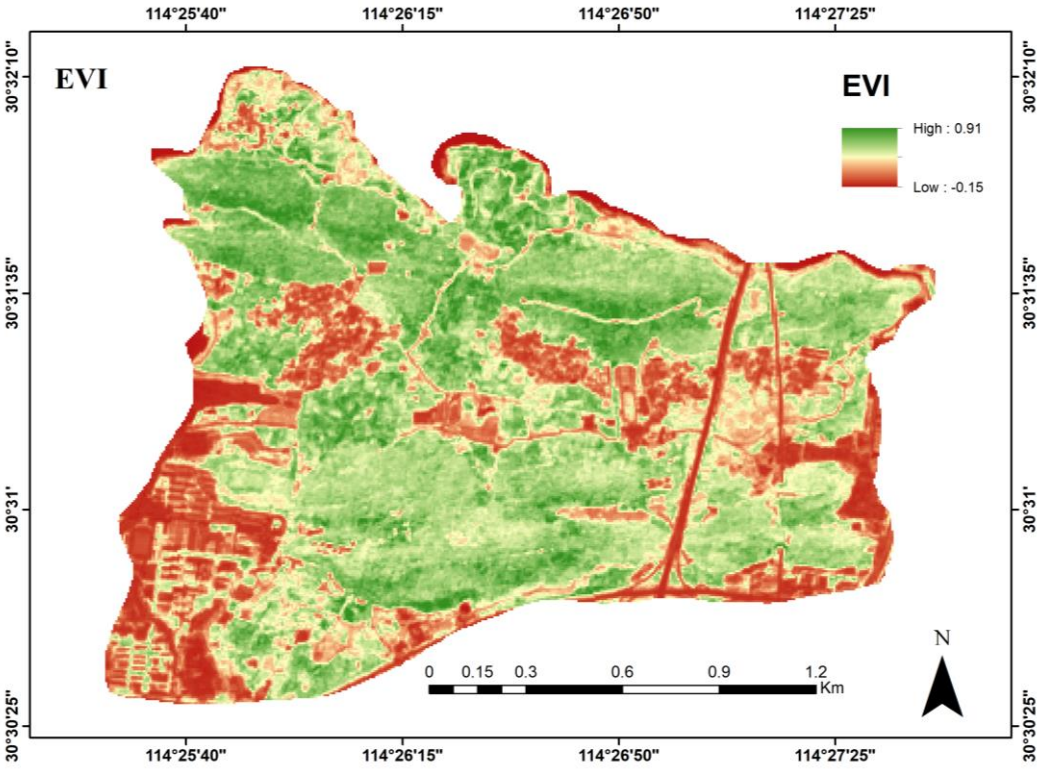

Figure 19 EVI visualization map of MFP

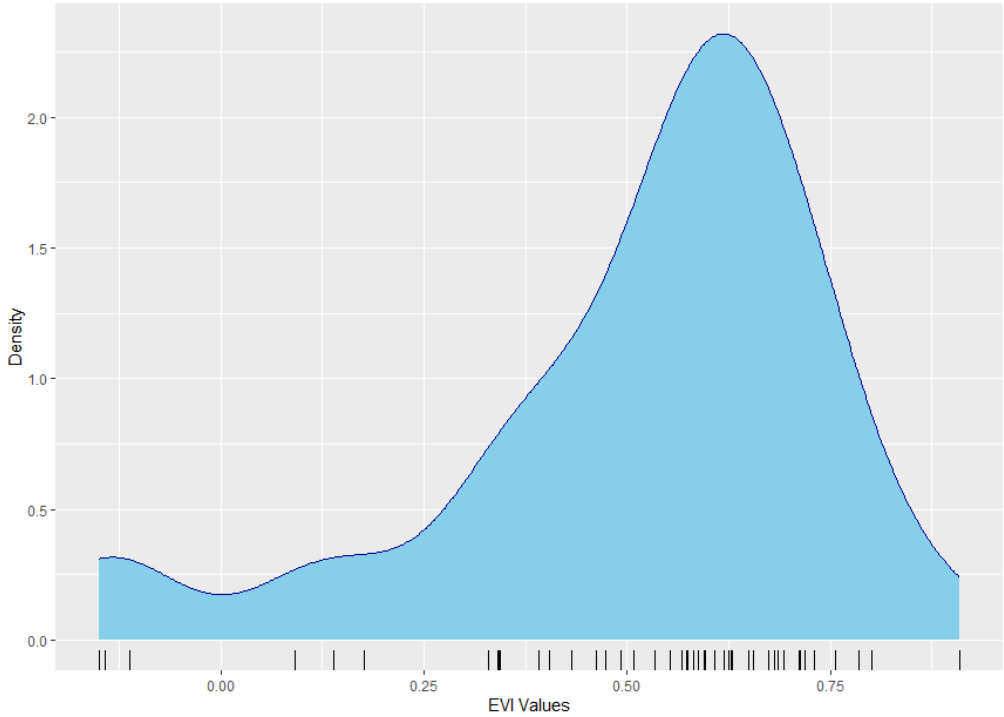

Figure 20 EVI density plot

4.2 Model evaluations, correlation, and AGB mapping of MFP

(Table 8) summarizes the performance of different models used to estimate AGB in the MFP. Three modeling approaches are evaluated: Stepwise Regression (SR), Random Forest (RF), and Support Vector Machine (SVM). Each model is tested with 11 different sets of predictor variables. The key findings are:

1) Overall Best Performing Set: The predictor set named "S1, Vis, and Bands" consistently produced the best models across all three methods (RF, SVM, and SR). This suggests this combination of variables is most effective for minimizing the error in AGB estimation.
2) Random Forest (RF): The "S1, Vis and Bands" set also achieved the best performance in the RF model, with the lowest Root Mean Square Error (RMSE) of 51.46 Mg ha^{-1} and an R-squared value of 0.33 (indicating a moderate positive relationship between predicted and actual values).
3) Support Vector Machine (SVM): The "S1, Vis, and Bands" set again performed best for SVM, resulting in an RMSE of 58.72 Mg ha^{-1} and an R-squared value of 0.19.
4) Stepwise Regression (SR): Likewise, the "S1, Vis and Bands" set yielded the lowest RMSE (58.29 Mg ha^{-1}) and an R-squared value of 0.11 for the SR model.

The "S1, Vis, and Bands" set of predictor variables seems to be the most informative combination for estimating AGB in the MFP study area, regardless of the modeling approach used.

The most influential variables for predicting the Above-Ground Biomass (AGB) are identified using a Random Forest (RF) model. This "S1, Vis, and Bands" model incorporated variables deemed most important by the RF algorithm. The ranking of these key variables for all three models (RF, Support Vector Machine (SVM), and Support Regression (SR)) is visualized in (Fig. 22, 23, and 24) respectively. Higher numerical values indicate greater importance within the model. To optimize performance, an iterative process that systematically removed variables with minimal contribution to the model's accuracy is employed. This refinement resulted in a streamlined "S1, Vis, and Bands" model containing only the most impactful variables, ultimately leading to the best possible AGB prediction.

Table 8 Model evaluations of the MFP

Predictor Variable Sets	RF		SVM		SR	
	RMSE (Mg ha^{-1})	R^2	RMSE (Mg ha^{-1})	R^2	RMSE (Mg ha^{-1})	R^2
S2 Bands	70.76	0.03	65.52	0.06	63.66	0.08
S2 Bands and S1	54.07	0.31	62.06	0.28	64.61	0.07
S2 Bands and S1 Texture	71.71	0.08	64.61	0.16	63.11	0.09
S2 Bands and VIs	68.11	0.04	65.54	0.03	60.96	0.15
S1	62.66	0.35	60.09	0.28	64.71	0.13
S1 and S1 Texture	55.42	0.24	64.59	0.19	71.04	0.18
S1, Vis and S2 Bands	51.46	0.33	58.72	0.19	58.29	0.11
S1 Texture	66.32	0.07	61.98	0.15	72.23	0.18
S1 Texture and Vis	68.51	0.16	66.77	0.08	69.01	0.05
Vis	65.02	0.06	69.67	0.22	61.84	0.18
All	54.44	0.22	66.38	0.23	69.61	0.09

The correlation heatmap offers a comprehensive view of the relationships between various remotely sensed parameters and Aboveground Biomass (AGB). (Fig. 21) shows the strength of these relationships. A statistical method called Spearman's correlation coefficient is used to assess these relationships. Several parameters exhibited moderate positive correlations, indicating their potential as valuable predictors in AGB estimation models. Notably, the backscatter coefficients from Sentinel-1 radar data (VH_Pol and VV_Pol), and the texture metrics derived from the Grey Level Co-occurrence Matrix (GLCM), such as (VH_cont, VH_ent, and VH_asm), also (S2_B4, S2_B8, and EVI) showed moderate positive correlations with AGB, suggesting their sensitivity to forest structural characteristics.

These findings highlight the importance of considering multiple data sources, including radar and optical data, as well as the inclusion of texture metrics that capture spatial patterns and heterogeneity beyond traditional vegetation index like NDVI which exhibited weaker positive correlations. Additionally, several other parameters exhibited weak but positive correlations, indicating their limited predictive power individually but potential to contribute to improved model

performance when combined with other relevant predictors. This comprehensive analysis provides valuable insights for developing robust predictive models that integrate complementary data sources for accurate biomass estimation in complex forest ecosystems.

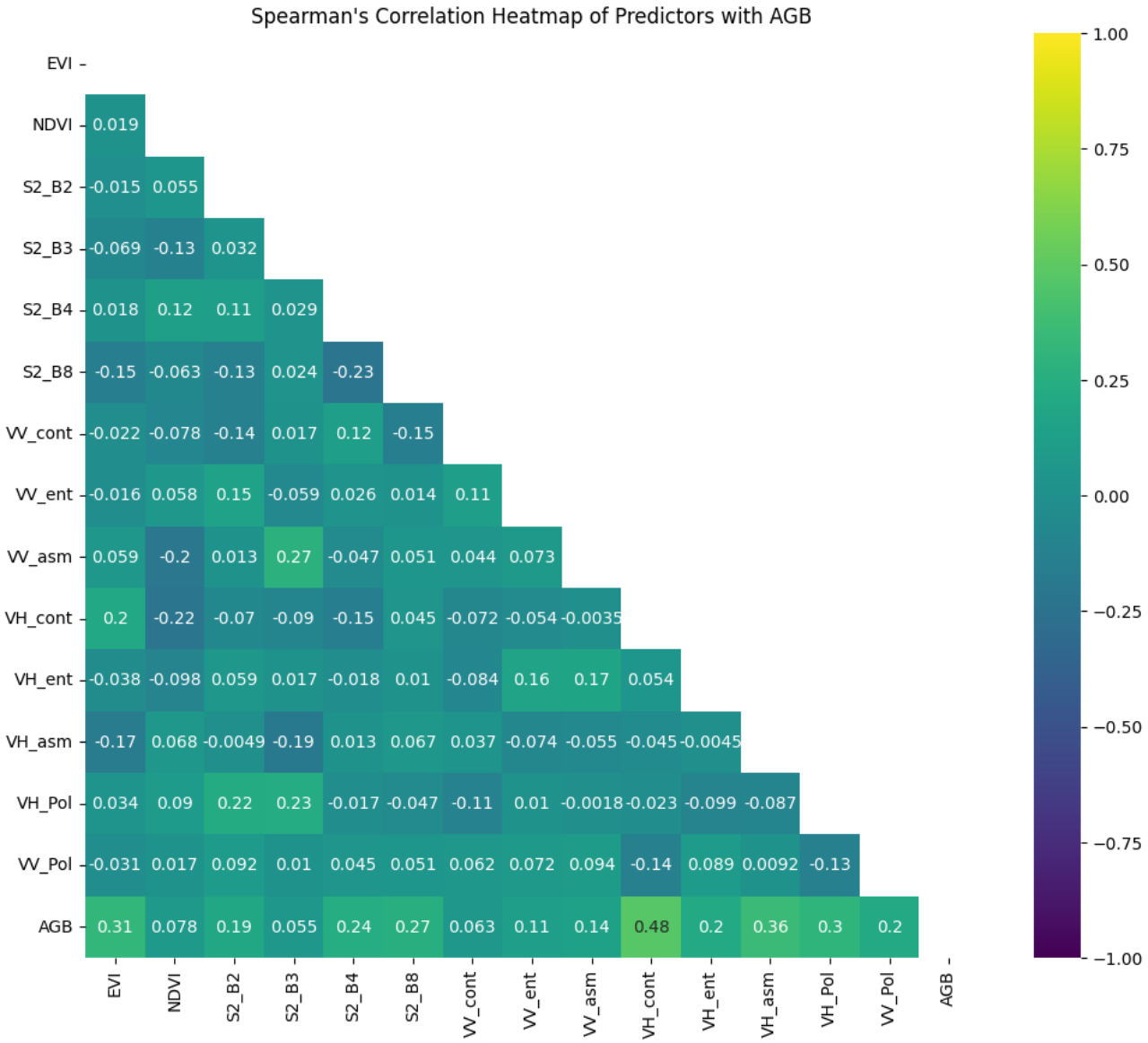

Figure 21 Correlation between selected predictor variables and AGB

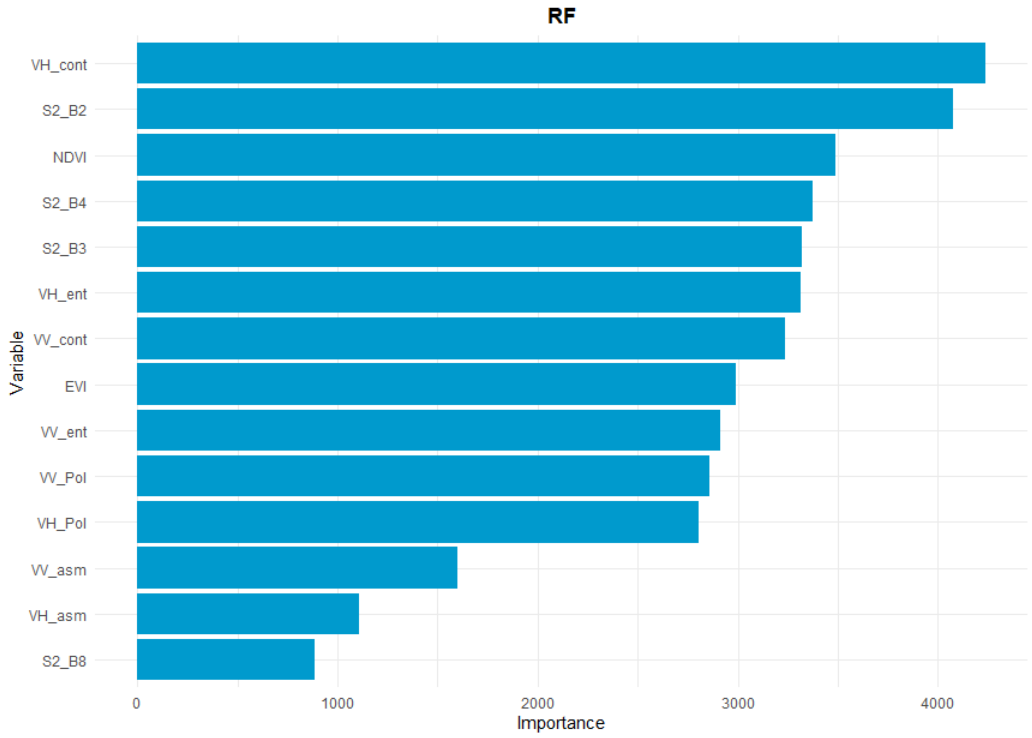

Figure 22 Ranking of the variables of the predictor set using RF algorithm

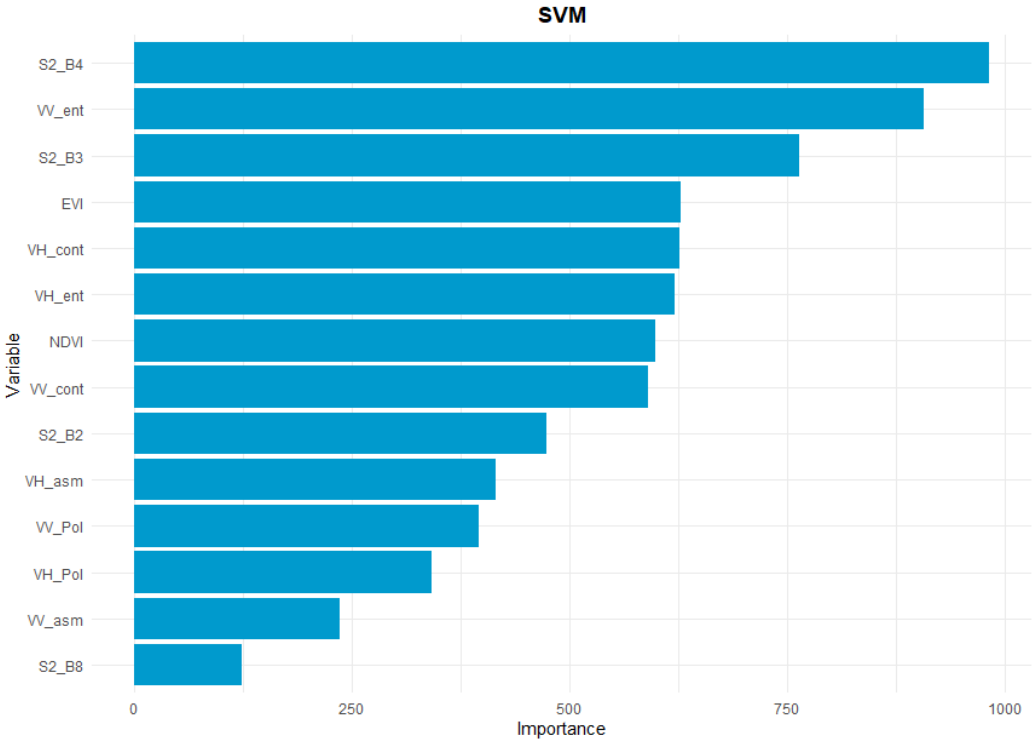

Figure 23 Variable importance using SVM

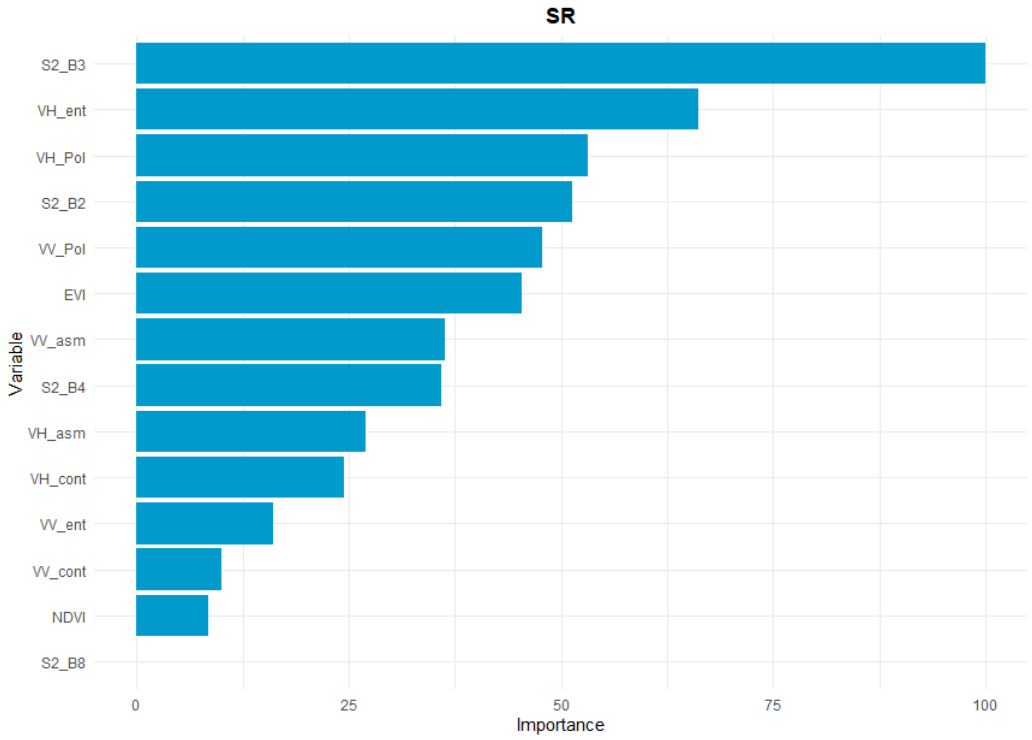

Figure 24 Variable importance using SR

As previously mentioned, the final RF model, which incorporates these top variables, achieves a Root Mean Square Error (RMSE) of 51.46 Mg ha^{-1} and an R-squared (R^2) value of 0.33, as detailed in (Table 7). This optimized RF model, constructed from the "S1, Vis, and Bands" predictor set, is subsequently employed for Aboveground Biomass (AGB) mapping using ground inventory data as validation. (Fig. 25) illustrates the generated AGB map, produced using the RF model. The map depicts AGB values ranging from a minimum of 47.37 Mg ha^{-1} to a maximum of 86.62 Mg ha^{-1}. In this visualization, areas with high AGB are depicted in green, while regions with lower AGB appear in red. Moderate AGB areas are represented by orange and yellow colors, providing a comprehensive view of biomass distribution across the study area. The blank portions of the map indicate non-forest areas, ensuring clarity in interpretation. This AGB map serves as a valuable tool for land management and conservation efforts, providing insights into biomass distribution. Additionally, the RF model's construction from the "S1, Vis, and Bands" predictor set further enhances its predictive power, leveraging a comprehensive range of input variables to refine its estimations.

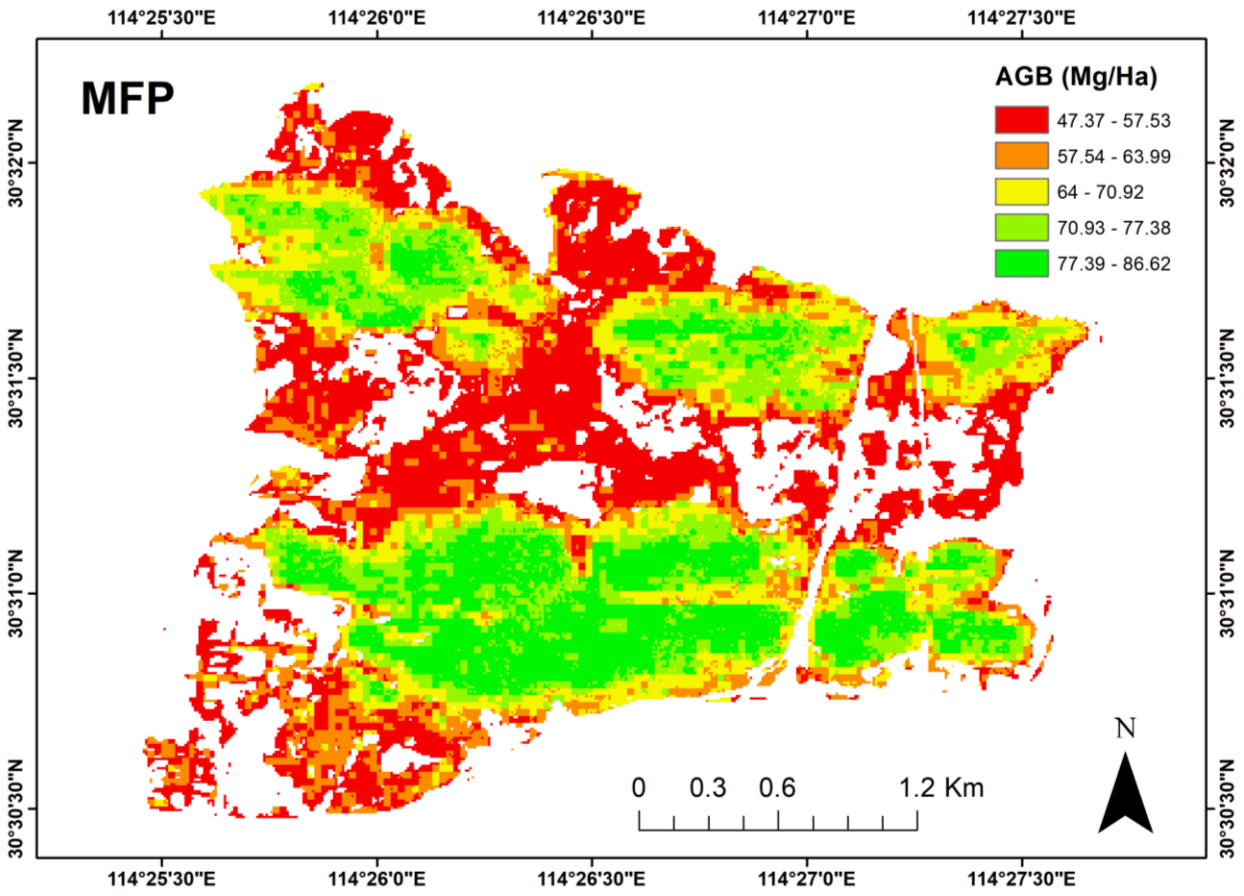

Figure 25 AGB map with field inventory validation data

To assess the effectiveness of different models for Above-Ground Biomass (AGB) prediction, I compared the predicted AGB values from each model with actual measurements obtained from field inventorying. The Random Forest (RF) model emerged as the clear leader, demonstrating a strong statistical association between its predictions and the observed AGB (Fig. 26). This is reflected in the high R-squared value of 0.87 and low RMSE=28.19 Mg/ha, indicating a close alignment between the RF model's estimates and the actual AGB, these metrics indicate the model's ability to accurately capture the variability in AGB across the study area. In contrast, both the Support Vector Machine (SVM) model (R-squared value of 0.51) and the Simple Regression (SR) model (R-squared value of 0.34) showed a moderate association between predicted and observed AGB (Fig. 27 and 28). This highlights the superior performance of the RF model, suggesting its effectiveness in accurately estimating AGB based on the chosen input variables.

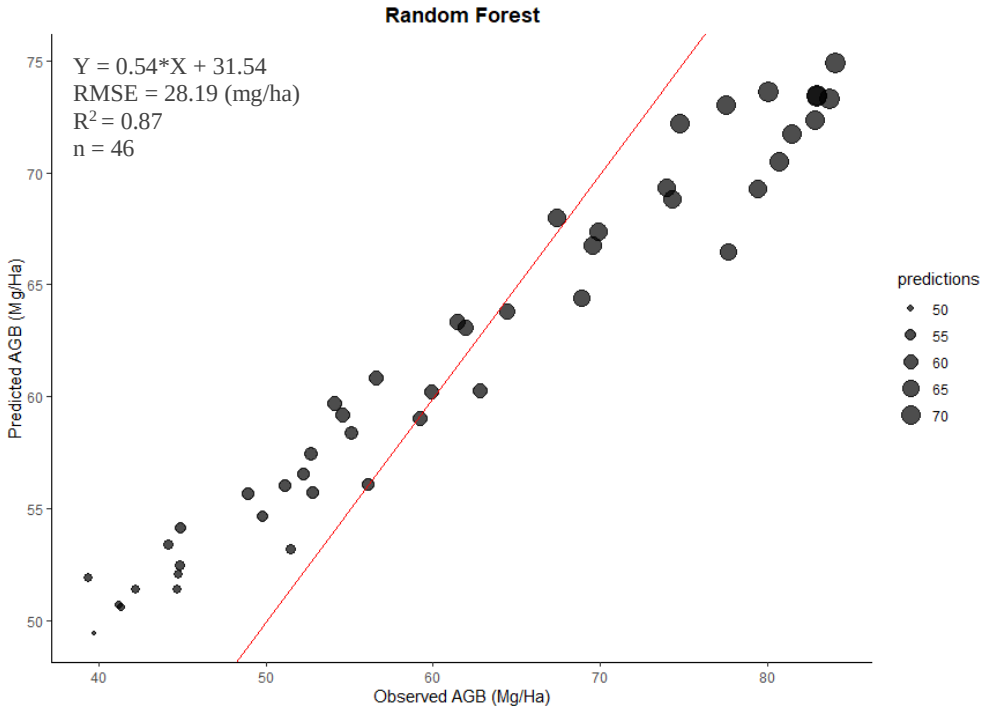

Figure 26 Predicted vs observed AGB using RF

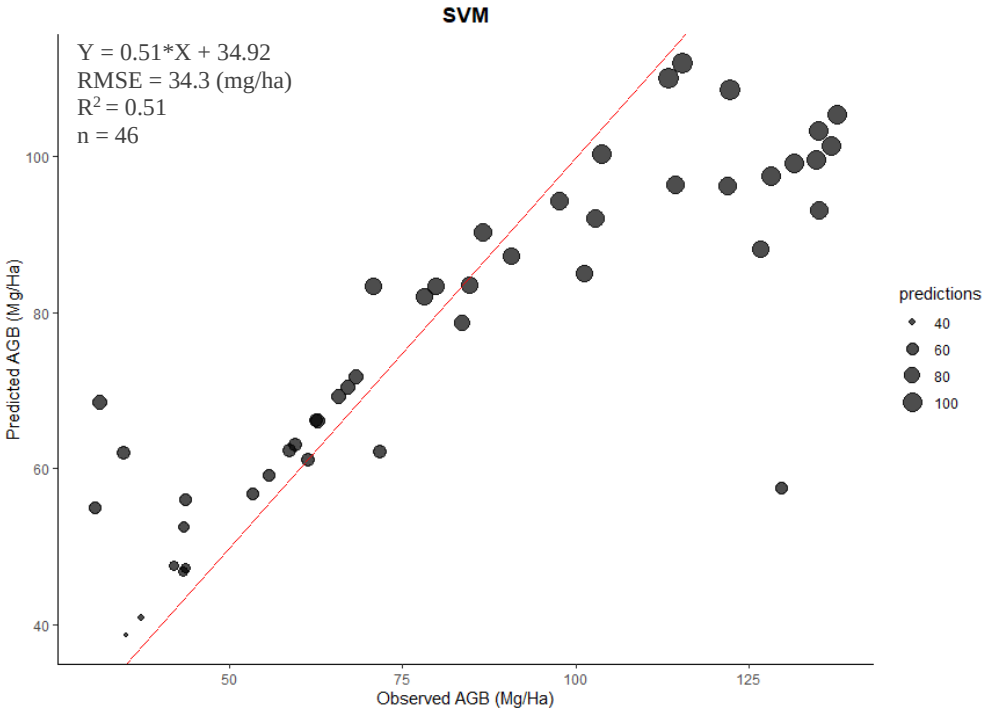

Figure 27 Predicted vs observed AGB using SVM

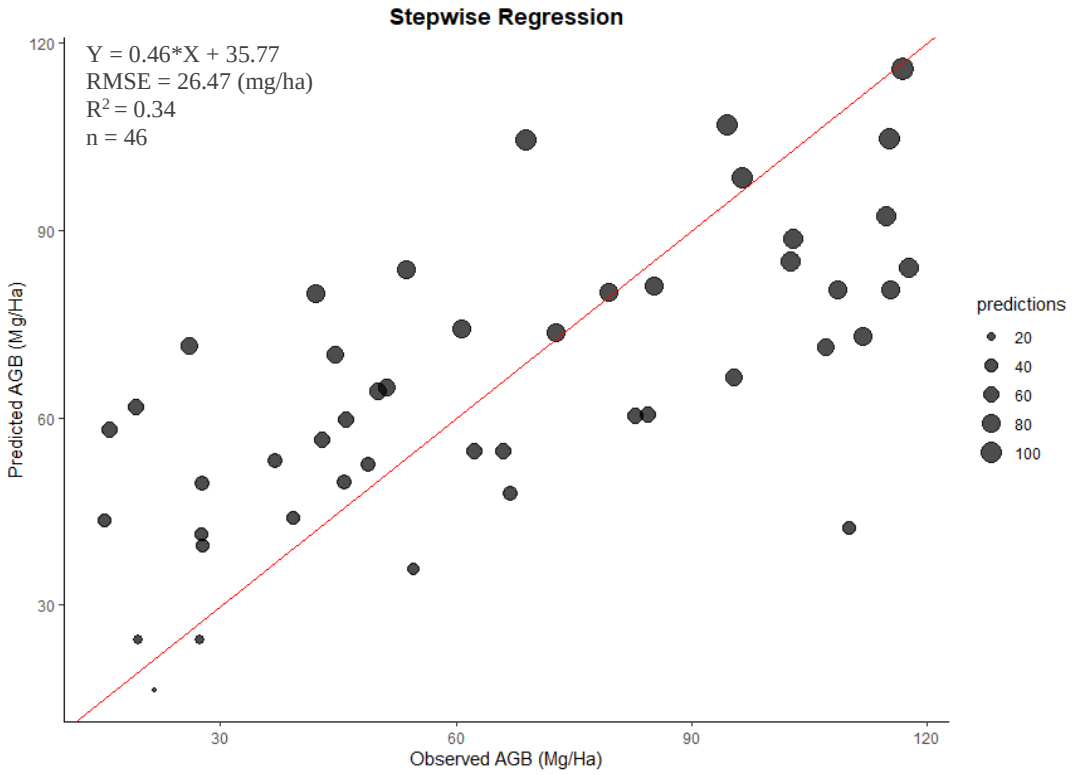

Figure 28 Predicted vs observed AGB using SR

4.2.1 AGB mapping using GEDI L4B as validation dataset

(Fig. 29) showcases the Aboveground Biomass (AGB) map generated by the Random Forest (RF) model, utilizing the "S1, Vis and Bands" predictor set. The predicted AGB values range from 50.37 Mg ha^{-1} to 90.62 Mg ha^{-1}, indicating spatial variations in biomass across the study area. This is particularly noteworthy considering the strong association ($R^2 = 0.83$) achieved by the RF model between predicted and observed AGB using GEDI L4B validation data. This surpasses the moderate associations observed with both the Support Vector Machine (SVM) model ($R^2 = 0.48$) and the Stepwise Regression (SR) model ($R^2 = 0.31$) for the same set of predictors. These findings highlight the effectiveness of the RF model in capturing the complex relationships between the chosen predictors and AGB distribution. Detailed results for all three models (RF, SVM, and SR) can be visualized in (Fig. 30, 31, and 32) respectively.

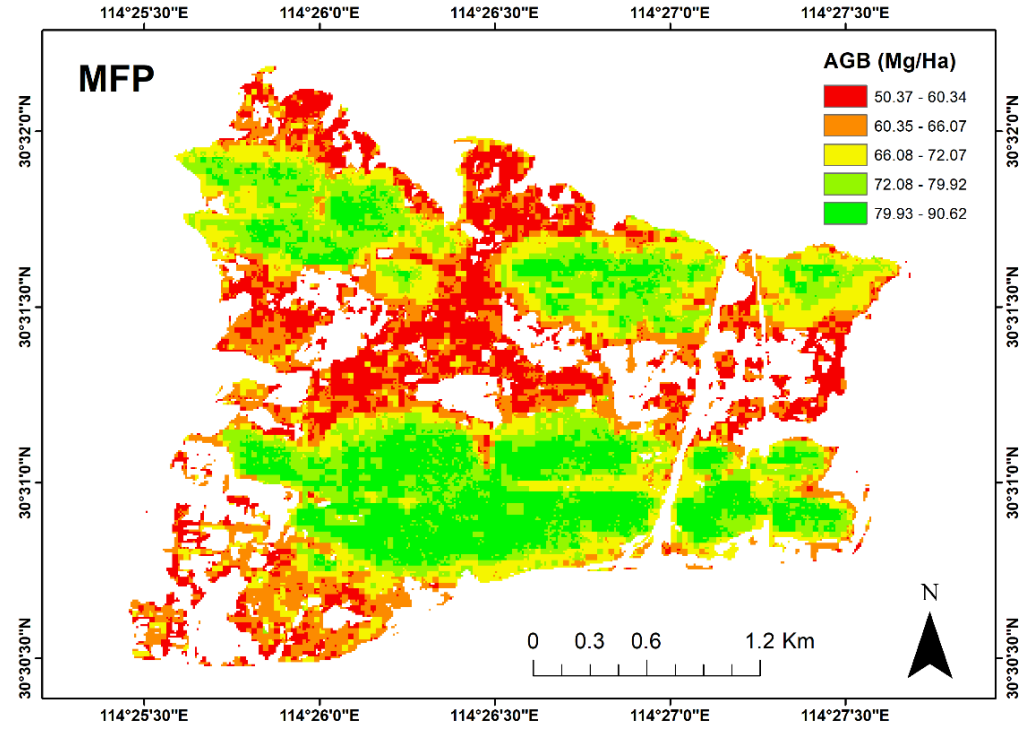

Figure 29 AGB map with GEDI L4B validation data

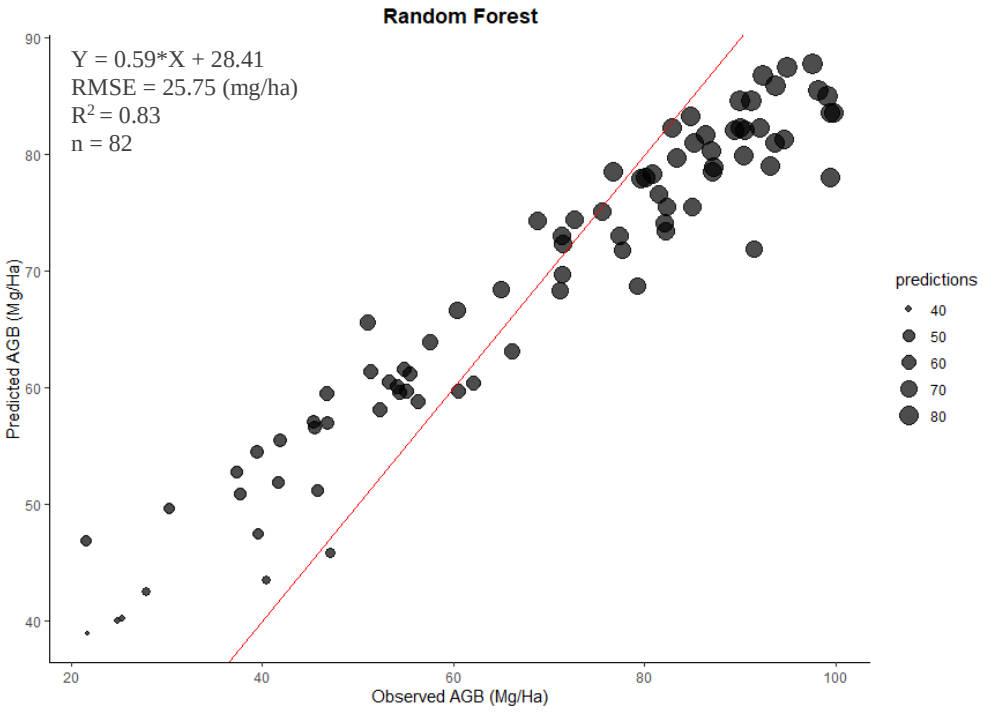

Figure 30 Predicted vs observed AGB using RF

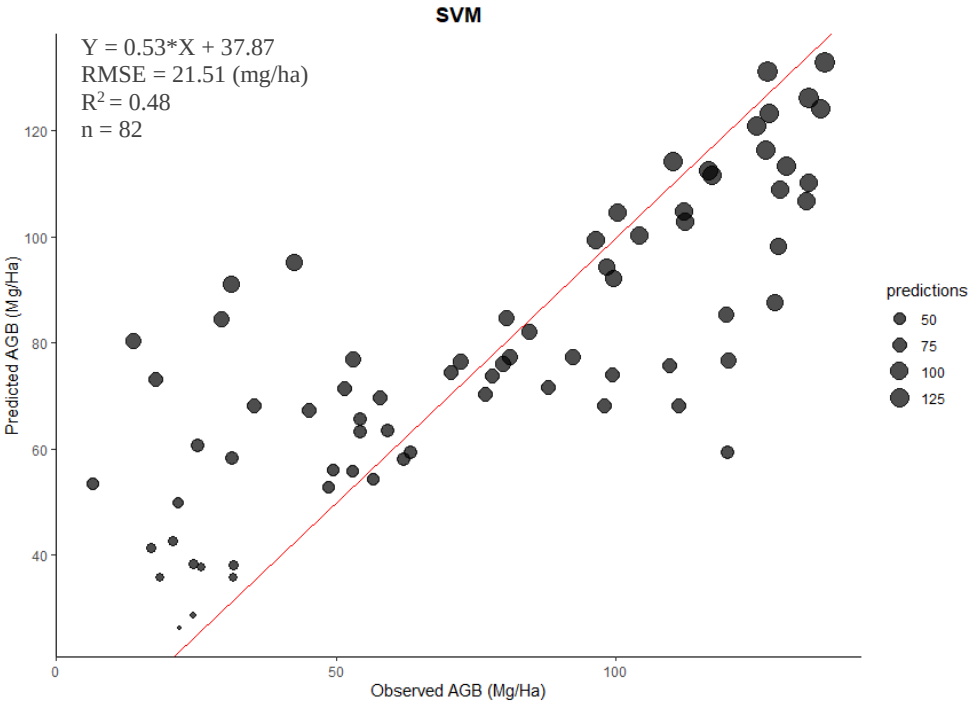

Figure 31 Predicted vs Observed AGB Using SVM

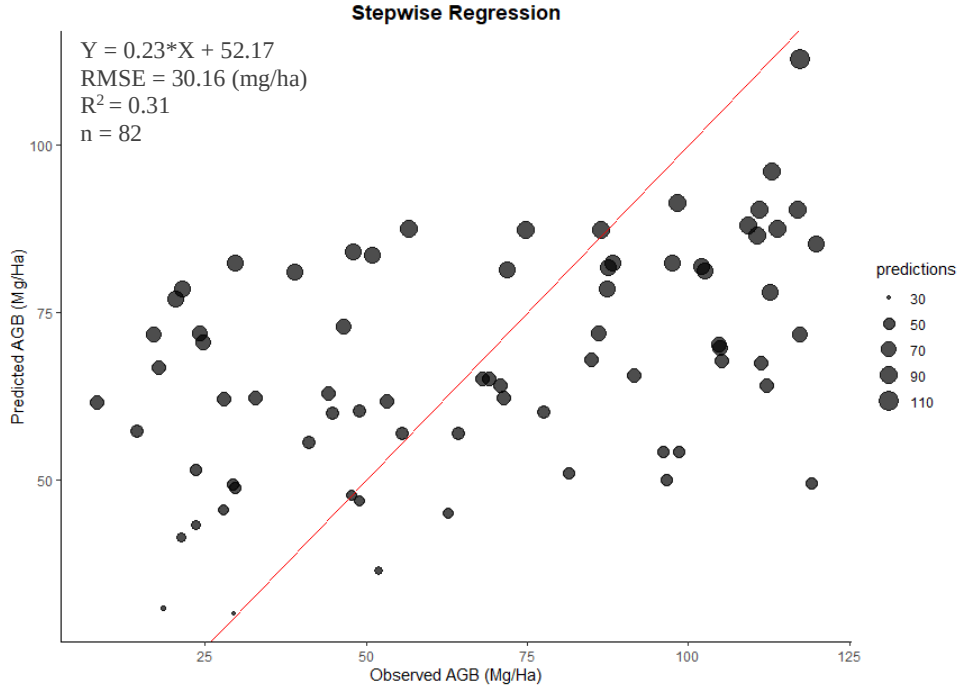

Figure 32 Predicted vs Observed AGB Using SR

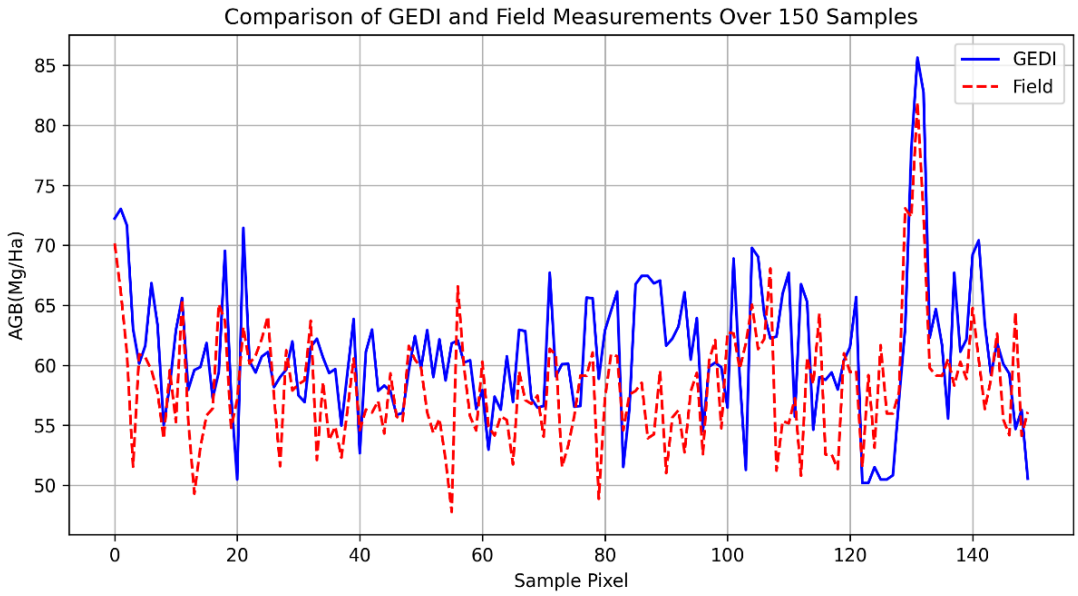

Figure 33 Trends of AGB mapped using GEDI L4B and field data

The comparison between GEDI L4B and field inventory predicted Above Ground Biomass (AGB) estimation reveals both similarities and differences (Fig. 33). Both datasets exhibit similar overall patterns, with peaks and troughs occurring at the same locations. This suggests they capture the same large-scale trends in AGB distribution. However, significant discrepancies exist between the two datasets at specific locations. Sometimes, GEDI L4B values are higher, and sometimes lower, than field inventory measurements. Field inventory data also shows more variation, with sharper peaks and deeper troughs. This could be because field data captures finer details missed by GEDI L4B due to its broader spatial coverage.

The comparison shows that models performed better using field inventory data for AGB estimation compared to GEDI L4B data. This change is caused due to differences in data quality, resolution, or other factors. GEDI L4B data offers a continuous and consistent view but may be affected by sensor limitations, atmospheric conditions, and processing assumptions. Field inventory data, although considered more accurate locally, may not be fully representative due to limited sampling and potential biases. Additionally, human error and measurement techniques can influence field measurements.

4.3 Summary of this chapter

This chapter presents the results of the biomass assessment and modeling analysis for the Maanshan Forest Park (MFP). The stand structure analysis of the MFP reveals a moderate level of biomass accumulation, with an average AGB of 33.77 Mg ha^{-1} and a high standard deviation of 51.63 Mg ha^{-1}, indicating variability across the sampled area. The mean tree volume is 37.82 m^3 ha^{-1}, and the average tree density is 142.90 trees ha^{-1}, suggesting a moderately dense forest. The analysis also shows a positive correlation between AGB and structural features, such as tree diameter, height, and volume.

The vegetation index (VI) analysis shows that the Normalized Difference Vegetation Index (NDVI) has a wide range from -0.31 to 0.83, with an average of 0.60 ± 0.19, indicating moderate vegetation health. The Enhanced Vegetation Index (EVI) has a bimodal distribution, with the majority of values (61.19%) falling in the 0.5-0.7 range. The chapter then presents the model evaluations and comparisons for the three modeling techniques: Random Forest (RF), Support Vector Machine (SVM), and Stepwise Regression (SR). The results show that the RF model with the "S1, Vis, and Bands" predictor set outperforms the other models, achieving the lowest RMSE of 51.46 Mg ha^{-1} and the highest R^2 of 0.33. The variable importance analysis for the RF, SVM, and SR models consistently identifies the "S1, Vis, and Bands" predictor set as the most influential, further confirming its superiority in AGB estimation. The final optimized RF model based on this predictor set is then used to generate the AGB map for the MFP, with values ranging from 47.37 Mg ha^{-1} to 86.62 Mg ha^{-1}.

The chapter also presents a comparison between the AGB estimates derived from the field inventory data and the GEDI L4B dataset. While the two datasets exhibit similar overall patterns, there are significant discrepancies at specific locations, with GEDI L4B sometimes overestimating and sometimes underestimating the AGB compared to the field measurements. The results in this chapter demonstrate the effectiveness of the integrated remote sensing and modeling approach in accurately estimating the aboveground biomass of the Maanshan Forest Park, providing valuable insights for forest management and carbon accounting efforts.

CHAPTER 5 CONCLUSION AND FUTURE WORK

In this chapter, I'm going to discuss a thorough evaluation of different remote sensing datasets, vegetation indices, validation datasets and advanced machine learning algorithms for estimating aboveground biomass (AGB) in the Maanshan Forest Park (MFP). The findings highlight the strengths and limitations of various approaches, emphasizing the potential for integrating multiple data sources to improve AGB estimation accuracy.

5.1 Conclusion

This study comprehensively evaluated the performance of different remote sensing datasets, vegetation indices, and advanced machine learning algorithms for estimating aboveground biomass (AGB) in the Maanshan Forest Park (MFP). The findings revealed valuable insights into the strengths and limitations of various approaches, as well as the potential for integrating multiple data sources to improve AGB estimation accuracy.

The study found a weak association between NDVI and aboveground biomass (AGB) for study area. Drawbacks of NDVI include saturation, which may not gauge increases in forest biomass. The Enhanced Vegetation Index (EVI) showed similar trends with weak associations with AGB. The study utilized C-band SAR data, showing a weak association with the aboveground biomass AGB. This result aligns with several researchers findings [53], while others have reported a strong association. SAR sensors have been used for AGB mapping for over three decades, with backscatter being a key measure related to AGB. The C-band sensitivity to leaves and smaller branches allows it to penetrate 1-2 meters into the canopy. However, it reaches saturation at a lower biomass level compared to longer wavelengths such as L-band and P-band, which can penetrate deeper into the canopy and are sensitive to larger branches and stems. The upcoming BIOMASS mission and NASA-ISRO NISAR mission are anticipated to offer valuable data for AGB mapping. The latter is particularly notable as the world's first SAR satellite operating in dual L-band and S-band wavelengths.

The comparison between GEDI L4B and field inventory based AGB estimates revealed both similarities and discrepancies. While the overall patterns were similar, with peaks and troughs occurring at the same locations, significant differences were observed at specific sample points.

Field inventory data exhibited sharper peaks and deeper troughs, potentially capturing finer-scale variations missed by GEDI L4B due to its broader spatial coverage and potential limitations in sensor characteristics, atmospheric conditions, and processing algorithms.

The study's results have practical implications for forest management and monitoring, demonstrating the potential of integrating remote sensing and machine learning techniques for AGB estimation. By improving our understanding of the factors influencing AGB distribution, this research contributes to more effective forest management strategies. The RF algorithm emerged as the top performer, outperforming SVM and SR models in estimating AGB using both field inventory data and GEDI L4B validation datasets. When utilizing field inventory data, the RF model achieved a significant statistical association ($R^2 = 0.87$) between predicted and observed AGB, with a root mean square error (RMSE) of 28.19 Mg ha^{-1}. Previous studies by [17], also reported the superior performance of RF in AGB prediction. RF's ability to handle complex non-linear relationships and a large set of variables makes it advantageous. The top predictor variables included C-band SAR data's VV and VH polarizations, as well as visible and shortwave infrared bands, underscoring their importance in AGB estimation.

For the GEDI L4B validation dataset, the RF model demonstrated a strong association ($R^2 = 0.83$) between predicted and observed AGB, with an RMSE of 25.75 Mg ha^{-1}, outperforming both SVM ($R^2 = 0.48$) and SR ($R^2 = 0.31$) models. This highlights the effectiveness of the RF algorithm in capturing complex relationships between predictor variables and AGB distribution.

The study also analyzed the vegetation indices NDVI and EVI, which showed a wide range of values across the MFP, reflecting variability in vegetation cover and health. However, the correlation between these indices and AGB was relatively weak, suggesting that they may not be the most reliable predictors of biomass in this particular study area. It is worth noting that the structural characteristics of the MFP, such as tree density, diameter at breast height (DBH), and tree height, exhibited strong positive associations with AGB. Stands with higher AGB tended to have larger trees (in terms of DBH and height) and greater total volume, underscoring the importance of these structural features in biomass accumulation.

5.2 Future work

Based on the comprehensive analysis and findings from this study, the following suggestions are proposed for the future work:

1) Data Generalizability: The study is conducted within a specific forest park, which may limit the generalizability of the results.
2) Model Refinement: Although the Random Forest model performed well, there is always scope for further optimization. Exploring additional variables or adjusting hyperparameters could potentially enhance the model's predictive capabilities.

Limitations:

Several limitations should be acknowledged and addressed in future research:

1) The study focused on a specific forest area (MFP), and the results may not be directly generalizable to other forest ecosystems with different structural and environmental characteristics. Further validation and testing in diverse forest types and regions would be necessary to assess the transferability of the findings.
2) The study relied on a limited number of field plots, which may not fully capture the spatial variability of AGB across the entire study area. Increasing the number and distribution of field plots could improve the representativeness of the ground truth data and enhance the accuracy of AGB models.

References

[1] A. R. B. Soutter and R. Mõttus, "Global warming versus 'climate change': a replication on the association between political self-identification, question wording, and environmental beliefs," *Journal of Environmental Psychology*, vol. 69, p. 101413, Jun. 2020, doi: 10.1016/j.jenvp.2020.101413.

[2] N. US Department of Commerce, "Global monitoring laboratory - carbon cycle greenhouse gases." Accessed: Mar. 14, 2024. Available: https://gml.noaa.gov/ccgg/trends/global.html

[3] P. Zhai, "Climate change and land: an ipcc special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems | IFPRI : International Food Policy Research Institute." Accessed: Mar. 14, 2024. Available: https://www.ifpri.org/publication/climate-change-and-land-ipcc-special-report-climate-change-desertification-land

[4] FAO, "Global forest resources assessment 2020: key findings." Accessed: Mar. 14, 2024. Available: https://reliefweb.int/report/world/global-forest-resources-assessment-2020-key-findings

[5] P. Vicharnakorn, "Carbon stock assessment using remote sensing and forest inventory data in savannakhet, lao pdr." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/6/6/5452

[6] T. Buchholz, "Mineral soil carbon fluxes in forests and implications for carbon balance assessments." Accessed: Mar. 14, 2024.
Available: https://onlinelibrary.wiley.com/doi/10.1111/gcbb.12044

[7] A. Rocha, S. C. Pereira, C. Viceto, R. Silva, J. Neto, and M. Marta-Almeida, "A consistent methodology to evaluate temperature and heat wave future projections for cities: a case study for lisbon," *Applied Sciences*, vol. 10, no. 3, Art. no. 3, Jan. 2020, doi: 10.3390/app10031149.

[8] S. Basheer *et al.*, "Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques," *Remote Sensing*, vol. 14, no. 19, Art. no. 19, Jan. 2022, doi: 10.3390/rs14194978.

[9] S. Fawzy, A. I. Osman, J. Doran, and D. W. Rooney, "Strategies for mitigation of climate change: a review," *Environmental Chemistry Letters*, vol. 18, no. 6, pp. 2069–2094, Nov. 2020, doi: 10.1007/s10311-020-01059-w.

[10] X. L. YUE, "Contributions of natural systems and human activity to greenhouse gas emissions - Science Direct." Accessed: Mar. 14, 2024.
Available: https://www.sciencedirect.com/science/article/pii/S1674927818300376

[11] L. Chen, "Forests | Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/1999-4907/9/10/582

[12] S. A. Mukul, Md. A. Halim, and J. Herbohn, "Forest carbon stock and fluxes: distribution, biogeochemical cycles, and measurement techniques," in *Life on Land*, W. Leal Filho, A. M. Azul, L. Brandli, A. Lange Salvia, and T. Wall, Eds., Cham: Springer International Publishing, 2020, pp. 1–16. doi: 10.1007/978-3-319-71065-5_23-1.

[13] S. Issa, "A review of terrestrial carbon assessment methods using geo-spatial technologies with emphasis on arid lands." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/12/12/2008

[14] M. Köhl, P. R. Neupane, and P. Mundhenk, "Redd+ measurement, reporting and verification – a cost trap? Implications for financing redd+mrv costs by result-based payments," *Ecological Economics*, vol. 168, p. 106513, Feb. 2020, doi: 10.1016/j.ecolecon.2019.106513.

[15] J. Geldmann, M. Barnes, L. Coad, I. D. Craigie, M. Hockings, and N. D. Burgess, "Effectiveness of terrestrial protected areas in reducing habitat loss and population declines," *Biological Conservation*, vol. 161, pp. 230–238, May 2013, doi: 10.1016/j.biocon.2013.02.018.

[16] M. Dumitrașcu, "Forests | Estimation of future changes in aboveground forest carbon stock in romania. a prediction based on forest-cover pattern scenario." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/1999-4907/11/9/914

[17] A. B. Debastiani, C. R. Sanquetta, A. P. D. Corte, N. S. Pinto, and F. E. Rex, "Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest," *Annals of Forest Research*, vol. 52, pp. 109–122, 2009, doi: 10.15287/afr.2018.1267.

[18] M. Hossain, "Development and evaluation of species-specific biomass models for most common timber and fuelwood species of bangladesh." Accessed: Mar. 14, 2024. Available: https://www.scirp.org/journal/paperinformation?paperid=97990

[19] Z. Asrat, T. Eid, T. Gobakken, and M. Negash, "Aboveground tree biomass prediction options for the Dry Afromontane forests in south-central Ethiopia," *Forest Ecology and Management*, vol. 473, p. 118335, Oct. 2020, doi: 10.1016/j.foreco.2020.118335.

[20] M. Sibanda, "Estimating biomass of native grass grown under complex management treatments using worldview-3 spectral derivatives." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/9/1/55

[21] A. Berninger, "Remote Sensing | SAR-Based Estimation of Above-Ground Biomass and Its Changes in Tropical Forests of Kalimantan Using L- and C-Band." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/10/6/831

[22] L. Duncanson *et al.*, "Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California," *Remote Sensing of Environment*, vol. 242, p. 111779, Jun. 2020, doi: 10.1016/j.rse.2020.111779.

[23] R. Dubayah *et al.*, "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography," *Science of Remote Sensing*, vol. 1, p. 100002, Jun. 2020, doi: 10.1016/j.srs.2020.100002.

[24] J. L. Hernández-Stefanoni, "Improving aboveground biomass maps of tropical dry forests by integrating lidar, alos palsar, climate and field data" *Carbon Balance and Management*, Accessed: Mar. 14, 2024.
Available: https://cbmjournal.biomedcentral.com/articles/10.1186/s13021-020-00151-6

[25] S. Sinha, C. Jeganathan, L. K. Sharma, and M. S. Nathawat, "A review of radar remote sensing for biomass estimation," *International Journal of Environmental Science and Technology*, vol. 12, no. 5, pp. 1779–1792, May 2015, doi: 10.1007/s13762-015-0750-0.

[26] G. Forkuor, J.-B. Benewinde Zoungrana, K. Dimobe, B. Ouattara, K. P. Vadrevu, and J. E. Tondoh, "Above-ground biomass mapping in west african dryland forest using sentinel-1 and 2 datasets - a case study," *Remote Sensing of Environment*, vol. 236, p. 111496, Jan. 2020, doi: 10.1016/j.rse.2019.111496.

[27] N. Jha *et al.*, "The real potential of current passive satellite data to map aboveground biomass in tropical forests," *Remote Sensing in Ecology and Conservation*, vol. 7, no. 3, pp. 504–520, 2021, doi: 10.1002/rse2.203.

[28] S. M. Ghosh and M. D. Behera, “Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest,” *Applied Geography*, vol. 96, pp. 29–40, Jul. 2018, doi: 10.1016/j.apgeog.2018.05.011.
[29] N. S. Skowronski, K. L. Clark, M. Gallagher, R. A. Birdsey, and J. L. Hom, “Airborne laser scanner-assisted estimation of aboveground biomass change in a temperate oak–pine forest,” *Remote Sensing Environment*, vol. 151, pp. 166–174, Aug. 2014, doi: 10.1016/j.rse.2013.12.015.
[30] C. Song, “Optical remote sensing of forest leaf area index and biomass - conghe song, 2013.” Accessed: Mar. 14, 2024. Available: https://journals.sagepub.com/doi/10.1177/0309133312471367
[31] X. Zhu and D. Liu, “Improving forest aboveground biomass estimation using seasonal landsat ndvi time-series,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 102, pp. 222–231, Apr. 2015, doi: 10.1016/j.isprsjprs.2014.08.014.
[32] J. Zhang, “A novel approach for estimation of above-ground biomass of sugar beet based on wavelength selection and optimized support vector machine.” Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/12/4/620
[33] E. Karimzadeh Jafari, H. Naghavi, K. Adeli, and H. Latifi, “A nondestructive, remote sensing-based estimation of the economic value of aboveground temperate forest biomass (case study: hyrcanian forests, nowshahr-iran),” *Journal of Sustainable Forestry*, vol. 39, no. 7, pp. 750–770, Oct. 2020, doi: 10.1080/10549811.2020.1723645.
[34] O. M. Opelele, Y. Yu, W. Fan, C. Chen, and S. K. Kachaka, “Biomass estimation based on multilinear regression and machine learning algorithms in the mayombe tropical forest, in the democratic republic of congo,” *Applied Ecology and Environmental Research*, vol. 19, no. 1, pp. 359–377, 2021, doi: 10.15666/aeer/1901_359377.
[35] J. Jin and J. Yang, “Effects of sampling approaches on quantifying urban forest structure,” *Landscape and Urban Planning*, vol. 195, p. 103722, Mar. 2020, doi: 10.1016/j.landurbplan.2019.103722.
[36] P. W. West, “The plane survey,” in *Tree and Forest Measurement*, P. W. West, Ed., Berlin, Heidelberg: Springer, 2009, pp. 125–134. doi: 10.1007/978-3-540-95966-3_12.
[37] S. L. Powell *et al.*, “Quantification of live aboveground forest biomass dynamics with landsat time-series and field inventory data: a comparison of empirical modeling approaches,” *Remote Sensing of Environment*, vol. 114, no. 5, pp. 1053–1068, May 2010, doi: 10.1016/j.rse.2009.12.018.
[38] P. W. West, “Tree height,” in *Tree and Forest Measurement*, P. W. West, Ed., Berlin, Heidelberg: Springer, 2009, pp. 17–21. doi: 10.1007/978-3-540-95966-3_4.
[39] P. W. West, “Stem diameter,” in *Tree and Forest Measurement*, P. W. West, Ed., Berlin, Heidelberg: Springer, 2009, pp. 11–15. doi: 10.1007/978-3-540-95966-3_3.
[40] M. Köhl, S. Magnussen, and M. Marchetti, "Sampling methods, remote sensing and gis multiresource forest inventory," in *Tropical forestry*. Berlin ; London: Springer, 2006.
[41] F. Chen *et al.*, “Normalized difference vegetation index continuity of the landsat 4-5 mss and tm: investigations based on simulation,” *Remote Sensing*, vol. 11, no. 14, Art. no. 14, Jan. 2019, doi: 10.3390/rs11141681.
[42] S. De Petris, F. Sarvia, and E. Borgogno-Mondino, “About tree height measurement: theoretical and practical issues for uncertainty quantification and mapping,” *Forests*, vol. 13, no. 7, Art. no. 7, Jul. 2022, doi: 10.3390/f13070969.

[43] Z. Szantoi and P. Strobl, "Copernicus sentinel-2 calibration and validation," *European Journal of Remote Sensing*, vol. 52, no. 1, pp. 253–255, Jan. 2019, doi: 10.1080/22797254.2019.1582840.
[44] C. Corbane, "A global cloud free pixel- based image composite from sentinel-2 data - sciencedirect." Accessed: Mar. 14, 2024. Available:https://www.sciencedirect.com/science/article/pii/S2352340920306314?via%3Dihub
[45] J. Wang *et al.*, "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*, vol. 154, pp. 189–201, Aug. 2019, doi: 10.1016/j.isprsjprs.2019.06.007.
[46] M. Mura *et al.*, "Exploiting the capabilities of the sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems," *International Journal of Applied Earth Observation and Geoinformation*, vol. 66, pp. 126–134, Apr. 2018, doi: 10.1016/j.jag.2017.11.013.
[47] N. Lamquin *et al.*, "An inter-comparison exercise of sentinel-2 radiometric validations assessed by independent expert groups," *Remote Sensing of Environment.*, vol. 233, p. 111369, Nov. 2019, doi: 10.1016/j.rse.2019.111369.
[48] M. Wessel, "Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on sentinel-2 data." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/10/9/1419
[49] L. Congedo, "Semi-automatic classification plugin documentation," *Journal of Open Source Software*, 2021.
[50] J. Chang and M. Shoshany, "Mediterranean shrublands biomass estimation using sentinel-1 and sentinel-2," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2016, pp. 5300–5303. doi: 10.1109/IGARSS.2016.7730380.
[51] G. V. Laurin, "Above-ground biomass prediction by sentinel-1 multitemporal data in central italy with integration of alos2 and sentinel-2 data." Accessed: Mar. 14, 2024. Available: https://www.spiedigitallibrary.org/journals/journal-of-applied-remote-sensing/volume-12/issue-01/016008/Above-ground-biomass-prediction-by-Sentinel-1-multitemporal-data-in/10.1117/1.JRS.12.016008.full#_=_
[52] D. Geudtner, R. Torres, P. Snoeij, A. Ostergaard, and I. Navas-Traver, "Sentinel-1 mission capabilities and sar system calibration," in *IEEE Radar Conference (RadarCon13)*, Apr. 2013, pp. 1–4. doi: 10.1109/RADAR.2013.6586141.
[53] N. Nuthammachot, A. Askar, D. Stratoulias, and P. Wicaksono, "Combined use of sentinel-1 and sentinel-2 data for improving above-ground biomass estimation," *Geocarto International*, vol. 37, no. 2, pp. 366–376, Jan. 2022, doi: 10.1080/10106049.2020.1726507.
[54] F. Filipponi, "Sentinel-1 grd preprocessing workflow," *Proceedings*, vol. 18, no. 1, Art. no. 1, 2019, doi: 10.3390/ECRS-3-06201.
[55] H. Laur *et al.*, "Derivation of the backscattering coefficient σo," no. 2, 2004.
[56] B. Dhruv, N. Mittal, and M. Modi, "Analysis of different filters for noise reduction in images," in *Recent Developments in Control, Automation & Power Engineering (RDCAPE)*, Oct. 2017, pp. 410–415. doi: 10.1109/RDCAPE.2017.8358306.
[57] A. Viña, A. A. Gitelson, A. L. Nguy-Robertson, and Y. Peng, "Comparison of different vegetation indices for the remote assessment of green leaf area index of crops," *Remote Sensing of Environment*, vol. 115, no. 12, pp. 3468–3478, Dec. 2011, doi: 10.1016/j.rse.2011.08.010.

[58] M. D. Steven, T. J. Malthus, F. Baret, H. Xu, and M. J. Chopping, "Intercalibration of vegetation indices from different sensor systems," *Remote Sensing of Environment*, vol. 88, no. 4, pp. 412–422, Dec. 2003, doi: 10.1016/j.rse.2003.08.010.

[59] P. L. Nagler, E. P. Glenn, T. Lewis Thompson, and A. Huete, "Leaf area index and normalized difference vegetation index as predictors of canopy characteristics and light interception by riparian species on the lower colorado river," *Agricultural and Forest Meteorology*, vol. 125, no. 1, pp. 1–17, Sep. 2004, doi: 10.1016/j.agrformet.2004.03.008.

[60] S. Taddeo, I. Dronova, and N. Depsky, "Spectral vegetation indices of wetland greenness: responses to vegetation structure, composition, and spatial distribution," *Remote Sensing of Environment*, vol. 234, p. 111467, Dec. 2019, doi: 10.1016/j.rse.2019.111467.

[61] A. Agapiou, "Remote Sensing | Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks." Accessed: Mar. 15, 2024. Available: https://www.mdpi.com/2072-4292/4/12/3892

[62] M. Guerini Filho, T. M. Kuplich, and F. L. F. D. Quadros, "Estimating natural grassland biomass by vegetation indices using sentinel 2 remote sensing data," *International Journal of Remote Sensing*, vol. 41, no. 8, pp. 2861–2876, Apr. 2020, doi: 10.1080/01431161.2019.1697004.

[63] E. L. Garroutte, A. J. Hansen, and R. L. Lawrence, "Using ndvi and evi to map spatiotemporal variation in the biomass and quality of forage for migratory elk in the greater yellowstone ecosystem," *Remote Sensing*, vol. 8, no. 5, Art. no. 5, May 2016, doi: 10.3390/rs8050404.

[64] X. Gao, A. R. Huete, W. Ni, and T. Miura, "Optical–biophysical relationships of vegetation spectra without background contamination," *Remote Sensing of Environment*, vol. 74, no. 3, pp. 609–620, Dec. 2000, doi: 10.1016/S0034-4257(00)00150-4.

[65] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the modis vegetation indices," *Remote Sensing of Environment*, vol. 83, no. 1, pp. 195–213, Nov. 2002, doi: 10.1016/S0034-4257(02)00096-2.

[66] P. C. Pandey, "Forest biomass estimation using remote sensing and field inventory: a case study of Tripura, India | *Environmental Monitoring and Assessment*." Accessed: Mar. 15, 2024. Available: https://link.springer.com/article/10.1007/s10661-019-7730-7

[67] L. Silvia, T. Alexander, K. Anna, and K. Polina, "Assessment of carbon dynamics in ecuadorian forests through the mathematical spatial model of global carbon cycle and the normalized differential vegetation index (ndvi)," *E3S Web Conference*, vol. 96, p. 02002, 2019, doi: 10.1051/e3sconf/20199602002.

[68] M. Buchhorn, "Influence of brdf on ndvi and biomass estimations of alaska arctic tundra - iopscience." Accessed: Mar. 15, 2024.
Available: https://iopscience.iop.org/article/10.1088/1748-9326/11/12/125002

[69] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, vol. 8, no. 2, pp. 127–150, May 1979, doi: 10.1016/0034-4257(79)90013-0.

[70] V. S. Thakare and N. N. Patil, "Classification of texture using gray level co-occurrence matrix and self-organizing map," in *International Conference on Electronic Systems, Signal Processing and Computing Technologies*, Jan. 2014, pp. 350–355. doi: 10.1109/ICESC.2014.66.

[71] G. D. T. GRASS Development Team, “Grass gis - bringing advanced geospatial technologies to the world.” Accessed: Mar. 15, 2024. Available: https://grass.osgeo.org/
[72] Y. O. Ouma and R. Tateishi, “Optimization of second-order grey-level texture in high-resolution imagery for statistical estimation of above-ground biomass,” *Journal of Environmental Informatics*, vol. 8, no. 2, Art. no. 2, Feb. 2015.
[73] B. Wang, H. Wang, and H. Qi, “Wood recognition based on grey-level co-occurrence matrix,” in *International Conference on Computer Application and System Modeling (ICCASM 2010)*, Oct. 2010, pp. V1-269-V1-272. doi: 10.1109/ICCASM.2010.5619388.
[74] T. M. Kuplich, “Relating SAR image texture and backscatter to tropical forest biomass | IEEE Conference Publication | IEEE Xplore.” Accessed: Mar. 15, 2024. Available: https://ieeexplore.ieee.org/document/1294615
[75] S. Eckert, “Remote Sensing | Improved Forest Biomass and Carbon Estimations Using Texture Measures from WorldView-2 Satellite Data.” Accessed: Mar. 15, 2024. Available: https://www.mdpi.com/2072-4292/4/4/810
[76] K. Iles, “A further neglected mean in: the mathematics teacher volume 70 issue 1 (1977).” Accessed: Mar. 15, 2024. Available: https://pubs.nctm.org/view/journals/mt/70/1/article-p27.xml
[77] P. W. West, *Tree and forest measurement*. Berlin, Heidelberg: Springer, 2009. doi: 10.1007/978-3-540-95966-3.
[78] D. M. Newbery, “M. s. philip measuring trees and forests.,” *Journal of Tropical Ecology*, vol. 11, no. 2, pp. 204–204, May 1995, doi: 10.1017/S0266467400008658.
[79] P. W. West, “Stem volume,” in *Tree and Forest Measurement*, P. W. West, Ed., Berlin, Heidelberg: Springer, 2009, pp. 23–32. doi: 10.1007/978-3-540-95966-3_5.
[80] P. W. West, “Measurements,” in *Tree and Forest Measurement*, P. W. West, Ed., Berlin, Heidelberg: Springer, 2009, pp. 5–10. doi: 10.1007/978-3-540-95966-3_2.
[81] P. W. West, “Stand measurement,” in *Tree and Forest Measurement*, P. W. West, Ed., Berlin, Heidelberg: Springer, 2009, pp. 65–89. doi: 10.1007/978-3-540-95966-3_8.
[82] J. Chave *et al.*, “Improved allometric models to estimate the aboveground biomass of tropical trees,” *Global Change Biology*, vol. 20, no. 10, pp. 3177–3190, 2014, doi: 10.1111/gcb.12629.
[83] B. Torres, “Structure and above ground biomass along an elevation small-scale gradient: case study in an evergreen andean amazon forest, ecuador | agroforestry systems.” Accessed: Mar. 15, 2024. Available: https://link.springer.com/article/10.1007/s10457-018-00342-8
[84] B. A. Tetemke, “Allometric models for predicting aboveground biomass of trees in the dry afromontane forests of northern ethiopia.” Accessed: Mar. 15, 2024. Available: https://www.mdpi.com/1999-4907/10/12/1114
[85] T. D. Pham, “Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, vietnam.” Accessed: Mar. 15, 2024. Available: https://www.mdpi.com/2072-4292/12/8/1334
[86] W. W. Chin, Chin, W.W. (1998). "The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), Modern methods for business research (pp. 295-336). Mahwah, NJ Lawrence Erlbaum Associates. - References - Scientific Research Publishing.” Accessed: Mar. 15, 2024.
Available: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=534264

[87] B. Ratner, "The correlation coefficient: its values range between +1/−1, or do they?" | *Journal of targeting, measurement and analysis for marketing*, Accessed: Mar. 15, 2024. Available: https://link.springer.com/article/10.1057/jt.2009.5
[88] K. Fawagreh, "Random forests: from early developments to recent advancements." Accessed: Mar. 15, 2024. Available: https://www.tandfonline.com/doi/full/10.1080/21642583.2014.956265
[89] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[90] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: from early developments to recent advancements," *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 602–609, Dec. 2014, doi: 10.1080/21642583.2014.956265.
[91] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, Jan. 2012, doi: 10.1016/j.isprsjprs.2011.11.002.
[92] S. Pandit, "Estimating above-ground biomass in sub-tropical buffer zone community forests, nepal, using sentinel 2 data." Accessed: Mar. 14, 2024. Available: https://www.mdpi.com/2072-4292/10/4/601
[93] A. T. N. Dang, S. Nandy, R. Srinet, N. V. Luong, S. Ghosh, and A. Senthil Kumar, "Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam," *Ecological Informatics*, vol. 50, pp. 24–32, Mar. 2019, doi: 10.1016/j.ecoinf.2018.12.010.
[94] J. M. B. Carreiras, "Estimating the above-ground biomass in miombo savanna woodlands (mozambique, east africa) using l-band synthetic aperture radar data." Accessed: Mar. 15, 2024. Available: https://www.mdpi.com/2072-4292/5/4/1524
[95] V. N. Vapnik, "The nature of statistical learning theory | springerlink." Accessed: Mar. 15, 2024. Available: https://link.springer.com/book/10.1007/978-1-4757-3264-1
[96] C. J.G.P.W, "Estimating grassland biomass using svm band shaving of hyperspectral data." Accessed: Mar. 15, 2024. Available: https://www.ingentaconnect.com/content/asprs/pers/2007/00000073/00000010/art00003;jsessionid=7ah04630eh8mm.x-ic-live-01
[97] A. Maxwell, "Implementation of machine-learning classification in remote sensing: an applied review." Accessed: Mar. 15, 2024. Available: https://www.tandfonline.com/doi/full/10.1080/01431161.2018.1433343
[98] F. Zhang and X. Yang, "Improving land cover classification in an urbanized coastal area by random forests: the role of variable selection," *Remote Sensing of Environment*, vol. 251, p. 112105, Dec. 2020, doi: 10.1016/j.rse.2020.112105.
[99] D. Domingo, "Comparison of regression models to estimate biomass losses and co2 emissions using low-density airborne laser scanning data in a burnt aleppo pine forest." Accessed: Mar. 15, 2024. Available: https://www.tandfonline.com/doi/full/10.1080/22797254.2017.1336067

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my Supervisor Professor Zhenfeng SHAO who has a kindness, patience and gave a good guidance to my work as well as any suggestion for study in Wuhan University and living in China. His knowledge and experience in remote sensing science are invaluable. This work can become a good dissertation by his improvements. Same with Dr. Nasir Mahmood and Dr. Iftikhar Ali, co-advisors who gave an importance assistance corresponding to the purpose of this research.

Thanks should also go to related staff from LIESMARS, Wuhan University who always gave a good supporting in knowledge and technical suggestion. Lastly, thanks to my parents for giving me an inspiration and warmth supporting as always.

Israr Ahmad

2024-05-24