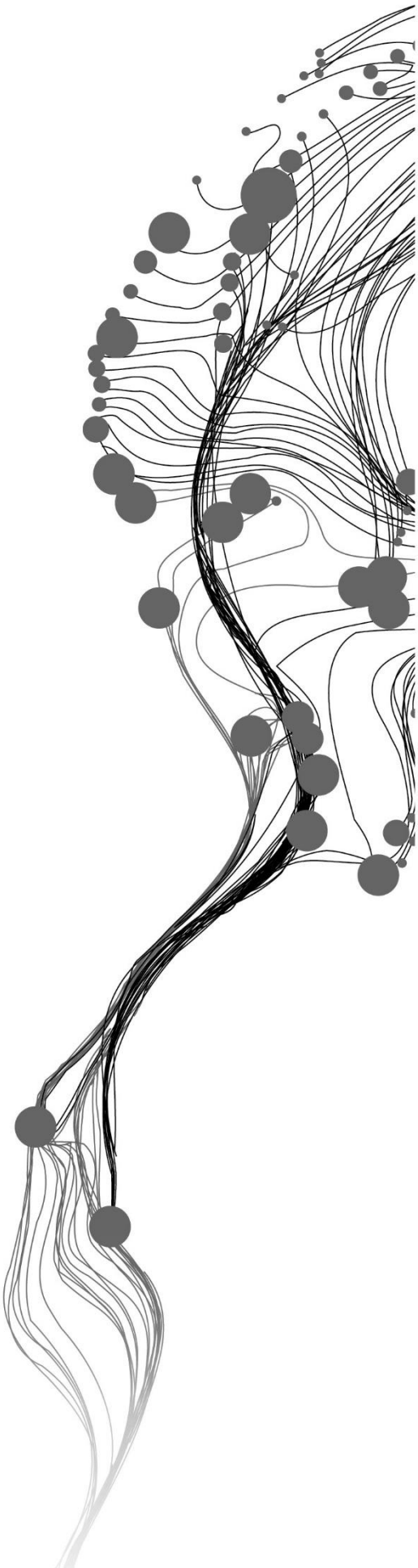
Integrating Single and Multi-Band Radar Data for Aboveground Biomass Assessment in Indonesia's Complex Tropical Landscape

ANCHAL CHAWLA
JUNE 2024

SUPERVISORS:
Dr. Michael Schlund
Dr. rer. nat. Florian Johannes Ellsäßer



Integrating Single and Multi-Band Radar Data for Aboveground Biomass Assessment in Indonesia's Complex Tropical Landscape

ANCHAL CHAWLA
Enschede, The Netherlands, June 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Dr. Michael Schlund
Dr. rer. nat. Florian Johannes Ellsäßer

THESIS ASSESSMENT BOARD:
Dr. Teijun Wang (Chair)
Dr. Stefan Erasmi (External Examiner, Thünen Institute of Farm Economics)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Forests are huge reserves of carbon. A reliable proxy for estimating the carbon stored in a forest is the Above Ground Biomass (AGB) of the vegetation. Accurate reporting of the change in forest carbon is an essential requirement under various international agreements and also for individual targets of countries. Radar remote sensing has the ability to penetrate the forest canopy and hence has been widely used for estimating AGB. Different radar bands, each with distinct wavelengths, are sensitive to various components of vegetation. Longer wavelengths (like L- band) are more sensitive to the trunk while shorter wavelengths (like X- band) are more sensitive to the leaves and branches. The landscape in Indonesia is rapidly changing due to several reasons. Hence, the long-standing primary forests are converted into other land use categories like plantations. Therefore, It becomes essential to estimate the carbon lost during this process. This study aimed at examining and comparing individual radar X-, C- and L- bands and an integration of these bands for predicting the AGB of a complex tropical landscape in Indonesia with heterogeneous vegetation. The models were created using linear regression incorporating all field plots across the four land use categories, as well as specific subsets: forest, oil palm plantation, and a combination of both forest and oil palm plantation plots. The training R^2 of the models using field plots of all four land use categories was consistently below 0.1 across all bands. When two land use categories were removed- rubber and shrubland, the R^2 increased by many folds, with the highest achieved when using a combination of all three bands together and a combination of L- and C- band. For the oil palm plantations, X- band performed the best (0.44), along with its combination with C- band. For forests, L-band proved to have the highest accuracy (0.20), along with its combination with C- band and using all 3 bands together. Moreover, a comparison of these model coefficients was carried out to understand the difference in the slopes and intercepts, using ANCOVA. Between the three bands, the model slopes did not differ significantly for different land use categories, except between X- and L-band for forests. Between models for forests and for oil palm plantations, the slopes were not significantly different, except when using X- band backscatter as the continuous variable(factor). The intercepts were different when using L- and C- bands as the factor. Further analysis revealed that pooling the land use categories with statistically different intercepts might introduce a persistent bias in the models. To check the performance of the models on unseen data, validation accuracies (R^2 and RMSE) were computed using the leave-one out cross validation methods. The highest validation R^2 was obtained for oil palm plantation field plots when using backscatter from X- and C- band (0.29). Relative RMSE showed that X- band had the least prediction errors for oil palm plantations, whereas L-band had the least errors for forests. A number of uncertainties in the models were discussed, some of them being the complexity of using different land use categories, non-linear relationships between AGB and backscatter and presence of outliers.

Keywords: AGB, Plantations, Radar, X- band, C-band, L- band

ACKNOWLEDGEMENTS

I would extend my thanks to various individuals who had supported me throughout towards the completion of my thesis. Firstly, my two supervisors- Dr. Michael Schlund and Dr. rer. nat. Florian Johannes Ellsäßer for their guidance. Their expertise in the area and their experience helped me overcome various difficulties that I faced during my thesis. It was an enriching experience to work with them.

I am also deeply thankful to my thesis chair - Dr. Teijun Wang, for giving me helpful and constructive feedback and always making me critically think about my thesis. I would also extend my gratitude to Drs. Raymond Nijmeijer who was always there to guide me through my study progress. Moreover, dr. I.C. van Duren, who always kept a check on the progress with thesis and gave useful insights. I appreciate her giving time to my doubts about the non-technical parts of the thesis.

I would really like to thank my family who were pillars of strength for me during my study. My friends that I made along the way have also helped me get through tough times and move forward.

Lastly, I feel extremely grateful to have studied at Faculty of Geo-Information Science and Earth Observation of the University of Twente, who provided me with the ITC excellence scholarship which aided me financially. The skills that I acquired at this university and the connections I made were priceless.

Anchal Chawla
June 2024
Enschede

TABLE OF CONTENTS

1. Introduction 9
   1.1. Background 9
   1.2. Problem statement 11
   1.3. Objectives, Research questions and Hypothesis 12
2. Study area and datasets 13
   2.1. Study area 13
   2.2. SAOCOM 1A L-band data 13
   2.3. Sentinel-1A C-band data 14
   2.4. TanDEM-X X-band data 14
   2.5. Field data 15
3. Methodology 17
   3.1. Research methodology 17
   3.2. SAR data pre-processing 18
   3.3. Backscatter extraction 18
   3.4. Developing regression model 19
   3.5. Comparison of models 20
   3.6. Accuracy assessment 23
4. Results 25
   4.1. Results from regression analysis 25
   4.2. Results of the comparison of models 30
   4.3. Assessment of predictive accuracy in developed models 33
5. Discussion 37
   5.1. Investigating the difference in model performance and coeffecients between different radar bands individually 37
   5.2. Investigating the difference in model coeffecients between forests and oil palm plantations 38
   5.3. Penalising model complexity for integration of bands 39
   5.4. Uncertainties in regression analysis 39
   5.5. Practical applications and future prospects 42
   5.6. Proposed recommendations 42
6. CONCLUSIONS 44
List of references 45
APPENDICES 52

LIST OF FIGURES

Figure 1- Location of the study area - Sumatra Island within Indonesia and Sarolangun Ban and Batang Hari regencies within Jambi province (right image). (Source: basemap from ESRI, Indonesia administrative subdivision from Bakosurtanal–RSGISforium-Bpk. Suwandito) 13
Figure 2 – Coverage of the three satellite images around the field plots. (Basemap from ESRI) 16
Figure 3- Flowchart of methodology explaining how the objectives were achieved. The same methodology was performed for predicting the AGB of different land use categories and their combinations. 17
Figure 4- a) Venn diagram representing the bands and their combinations which will be used as the independent variable. b) Representing the different land uses and their combinations present in the study. 19
Figure 5- Methodological steps followed for performing ANCOVA. 21
Figure 6- Leave-one out cross validation (Cha et al., 2020) 23
Figure 7- Box plot showing the range of AGB values across different land use categories. 24
Figure 8- Heat map visualising the significance levels of various models created with different bands and their combinations for different land use categories. For forests, all models with L-band and for oil palm plantations all models with X- band have a significant slope. 26
Figure 9- Regression plots showing relationship between log (AGB) of all four land use categories and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter. 27
Figure 10- Regression plots showing relationship between log (AGB) of oil palm plantations and forests and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter 28
Figure 11- Regression plots showing relationship between log (AGB) of oil palm plantations and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter 29
Figure 12- Regression plots showing relationship between log (AGB) of forests and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter 30
Figure 13- Comparison of the regression lines between models created with X-, C- and L-band backscatter for AGB of a) all 4 land use categories b) forests and oil palm plantations c) oil palm plantations and d) forests. 32
Figure 14- comparison of the regression lines between models created for predicting the AGB of forests and AGB of oil palms using a) X- band, b) C- band and c) L-band data. 33
Figure 15- Comparison of the relative RMSE between models created for predicting AGB of different land use categories using different bands and their combinations. 34
Figure 16- Comparison of the R^2 between models created for predicting AGB of different land use categories using different bands and their combinations. 35
Figure 17- Comparison of AIC values between models created for predicting AGB of different land use categories using different bands and their combinations. 36
Figure 18- Comparison of the training and validation R^2 for a) all four land use categories b) forests and oil palm plantations c) oil palm plantations and d) forests. These graphs show the overfitting of the models. 38
Figure 19- Plot S27 representing the spaces without vegetation in the plot. The red circle depicts the original extent of the field plot and the black square is the extent of the field plot created for the backscatter extraction from the pixels. 40
Figure 20- Elaeis *guineensis* oil palm plantation with horizontal structured branches (CABI, 2024). 41
Figure 21- Showing the cloud disturbance (outlined in red) present in the TanDEM-X image 42

LIST OF TABLES

Table 1- Comparison of the characteristics of the X-, C- and L-band data 14
Table 2- Representing the different bands (covariate) between which the ANCOVA was performed for different land uses and their combinations, with the backscatter as the factor..... 22
Table 3 – Representing the two land use categories (covariate) between which ANCOVA was performed with the backscatter from different bands as the factor. 22
Table 4- The training R^2 and model coefficients with significance levels marked as * (lightest blue) for p-values < 0.05, ** (light blue) for < 0.01, and *** (dark blue) for < 0.001. Coefficients without any * symbol are not statistically significant. 25
Table 5- The % confidence that the slopes are statistically different between different bands. Where, n.s indicates no significant difference between slopes, 95% (lightest blue shade) ,99% (light blue) and 99.9% (dark blue) represent the respective confidence that slopes are statistically different from each other for the bands. 30
Table 6- The % confidence that the intercepts are statistically different between different bands. Where, n.s means no significant difference between slopes, 95% (lightest blue shade) ,99% (light blue) and 99.9% (dark blue) represent the respective confidence that slopes are statistically different from each other for the bands. 31
Table 7- The % confidence that the slopes and intercepts are statistically different between models created with forests and oil palm plantations data. Where, n.s means no significant difference between slopes, 95% (lightest blue shade) ,99% (light blue) and 99.9% (dark blue) represent the respective confidence that slopes are statistically different from each other for the bands. 32

1. INTRODUCTION

1.1. Background

Climate change is the biggest threat to our environment in recent times (Deep, 2023; Fanelli, 2013). Human-induced deforestation is a significant contributor to climate change (IPCC, 2014). This is based on the fact that deforestation increases the level of carbon dioxide in the atmosphere, amplifying the greenhouse effect (Achard et al., 2004; Hansen et al., 2013). Forests have the capacity to serve as either carbon dioxide sources or sinks (IPCC, 2014). However, when they are cut down for reasons like land use change, their role shifts from being carbon sinks to becoming carbon sources (Gibbs et al., 2007). About a third of total carbon emissions in the past 250 years can be attributed to anthropogenic deforestation, forest degradation and land conversions activities (Smith et al., 2014). A fundamental aspect of comprehending the global carbon cycle involves understanding the storage of carbon within vegetation as carbon sinks. (Falkowski et al., 2000).

Accurate assessment of the changes in forest carbon is also essential for international agreements (Goetz & Dubayah, 2011). The Reducing Emissions from Deforestation and Forest Degradation (REDD+) initiative was started by the United Nations Framework Convention on Climate Change (UNFCCC) with the aim to reduce carbon emissions from forests and instead convert them in effective sinks (UNEP 2018; Herold et al., 2019). Within this initiative, partner countries are also expected to engage in monitoring, measuring, and reporting on their forest resources (UNEP 2018). REDD+ entails measuring changes in forest extent, evaluating forest quality and quantification of carbon stocks.

In view of Indonesia, the contribution of the forestry sector is potentially immense in achieving Indonesian government's target of reducing overall greenhouse gas emissions by 26% (Verchot et al., 2010). The implementation of REDD+ in Indonesia faces difficulties such as consistently monitoring and collecting information on the status of forests. It also highlights the significance of developing a carbon inventory system to accurately monitor carbon emissions and reduction (Verchot et al., 2010). In addition, there is an ever-increasing demand for oil palm plantations globally. However, oil palm plantations are considered to harm the ecosystem and environment where they are planted (Shigetomi, Ishimura, & Yamamoto, 2020). In the last 25 years, the rapid growth of oil palm plantations has resulted in conversion of about 3 million hectares of tropical forests in Southeast Asia (Miettinen et al., 2016). Large scale oil palm plantations managed by big companies were considered responsible for around 23% of deforestation in Indonesia between 2001 and 2016 (Austin et al., 2019). While there is a reasonable understanding of the carbon emissions produced from the soil organic matter oxidation after converting peat swamp forests to Oil palm plantations, the overall net carbon emission resulting from this conversion over the entire life cycle of the plantation is still not well-defined or accurately measured (Tonks et al., 2017).

The aboveground biomass (AGB) is a reliable proxy of the carbon stored in a forest and is utilized for monitoring forests widely (Larocque et al., 2014). It consists of all organic matter above the soil surface of a tree like stem, branches, trunk etc. (Ravindranath & Ostwald, 2007). AGB is hence an important climate indicator in this regard (Herold et al., 2019). Approximately half of Indonesia's land cover is forest, which stores about 60 Gt of carbon, out of which 6 Gt is stored in the form of aboveground biomass (FAO 2009). In addition to carbon, the vegetation biomass is intricately linked

to several vital ecosystem processes, such as the cycling of water (Reichstein & Carvalhais, 2019). Hence, accurately estimating AGB is a requirement for development of effective forest management strategies (Urbazaev et al., 2016). AGB can be estimated through on-field measurements or by remote sensing techniques (UNEP 2018). The most accurate estimation of biomass is through field measurements. Nonetheless, some measurement techniques involve destruction or cutting down of trees, which goes against the principle of REDD+ (Stovall et al., 2017). In contrast to destructive measurements, allometric equations can be used where certain parameters are derived through ground-based measurements of trees in sample plots like Diameter at Breast Height (DBH) and tree height. These parameters are then used to estimate the AGB values of the trees via allometric equations and finally on plot level (Brown, 1997). However, the acquisition of such measurements in the field involves substantial time and labour (Segura & Kanninen, 2005). Remote sensing demonstrates the ability to regularly monitor the Earth's surface and its characteristics over large regions. This makes it an important technique for biomass estimation (Lu et al., 2012).
Many previous studies have used optical imagery for AGB estimation (Li et al., 2020; Urbazaev et al., 2016). However, optical imagery is not well suited for biomass estimation due to reasons like effects of cloud cover, especially in tropical regions. This limits the number of acquisitions available (Asner, 2010; Luckman et al., 2000). Moreover, in regions with high biomass problem of data saturation occurs when using optical imagery (Lu et al., 2012). LiDAR (light detection and ranging) technique is considered to accurately estimate AGB. This is due to its power to capture 3-dimensional information of a landscape accurately (Duncanson et al., 2010). On the other hand, use of LiDAR equipment is expensive, and it can only provide small to moderate spatial coverage (Gibbs et al., 2007).

Synthetic Aperture Radar (SAR) data is widely used for biomass estimation due to its strong correlation with AGB and advantages like resilience to weather impacts and ability to penetrate vegetation structures (Le Toan et al., 2011). As AGB increases, the backscatter, which measures the radar signal strength also increases until it reaches a saturation point (Mitchard et al., 2009a). Generally, linear, or logarithmic models have been used to describe the relationship between SAR backscatter and AGB (Yu & Saatchi, 2016; Schlund et al., 2018). Vegetation mainly causes volume scattering, where multiple scattering often changes the polarisation of the emitted and received signal (Vreugdenhil et al., 2020). Hence, cross-polarisation bands like Horizontal transmit, Vertical receive (HV) and Vertical transmit, Horizontal receive (VH) are considered generally more sensitive to vegetation. Longer wavelength radars (like L- and P-band) can derive more vertical information of vegetation due to their potential penetration into the vegetation and are considered more suitable for measuring higher biomass, whereas shorter wavelengths are considered more appropriate for measuring lower biomass (Patenaude et al., 2005). Shorter wavelengths like C- and X-band could also have benefits for AGB estimation and monitoring. One of the key benefits of utilizing Sentinel-1 images lies in their short revisit time, where the frequency utilized by the Sentinel-1 mission of the European Space Agency (ESA) is in the C-band. Additionally, these images are available free of cost presenting a user-friendly and cost-effective option (Argamosa et al., 2018). X-band SAR particularly interacts more with the vegetation characterized by small leaf and branch components, along with slender trunks (Mitchard et al., 2009b). Hence, X-band SAR has also proved to be promising in mapping lower biomass regions (Solberg et al., 2010).

Some studies have previously used a combination of radar bands for AGB retrieval (Englhart et al., 2011; Schlund & Davidson, 2018; Sivasankar et al., 2018). Naidoo et al. (2015), used individual X-, C- and L-band SAR as well as their combination for modelling various structural parameters such as

woody AGB in African savannah. They discovered that a combination of all three bands had the highest accuracy in AGB estimation, although this is not most convenient. Secondary to this, using L-band alone showed a better estimation than using both X- and C-band individually and in combination. A study by Koh & Wilcove (2008) quantified the extent of primary forests converted to oil palm plantations, but quantification of the carbon emissions due to conversion from primary forests to oil palm plantations has not been performed. Another study by Asari et al. (2017) used L-band SAR acquisitions to map the AGB of oil palm plantations and forests in Malaysia. Although this was done accurately for forests, they were unable to obtain similar accurate retrievals for oil palm plantations.

The X-band experiences pronounced scattering when it interacts with the leaves, which proves to be relevant in getting information about the outer layer of trees. On the other hand, C-band interacts mostly with branch elements and L-band with the trunk of the tree (Roy et al., 2021). Hence, the integration of two or three radar frequencies can enhance the understanding about different terrain features as different frequencies are sensitive to distinct surface features (Notarnicola et al., 2007). Schmullius & Evans (2010), implied that combining high frequency SAR (X- or C-band) with lower frequencies (L- or P-band) is essential for monitoring different parameters across forestry. Hence, in the complex landscape of Indonesian forests which consist of both high biomass forest regions and low biomass plantations, it could be assumed that combining different SAR frequencies may lead to more accurate AGB estimation.

1.2. Problem statement

To fulfil Indonesian government's target of greenhouse gas reduction and achieving REDD+ goals, it is imperative to establish a suitable forest inventory system for Indonesia. Monitoring forest resources helps to track the carbon stock changes at a smaller scale. This, in turn, facilitates the assessment of advancements in local policies and lays the groundwork for forthcoming regulations.

The limitations of using a single SAR band have been highlighted in the preceding section. The literature review suggests that some studies achieved a higher accuracy when a combination of bands was used. Furthermore, using only the L-band wasn't effective in measuring the amount of AGB in land areas with lower vegetation like oil palm plantations. (Asari et al. ,2017). It is worth noting that these studies have been conducted in varying landscapes (Naidoo et al. ,2015; Englhart et al., 2011; Schlund & Davidson, 2018; Sivasankar et al., 2018). The synergistic capabilities of shorter and longer SAR frequency sensor technologies within the diverse and intricate tropical landscapes of Indonesia including forests and cash crops like rubber and oil palm have not been evaluated.

Therefore, this study aims to overcome the limitations of using SAR bands individually. This will be done by analysing the accuracy of the AGB retrieval for the landscape in Indonesia with different combinations of SAR X-, C- and L-bands as compared to individual band retrievals. The insights from this study can help to understand how the choice of SAR wavelength (X-, C-, or L-band) impact the accuracy of AGB estimation in different land cover types within the Jambi landscape. Moreover, it can assess whether the combination of SAR bands improves AGB estimation compared to using individual bands. In this process, the method with highest accuracy for AGB retrieval of the whole region as well as for oil palm plantations will be identified.

1.3. Objectives, Research questions and Hypothesis

The aim of this study is to use individual Synthetic Aperture Radar (SAR) X-, C- and L-bands and their combinations for AGB estimation in the complex landscape of Jambi, Indonesia. Additionally, this study will encompass AGB estimation specifically for both oil palm plantations and forested areas within the area. Furthermore, we will assess the accuracy of these AGB estimations and conduct a comparative analysis of models developed using different bands and their combinations.

Objective 1: To assess the relationship between AGB and backscatter of SAR X-, C- and L-bands and their combinations.
RQ1: For the whole landscape, what is the relationship between backscatter of SAR X-, C- and L-bands and their combinations with the field-measured AGB?
RQ2: Specifically for the oil palm plantations, what is the relationship between backscatter of SAR X-, C- and L-bands and their combinations with the field-measured AGB?
RQ3: Specifically for the forests, what is the relationship between backscatter of SAR X-, C- and L-bands and their combinations with the field-measured AGB?
Hypothesis: - There is a significant logarithmic relationship between AGB and the backscatter values obtained from SAR X-, C-, and L-bands, both individually and in various combinations.

Objective 2: To compare the developed AGB-backscatter relationships between individual SAR bands and land uses.
RQ4: For the whole landscape, do the AGB-backscatter relationships significantly differ across the X-, C-, and L-bands?
RQ5: Do the AGB-backscatter relationships significantly differ between the forests and oil palm plantations across the X-, C-, and L-bands?
Hypothesis: - The AGB-backscatter relationship does not differ for the whole landscape across SAR X-, C- and L-bands and their combinations.
Hypothesis: - The AGB-backscatter relationship does not differ between the forests and oil palm plantations across SAR X-, C- and L-bands and their combinations.

Objective 3: To compare AGB estimation accuracies for X-, C-, and L-bands and their combinations.
RQ6: For the whole landscape, Do the AGB estimation accuracies differ across SAR X-, C-, and L-bands and their combinations?
RQ7: For the Forests, Do the AGB estimation accuracies differ across SAR X-, C-, and L-bands and their combinations?
RQ 8: For the oil palm plantations, Do the AGB estimation accuracies differ across SAR X-, C-, and L-bands and their combinations?
Hypothesis: - The accuracy of Above-Ground Biomass (AGB) estimation will vary between individual SAR bands and their combinations across the entire landscape, for forests and oil palm plantation.

2. STUDY AREA AND DATASETS

2.1. Study area

The forest cover of Sumatra island underwent a significant decline, plummeting from 71% in 1950 to just 30% by 2010. Within Sumatra, the province of Jambi experienced an even more pronounced reduction, with forest loss exceeding 40% from 1990 to 2010 (Margono et al., 2012). For this study, the Indonesian province of Jambi has been selected. It is positioned at the heart of Sumatra island. The study area is situated between 1.233°S to 2.769°S latitude and 101.2°E to 104.733°E longitude (Rustiadi et al., 2022). The province has a land area of 49,000 km². The study is conducted in the Sarolangun Ban and Batang Hari regencies of Jambi (Figure 1). Overall, This region experiences a tropical and humid climate with a consistently moderate temperature year-round, averaging around 26 degree Celsius (Rustiadi et al., 2022). In addition, rainfall is abundant, but there are intermittent dry spells, notably in July. Acrisols are the prevailing soil type in this area. Regarding the vegetation, the study area is mostly covered by tropical rainforest, rubber, oil palm plantations and shrubland. The prevalent plantation type in the region includes Hevea brasiliensis monoculture (rubber) and Elaeis guineensis plantations (oil palm) (Kotowska et al., 2015).
The Jambi province secured the fourth position within Sumatra regarding the overall expanse of oil palm plantations. These plantations are the major source of economic growth in the region. A mix of smallholder plantations, and large-scale company-owned plantations of these cash crops can be found in the study area. Due to the rapid expansion of oil palm plantations, the region has been facing several challenges like illegal logging caused by inconsistent policies (Rustiadi et al., 2022).

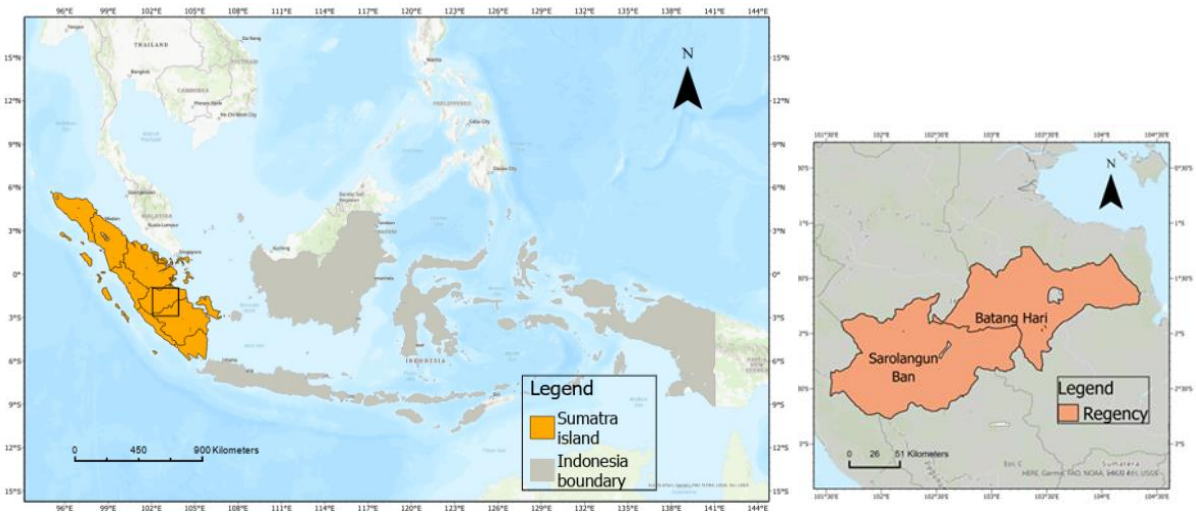


Figure 1- Location of the study area - Sumatra Island within Indonesia and Sarolangun Ban and Batang Hari regencies within Jambi province (right image). (Source: basemap from ESRI, Indonesia administrative subdivision from Bakosurtanal–RSGISforium-Bpk. Suwandito)

2.2. SAOCOM 1A L-band data

SAOCOM 1A (Satélite Argentino de Observación COn Microondas) was launched by Argentina's space agency, Comisión Nacional de Actividades Espaciales (CONAE) on 8 October 2018. It is equipped with a

L-band sensor, and it has been found useful for forest studies (Seppi, López-Martinez, & Joseau, 2022). Three sets of dual-polarized images (HH and HV) were acquired, providing complete coverage of the field plots (Figure 2). However, for this study, only the cross-polarized HV band was employed, as it exhibits greater sensitivity to volume scattering induced by vegetation (Vreugdenhil et al., 2020). The processing level of the images corresponded to Single Look Complex (SLC), and the acquisition mode was StripMap (SM), this corresponded to a ground range and azimuth resolution of 10 m by 5 m, respectively (Table 1). The SLC product consisted of both phase and amplitude information and had undergone radiometric calibration but without any geometric adjustments. The acquisition dates were 25th October 12th and 28th December 2022.

2.3. Sentinel-1A C-band data

Sentinel-1A satellite was launched by the European Space Agency (ESA) in 2014. It is equipped with a C-band SAR sensor. Two images were acquired to cover the field plots in the study area (Figure 2). The dataset used in this study includes two polarization channels known as VV and VH. For this study, just the cross-polarised VH product was utilised (Vreugdenhil et al., 2020). The Ground Range Detected (GRD) product in the Interferometric Wide (IW) swath mode was acquired for the purpose of analysis (Table 1), which corresponds to a pixel spacing of 10 m by 10 m. Moreover, the ground range and azimuth resolution of the image is 20 m by 22 m. GRD data was selected as the phase information is not of relevance to this study and amplitude data is only required. At the GRD processing level, multi looking is performed to account for the speckle in the image. The sensing date was 22nd December 2022. This date corresponds to the dry season in Indonesia, which minimizes potential moisture-related interference with SAR backscatter. Sentinel-1 data with similar parameters has been frequently used in forestry applications (Roy et al., 2021; Zhang et al., 2023b). The data was acquired from European Space Agency's scientific Copernicus data hub.

2.4. TanDEM-X X-band data

TerraSAR-X and TanDEM-X satellites were launched on June 15, 2007, and on June 21, 2010, respectively, as a collaborative effort between the German Aerospace Center (DLR) and Airbus Defence and Space. TerraSAR-X and TanDEM-X fly in close formation to capture interferometric SAR data. This study utilized three TanDEM-X single polarisation (HH) images, which collectively covered the field plots within the designated study area (Figure 2). The processing level and operational mode of the acquired data was Single-look Slant Range Complex (SSC) and StripMap (SM) respectively (Table 1). The products at SSC processing level are not geocoded and exhibit uniform pixel spacing in both azimuth and slant range. The resolution measures approximately 3.3 meters in the acquired images (Airbus, 2015). The SM mode has a higher resolution and narrower swath width than other modes like ScanSAR and Wide ScanSAR, but lower resolution and swath width than the spotlight mode (Airbus, 2015). This provides a good balance between spatial resolution and ground coverage. The TanDEM-X images from 27th February 2018, 9th May 2018, and 25th February 2019 were acquired (Table 1).
The radar images acquired from different satellites for different bands, exhibit distinct characteristics and are accompanied by specific sets of metadata (Table 1).

Table 1- Comparison of the characteristics of the X-, C- and L-band data

Property	**X-band**	**C-band**	**L-band**
Satellite	TanDEM-X	Sentinel-1A	SAOCOM-1A

Observation date	27 February 2018, 9 May 2018, and 25 February 2019	22 December 2022	25 October, 12 and 28 December 2022
Polarization	HH	VV and VH	HH and HV
Orbit direction/PASS	Ascending	Descending	Ascending
Ground range and azimuth	3.3*3.3 m	20*22 m	10*5 m
Swath width	30km	250 km	55 km
wavelength	3.11 cm	5.54 cm	23.51 cm
instrument mode	Strip Map (SM)	Interferometric Wide Swath (IW)	Strip map (SM)
Observation direction	Right	Left	Right
Processing level	Single Look Slant Range Complex (SSC)	Ground Range Detected (GRD)	Single-Look Complex (SLC)

2.5. Field data

Data were gathered in the field from May 2021 to April 2022 during an extensive data collection in the framework of the EFForTS (Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems) project. The field plots were circular shaped with a radius of 17.84 metres, accounting for an area of approximately 1000 m² per plot. Data on all trees with Diameter at Breast Height (DBH) more than or equal to 10 cm were collected. The DBH was measured at 1.3 metres above ground level. The height of the trees was recorded using a Vertex III height meter (Haglöf, Långsele, Sweden). Oil palm height was determined as the distance from the trunk's base to the lowest point of the youngest leaf. A total number of 103 field plots were used in this study, with 27, 30, 25 and 21 plots for oil palm, forests, rubber plantations and shrubland each. The spatial distribution of these plots is displayed in Figure 2.
For the conversion of these biometric information to AGB, allometric equations were applied.
For trees in shrubland, rubber plantations and forest plots, the allometric equation by Chave et al. (2014) was used.

$$\mathrm{AGB} = 0.0673 \times \rho\, \mathrm{D}\, H^{0.976}$$

Where,
AGB = Aboveground Biomass of trees and rubber plantations (in Kg per tree)
ρ = Wood density (in g cm^{-3})
D= Diameter at Breast Height (in cm)
H= Total tree height (in m)

The wood density models employed in the study by Kotowska et al. (2015) were utilized.
An allometric equation used by Khasanah et al. (2015) was applied for AGB estimation of oil palms. This equation equates the oil palm height to its biomass considering the soil type as well.

$$\mathrm{AGB} = 0.0939 \times \mathrm{H} + 0.0951$$

Where,
AGB = Aboveground biomass of oil palm plantations (in Mg)
H = Total tree height (in m)
Lastly, the AGB of all the trees in a plot were added and normalized to tons ha^{-1}.

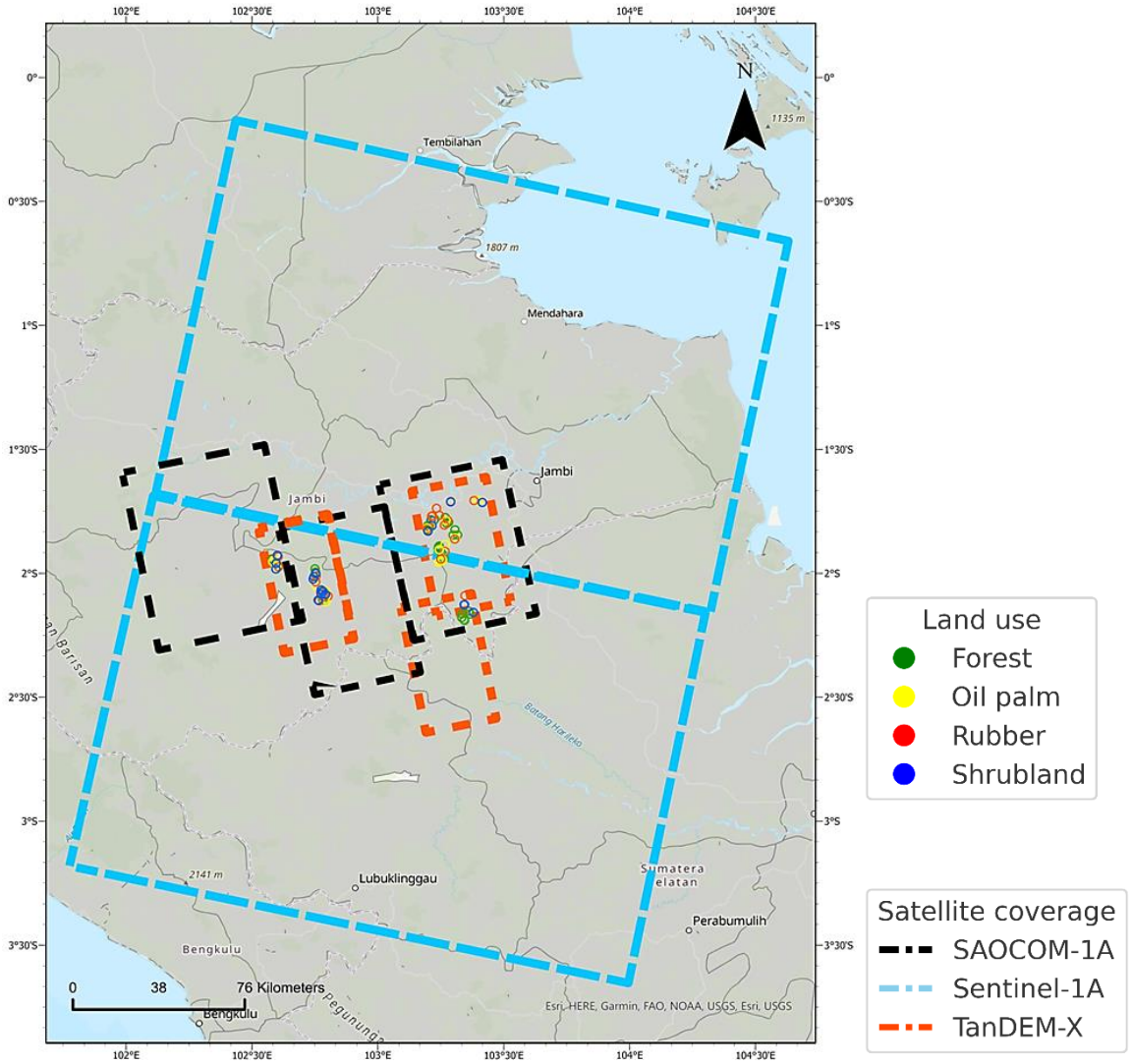


Figure 2 – Coverage of the three satellite images around the field plots. (Basemap from ESRI)

3. METHODOLOGY

3.1. Research methodology

Figure 3 explains the sequence of methods that were followed to fulfil the final objectives and answer the corresponding research questions.

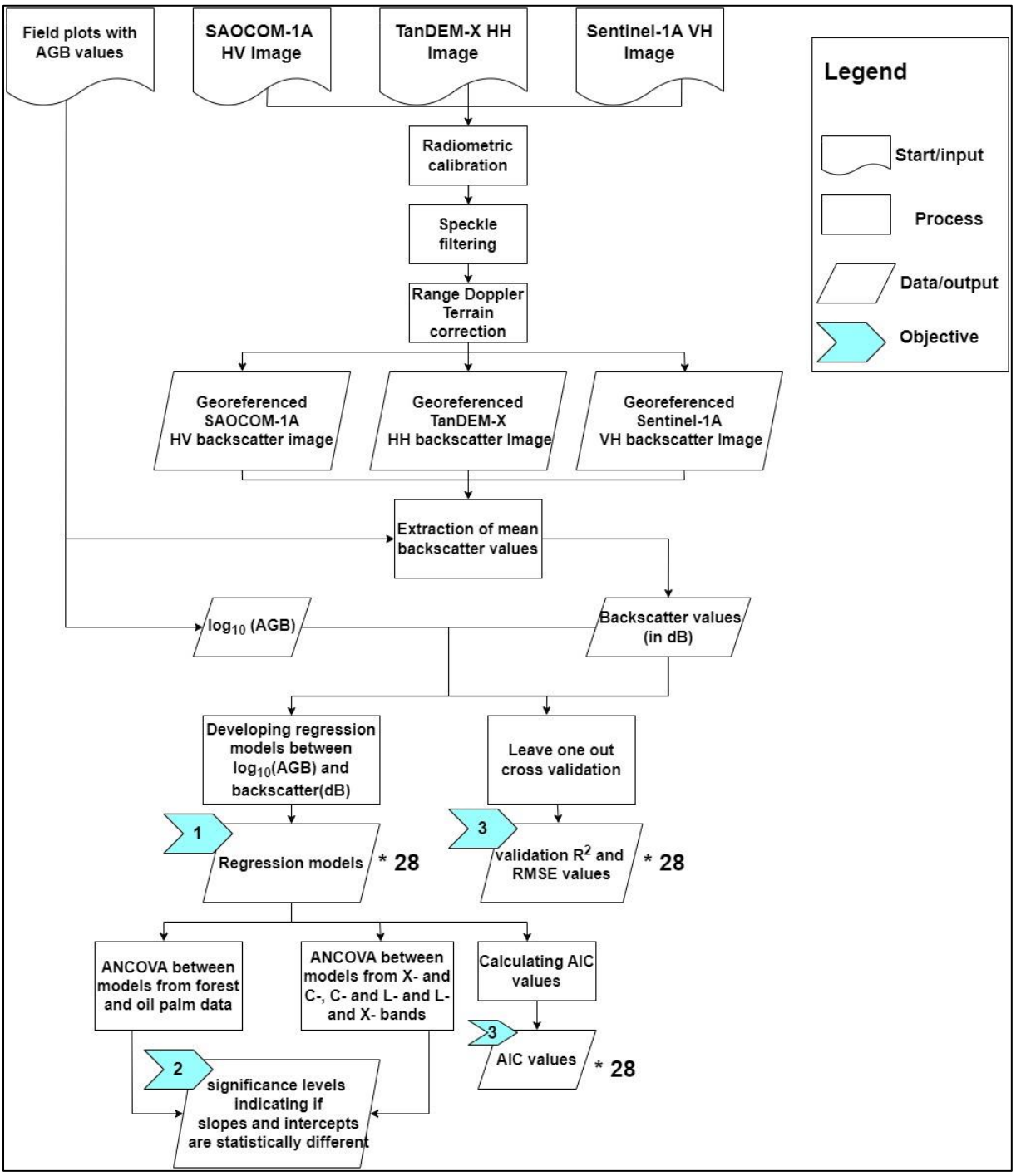


Figure 3- Flowchart of methodology explaining how the objectives were achieved. The same methodology was performed for predicting the AGB of different land use categories and their combinations.

3.2. SAR data pre-processing

The X-, C- and L-band images required certain pre-processing steps, before they could be utilised. As mentioned in section 2.2, only the cross-polarization bands of C- and L-band were utilised. Appendix B also provides a comparison between the models developed with co-polarized bands. This process involved performing a radiometric calibration, speckle filtering and geometric correction to obtain the final geocoded image with backscatter values. The pre-processing was conducted using the Sentinel Application Platform (SNAP) software, developed by the European Space Agency (ESA). SNAP is an open-source software that offers comprehensive SAR preprocessing capabilities for missions such as Sentinel-1, SAOCOM-1A, and TanDEM-X.

As a first step, Radiometric calibration was performed. This converted the raw Digital Number (DN) values of the image into physical units (σ^0), that represented the backscatter in order to perform a quantitative analysis. After calibration, backscatter coefficients for the three different images were obtained.
This step was followed by speckle filtering, which was required to reduce the salt and pepper effect that arises in radar images due to inherent speckle noise. The aim of despeckling is to reduce speckle noise while retaining the details in the image (Painam & Manikandan, 2023). Some available filters in the SNAP software are Enhanced Frost, Mean, Kuan, Median, Frost, Lee, Wiener, and Gamma MAP filter. A kernel of certain window size is applied over the imagery to reduce the noise. For this study, a Lee filter with a window size of 7*7 (Lee, 1980) was used to reduce the noise in the image to make it more interpretable. A 7*7 window size provides a good trade-off between speckle reduction and edge preservation without compromising the spatial resolution of the images with 10*10 metre pixel size (Rubel et al., 2021). The Lee filter employs a spatial filtering technique that is relatively straightforward to implement, thereby minimizing the computational demands involved (Liu et al., 2021).

The images' orientation were distorted, a result of how the sensor acquires the data. To rectify this, a geocoding process was performed. The topographic variations of the earth's surface need to be accounted for, due to which we used a Range-Doppler Terrain correction in SNAP. Shuttle Radar Topography Mission (SRTM) data provides accurate topographic details, which plays a crucial role in geometric correction. The SRTM 3-sec DEM which is directly downloadable via SNAP was used for this purpose. In this step, the images were also resampled to a pixel spacing of 10 by 10 metres. This corresponds to the coarsest pixel resolution present in the three imageries. This was done to prevent oversampling of the Sentinel-1 data. Having the same pixel spacing for all images was important for consistency of analysis.
For the SAOCOM-1A images, a manual georeferencing was also conducted. This was done keeping in consideration the geo-location uncertainty of the satellite, which is about $\pm$70 metres. A visual inspection was done to identify any potential geo-location errors and it was found that the images were off by a few pixels. This potentially meant that the field plots were located in a completely different location on the imagery than on the ground. As a consequence, ground control points were collected on the software QGIS using a very high resolution base imagery from google satellite images Polynomial degree 3 transformation was used along with a nearest neighbour resampling. Further, the raster georeferencer was utilised for georeferencing with a mean error between 1.5-3 metres for all three images.

3.3. Backscatter extraction

The backscatter values were extracted for all the field plot polygons using the zonal statistics tool in QGIS. The diameter of the circular field plots was 35.68 metres. It was difficult to extract backscatter information on a plot which was circular, as the pixels of the image were square. Considering the resolution of the images to be 10 metres, a pixel window of 4*4 was created to fit around the plots. The mean value of backscatter coefficients was extracted to ensure consistency in case of noisy data. After the extraction of the mean backscatter values that were in a linear scale, the SAR data was converted to Decibels (dB).

$$\sigma^0 = 10 * log_{10}(\sigma^0[linear])$$

Where,
σ^0 = Backscatter (in dB)
$\sigma^0[linear]$ = Backscatter (in linear units)

3.4. Developing regression model

The next step involved deriving regression models using backscatter values obtained from all field plots. Parametric regression models have been extensively used for AGB estimation (Lu et al., 2012; Schlund et al., 2018). These models were developed separately for single, combination of two and all bands (as depicted in Figure 4a, for the 4 land use classes, for data from just forests and oil palm plantations as well as for two classes separately - oil palm plantations and forests (Figure 4b). This added up to a total of 28 models.

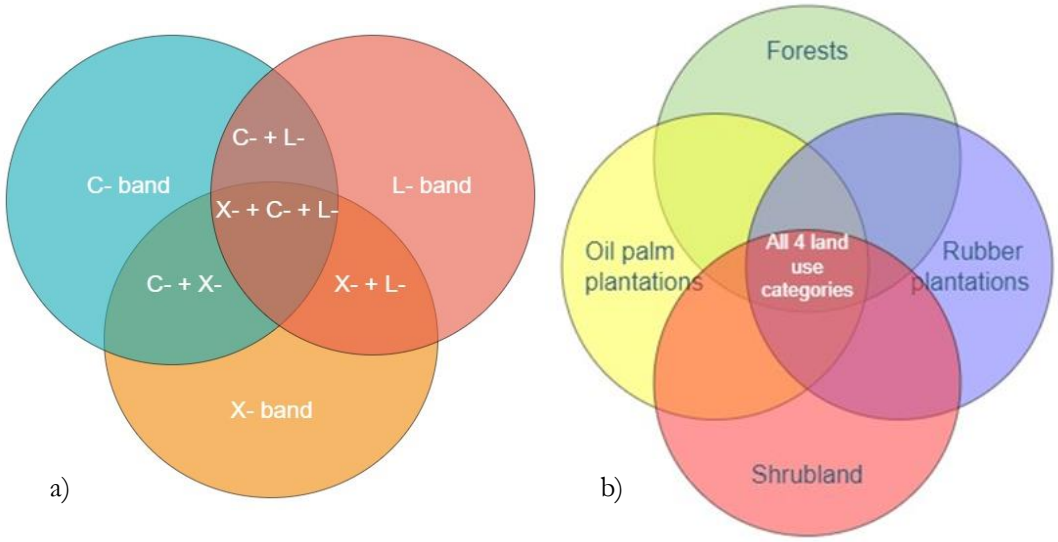


Figure 4- a) Venn diagram representing the bands and their combinations which will be used as the independent variable. b) Representing the different land uses and their combinations present in the study.

All the 103 data points were used for developing these regression models for all 4 land use categories, oil palm plantations and forests. A log-log relationship between AGB and SAR backscatter has been extensively utilized (Schlund et al., 2018). A linear regression was applied to establish a log-log relationship between the $\log_{10}$(AGB) and backscatter coefficients (in dB), resulting in a forward model. This forward model was inverted to derive AGB values by computing the antilog, a process commonly known as the backward model. For creating models for combinations of bands, multiple linear regression was applied, with the backscatter of bands as the independent variables. The equation corresponding to the linear regression model is provided below.

$$\log_{10}(AGB) = \beta_0 \sigma^0 + \beta_1$$

Where,
$\log_{10}$ = Logarithmic value of AGB
β_0 = Model coefficient for the corresponding backscatter
σ^0 = Backscatter coefficient (in dB)

β_1 = Intercept

3.5. Comparison of models

A comparison of the slopes (β_0) and intercepts (β_1) of the created models was done to understand if the established relationships between AGB and backscatter for different land use categories was significantly different. An Analysis of Covariance (ANCOVA) is a suitable method for this purpose. It is used in cases where one independent variable is categorical (covariate), and the other one is continuous (factor) (Khammar et al., 2020). It measures the impact of the categorical variable, across 2 categories, on the relationship between the factor and the dependent variable (Faraway, 2006, Khammar et al., 2020). An interaction term was introduced in the created model. This variable represents the combined effect of two variables on the dependent variable and is crucial for assessing whether the choice of radar band or land use category influences the relationship between AGB and backscatter. P-values were calculated to evaluate the null hypothesis that the interaction terms have no effect. Commonly used significance levels of 0.05, 0.01, and 0.001 to interpret the strength of evidence against the null hypothesis were applied. For instance, a p-value less than 0.05 indicates statistical significance at the 95% confidence level, while a p-value less than 0.01 corresponds to a significance level of 99%, and a p-value less than 0.001 indicates significance at the 99.9% confidence level. These thresholds guide the interpretation of the statistical significance of the interaction terms in our analysis.

A significant p-value indicated strong evidence against the null hypothesis, suggesting a statistically significant interaction between factors. For example, a p-value less than 0.05 indicates less than a 5% probability that the observed interaction could occur if the null hypothesis were true, leading us to reject the null hypothesis in favour of the alternative—that there is a meaningful interaction affecting the relationship. Figure 5 explains the sequence of process with the example of comparison of two particular bands. The same steps were followed for other combinations of bands and the two land use categories.

H_0– There is no statistically significant difference between the slopes of the two bands or land uses.
H_1- There is a statistically significant difference between the slopes of the two bands or land uses.

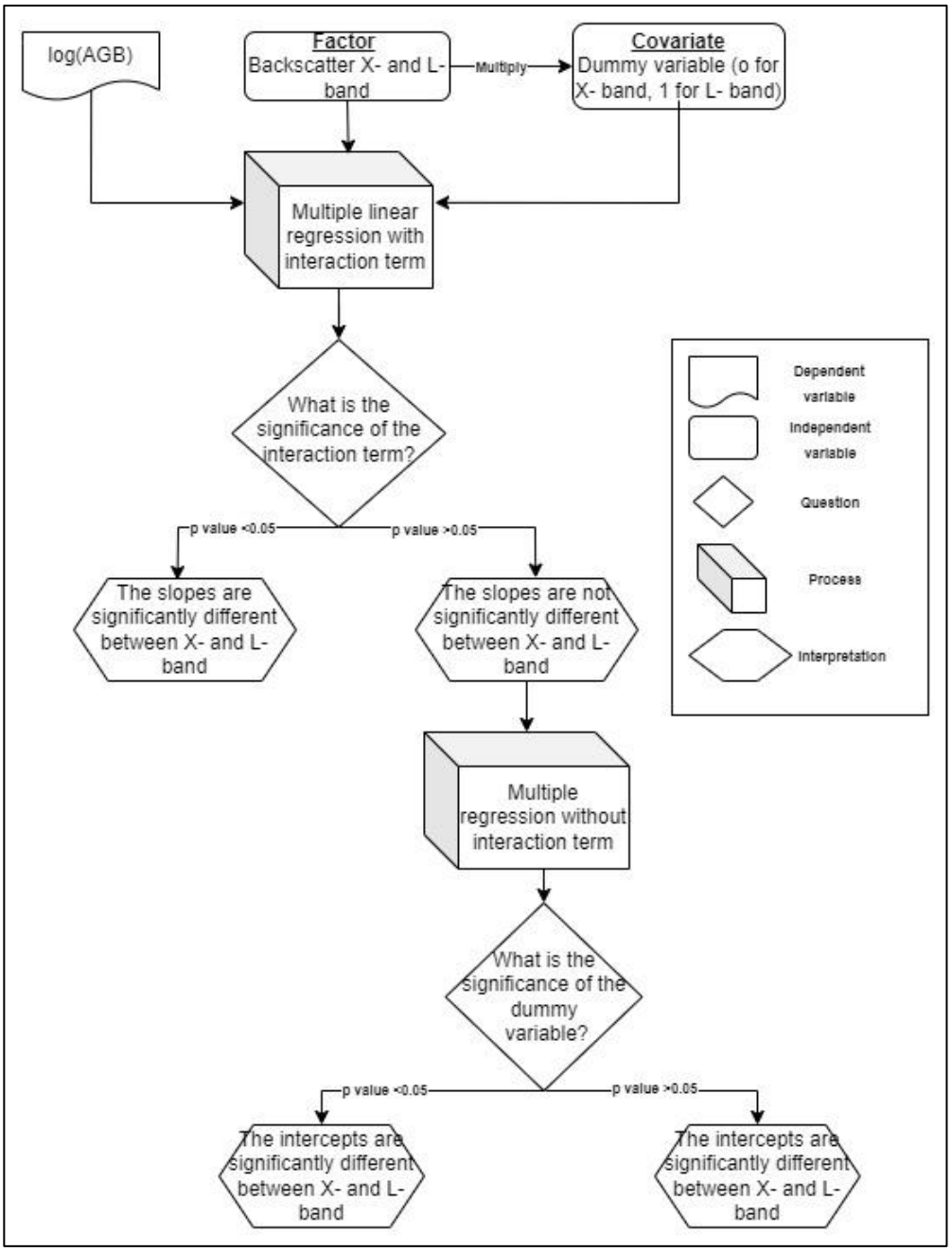


Figure 5- Methodological steps followed for performing ANCOVA.

If the interaction was significant between the factor and covariate, it was interpreted that the slopes of the bands/land uses were significantly different (Figure5). For such models, ANCOVA could not be used to test the difference in intercepts. If the interaction term was not significant, it was interpreted that the slopes of the model between different bands/land uses is not significantly different. For models like these, a simpler model without the interaction term was created as seen in Figure 5. The significance level of the covariate was tested and if it was significant, it was interpreted that the intercepts between the two categories were significantly different. On the other hand, an insignificant covariate meant that the intercepts were not significantly different between the two categories.

Firstly, to find out if the models created using different bands had significantly different slopes across forests, oil palms and the 4 land use categories, an additional categorical variable was added representing the

presence or absence of a particular band. Since, in each model there are two bands, a single dummy represented one band, where 1 indicated presence of that band and 0 indicated the absence. This makes the other band as the reference category (Schlund et al., 2018). Table 2 shows the combination of bands between which the differences in slopes and intercepts was tested. In total, twelve pairs of slopes and intercepts were tested. A similar method was used to compare model coefficients across different land use categories, which is forests versus oil palms as can be seen in Table 3. Appendix D shows the same analysis done between different combinations of the four land uses. Land use was treated as a covariate, and the backscatter from one of the bands served as the factor. A similar approach was used as above, where a categorical variable representing the presence or absence of a land use category was introduced. A similar equation was established with the interaction term with the covariate representing the two land use categories. The procedure to establish whether the coefficients were significantly different or not was the same as above.

Table 2- Representing the different bands (covariate) between which the ANCOVA was performed for different land uses and their combinations, with the backscatter as the factor

Combination of bands (covariate)	Land use			
X- + C-	All 4 land use categories	Forests and oil palm plantations	Forests	Oil palm plantations
C- + L-				
X- + L-				

Table 3 – Representing the two land use categories (covariate) between which ANCOVA was performed with the backscatter from different bands as the factor.

Bands	Land use (covariate)
X-	Forests and oil palm plantations
C-	
L--	

The model equation showing the interaction term was as follows.

$$\log_{10}(AGB) = \beta_0 \sigma^0 + \beta_2 BAND + \beta_3 BAND * \sigma^\circ + \beta_1$$

Where,

$\log_{10}(AGB)$ = Logarithmic value of AGB

β_0 = Slope for the factor

σ^0 = Backscatter coefficient (in dB)

β_1 = Intercept

β_2 = Slope for the covariate

BAND = Categorical variable representing presence or absence of a band or land use

$\beta_{3=}$ Slope for the interaction term

3.6. Accuracy assessment

3.6.1. Leave-one out cross validation

Accuracy assessment was performed to assess how well the three single radar bands and their combinations performed in estimating AGB for all 4 land use categories, oil palm plantations and forests on unseen or independent data. For this purpose, a leave-one out cross validation approach was used. This approach is common to use when there is insufficient data for creating a training-validation split (Webb et al., 2010). In this approach, each observation is used for validation. For the first step, all except one observation is used for model training and then the error metric is calculated on the one data point, as can be seen in Figure 6 (Cha et al., 2020). This is continued till all the observations have been used for validation. An average for the error metric will be taken as the final value. This provides the validation accuracy of the models.

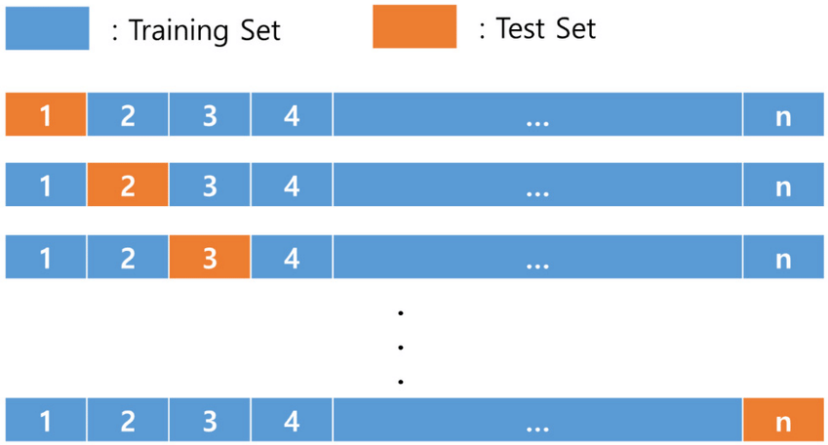


Figure 6- Leave-one out cross validation (Cha et al., 2020)

In this case, there were 103 data points, whereas individually within the land use category oil palm plantations and forests, there are 27 and 30 observations respectively. This was insufficient to create a training-validation split. The performance of 21 models using the coefficient of determination (R^2) and Root Mean Square Error (RMSE) was evaluated. This metric provided insights into how well the models perform on unseen data. Hence, the R^2 and RMSE values for 28 models was obtained. The logarithmic value of AGB was used as the dependent variable. Then, for each iteration the prediction in logarithmic value was converted through an antilog to obtain the prediction in tons/ha. After this, the RMSE values in the original AGB units from all iterations were averaged to obtain the absolute RMSE in tons/ha. A challenge was to compare the RMSE values of different land use types with each other. This was because the range of AGB values for all land use categories were different as can be seen in the box plot in Figure 7. An approach for the comparison of the RMSE values across land use categories was developed to rule out this bias. The relative RMSE was calculated by firstly computing the standard deviation of the RMSE values across all iterations and then dividing the average RMSE of the model with the standard deviation (Moriasi et al., 2007). This ensured that the RMSE is scaled according to the variability of the AGB values of that land use category. The relative RMSE values less than 1 would indicate that the model's average prediction error is less than the variability of the data (standard deviation). Whereas values greater than 1 would suggest that the model's average prediction error exceeds the variability of the data. A model with lower relative and absolute RMSE indicated a better fit between the predicted and actual values.

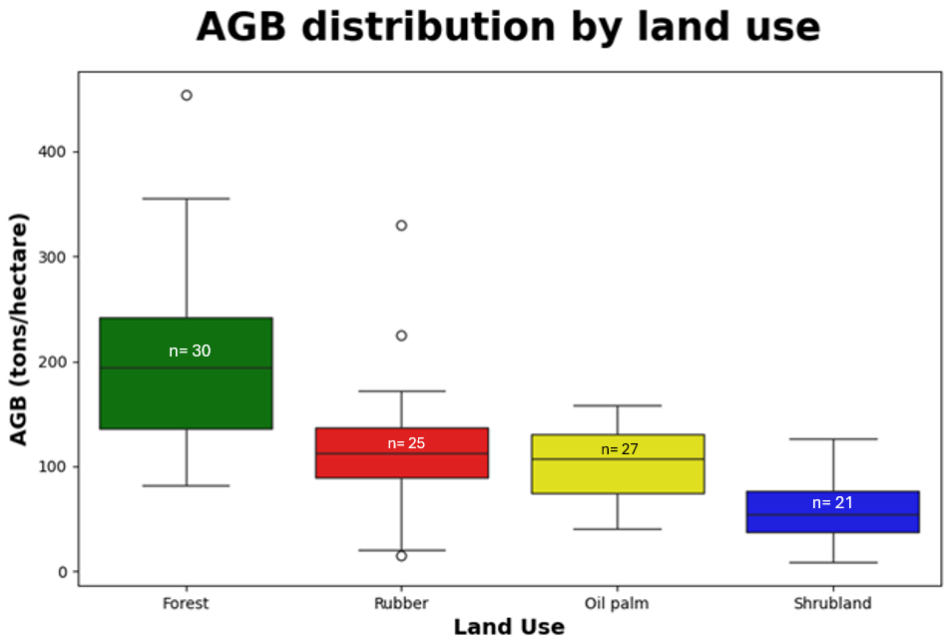


Figure 7- Box plot showing the range of AGB values across different land use categories.

Moreover, R^2 could not be calculated for each iteration as there was only one data point. Hence, the variability could not be explained by using just one data point. Therefore, all individual predictions across all iterations were compared to the true values to calculate the R^2. We compared these performance metrics of our models between X-, C- and L-bands and their combinations for the 4 land use categories. Further, this comparison was done between the accuracies obtained for the developed models of oil palm plantations and forests. A model with higher R^2 value meant that the particular band or band combination was able to explain more variability in AGB than a model with lower R^2 value.

The training R^2 explains how well the model explains the variance of the dataset it was used to be created. This can be overly optimistic and hence validation R^2 was used to assess the performance of the model on unseen data. Comparing the training and validation R^2 can give an indication if the model is overfitting. Overfitting occurs when a model performs well on the data it was trained on but not on unseen independent data.

3.6.2. Akaike Information Criterion (AIC)

For the purpose of evaluation of the Goodness of Fit (GoF) in regard to the complexity of the model, several methods could be used. One of the most common measures used for this purpose is Akaike Information Criterion (AIC). AIC provides an assessment of how well the developed model fits on the training data. AIC provides a trade-off between fit of data and complexity of the model (Kuha, 2004). A penalization is introduced for models with more parameters. This is essential to compare the models developed by single, dual, and triple bands. A model using all three radar bands might explain the variation in AGB better than a model using a single band, but at the same time it increases the complexity. In this case, the complexity is related to the acquisition cost, computational power, and time. In terms of comparison of AIC values, differences of more than 2 are considered evidence of a better fit, and differences of ten or more are considered strong evidence (Burnham and Anderson, 2004). AIC values can be compared for models fitted on the same dataset, with lower values representing a better fitted model while balancing the complexity. Hence, to compare the AGB-backscatter relationships developed using single bands, combination of two bands and all three bands, AIC was used. A lower AIC meant the model was a better fit to the provided data than a model with a higher AIC.

4. RESULTS

4.1. Results from regression analysis

The model parameters established as a result of the regression analysis between log (AGB) as the dependent variable and backscatter of the three different radar bands and their combinations as independent variables can be seen in Table 4. The significance levels of the model intercepts and slopes along with the training accuracies is also mentioned in Table 4.

Table 4- The training R^2 and model coefficients with significance levels marked as * (lightest blue) for p-values < 0.05, ** (light blue) for < 0.01, and *** (dark blue) for < 0.001. Coefficients without any * symbol are not statistically significant.

Band/band combinations	Land use	Intercept	Slope X-	Slope C-	Slope L-	Training R^2
X-	Rubber, Shrubland, Oil palm and Forest	2.28***	0.04**			0.07
C-		2.38***		0.02		0.01
L-		2.14***			0.01	0.003
X- + C-		2.44 ***	0.03*	0.01		0.07
C- + L-		2.40***		0.02	0.001	0.01
L- + X-		2.47***	0.04**		0.01	0.07
X- + C- + L-		2.50***	0.04*	0.003	0.01	0.07
X-	Forests and oil palm plantations	2.38***	0.03*			0.10
C-		3.24***		0.08***		0.22
L-		3.28***			0.08***	0.33
X- + C-		3.22***	0.02	0.07**		0.23
C- + L-		3.51***		0.03	0.06**	0.36
L- + X-		3.45***	0.03*		0.07**	0.4
X- + C- + L-		3.51***	0.03*	0.01	0.07***	0.4
X-	Oil palm plantations	2.35***	0.05***			0.41
C-		2.74***		0.05		0.05
L-		2.16***			0.01	0.005
X- + C-		1.75**	0.06***	-0.04		0.44
C- + L-		2.81**		0.05	0.005	0.05
L- + X-		2.34***	0.05***		-0.0002	0.41
X- + C- + L-		1.79*	0.06***	-0.04	0.003	0.44
X-	Forests	2.20***	-0.009			0.007
C-		2.65***		0.03		0.05
L-		2.97***			0.05*	0.19

X- + C-		2.63***	-0.02	0.04		0.09
C- + L-		3.00***		0.003	0.05*	0.20
L- + X-		2.92***	-0.006		0.05*	0.20
X- + C- + L-		2.96***	-0.01	0.009	0.046	0.20

Among the models predicting AGB for all four land use categories, those utilizing solely the X-band exhibited the highest statistical significance, with p-values less than 0.01 (Figure 8). Additionally, models combining the X-band with both the C- and L-bands, as well as models incorporating all three bands, also yielded significant results, with p-values less than 0.05. All models predicting the AGB of combined forest and oil palm data exhibited statistical significance, with p-values below 0.001 across all models, except for the model utilizing solely the X-band as the predictor (Figure 8).

The significance of models predicting the AGB of forests using either the L-band alone or it's combination with X- and C-bands was confirmed, with p-values less than 0.05 for all three models. This indicates strong statistical evidence against the null hypothesis, suggesting that these models are statistically significant predictors of forest AGB (Figure 8).

Similarly, for models created for AGB of oil palm plantations, models incorporating just the X-band and also its combination with the C-band had highly significant p-values below 0.001, indicating strong evidence against the null hypothesis. Additionally, models including the X-band with the L-band and all three bands together also showed a strong significance, with p-values below 0.01. Overall, all the models produced for different land use categories were significant when using a combination of X- and L-band as predictors.

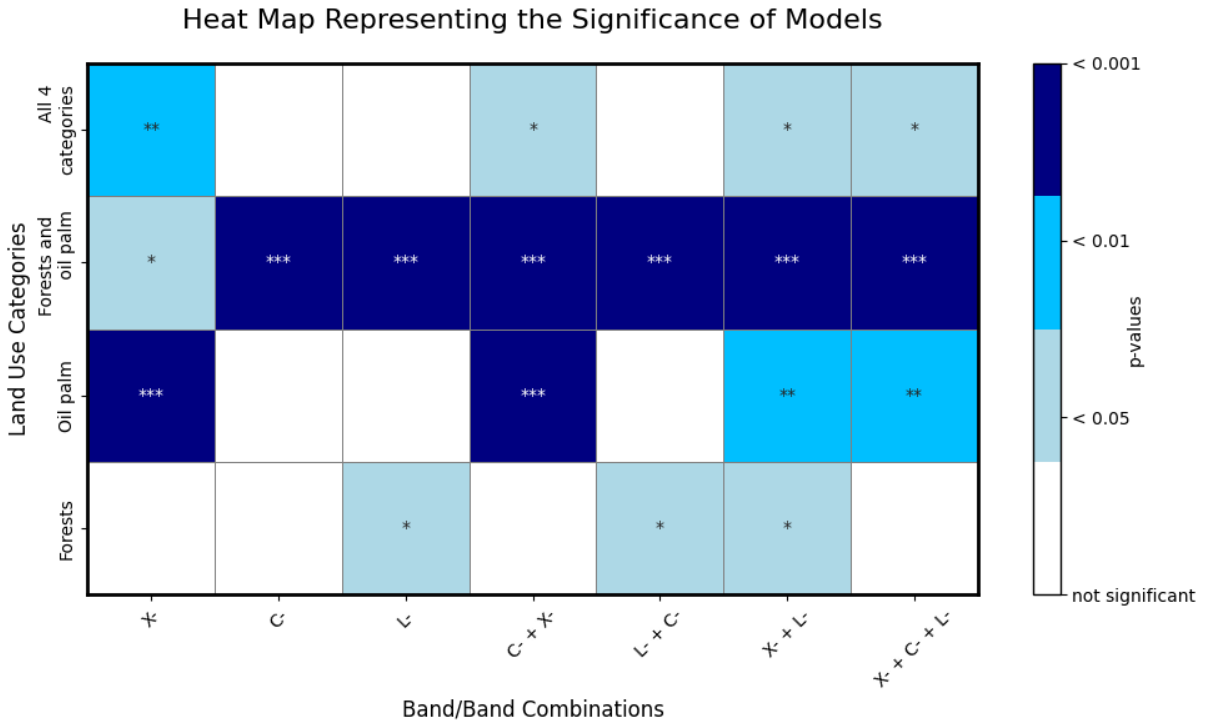


Figure 8- Heat map visualising the significance levels of various models created with different bands and their combinations for different land use categories. For forests, all models with L-band and for oil palm plantations all models with X- band have a significant slope.

In statistical analysis of the models developed for the four land use categories, the R^2 for the training data was consistently below 0.1 across all bands and their combinations (Table 4). This indicated a generally poor fit of the models to the training data of all 4 land use types. Figure 9 provides a comparison between the regression models developed using single C- band VH, L-band HV and X- band HH backscatter. Out of the individual three bands, X- band had the highest R^2 of 0.07. L-band had the least training R^2 among the three individual bands and also among the combination of bands. X-band was consistently the only significant predictor, with p-values indicating statistical significance (Table 4).

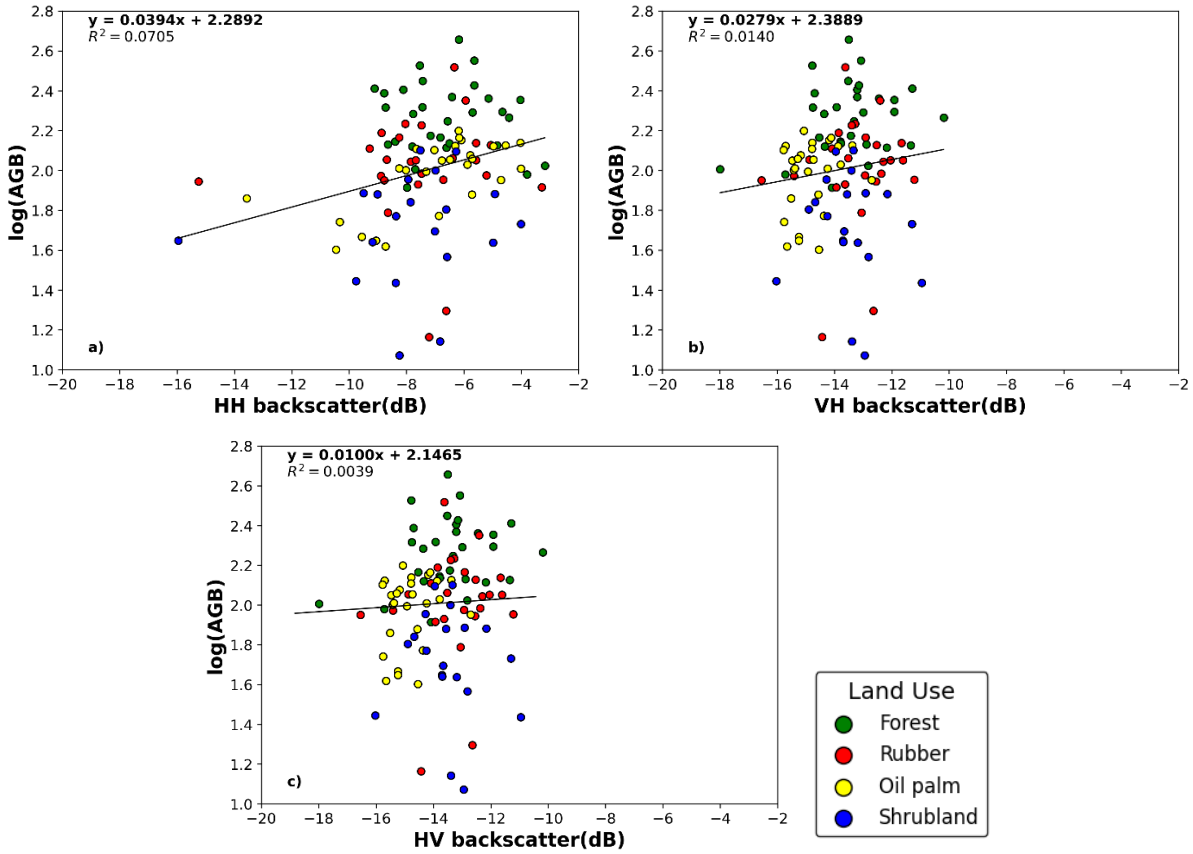


Figure 9- Regression plots showing relationship between log (AGB) of all four land use categories and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter.

The results of linear regression performed using just the data of the forests and oil palm land use category showed improvement in some bands than using the data of all four land use categories together as can be seen in Table 4. The highest R^2 of 0.40 was achieved when using all the three bands together and also when using a combination of X- and L- band. On the other hand, the lowest R^2 of 0.09 was achieved when using just the X- band. Figure 10 provides a comparison between the regression using the three bands individually. L-band had the highest R^2 of 0.33 whereas X- band had the lowest corresponding to 0.09. L-band was a significant predictor in all the models that were created with it either individually or with a combination with other bands with p values below 0.001 and 0.01. In models utilizing X-band exclusively, as well as in combination with L-band and a combination of all three bands, X-band exhibited notable significance. Moreover, C-band demonstrated statistical significance as a predictor in models using solely C-band backscatter and in models incorporating both C-band and X-band data. Hence, when using the three bands individually, all three bands were established as significant predictors.

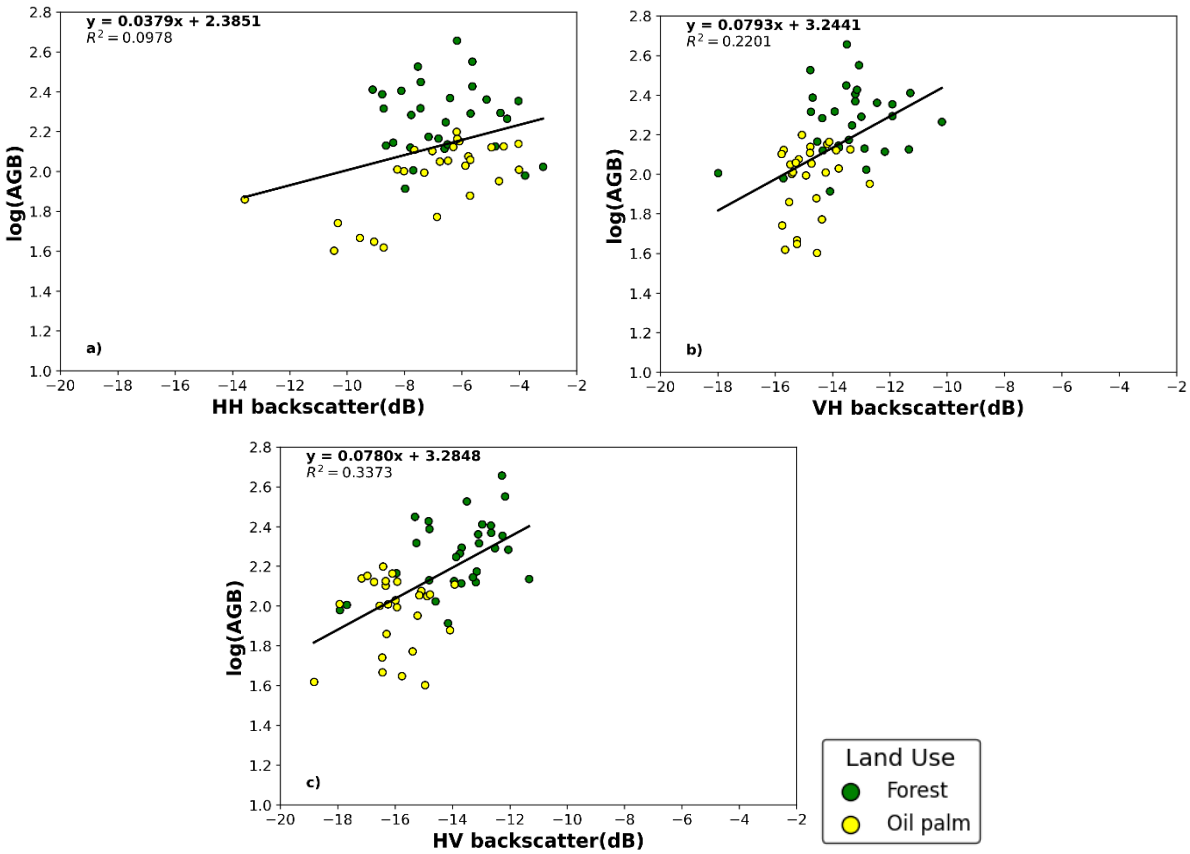


Figure 10- Regression plots showing relationship between log (AGB) of oil palm plantations and forests and a) X-band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter

In the specific case of oil palm plantations, the model performance varied more significantly across different radar bands. The highest training R^2 achieved was 0.44, utilizing backscatter data from all three bands (X-, C-, and L-band) and separately from a combination of just the X- and C-bands. Conversely, the use of solely the L-band yielded a notably low R^2 of 0.005. A comparison between the models created with the three bands individually can be seen in Figure 11. Within the three single bands, X- band yielded the highest R^2 of 0.41. Statistical analysis identified the X-band as a consistently significant predictor for estimating oil palm plantation AGB (Table 4).

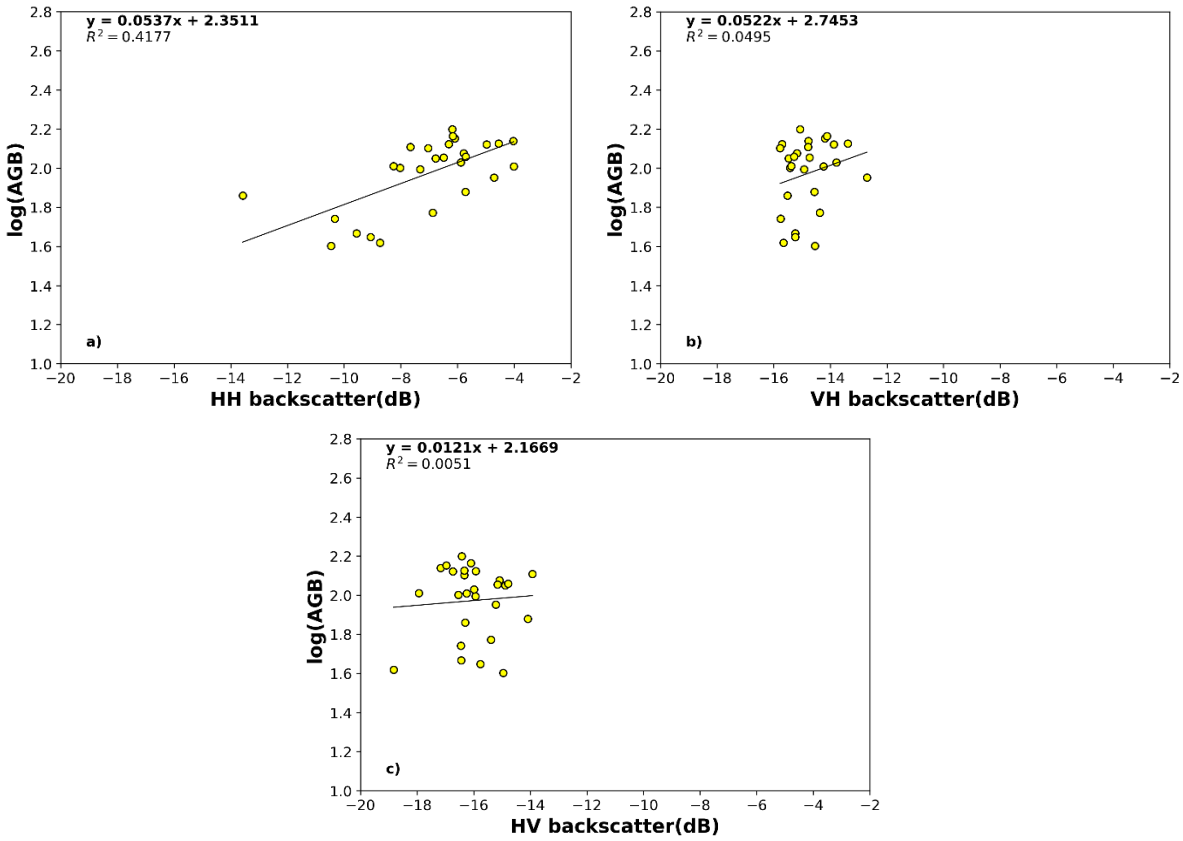


Figure 11- Regression plots showing relationship between log (AGB) of oil palm plantations and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter

Regarding forest land cover, the optimal model performance was an R^2 of 0.20. This highest R^2 was observed in models using combined backscatter data from X- and L-bands, L- and C-bands, as well as from a combination of all three bands. From Figure 12, it can be concluded that within the three individual bands, the highest accuracy of 0.19 was obtained with L-band backscatter. This means that 19% of variability in forest AGB can be explained with the help of L- band. A negative slope is seen between X- band backscatter and forest AGB along with the lowest R^2 of 0.007. The L-band backscatter was identified as the only statistically significant predictor in these models, with its significance demonstrated by a p-value less than 0.05 (Table 4).

From the analysis presented in Table 4, it is evident that the intercepts for all models were statistically significant. In these models, the intercept represents log (AGB) when backscatter from the radar bands is zero. The significance of the intercept indicates that AGB is detected even in the absence of backscatter. The p-values for the intercepts vary, falling below 0.05, 0.01, and even 0.001, demonstrating varying degrees of strong statistical evidence against the null hypothesis at these thresholds. This suggests a significant baseline level of AGB across different models.

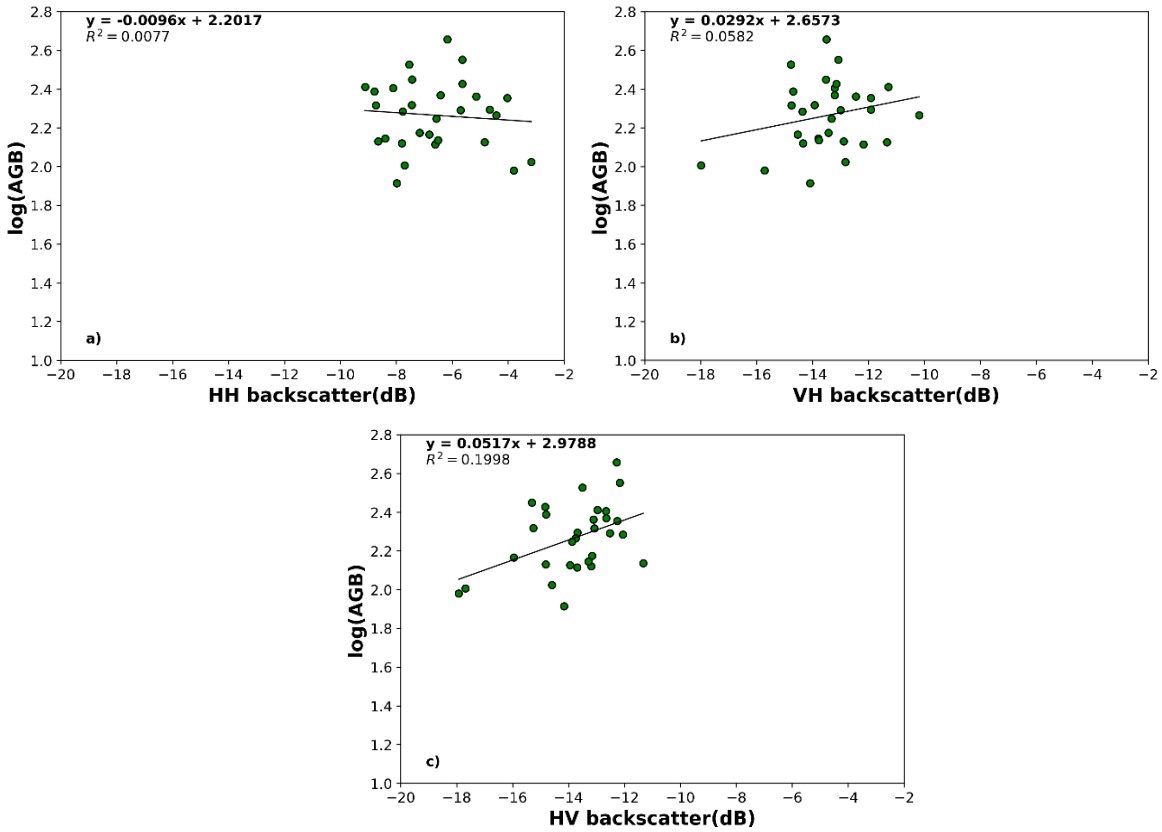


Figure 12- Regression plots showing relationship between log (AGB) of forests and a) X- band HH backscatter, b) C- band VH backscatter and c) L-band HV backscatter

4.2. Results of the comparison of models

A comparison of the model coefficients using ANCOVA was performed where firstly the model slopes were compared using an interaction term. A significant interaction was noted only between X- band and L-band for forests (Table 5). This indicates a significant difference in the slopes between X- and L- bands for forest data. The interaction term was not significant for other land use categories, indicating no significant difference between the slopes of different bands for these land uses (Table 5).

Table 5- The % confidence that the slopes are statistically different between different bands. Where, n.s indicates no significant difference between slopes, 95% (lightest blue shade) ,99% (light blue) and 99.9% (dark blue) represent the respective confidence that slopes are statistically different from each other for the bands.

Bands	Four land use categories	Forests and oil palms	Oil palm plantations	Forests
X- + C-	n.s	n.s	n.s	n.s
C- + L-	n.s	n.s	n.s	n.s
L- + X-	n.s	n.s	n.s	95%

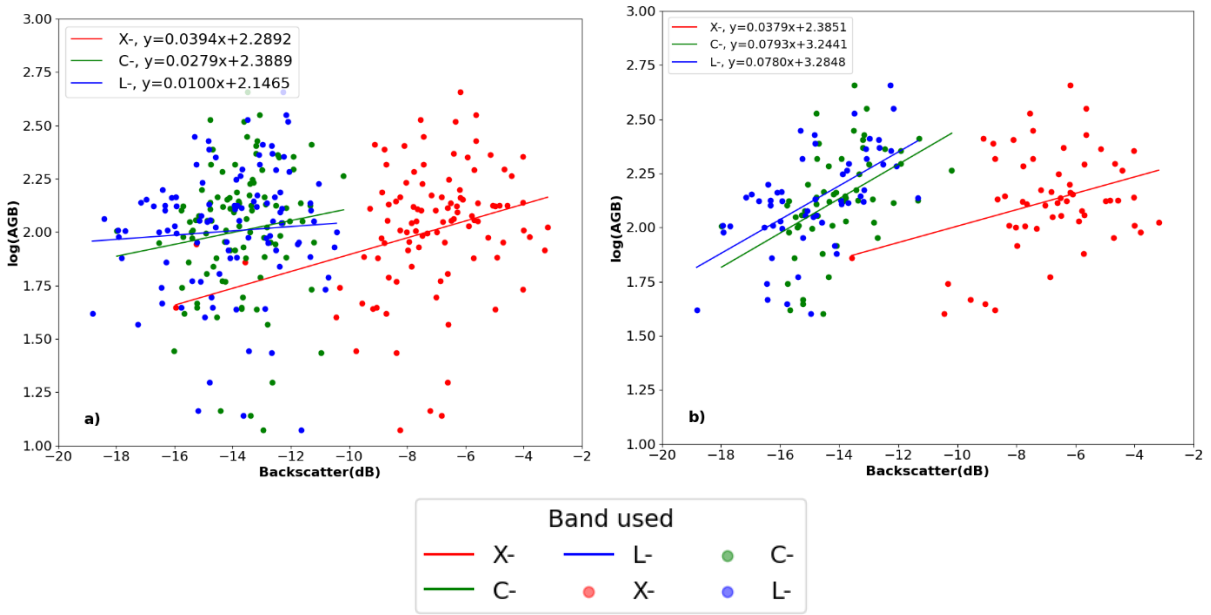
As a next step, for the models which didn't have a significant difference between the slopes, a significant difference for intercepts was checked by assessing the significance of the categorical variable in a model without the interaction term (Table 6). For models where the interaction term was significant, implying a significant difference in slopes, a similar method could not be used. A significantly different intercept was observed for the estimation of oil palm plantation AGB and AGB of all four land use categories between

X- and C- band, and X- and L-band (Table 6). For other models, the categorical variable (representing two bands) was not significant, with p-value more than 0.05. Hence, the null hypothesis was accepted, signifying the intercepts are not significantly different between different bands.

Table 6- The % confidence that the intercepts are statistically different between different bands. Where, n.s means no significant difference between slopes, 95% (lightest blue shade) ,99% (light blue) and 99.9% (dark blue) represent the respective confidence that slopes are statistically different from each other for the bands.

Bands	**Four land use categories**	**Forests and oil palm plantations**	**Oil palm plantations**	**Forests**
X- + C-	99%	n.s	99.9%	n.s
C- + L-	n.s	n.s	n.s	n.s
L- + X-	95%	n.s	99%	-

Figure 13 displays the regression lines generated using different radar bands, facilitating a comparison of their slopes and intercepts. When models were constructed using both forest and oil palm data (Figure 13b), the regression lines for models with the C-band and L-band as independent variables appear visually parallel. From Table 5, it can be seen that statistically these two models have a similar slope. However, for all other land use categories, the regression lines for models created with C-band and L-band backscatter do not overlap and are not parallel. Statistically, as seen from Table 5, there is no significant difference in the slopes of these models. Additionally, the regression lines for X- and L-band models for forests diverge into different directions (Figure 13b). These two models also have statistically different slopes as can be seen from Table 5.



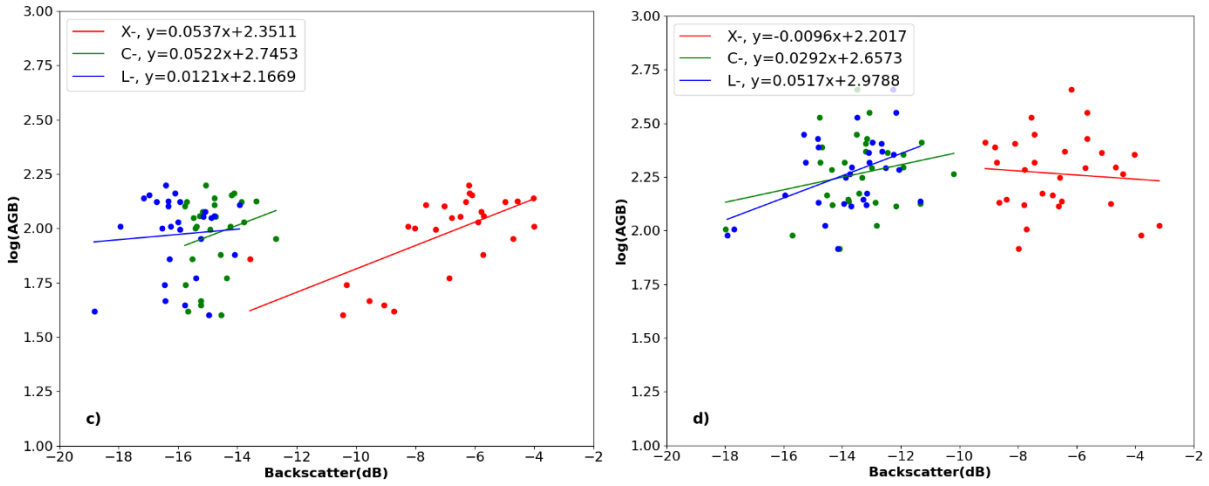


Figure 13- Comparison of the regression lines between models created with X-, C- and L-band backscatter for AGB of a) all 4 land use categories b) forests and oil palm plantations c) oil palm plantations and d) forests.

To examine if the slopes of models created with forests and oil palm plantation data are significantly different, the two land use categories were introduced as the categorical variable. Table 7 outlines the confidence levels associated with the corresponding significance levels for the difference in slopes and intercepts between forests and oil palm plantations. Analysis revealed that only when the model was created with X- band backscatter, there was a significant difference in slopes between forests and oil palm plantation models. Moreover, models constructed with the L- and C-band individually exhibited confidence levels indicating a significant difference in intercepts between forests and oil palm plantations for these models.

Table 7- The % confidence that the slopes and intercepts are statistically different between models created with forests and oil palm plantations data. Where, n.s means no significant difference between slopes, 95% (lightest blue shade) ,99% (light blue) and 99.9% (dark blue) represent the respective confidence that slopes are statistically different from each other for the bands.

Band	**% confidence- slopes**	**% confidence- intercepts**
X- band	99%	-
C- band	n.s	99.9%
L- band	n.s	99%

Figure 14 demonstrates the visual comparison of regression lines between models created for forest and oil palm plantation data using the three bands individually. In models utilizing X-band data for forests, the slope parameter shows a negative value. This indicates a notable difference in the slope of the regression lines between forests and oil palm plantations, with the lines diverging distinctly. Table 7 confirms this statistical difference between the model slopes. The regression lines for the C-band data exhibited a slightly parallel trend, suggesting some similarity in the relationship between AGB and C- band backscatter across forests and oil palm plantations. However, for the L-band data, there was no parallel trend observed between the regression lines. Notably, the regression line for forests appears steeper than that for oil palm plantations,

indicating a more pronounced relationship between the AGB and L-band backscatter in forested areas compared to oil palm plantations.

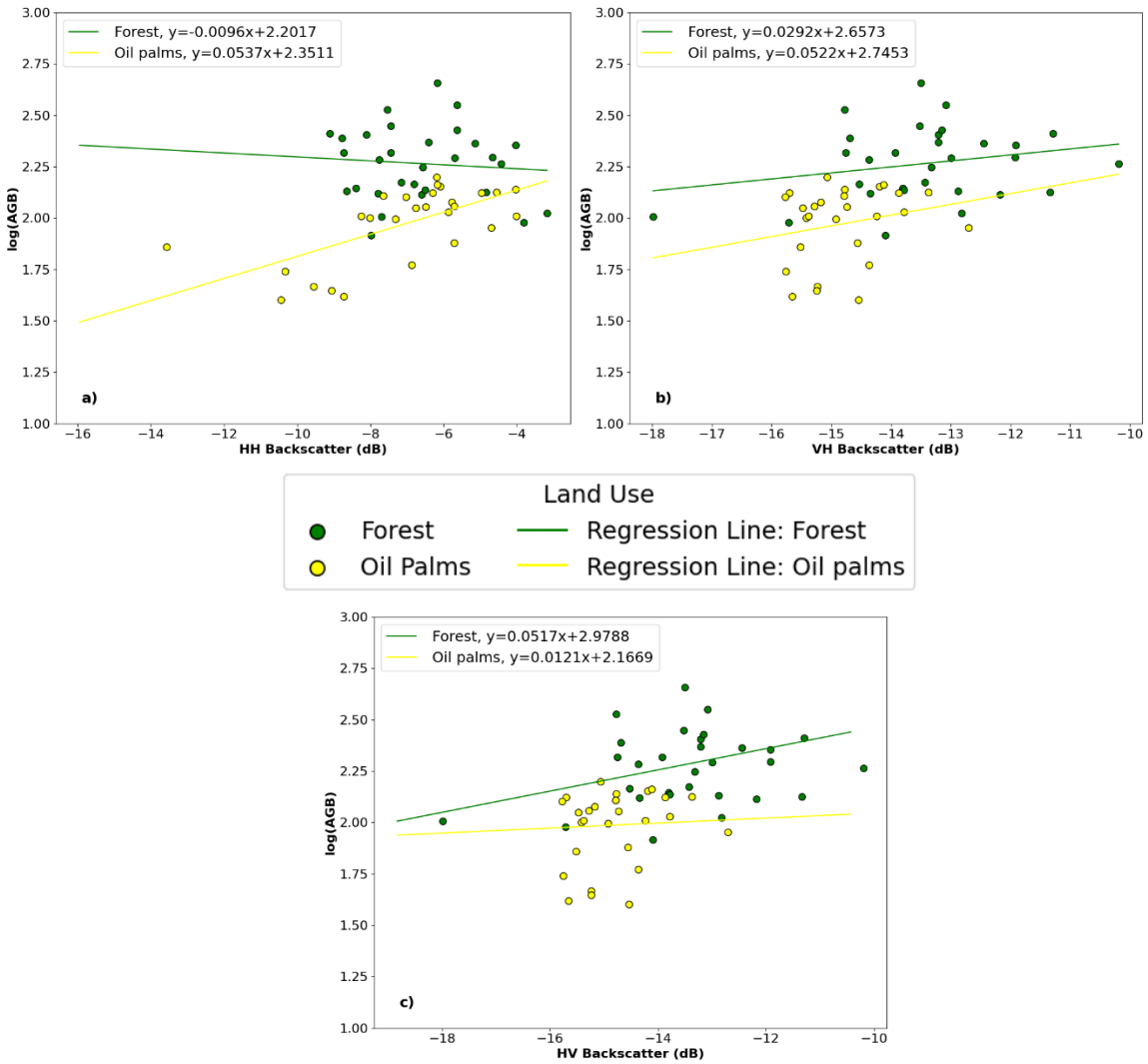


Figure 14- comparison of the regression lines between models created for predicting the AGB of forests and AGB of oil palms using a) X- band, b) C- band and c) L-band data.

4.3. Assessment of predictive accuracy in developed models

4.3.1. RMSE analysis for model validation

The RMSE values were calculated in absolute terms and also in relative terms keeping in consideration the variation in AGB in the data of that particular land use type. Refer to appendix A for the exact values of the relative and absolute RMSE. Figure 15 provides a comparison of the relative RMSE between models created for predicting AGB of different land use types using different bands and their combinations. The values of relative RMSE do not have units.

When incorporating all four land use categories simultaneously, the relative RMSE remained consistently below 1 across all models. Notably, the lowest relative RMSE of 0.69 was observed when utilizing only the X-band data, indicating the most accurate predictions within this land use category. Conversely, when

focusing solely on forest and oil palm data, employing only the X-band data resulted in the highest relative RMSE of 0.72 among all models, reflecting the highest prediction error within this specific land use category. Generally, the prediction errors for models created with forest and oil palm data were lower than the errors in models created with other land use types, except when using solely the X- band backscatter.

For forests, the relative RMSE was also lower than 1 for all bands and their combinations. The lowest relative RMSE of 0.72 was achieved when using just the L-band whereas the highest RMSE of 0.81 was obtained when using just the X- band. This means that size of the prediction error was 72% of the standard deviation of the observed values of forest AGB (in tons/ha). As the value is below one, it suggests that the error of the model is smaller than the data's standard deviation. For the oil palm plantations, the model created with using a combination of L- and C- had the highest relative RMSE of 1, signifying that the prediction error is equal to the standard deviation of the observed values. The lowest RMSE of 0.71 was achieved using just the X- band.

In summary, using X- band backscatter resulted in the most accurate predictions for models created with oil palms and all the 4 land use categories together, whereas it resulted in the least accurate predictions for forests and using forests and oil palm data together. L- band, led to the lowest prediction errors for forests and highest for models created with the 4 land use categories together.

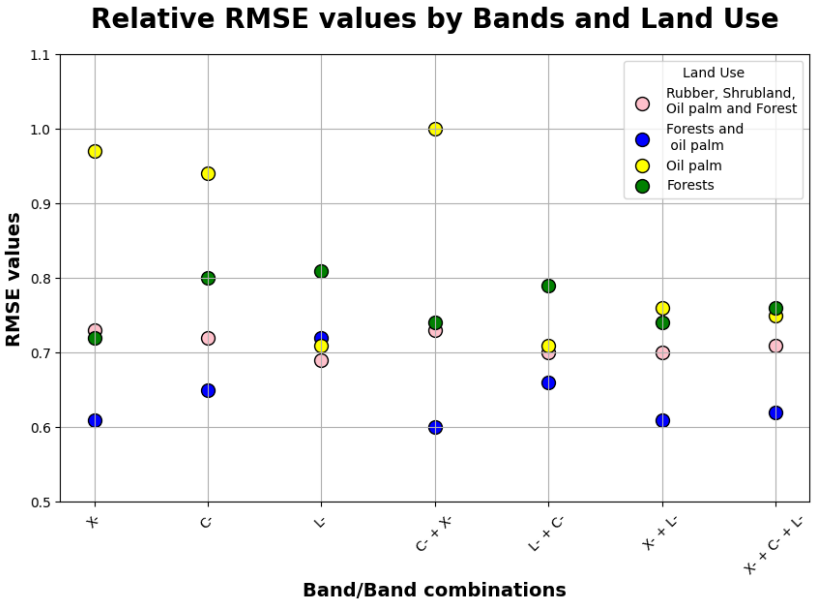


Figure 15- Comparison of the relative RMSE between models created for predicting AGB of different land use categories using different bands and their combinations.

4.3.2. Model fit assessment through R^2

The validation R^2 computed using the leave one out cross validation method indicated the variability that was explained with the models created on unseen data. Figure 16 provides a comparison between the validation R^2 of models created for different land use categories using different bands and band combinations.

For all the models created using the data from all 4 land use categories, the R^2 was consistently below 0. This indicated that the model was performing worse than a model that simply predicts the mean of the dependent variable- AGB for all observations.

Some improvement in the R^2 values were noticed when two land use categories-forests and oil palms were used to create the models. The highest R^2 value of 0.26 was obtained when using a combination of L- and C- band, which also corresponds to the lowest relative RMSE value of 0.60 for this particular land use

category. The lowest R^2 of -0.08 for this category was obtained using just the X- band, which also corresponds to the highest RMSE of 0.72 for this category.
For forest data, all models had a R^2 below 0 except the model created using L-band backscatter, having an R^2 of 0.02. For oil palm plantations, the highest R^2 was of 0.29 was obtained when using a combination of X- and C- band. The lowest R^2 of -0.33 was obtained when using a combination of the L- and C- band.
Generally, the validation R^2 of all models were at low levels signifying that the relationship between AGB and backscatter could not be properly captured by the models and that the models are not generalizing well to new unseen data.

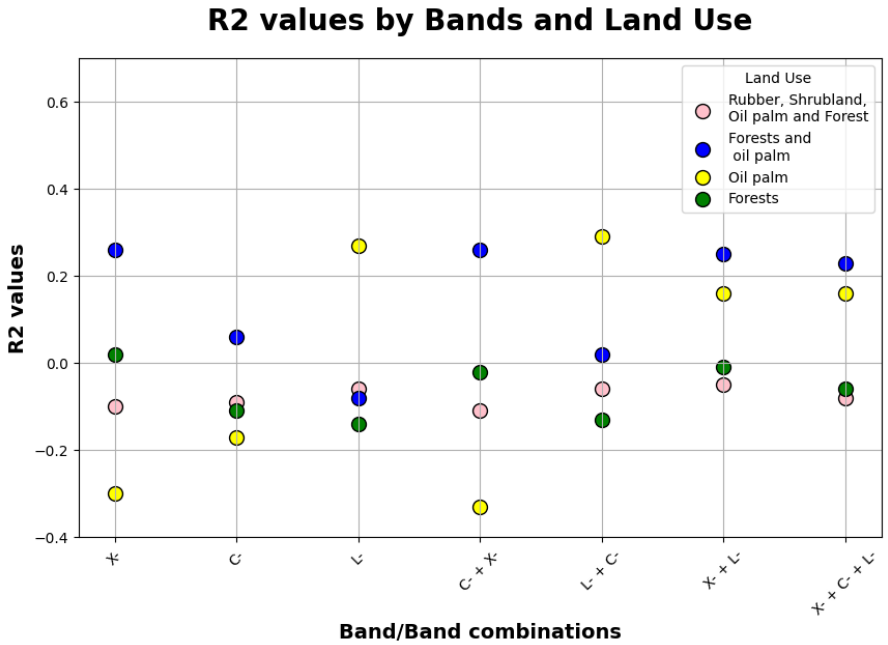


Figure 16- Comparison of the R^2 between models created for predicting AGB of different land use categories using different bands and their combinations.

4.3.3. AIC: A measure of model fit and complexity

An assessment of the AIC values was performed to compare models created with different bands or band combinations to predict the AGB of a particular land use category or several land use categories combined (Figure 17). The comparison was done between models created for the same land use category.
When using the data for all 4 land use categories together, the highest AIC value of 61.08 was achieved when using all three bands together. This increases the complexity of the model and hence the model gets penalised for this. On the other hand, the lowest AIC value of 57.09 was obtained when using just the C-band. According to the rule of thumb mentioned in section 3.6.2, a difference of more than 2 in AIC provides evidence of better performance of the model with lower AIC value. Hence, there is substantial evidence that the model with C- band performs better than model with all three bands, taking into consideration the complexity as well. For the other models, the difference between AIC values was not significant to make conclusions.
The AIC values of models created for AGB estimation of just two land use categories-forests and oil palms were compared. Using L-band individually along with its combination with X- and C- band proved to have a better fitted model than utilising just the X- band or it's combination with C- band (Figure 17) This was established by a difference of AIC value larger than 10 between these models. The highest AIC of -0.42 was obtained using the X- band and the lowest of -24.70 using a combination of X- and L- bands.

Within the forest land use category, L-band individually, along with its combination with C- and X- band showed substantial evidence of a better fitted model than a model with X- band and its combination with C- band, with a difference of more than 2 in the AIC values (Figure 17).
For the oil palm plantations, there was evidence based on AIC values that the models created using C- and X- band individually performed better than the model created using a combination of all three bands. For other models, the difference between AIC values was not greater than 2, hence no conclusions about their performance could be made (Figure 17).

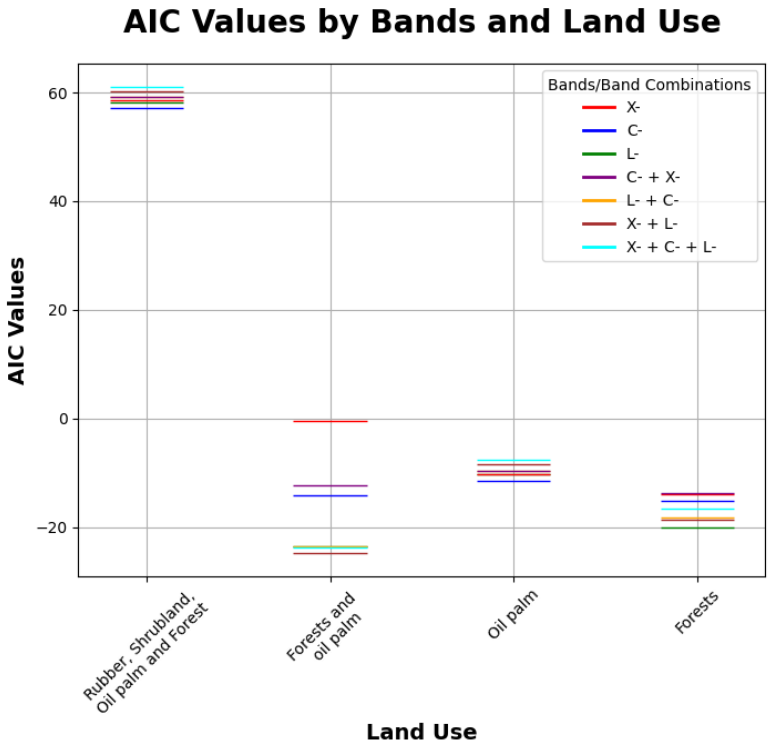


Figure 17- Comparison of AIC values between models created for predicting AGB of different land use categories using different bands and their combinations.

5. DISCUSSION

5.1. Investigating the difference in model performance and coeffecients between different radar bands individually

This study found significant differences in the slopes of models predicting AGB using X-band and L-band backscatter, particularly for forests. Moreover, The band with the highest accuracy and the only one with significant slopes for estimating the AGB of forests was L-band and for oil palm plantations was X- band.

These differences can be attributed to the different penetration capabilities of the two bands in forests. L-band, with its longer wavelength, can penetrate deeper into forest canopies and interact with larger vegetation elements such as trunks and large branches, leading to more accurate AGB estimates (Patenaude et al., 2005). This observation aligns with previous research, which has demonstrated the superior performance of L-band SAR in high-biomass areas due to reduced signal attenuation at higher biomass levels (Johannsen et al., 2015; Saatchi et al., 2011). Conversely, X-band SAR, with its shorter wavelength, primarily interacts with smaller vegetation elements like leaves and small branches and is more sensitive to vegetation with lower biomass (Mitchard et al., 2009b). A visible difference between the ranges of biomass of these land use categories can be seen from Figure 7. This results in lower sensitivity and accuracy of X-band for AGB estimation in dense forests, whereas a higher accuracy for oil palm plantations. More reasons for the better performance of X-band for oil palm plantations is discussed in Section 5.4.2. Furthermore, this study found that the X-band model for forests was not statistically significant, suggesting that X-band backscatter is less effective for high-biomass regions. This is consistent with other studies highlighting the limitations of X-band SAR in densely vegetated areas (Shivamogga et al., 2019). This can be attributed to the saturation effect of lower wavelength bands for high biomass (Asner & Mascaro, 2010).

Moreover, for some land use categories, the intercept of the model using solely X-band differs from the intercept of the two models that only use L- or C-band. This suggests that X-band responds differently from the other two bands to non-vegetation elements, meaning when AGB is zero. In case of no vegetation, the backscatter responds to soil properties like moisture and terrain roughness (Yu & Saatchi, 2016). This can be attributed to difference in wavelength between the three bands.

When all four land use categories were combined, the training accuracies obtained were below 0.1 and validation accuracies below 0. The potential reasons for this are discussed in Section 5.2. Hence, any conclusions about the band with highest performance for predicting the AGB of all four land use categories cannot be made. This also shows that the developed models did not generalize well on unseen or independent data. This is often an indication of overfitting (Figure 18)- a case when the training accuracy is substantially higher than the validation accuracy (Salman & Liu, 2019). These findings highlight the importance of selecting appropriate radar bands for accurate AGB estimation, supporting effective land management and conservation strategies. All in all, the variation in penetration capabilities across different biomass ranges of the L-, C- and X-band lead to statistically distinct model coefficients and accuracies of the models. For that reason, it is important to select the appropriate radar wavelength for different vegetation types to improve AGB estimation accuracy (Burt et al., 2020).

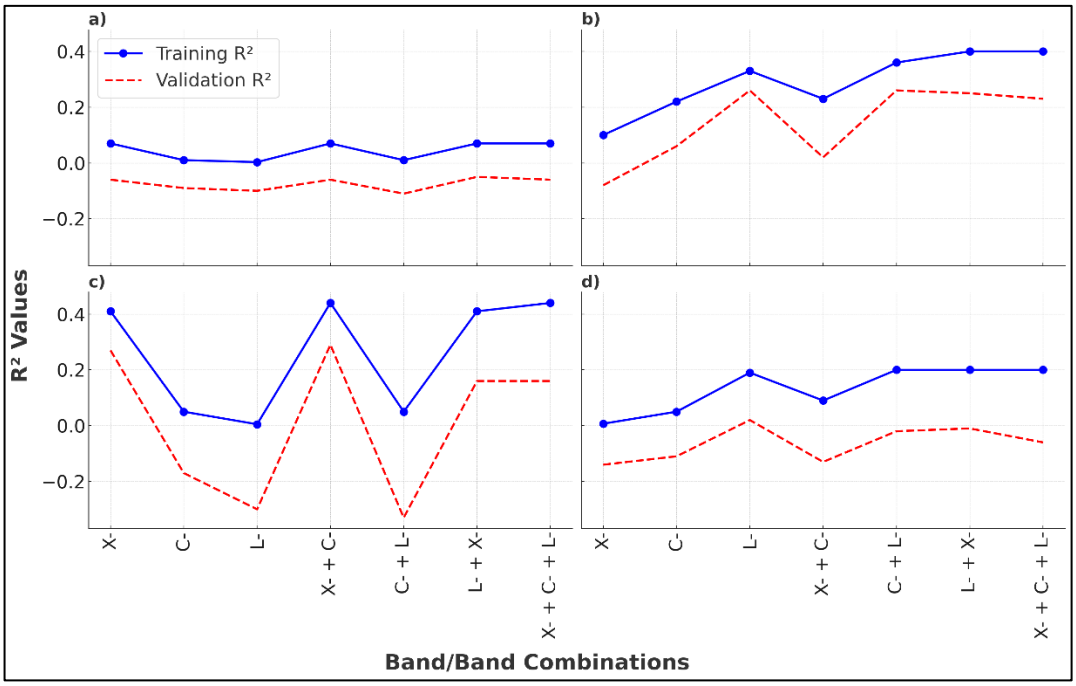


Figure 18- Comparison of the training and validation R^2 for a) all four land use categories b) forests and oil palm plantations c) oil palm plantations and d) forests. These graphs show the overfitting of the models.

5.2. Investigating the difference in model coeffecients between forests and oil palm plantations

The models created to predict AGB using X-band backscatter showed significantly different slopes for forests and oil palms. Additionally, the intercepts of the four models, each combining either forest or oil palm plantation with L-band or C-band, differ from one another.

Firstly, As mentioned in section 5.1, this difference can be attributed to the distinct biomass ranges of these two land use categories. Secondly, the structural differences between these land use categories could alter the backscatter that is received by the same radar band. Oil palm plantations have a more uniform canopy height compared to the heterogeneous and multi-layered canopy structure of natural forests (Foster et al., 2011; Luskin et al., 2017). As AGB is also sensitive to the structure of vegetation (Dobson et al., 1995), this could also be a potential reason for the varied performance of the different bands for different land use categories. The uniform structure and lower biomass of oil palms result in a stronger and more consistent response in X-band backscatter, leading to a steeper slope in the AGB model. In contrast, the varied structure and higher biomass of forests result in a weaker and more variable X-band backscatter response, contributing to a flatter slope (Shivamogga et al., 2019).

A study by Woodhouse et al. (2012) presented that combining different forest types for predicting AGB should be approached with caution due to potential biases. Similarly, combining different land use categories for AGB estimation should be avoided as they represent distinct groups with different biomass ranges, which could lead to inaccurate predictions or bias. The main issue is that combining two distinct groups for linear regression can lead to a model that reflects the relationship between the groups rather than accurately capturing the unique characteristics of each group. In addition, interpreting the differences in the slopes and intercepts of the models presents a challenge. Many of the models and their slopes are not statistically significant, which could introduce bias or inaccuracies in the conclusions derived from these slopes. However, it is noteworthy that all the intercepts of the models were statistically significant.

5.3. Penalising model complexity for integration of bands

In the case of all land use categories and their combinations, one of the highest training accuracies was achieved when using a combination of all the three bands. However, Naidoo et al. (2015) emphasized that while high accuracy is achieved with the use of three bands, it is important to evaluate the individual contributions of each band to the model's performance. It was noted that in this case, only one or two bands had significant slopes (Table 4). This signifies that although a higher accuracy was being achieved, not all the bands contributed significantly to explaining the variance of the model. The time and cost associated with the use of all the three bands needs considerable attention (Taboga, 2024). The C- band used in this study is available publicly, whereas the X- and L-band are not. When penalising the models for their complexity, a combination of all the three bands provides lowest performance for oil palms and all four land use categories. Therefore, the best performing models in terms of accuracy might not be the same when the model is penalised for added variables (Claeskens & Hjort, 2012). With limited resources, the integration of multiple bands should be considered critically.

5.4. Uncertainties in regression analysis

The linear regression models of this study might be influenced by the difference in polarisations between L- and C- bands and X- band, geo-location errors (Urbazaev et al., 2018) and cloud related disturbances in TanDEM-X imagery. Other potential reasons for errors in the AGB estimation have been widely discussed (Asner & Mascaro, 2010). For instance, a study by Zhang et al. (2023b) found that while volume scattering is the dominant mechanism in forested areas, double bounce scattering still contributes significantly to the overall backscatter, especially in dense forest environments. This can introduce errors in AGB estimation if not properly accounted for.

5.4.1. Outliers

The results obtained from the regression analysis using the data from all four land use categories had training accuracies below 0.1 and validation accuracies below 0 (Figure 16 and table 4). When two land use categories- rubber and shrubland were removed, the model performance increased by many folds. To examine the two land categories as potential outliers, an examination of the orthophotos of the study area was done to further investigate the reason. In the case of shrubland, it was noticed that some plots had visible spaces between the vegetation representing ground or bare soil. Figure 19 shows an example of such plot from the acquired orthophoto.

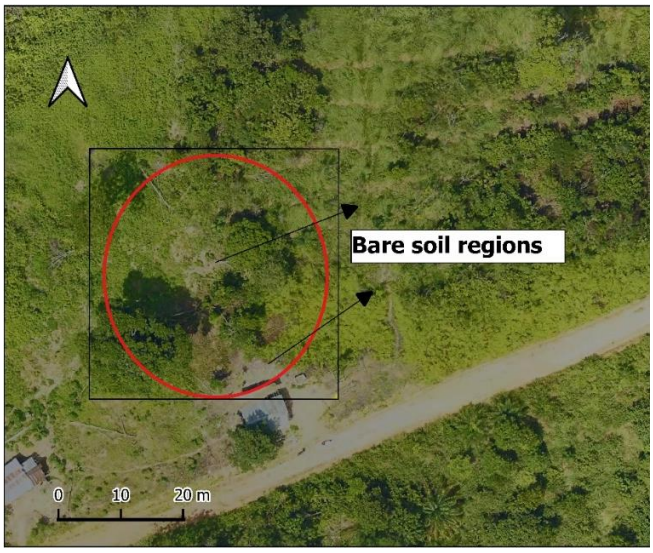


Figure 19- Plot S27 representing the spaces without vegetation in the plot. The red circle depicts the original extent of the field plot and the black square is the extent of the field plot created for the backscatter extraction from the pixels.

The description of shrubland plots provided under the project was "Unused fallow land, including grasslands, which is in transition to a forest, will be defined with DBH size distribution. In practice, shrubland can contain up to 50 % tree canopy closure above the lowest vegetation layer. Each chosen shrubland plot has at least one tree with a canopy starting at least 1.5 metre height and at least a canopy diameter of 2 metres". This potentially meant that the backscatter in these plots were not just from the biomass of vegetation, but also from bare soil or ground. Empty spaces of ground between less dense vegetation in plots could mean that the backscatter is representing not just the biomass of the vegetation but also the ground or soil.

Moreover, the field plots consisted of two varieties of rubber plantations- Jungle rubber, which is grown wild and monoculture, which is planted by humans in a systematic manner (Agustina & Oktavia, 2021). The rubber plantation plots were defined under the project as "Any rubber plantation from monocultures down to jungle rubber with as little as 10% rubber tree stems out of the total stem number in a 2500 m^2 area". In this case, the plots with jungle rubber contain a mix of jungle rubber along with other trees. This heterogeneity in the vegetation structure of the plots could potentially not be captured by a simple linear model. When combining vegetation with variable structure, the complexity of the model requires to capture the variability of AGB increases. (Zhang et al., 2020). Hence, various non-parametric methods have been tested to predict AGB (Zhang et al., 2020, Li et al., 2020).

5.4.2. Polarisations

The study was performed using two cross polarisations from L- and C- band, and one co-polarisation from X- band. Cross polarisations respond better to volume scattering, which is characteristic of vegetation. On the other hand, co-polarisations do not perform similarly (Vreugdenhil et al., 2020). Hence, a comparison between different bands would introduce a bias. Due to lack of available cross polarisation bands for X-band data, this study could not perform a comparison between cross polarisations. Appendix B provides a comparison between co-polarisations of the three bands. The training accuracy of these were lower than when cross polarisations of C- and L-band were utilised. This is in line with the assumption that was based on previous studies that cross polarisations perform better for AGB estimation than co-polarised bands (Vreugdenhil et al., 2020). Furthermore, Patel et al. (2006) found that due to the horizontal canopy structure

of the vegetation in their study, HH polarisation of L- and C- band was sensitive to the plant density. The structure of the vegetation can also impact it's sensitivity to a polarisation. The type of oil plantation used in this study is also seen to have horizontal structured branches as can be seen in Figure 20 (CABI, 2024). This characteristic likely explains the effective performance of X-band HH polarization in capturing the plantation's horizontal characteristics.

Figure 20- Elaeis *guineensis* oil palm plantation with horizontal structured branches (CABI, 2024).

5.4.3. Temporal differences between ground data and satellite imagery

The field data collection was conducted from 2021-22, whereas the TanDEM-X imagery was already acquired in 2018-2019. Hence, there was a gap of two to three years between the field plots and the satellite imagery. This could introduce possible biases in the models if the vegetation growth was higher than the error margins of the model. An additional analysis using available AGB data from 2012 and 2019 field plots to calculate the growth rate of the vegetation was performed. This is presented in Appendix E. A further analysis into this was performed by using additional Canopy Height Model (CHM) using LiDAR data to calculate the annual change in tree height (Appendix F). As the growth rate was within the error margins, possible interference from this temporal difference can be rejected. Additionally, although matured trees in a forest do accumulate carbon, the growth rate of biomass is not substantial in a few years to impact AGB estimation (Köhl, Neupane, & Lotfiomran, 2017).

5.4.4. Quality of satellite images

In a manual revision of the geo-location of the SAOCOM imagery, it was found that the location was off by a few pixels (30-40 metres), a manual georeferencing was performed after which the mean error was reduced to 1-3 metres. Earlier studies have also mentioned the effect of geo-location errors of satellite imagery and the potential errors it can introduce for estimating the AGB (Urbazaev et al., 2018). As the georeferencing was performed manually, there were still chances of a geo location error or the plots not perfectly overlapping the satellite imagery. For the circular field plots, a square region was chosen keeping in consideration the shape of the pixels. Figure 19 shows the circular plots (in red outline) and the corresponding square plots (in black outline). Hence, the square plot introduced regions that were not present in the field plot on ground. This could interfere with the backscatter signal.
Moreover, for the TanDEM-X imagery, a disturbance due to cloud interference was spotted as seen in Figure 21. Three field plots belonging to oil palm, rubber and shrubland land use category were located on this patch. As the cloud is absorbing the signal, the signal was attenuated for these three plots.

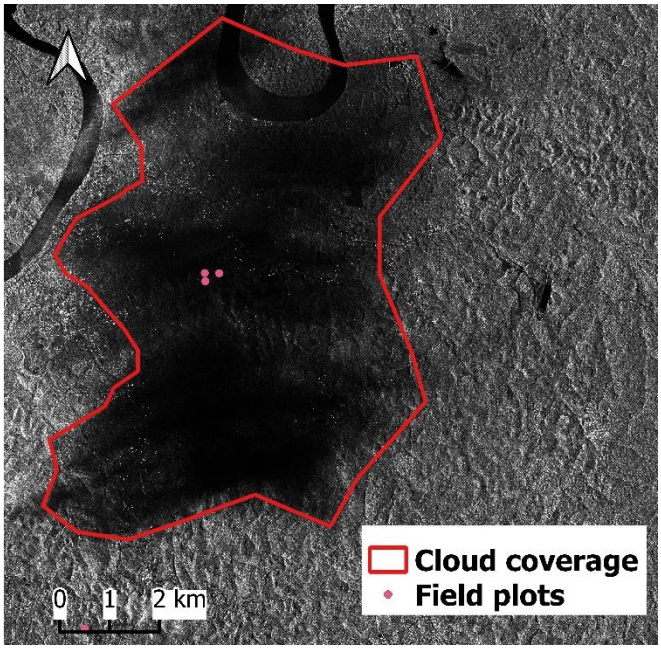


Figure 21- Showing the cloud disturbance (outlined in red) present in the TanDEM-X image

5.5. Practical applications and future prospects

The study aimed at comparing different radar bands and their combinations for estimating AGB of different land use categories together as well as individually. The results of the study indicated the importance of certain bands for estimating the AGB of a particular land use category over others. It also signified that although using a combination of bands might give better accuracies, it also increases model complexity much more than the increase in accuracy. These results could potentially help the Indonesian government to achieve their target of reducing carbon emissions and well as for REDD+ MRV. A quantification of the carbon lost when primary forests are converted to oil palm plantations could also be done. In this view the study provides information about the best band/band combination for forest and oil palm plantation AGB, while also considering the complexity of more bands.

The analysis also revealed if pooling different land use categories is reasonable to create a linear model to predict AGB. It was found that for land use categories with similar AGB ranges like rubber and oil palm plantations, there is no significant difference between the model coefficients. For other land use categories, with different AGB ranges, a significant difference in model coefficients introduces a bias. Hence, the variability in the AGB couldn't be captured by a linear model. This is important if accurate models need to be created for complex landscapes such as the one in this study. A landscape with varied vegetation types might need a non-linear model to capture the variability in biomass.

5.6. Proposed recommendations

Proposed recommendations to enhance the applicability of this study are:

1. The studies focusing on an integration of radar bands or a comparison between radar bands for AGB estimation should utilise the same polarisation for consistency. Different radar polarisations respond differently to vegetation. This might lead to biases in the comparison of bands.
2. Integration of several other remote sensing technologies such as LiDAR and optical imagery can be tested. Moreover, inclusion of several variables like tree height and vegetation indices can help to mitigate the drawback of saturation that is faced with using just radar data (Tian et al., 2023).

Combining radar with other remote sensing technologies can improve the robustness of biomass estimates across diverse forest types and conditions. Data from multiple remote sensing sources can increase the accuracy of AGB estimation.

3. Future studies can dive into different parametric and non-parametric methods that can be applied to capture the variability of AGB when different types of vegetation are utilised. These models can range from machine learning models such as random forest, SVM to deep learning methods such as gradient boosting and neural networks. For estimation of AGB at a regional scale, these models could capture the vegetation complexity well.
4. An effect of a combination of different vegetation types including plantations could be studied. Plantations have a more homogeneous structure than wild vegetation, making these categories different. The need for separate models for estimating the AGB of plantations can be explored.

6. CONCLUSIONS

In order to achieve different targets for reducing carbon emissions from deforestation, an accurate measurement of the carbon lost when one land use category is converted to another is essential. Moreover, to meet targets of international agreements and national targets, carbon quantification of forests is essential. As AGB is a reliable proxy for the carbon stored in a forest, accurate methods to estimate AGB needs to be established. This study examined the role of different radar bands, which have proved to have high accuracy in AGB estimation. An effect of the integration of these bands was also examined. This study revealed the bands that have the most optimal performance for a specific land use category. L-band had indeed the best performance along with significant slopes out of the three bands for high biomass vegetation of forests. For the lower biomass vegetation of oil palm plantations, X- band proved to have the best performance. The accuracies obtained for models when the four land use categories were extremely low. Investigation into the possible reasons revealed presence of two outlier land use categories, oversimplification of linear models for the analysis and different biomass ranges of the vegetation types. Hence, the analysis was performed again by excluding the two outlier land use categories. The accuracies achieved when forests and oil palms were combined increased by many folds, With L-band proving to give the most accurate model within the three bands.

The investigation also uncovered an important finding about the use of different radar bands for AGB estimation. A major contribution of the study could be to show that although integration of different radar bands improves the model performance, the increased costs of acquiring different bands or increase in complexity overshadows the increase in the accuracy of the model. In practical terms, when estimating the AGB of a big landscape, acquiring multiple radar bands might not be optimal if similar results could be obtained from a single radar band. A combination of the three bands produced one of the highest training accuracies, but the slopes of just one or two bands were significant. This indicated the redundance of the other bands in the analysis. AIC values revealed that using single bands gives the highest accuracy when the models are penalised for their complexity.

The study also examined the efficiency of integration of different vegetation types, with different biomass ranges for estimating the AGB. The model coefficients, which is in most cases intercepts and in one case slope for these vegetation types were statistically different. Hence, an integration of these land use categories could introduce a bias in the analysis. It was found also that vegetation with similar biomass ranges like oil palm and rubber plantations had coefficients which were not statistically different. This makes it acceptable to pool them in order to develop a model for AGB estimation.

In general, discretion should be used when integrating vegetation of different biomass ranges in a linear model and alternative complex models could be utilised. The sensitivity of different bands to different land use categories along with the increasing costs developed for acquiring more bands should be considered.

LIST OF REFERENCES

Achard, F., Eva, H. D., Mayaux, P., Stibig, H.-J., & Belward, A. (2004). Improved estimates of net carbon emissions from land cover change in the tropics for the 1990s. *Global Biogeochem. Cycles*, 18(2). https://doi.org/10.1029/2003GB002142

Agustina, D. S., & Oktavia, F. (2021). Rubber agroforestry system in Indonesia: Past, present, and future practices. *E3S Web of Conferences*, 305, 02005. https://doi.org/10.1051/e3sconf/202130502005

Airbus. (2015). *TerraSAR-X Image Product Guide Basic and Enhanced Radar Satellite Imagery* (Issue 2.1). Airbus Defence and Space. Retrieved on 10 October 2023 https://www.intelligenceairbusds.com/files/pmedia/public/r459_9_20171004_tsxX-airbusds-ma-0009_tsX-productguide_i2.01.pdf

Argamosa, R. J. L., Blanco, A. C., Baloloy, A. B., Candido, C. G., Dumalag, J. B. L. C., Dimapilis, L. L. C., & Paringit, E. C. (2018). Modelling above ground biomass of mangrove forest using Sentinel-1 imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-3, 13-20. https://doi.org/10.5194/isprs-annals-IV-3-13-2018

Asari, N., Suratman, M. N., & Jaafar, J. (2017). Modelling and mapping of above ground biomass (AGB) of oil palm plantations in Malaysia using remotely sensed data. *International Journal of Remote Sensing*, 38(16), 4741–4764. https://doi.org/10.1080/01431161.2017.1325533

Asner, G. P. (2010). Cloud cover in Landsat observations of the Brazilian Amazon. *International journal of remote sensing*, 22(18), 3855-3862. https://doi.org/10.1080/01431160010006926

Asner, G. P., & Mascaro, J. (2010). Forest aboveground biomass mapping and estimation across multiple spatial scales: a review of the role of remote sensing. *Remote Sensing of Environment, 114*(7), 1371-1385. https://doi.org/10.1016/j.rse.2010.01.024

Austin, K. G., Schwantes, A., Gu, Y., & Kasibhatla, P. S. (2019). What causes deforestation in Indonesia? *Environmental Research Letters*, 14(2), 024007. https://doi.org/10.1088/1748-9326/aaf6db

Brown, S. (1997). Estimating biomass and biomass change of tropical forests: a primer. *In FAO Forestry Paper* (Vol. 134, Issue August). http://www.fao.org/docrep/W4095E/W4095E00.htm

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261-304. https://doi.org/10.1177/0049124104268644

Burt, A., Calders, K., Cuni-Sanchez, A., Gómez-Dans, J., Lewis, P., Lewis, S. L., Malhi, Y., Phillips, O. L., & Disney, M. (2020). Assessment of bias in pan-tropical biomass predictions. *Frontiers in Forests and Global Change*, 3, Article 12. https://doi.org/10.3389/ffgc.2020.00012

CABI. (2024). *Integrated crop management (ICM)*. CABI Digital Library. https://doi.org/10.1079/cabicompendium.20295

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C. M., & Saldarriaga, J. G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology, 20(10), 3177-3190. https://doi.org/10.1111/gcb.12629

Claeskens, G., & Hjort, N. L. (2012). Akaike's information criterion. In *Model Selection and Model Averaging* (pp. 22-69). Cambridge University Press. https://doi.org/10.1017/CBO9780511790485.003.

Deep, G. (2023). Climate change: The biggest challenge of this era. International Journal of Science and Research Archive, 08(01), 499–506. https://doi.org/10.30574/ijsra.2023.8.1.0086

Dobson, M. C., Ulaby, F. T., Pierce, L. E., Sharik, T. L., Bergen, K. M., Kellndorfer, J., Kendra, J. R., Li, E., Lin, Y. C., Nashashibi, A., Sarabandi, K., & Siqueira, P. (1995). Estimation of forest biophysical characteristics in Northern Michigan with SIR-C/X-SAR. *IEEE Transactions on Geoscience and Remote Sensing, 33*(4), 877-895. https://doi.org/10.1109/36.406674

Duncanson, L. I., Niemann, K. O., & Wulder, M. A. (2010). Estimating forest canopy height and terrain relief from GLAS waveform metrics. Remote Sensing of Environment, 114(1), 138–154. https://doi.org/10.1016/J.RSE.2009.08.018

Englhart, S., Keuck, V., & Siegert, F. (2011). Aboveground biomass retrieval in tropical forests - The potential of combined X- and L-band SAR data use. *Remote Sensing of Environment*, 115(5), 1260–1271. https://doi.org/10.1016/j.rse.2011.01.008

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore III, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., & Steffen, W. (2000). The global carbon cycle: A test of our knowledge of Earth as a system. *Science*, 290(5490), 291-296. https://doi.org/10.1126/science.290.5490.291

Fanelli, C. (2013). Climate Change: 'The Greatest Challenge of Our Time'. *Alternate Routes: A Journal of Critical Social Research*, 25. Retrieved from https://alternateroutes.ca/index.php/ar/article/view/20592

FAO. (2009). State of the world's forests, 2009. *Food and Agriculture Organization*. Retrieved from https://www.fao.org/3/i0350e/i0350e00.htm

Faraway, J. J. (2006). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall. https://doi.org/10.1007/s00180-009-0152-1

Foster, W. A., Snaddon, J. L., Turner, E. C., Fayle, T. M., Cockerill, T. D., Ellwood, M. F., Broad, G. R., Chung, A. Y., Eggleton, P., & Ewers, R. M. (2011). Establishing the evidence base for maintaining biodiversity and ecosystem function in the oil palm landscapes of South East Asia. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1582), 3277-3291. https://doi.org/10.1098/rstb.2011.0041

Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2(4), 045023. https://doi.org/10.1088/1748-9326/2/4/045023

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management*, 2(3), 231-244. https://doi.org/10.4155/CMT.11.18

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853. https://doi.org/10.1126/SCIENCE.1244693

Herold, M., Carter, S., Avitabile, V., Espejo, A. B., Jonckheere, I., Lucas, R., McRoberts, R. E., Næsset, E., Nightingale, J., Petersen, R., Reiche, J., Romijn, E., Rosenqvist, A., Rozendaal, D. M. A., Seifert, F. M., Sanz, M. J., & De Sy, V. (2019). The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surveys in Geophysics*, 40(4), 757–778. Springer. Netherlands. https://doi.org/10.1007/s10712-019-09510-6

Intergovernmental Panel on Climate Change. (2014). Climate Change 2014: Synthesis Report. *Working Group III Contribution to the IPCC Fifth Assessment Report* (Report No. IPCC-AR5-SYR). Cambridge University Press. https://doi.org/10.1017/cbo9781107415416

Johannsen, V. K., Saatchi, S., Fensholt, R., & Madsen, K. N. (2015). L-band SAR backscatter related to forest cover, height and aboveground biomass at multiple spatial scales across Denmark. *Remote Sensing*, 7(4), 4442-4472. https://doi.org/10.3390/rs70404442

Khammar, A., Yarahmadi, M., & Madadizadeh, F. (2020). What is analysis of covariance (ANCOVA) and how to correctly report its results in medical research? *Iranian Journal of Public Health, 49*(5), 1016-1017. PMID: 32953697; PMCID: PMC7475615. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7475615/

Khasanah, N., van Noordwijk, M., & Ningsih, H. (2015). Aboveground carbon stocks in oil palm plantations and the threshold for carbon-neutral vegetation conversion on mineral soils. In S. Wich (Ed.), *Cogent Environmental Science* (Vol. 1, No. 1, Article 1119964). Cogent OA. https://doi.org/10.1080/23311843.2015.1119964

Koh, L. P., & Wilcove, D. S. (2008). Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters*, 1(2), 60–64. https://doi.org/10.1111/J.1755-263X.2008.00011.X

Köhl, M., Neupane, P. R., & Lotfiomran, N. (2017). The impact of tree age on biomass growth and carbon accumulation capacity: A retrospective analysis using tree ring data of three tropical tree species grown in natural forests of Suriname. *PLOS ONE, 12*(8), 1-17. https://doi.org/10.1371/journal.pone.0181187

Kotowska, M. M., Leuschner, C., Triadiati, T., Meriem, S., & Hertel, D. (2015). Quantifying above- and belowground biomass carbon loss with forest conversion in tropical lowlands of Sumatra (Indonesia). *Global Change Biology*, 21(9), 3620-3634. https://doi.org/10.1111/gcb.12979.

Kuha, J. (2004). AIC and BIC. Sociological Methods & Research, 33(2), 188-229. https://api.semanticscholar.org/CorpusID:123129645

Larocque, G. R., Barbosa, J. M., Broadbent, E. N., & Bitencourt, M. D. (2014). Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. International Journal of Forestry Research, 2014, 715796. https://doi.org/10.1155/2014/715796

Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., & Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115(11), 2850– 2860. https://doi.org/10.1016/J.RSE.2011.03.020

Lee, J.-S. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(2), 165-168. https://doi.org/10.1109/TPAMI.1980.4766994

Li, Y., Li, M., Li, C., & Liu, Z. (2020). Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports* 2020 10:1, 10(1), 1–12. https://doi.org/10.1038/s41598-020-67024-3

Liu, Y., Huang, G., Lu, W., Peng, H., & Wang, J. (2021). An Optimized Lee Filter Denoising Method Based on EIP Correction. In S. Shi, L. Ye, & Y. Zhang (Eds.), *Artificial Intelligence for Communications and Networks* (pp. 424-436). Springer International Publishing. https://doi.org/10.1007/978-3-030-69066-3_37

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G., & Saah, D. (2012). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012, 1–16. https://doi.org/10.1155/2012/436537

Luckman, A., Baker, J., & Wegmüller, U. (2000). Repeat-Pass Interferometric Coherence Measurements of Disturbed Tropical Forest from JERS and ERS Satellites. *Remote Sensing of Environment*, 73(3), 350–360. https://doi.org/10.1016/S0034-4257(00)00110-3

Luskin, M. S., Ickes, K., Yao, T. L., Davies, S. J., & Pallzieri, Y. M. (2017). Cross-boundary subsidy cascades from oil palm degrade distant tropical forests. *Nature Communications, 8*(1), 2231. https://doi.org/10.1038/s41467-017-01920-7

Margono, B. A., Turubanova, S., Zhuravleva, I., Potapov, P., Tyukavina, A., Baccini, A., Goetz, S. J., & Hansen, M. C. (2012). Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environmental Research Letters*, 7(3), 034010. https://doi.org/10.1088/1748-9326/7/3/034010

Miettinen, J., Shi, C., & Liew, S. C. (2016). Land cover distribution in the peatlands of Peninsular Malaysia, Sumatra and Borneo in 2015 with changes since 1990. *Global Ecology and Conservation, 6*, 67-78. https://doi.org/10.1016/j.gecco.2016.02.004

Mitchard, E. T. A., Saatchi, S. S., Gerard, F. F., Lewis, S. L., & Meir, P. (2009a). Measuring woody encroachment along a forest-savanna boundary in Central Africa. *Earth Interactions*, 13(8), 1–29. https://doi.org/10.1175/2009EI278.1

Mitchard, E. T. A., Saatchi, S. S., Woodhouse, I. H., Nangendo, G., Ribeiro, N. S., Williams, M., Ryan, C. M., Lewis, S. L., Feldpausch, T. R., & Meir, P. (2009b). Using satellite radar backscatter to predict above- ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, 36(23). https://doi.org/10.1029/2009GL040692

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885-900. https://doi.org/10.13031/2013.23153

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G., & Leblon, B. (2015). Savannah woody structure modelling and mapping using multi-frequency (X-, C- and L-band) Synthetic Aperture Radar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105. https://doi.org/10.1016/j.isprsjprs.2015.04.007

Notarnicola, C., Posa Notarnicola, F. C., & Posa, F. (2007). Combination of X, C, and L band SAR images for retrieval of surface parameters (Vol. 6746, pp. 84-91). *Proceedings of SPIE - The International Society for Optical Engineering*, Florence, Italy. https://doi.org/10.1117/12.753797

Nurdiana, A., Setiawan, Y., Pawitan, H., Prasetyo, L. B., & Permatasari, P. A. (2016). Land Changes Monitoring Using MODIS Time-series Imagery in Peat Lands Areas, Muaro Jambi, Jambi Province, Indonesia. *Procedia Environmental Sciences*, 33, 443–449. https://doi.org/10.1016/j.proenv.2016.03.095

Painam, R. K., & Manikandan, S. (2023). BEMD based adaptive Lee filter for despeckling of SAR images. *Advances in Space Research, 71*(8), 3140-3149. https://doi.org/10.1016/j.asr.2022.12.009

Patel, P., Srivastava, H. S., Panigrahy, S., & Parihar, J. S. (2006). Comparative evaluation of the sensitivity of multi-polarized multi-frequency SAR backscatter to plant density. *International Journal of Remote Sensing, 27*(2), 293-305. https://doi.org/10.1080/01431160500214050

Patenaude, G., Milne, R., & Dawson, T. P. (2005). Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science & Policy*, 8(2), 161–178. https://doi.org/10.1016/J.ENVSCI.2004.12.010

Ravindranath, N. H., & Ostwald, M. (2007). *Carbon inventory methods: handbook for greenhouse gas inventory, carbon mitigation and roundwood production projects* (Vol. 29). Springer Science & Business Media. https://doi.org/10.1007/978-1-4020-6547-7

Reichstein, M., & Carvalhais, N. (2019). Aspects of Forest Biomass in the Earth System: Its Role and Major Unknowns. *Surveys in Geophysics*, 40(4), 693–707. https://doi.org/10.1007/s10712-019-09551-x

Roy, S., Mudi, S., Das, P., Ghosh, S., Shit, P. K., Bhunia, G. S., & Kim, J. (2021). Estimating Above Ground Biomass (AGB) and Tree Density using Sentinel-1 Data. In P. K. Shit, H. R. Pourghasemi, P. Das,

& G. S. Bhunia (Eds.), Spatial Modeling in Forest Resources Management: Rural Livelihood and Sustainable Development (pp. 259-280). Springer International Publishing. https://doi.org/10.1007/978-3-030-56542-8_11

Rubel, O., Lukin, V., Rubel, A., & Egiazarian, K. (2021). Selection of Lee filter window size based on despeckling efficiency prediction for Sentinel SAR images. *Remote Sensing, 13*(18), 1887. https://doi.org/10.3390/rs13101887

Rustiadi, E., Mulya, S., Pribadi, D., Saad, A., Supijatno, S., Iman, L. O., Pravitasari, A., Ermyanyla, M., & Nurdin, M. (2022). Study of oil palm plantation on peatland under spatial policies in Jambi Province, Indonesia. *IOP Conference Series: Earth and Environmental Science, 1025*, 012004. https://doi.org/10.1088/1755-1315/1025/1/012004

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., ... & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Nature Climate Change*, 1(3), 182-185. https://doi.org/10.1038/nature09985

Salman, S., & Liu, X. (2019). Overfitting mechanism and avoidance in deep neural networks. *arXiv*. https://arxiv.org/abs/1901.06566

Schlund, M., & Davidson, M. W. J. (2018). Aboveground forest biomass estimation combining L- and P-Band SAR acquisitions. *Remote Sensing*, 10(7). https://doi.org/10.3390/rs10071151

Schlund, M., Scipal, K., & Quegan, S. (2018). Assessment of a Power Law Relationship Between P-Band SAR Backscatter and Aboveground Biomass and Its Implications for BIOMASS Mission Performance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(10), 3538-3547. https://doi.org/10.1109/JSTARS.2018.2866868

Schmullius, C. C., & Evans, D. L. (2010). Review article Synthetic aperture radar (SAR) frequency and polarization requirements for applications in ecology, geology, hydrology, and oceanography: A tabular status quo after SIR-C/X-SAR. *International Journal of Remote Sensing*, 18(13), 2713-2722. https://doi.org/10.1080/014311697217297

Segura, M., & Kanninen, M. (2005). Allometric Models for Tree Volume and Total Aboveground Biomass in a Tropical Humid Forest in Costa Rica1. *Biotropica*, 37(1), 2–8. https://doi.org/10.1111/J.1744-7429.2005.02027.X

Seppi, S. A., López-Martinez, C., & Joseau, M. J. (2022). Assessment of L-Band SAOCOM InSAR Coherence and Its Comparison with C-Band: A Case Study over Managed Forests in Argentina. *Remote Sensing*, 14(22), 5652. https://doi.org/10.3390/rs14225652

Shigetomi, Y., Ishimura, Y. & Yamamoto, Y. Trends in global dependency on the Indonesian palm oil and resultant environmental impacts. Sci Rep 10, 20624 (2020). https://doi.org/10.1038/s41598-020-77458-4

Shivamogga, R., Musthafa, A., & Gonsamo, A. (2019). The role of time-series L-band SAR and GEDI in mapping sub-tropical above-ground biomass. *Frontiers in Forests and Global Change*, 2, Article 12. https://doi.org/10.3389/ffgc.2019.00012

Sivasankar, T., Lone, J. M., K.K., S., Qadir, A., & P.L.N., R. (2018). The potential of multi-frequency multipolarized ALOS-2/PALSAR-2 and Sentinel-1 SAR data for aboveground forest biomass estimation. *International Journal of Engineering and Technology*, 10(3), 797–802. https://doi.org/10.21817/IJET/2018/V10I3/181003095

Smith, P., Bustamante, M., Ahammad, H., Clark, H., Dong, H., Elsiddig, E. A., Haberl, H., Harper, R., House, J. I., Jafari, M., Masera, O., Mbow, C., Ravindranath, N. H., Rice, C. W., Abad, C. R., Romanovskaya, A., Sperling, F., & Tubiello, F. N. (2014). Agriculture, forestry and other land use (AFOLU). In Climate change 2014: mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 811-

922). Cambridge University Press. https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_chapter11.pdf

Solberg, S., Astrup, R., Gobakken, T., Næsset, E., & Weydahl, D. J. (2010). Estimating spruce and pine biomass with interferometric X-band SAR. *Remote Sensing of Environment*, 114(10), 2353–2360. https://doi.org/10.1016/J.RSE.2010.05.011

Stovall, A. E. L., Vorster, A. G., Anderson, R. S., Evangelista, P. H., & Shugart, H. H. (2017). Non-destructive aboveground biomass estimation of coniferous trees using terrestrial LiDAR. *Remote Sensing of Environment*, 200, 31-42. https://doi.org/10.1016/j.rse.2017.08.013

Taboga, M. (2024). *Linear regression - Model selection criteria*. Statlect. Retrieved from https://www.statlect.com/fundamentals-of-statistics/linear-regression-model-selection-criteria

Tian, L., Wu, X., Tao, Y., Li, M., Qian, C., Liao, L., & Fu, W. (2023). Review of remote sensing-based methods for forest aboveground biomass estimation: Progress, challenges, and prospects. *Forests, 14*(6), 1086. https://doi.org/10.3390/f14061086

Tonks, A. J., Aplin, P., Beriro, D. J., Cooper, H., Evers, S., Vane, C. H., & Sjögersten, S. (2017). Impacts of conversion of tropical peat swamp forest to oil palm plantation on peat organic chemistry, physical properties and carbon stocks. *Geoderma*, 289, 36-45. https://doi.org/10.1016/j.geoderma.2016.11.018

UNEP. (2018). Advancing on REDD+ Module 1: National Forest Monitoring Systems for REDD+. In REDD+ *Academy Learning Journals* (Issue August). United Nations Collaborative Programme on Reducing Emissions from Deforestation and Forest Degradation (UN-REDD). https://www.unredd.net/documents/globalprogramme- 191/redd-academy-3509/redd-academy-learning-journals/english/17232- fundamentals-on-redd- module-1-climate-change-and-the-role-of-forests-1.html

Urbazaev, M., Thiel, C., Migliavacca, M., Reichstein, M., Rodriguez-Veiga, P., & Schmullius, C. (2016). Improved multi-sensor satellite-based above ground biomass estimation by selecting temporally stable forest inventory plots using NDVI time series. *Forests*, *7*(8). https://doi.org/10.3390/f7080169

Urbazaev, M., Thiel, C., Cremer, F., & et al. (2018). Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon Balance and Management, 13*(5). https://doi.org/10.1186/s13021-018-0093-5

Verchot, L. V., Petkova, E., Obidzinski, K., Atmadja, S., Yuliani, L., Dermawan, A., Murdiyarso, D., & Amira, S. (2010). *Reducing forestry emissions in Indonesia*. Bogor, Indonesia: Center for International Forestry Research (CIFOR). https://cgspace.cgiar.org/handle/10568/20491

Vreugdenhil, M., Navacchi, C., Bauer-Marschallinger, B., Hahn, S., Steele-Dunne, S., Pfeil, I., Dorigo, W., & Wagner, W. (2020). Sentinel-1 Cross Ratio and Vegetation Optical Depth: A Comparison over Europe. Remote Sensing, 12(20), 3404. https://doi.org/10.3390/rs12203404

Yu, Y., & Saatchi, S. (2016). Sensitivity of L-Band SAR Backscatter to Aboveground Biomass of Global Forests. *Remote Sensing*, 8(6), 522. https://doi.org/10.3390/rs8060522

Webb, G., Sammut, C., Perlich, C., Horváth, T., Wrobel, S., Korb, K., Noble, W., Leslie, C., Lagoudakis, M., Quadrianto, N., Buntine, W., Getoor, L., Namata, G., Jin, J., Ting, J.-A., Vijayakumar, S., Schaal, S., & De Raedt, L. (2010). Leave-one-out cross-validation. In *Encyclopedia of Machine Learning* (pp. 600-601). Springer. https://doi.org/10.1007/978-0-387-30164-8_469

Woodhouse, I., Mitchard, E., Brolly, M., Maniatis, D., & Ryan, C. (2012). Radar backscatter is not a 'direct measure' of forest biomass. *Nature Climate Change*, 2, 556-557. https://doi.org/10.1038/nclimate1601

Zhang, H., Wang, C., Zhu, J., Fu, H., Han, W., & Xie, H. (2023a). Forest aboveground biomass estimation in subtropical mountain areas based on improved water cloud model and PolSAR decomposition using L-band PolSAR data. *Forests, 14*(12), 2303. https://doi.org/10.3390/f14122303

Zhang, L., Zhang, X., Shao, Z., Jiang, W., & Gao, H. (2023b). Integrating Sentinel-1 and 2 with LiDAR data to estimate aboveground biomass of subtropical forests in northeast Guangdong, China. *International Journal of Digital Earth*, 16(1), 158-182. https://doi.org/10.1080/17538947.2023.2165180

APPENDICES

Appendix A: RMSE, validation R^2 and AIC values for different categories

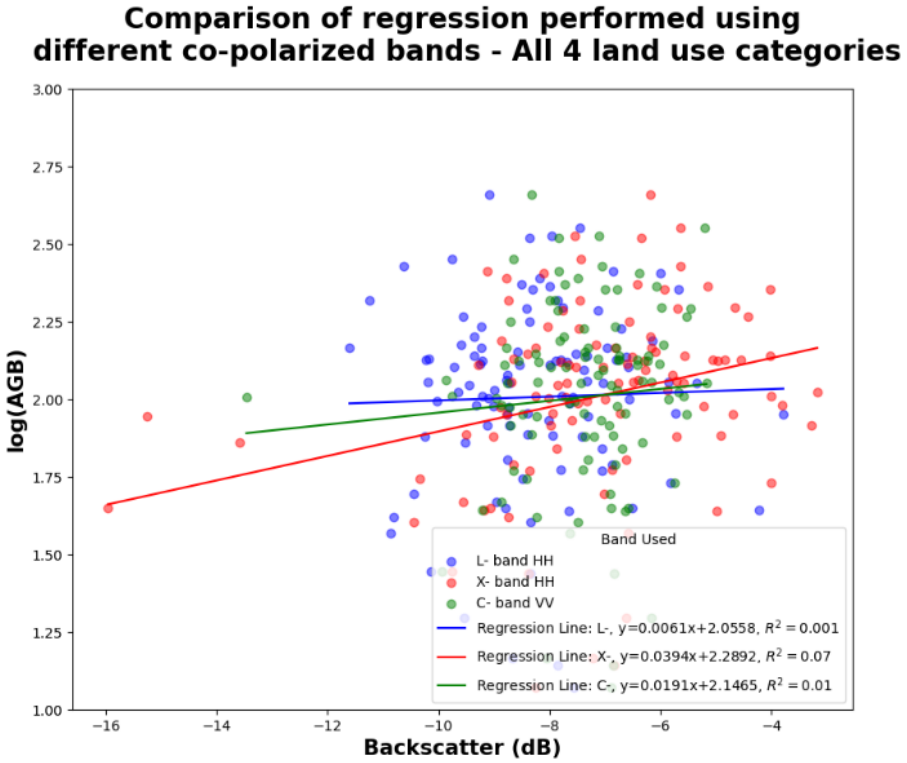
Bands	Land use	RMSE (tons/ha)	Relative RMSE
X-	Rubber, Shrubland, Oil palm and Forest	54.89	0.69
C-	Rubber, Shrubland, Oil palm and Forest	57.23	0.72
L-	Rubber, Shrubland, Oil palm and Forest	57.78	0.73
C- + X-	Rubber, Shrubland, Oil palm and Forest	55.45	0.70
L- + C-	Rubber, Shrubland, Oil palm and Forest	57.81	0.73
X- + L-	Rubber, Shrubland, Oil palm and Forest	55.44	0.70
X- + C- + L-	Rubber, Shrubland, Oil palm and Forest	55.92	0.71
X-	Forests and oil palm plantations	59.66	0.72
C-	Forests and oil palm plantations	53.74	0.65
L-	Forests and oil palm plantations	50.74	0.61
C- + X-	Forests and oil palm plantations	54.87	0.66
L- + C-	Forests and oil palm plantations	49.8	0.60
X- + L-	Forests and oil palm plantations	50.42	0.61
X- + C- + L-	Forests and oil palm plantations	51.00	0.62
X-	Oil palm plantations	25.26	0.71
C-	Oil palm plantations	33.36	0.94
L-	Oil palm plantations	34.33	0.97
C- + X-	Oil palm plantations	25.11	0.71
L- + C-	Oil palm plantations	35.36	1.00
X- + L-	Oil palm plantations	26.88	0.76
X- + C- + L-	Oil palm plantations	26.82	0.75
X-	Forests	68.68	0.81
C-	Forests	67.52	0.80
L-	Forests	60.82	0.72

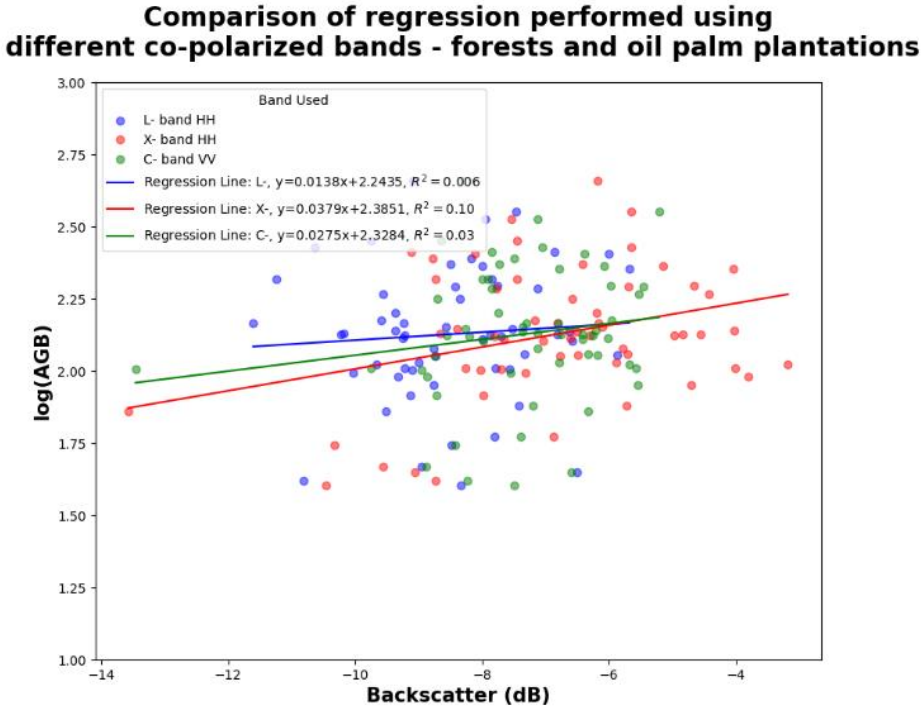
C- + X-	Forests	66.85	0.79
L- + C-	Forests	62.49	0.74
X- + L-	Forests	62.62	0.74
X- + C- + L-	Forests	63.9	0.75

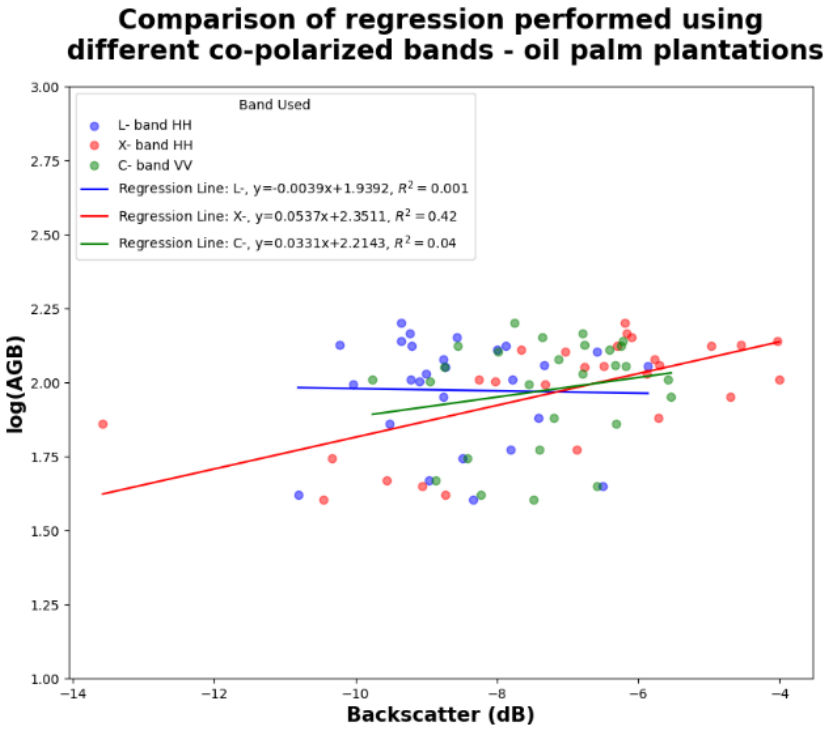
Bands	Land use	Validation R^2
X-	Rubber, Shrubland, Oil palm and Forest	-0.06
C-	Rubber, Shrubland, Oil palm and Forest	-0.09
L-	Rubber, Shrubland, Oil palm and Forest	-0.10
C- + X-	Rubber, Shrubland, Oil palm and Forest	-0.06
L- + C-	Rubber, Shrubland, Oil palm and Forest	-0.11
X- + L-	Rubber, Shrubland, Oil palm and Forest	-0.05
X- + C- + L-	Rubber, Shrubland, Oil palm and Forest	-0.06
X-	Forests and oil palm plantations	-0.08
C-	Forests and oil palm plantations	0.06
L-	Forests and oil palm plantations	0.26
C- + X-	Forests and oil palm plantations	0.02
L- + C-	Forests and oil palm plantations	0.26
X- + L-	Forests and oil palm plantations	0.25
X- + C- + L-	Forests and oil palm plantations	0.23
X-	Oil palm plantations	0.27
C-	Oil palm plantations	-0.17
L-	Oil palm plantations	-0.30
C- + X-	Oil palm plantations	0.29
L- + C-	Oil palm plantations	-0.33
X- + L-	Oil palm plantations	0.16
X- + C- + L-	Oil palm plantations	0.16
X-	Forests	-0.14
C-	Forests	-0.11
L-	Forests	0.02
C- + X-	Forests	-0.13
L- + C-	Forests	-0.02
X- + L-	Forests	-0.01
X- + C- + L-	Forests	-0.06

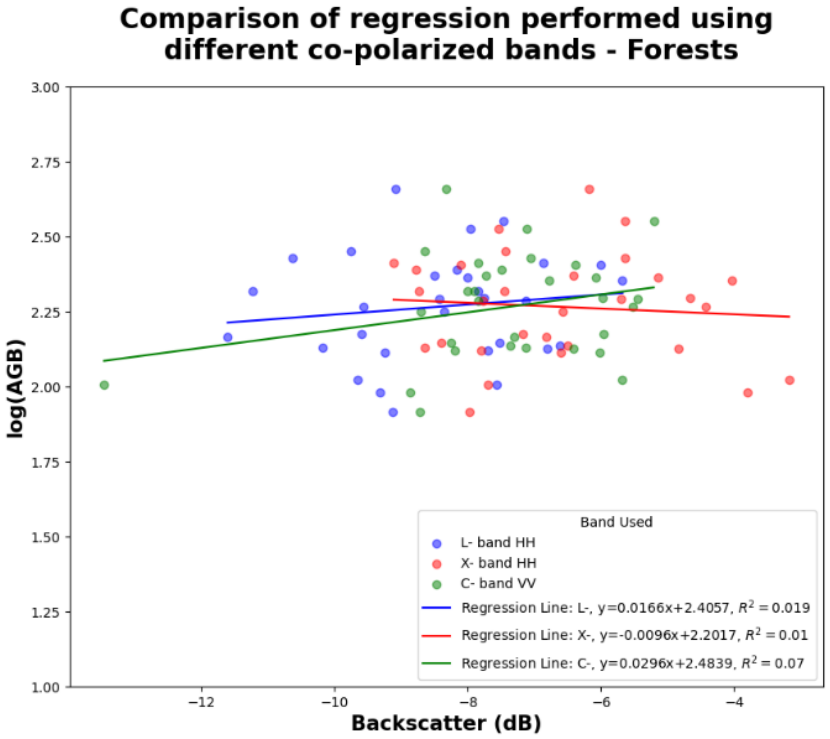
Bands	Land use	AIC values
X-	Rubber, Shrubland, Oil palm and Forest	58.47
L-	Rubber, Shrubland, Oil palm and Forest	58.15
C-	Rubber, Shrubland, Oil palm and Forest	57.09
C- + X-	Rubber, Shrubland, Oil palm and Forest	59.08
L- + C-	Rubber, Shrubland, Oil palm and Forest	59.09
X- + L-	Rubber, Shrubland, Oil palm and Forest	60.14
X- + C- + L-	Rubber, Shrubland, Oil palm and Forest	61.08
X-	Forests and oil palm plantations	-0.42
L-	Forests and oil palm plantations	-23.53
C-	Forests and oil palm plantations	-14.25
C- + X-	Forests and oil palm plantations	-12.41
L- + C-	Forests and oil palm plantations	-23.58
X- + L-	Forests and oil palm plantations	-24.70
X- + C- + L-	Forests and oil palm plantations	-23.75
X-	Oil palm plantations	-10.25
L-	Oil palm plantations	-10.37
C-	Oil palm plantations	-11.60
C- + X-	Oil palm plantations	-9.61
L- + C-	Oil palm plantations	-9.63
X- + L-	Oil palm plantations	-8.51
X- + C- + L-	Oil palm plantations	-7.68
X-	Forests	-14.05
L-	Forests	-20.16
C-	Forests	-15.28
C- + X-	Forests	-13.72
L- + C-	Forests	-18.18
X- + L-	Forests	-18.69
X- + C- + L-	Forests	-16.69

Appendix B: Comparison of regression between co-polarizations of bands.






Comparison of regression performed using different co-polarized bands - oil palm plantations
Band Used
L- band HH
X- band HH
C- band VV
Regression Line: L-, y=-0.0039x+1.9392, R^2 = 0.001
Regression Line: X-, y=0.0537x+2.3511, R^2 = 0.42
Regression Line: C-, y=0.0331x+2.2143, R^2 = 0.04
log(AGB)
Backscatter (dB)


Comparison of regression performed using different co-polarized bands - Forests
Band Used
L- band HH
X- band HH
C- band VV
Regression Line: L-, y=0.0166x+2.4057, R^2 = 0.019
Regression Line: X-, y=-0.0096x+2.2017, R^2 = 0.01
Regression Line: C-, y=0.0296x+2.4839, R^2 = 0.07
log(AGB)
Backscatter (dB)

Appendix C: Residual Standard Errors of the models.

Bands	Land use	RSE (tons/ha)
L-	Rubber, Shrubland, Oil palm and Forest	78.15
C-	Rubber, Shrubland, Oil palm and Forest	78.04
X-	Rubber, Shrubland, Oil palm and Forest	77.00
L- + C-	Rubber, Shrubland, Oil palm and Forest	78.11
C- + X-	Rubber, Shrubland, Oil palm and Forest	76.91
X- + L-	Rubber, Shrubland, Oil palm and Forest	76.20
X- + C- + L-	Rubber, Shrubland, Oil palm and Forest	76.55
X-	Forests and oil palm plantations	81.22
C-	Forests and oil palm plantations	75.06
L-	Forests and oil palm plantations	67.52
C- + X-	Forests and oil palm plantations	75.73
L- + C-	Forests and oil palm plantations	67.70
X- + L-	Forests and oil palm plantations	67.19
X- + C- + L-	Forests and oil palm plantations	67.74
X-	Oil palm plantations	27.85
C-	Oil palm plantations	35.24
L-	Oil palm plantations	36.03
C- + X-	Oil palm plantations	27.7
L- + C-	Oil palm plantations	35.96
X- + L-	Oil palm plantations	28.37
X- + C- + L-	Oil palm plantations	28.33
X-	Forests	85.48
C-	Forests	84.56
L-	Forests	78.64
C- + X-	Forests	85.40
L- + C-	Forests	80.00
X- + L-	Forests	85.01
X- + C- + L-	Forests	81.50

Appendix D: Comparison of slopes and intercepts between other land use categories

1. showing % confidence that the slopes and intercepts are statistically different between models created with **forests and shrubland** data.

Band	**% confidence- slopes**	**% confidence- intercepts**
X- band	n.s	99.9%
C- band	n.s	99.9%
L-- band	n.s	99.9%

2. showing % confidence that the slopes and intercepts are statistically different between models created with **forests and rubber** data.

Band	**% confidence- slopes**	**% confidence- intercepts**
X-- band	n.s	99.9%
C- band	n.s	99.9%
L-- band	n.s	99.9%

3. showing % confidence that the slopes and intercepts are statistically different between models created with **oil palm and rubber** data.

Band	**% confidence- slopes**	**% confidence- intercepts**
X- band	n.s	n.s
C- band	n.s	n.s
L-- band	n.s	n.s

4. showing % confidence that the slopes and intercepts are statistically different between models created with **rubber and shrubland** data.

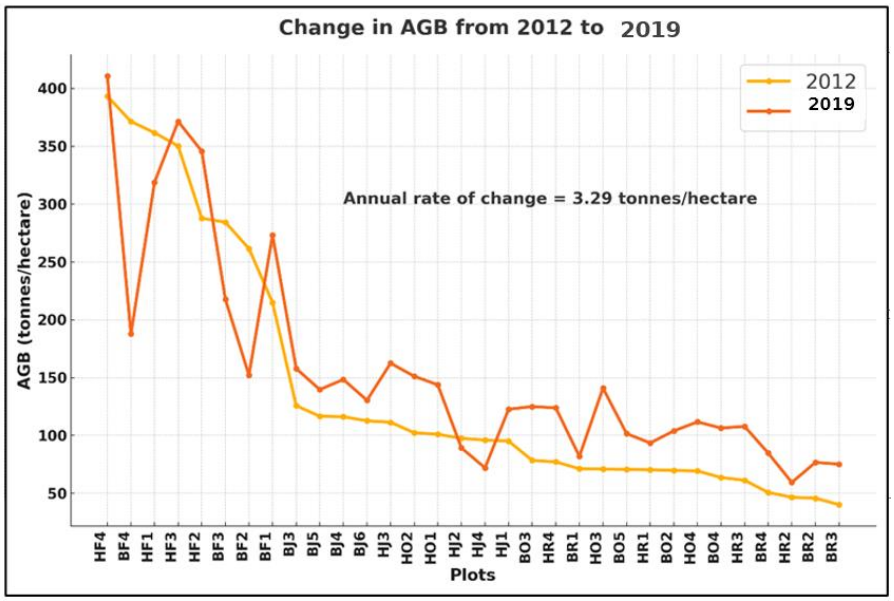
Band	**% confidence- slopes**	**% confidence- intercepts**
X-- band	n.s	99.9%
C- band	n.s	99.9%
L-- band	n.s	99.9%

5. showing % confidence that the slopes and intercepts are statistically different between models created with **oil palm and shrubland** data.

Band	% confidence- slopes	% confidence- intercepts
X-- band	n.s	99%
C- band	n.s	99.9%
L-- band	n.s	99.9%

Appendix E: AGB change in the field plots from 2012 to 2019

AGB data for the field plots were also collected in 2012 and 2019. The plot size was different than the one in 2021. The plots from these two years were selected due to the consistent plot size of these two years. The common plots were joined from 2012 and 2019 to calculate the annual growth rate of the plots.
To determine if the temporal inconsistency caused due to the difference in the acquisition of the satellite imagery and ground truth data is significant, firstly, the growth rate of the oil palm, rubber and forest vegetation was calculated by using the AGB plot data from 2012 and 2019. The growth rate was 3.29 tons/ha per year. Hence in two years, the growth was 6.58 tons/year. To determine whether this growth is substantial to affect the model's performance, the Residual Standard Errors (RSE) of the models were compared to the growth rate. The RSE of the models varied between 78.15 and 76.20 tons/ha using data for all land use categories together, 27.7 to 36.03 and 85.48 to 78.64 for models created for predicting AGB of oil palm plantations and forests respectively (Appendix C).



Appendix F: Tree height change in the plots from 2020 to 2022