

https://dspace.org/

Modelling thicket above-ground biomass using LiDAR and SAR data, in the Addo Elephant National Park

Mkheswa, Bongani

2021
https://hdl.handle.net/10539/33103

Downloaded from a DSpace repository. https://dspace.org/

MODELLING THICKET ABOVE-GROUND BIOMASS USING LIDAR AND SAR DATA, IN THE ADDO ELEPHANT NATIONAL PARK

Bongani Mkheswa



A research report submitted to the Faculty of Science, University of the Witwatersrand, in partial fulfilment of the requirement for the degree of

Master of Science in Geographical Information Systems and Remote Sensing

Johannesburg, 2021

Supervisors:

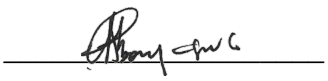
Dr Russell Main: Advanced Agriculture and Food Cluster, CSIR, Pretoria, South Africa

Dr Laven Naidoo: Advanced Agriculture and Food Cluster, CSIR, Pretoria, South Africa

Dr Elhadi Adam: University of the Witwatersrand, Johannesburg School of Geography, Archaeology and Environmental Studies

DECLARATION

I declare that this research report is my own unaided work. It is being submitted for the degree of Master of Science (Geographical Information Systems and Remote Sensing) in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.



Bongani Mkheswa

2288647

September 2021 in Johannesburg

ACKNOWLEDGMENTS

I would like to thank my supervisors for giving their guidance and encouragement. I thank the Council of Scientific and Industrial Research (CSIR) for providing LiDAR and field data used in this research. I would also like to thank JAXA for providing ALOS PALSAR 30m global mosaic data and The European Space Agency for providing Sentinel-1 SAR data. I also appreciate the opportunity that I was given by the University of Wits to further my studies. Above all, I am thankful to my family for all the support.

ABSTRACT

Due to the ecological effects arising from climate change, it is necessary to quantify terrestrial carbon stocks at various spatial and temporal scales for mitigation purposes. Whereas deforestation is known to increase carbon dioxide emissions and negatively impact the fight against global climate change, studies have also shown that increased carbon results in woody thickening and the transformation from open savanna to bush encroached thickets.

Hence, this study sought to quantitatively model the Above Ground Biomass (AGB) within the Addo Elephant National Park in South Africa. The dominant vegetation type of the park is the unique and threatened Albany thicket. Field-based AGB measurements and LiDAR metrics were used to derive a local AGB model that represented the Addo thicket vegetation type, and could then be upscaled using regional-scale satellite-based Synthetic Aperture Radar (SAR) backscatter values. Furthermore, this study sought to determine the performance of Sentinel-1 C band SAR data versus ALOS PALSAR-2 L band SAR data in retrieving the thicket AGB. Specific objectives to the study were to use field-collected AGB plots and airborne LiDAR data to develop and apply a LIDAR-based thicket AGB model to the LiDAR dataset; to use the LiDAR-based AGB product as calibration and validation data in exploring different modelling techniques for estimating regional scale SAR-based AGB and to investigate the effects of SAR wavelength in retrieving thicket AGB of the Albany thicket biome by comparing Sentinel-1 multi-temporal model against the Global mosaic ALOS PALSAR-2 L band model.

The coefficient of determination (R^2), the Root Mean Square Error (RMSE) and the relative Root Mean Square Error (rRMSE) were used to evaluate and compare the performance of the models. Field-based AGB measurements were successfully upscaled using LiDAR metric and the best Simple Linear Regression (SLR) model established was $\widehat{\boldsymbol{AGB}} = \mathbf{0.258}(\boldsymbol{CC} * \boldsymbol{MH})$ with an R^2 of 89%, RMSE of 3.779t/ha and rRMSE of 35.6%. Then Simple linear regression and Random Forest modelling techniques were used to derive a regional AGB model. The Random Forest model yielded respectable results (R^2 of 67% and rRMSE of 38.4%) and outperformed the Simple linear regression model. Furthermore, the time series of Sentinel-1 C band data showed higher retrieval results (give best result, R^2 & rRMSE) than the one-off ALOS PALSAR-2 L band data (give best result, R^2, rRMSE) in retrieving AGB using both Random Forest and Simple Linear Regression techniques.

Importantly, the study demonstrated the effective use of machine learning predictive mapping techniques for thicket AGB using freely available Sentinel-1 C band remote sensing data by taking advantage of its relatively high spatial and temporal resolution.

Keywords: Above-ground biomass, Random Forest, Albany thicket, Synthetic Aperture Radar

TABLE OF CONTENTS

DECLARATION ..... **i**
ACKNOWLEDGMENTS ..... **ii**
ABSTRACT ..... **iii**
CHAPTER 1: INTRODUCTION **1**
1.0 Introduction 1
1.1 Problem Statement 3
1.2 Aim of the study 4
1.3 Objectives 4
CHAPTER 2: LITERATURE REVIEW **5**
2.0 Global warming and Climate change 5
2.1 The Carbon Cycles 6
2.2 The Slow Cycle 7
2.3 The Fast Carbon Cycle 7
2.4 Biomass 8
2.5 Methods of estimating forest above-ground biomass 10
2.5.1 Harvest method 10
2.5.2 Allometric Equations 10
2.5.3 Remote Sensing based methods 11
2.6 SAR signal wavelengths 13
2.7 SAR Polarizations 14
2.8 LiDAR 15
2.9 Modelling Techniques 16
2.10 Simple linear regression model (SLR) 16
2.11 Random Forest machine learning algorithm (RF) 17
CHAPTER 3: MATERIALS AND METHODS **19**
3.0 Study area 19
3.1 Synthetic Aperture Radar (SAR) data pre-processing steps 20

3.2 Subsetting 21
3.3 Radiometric calibration 21
3.4 Speckle reduction 21
3.5 Terrain Correction 22
3.6 Convert from sigma0 to Direct Backscatter (dB) 22
3.7 Data acquisition 22
3.7.1 Field data 22
3.7.2 LiDAR data 26
3.7.3 Sentinel-1 26
3.7.4 ALOS PALSAR-2 imagery 27
3.8 Methodology 29
3.9 Field plot data and LiDAR data 31
3.10 Calculation of Field Plot AGB 31
3.11 LiDAR Metrics and AGB model 33
3.12 Validation method for small samples 33
3.13 Calculation of the prediction model 34
3.14 Hypothesis Test of each model 34
3.15 Hypothesis Test of the Intercept B_0 35
3.16 Hypothesis Test of the coefficient of CC*MH for each model 35
3.17 Final LiDAR AGB model 36
3.18 Hypothesis Test of Final LiDAR AGB model 36
3.19 Hypothesis of the Final LiDAR AGB model intercept 36
3.20 Hypothesis of the Final LiDAR AGB model β_1 36
3.21 Sar-based predictors variable selection 37
3.22 Data Cleaning 37
3.23 Histogram 38
3.23 Boxplots 39
3.24 QQ plots 40

3.25 Analysis of Predictor variables and SAR AGB Model 40
3.26 Mapping procedure and interpretation 41
3.27 Modelling LiDAR AGB and SAR variables 41
3.28 Model Selection 42
3.29 Comparison of Sentinel-1 C band and ALOS PALSAR-2 L band 42
4.0 LiDAR AGB model 43
4.1 Exploratory data analysis 44
4.2 LiDAR AGB Summary statistics 45
4.3 LiDAR AGB Model Validation 45
4.4 LiDAR-derived Addo Thicket AGB map 46
4.5 Goodness of fit Sentinel-1 C and ALOS PALSAR-2 L Models 47
4.5.1 Sentinel Simple Linear Regression model 47
4.5.2 ALOS-PALSAR Simple linear Regression model 48
4.5.3 Random Forest Regression Model 48
4.5.4 ALOS PALSAR-2 Random Forest 49
4.6 Final Best Model Selection 50
4.7 Signal Saturation 51
4.8 Results: Sentinel-1 VH C-band Addo Thicket AGB map 52
4.9 Sentinel 1 SAR Results 53
4.10 Fence line effect 54
4.11 Unsampled land cover 54
4.12 Topographical influence 54
CHAPTER 5: DISCUSSIONS CONCLUSIONS AND RECOMMENDATIONS 56
5.0 Discussions 56
5.0.1 LiDAR model 56
5.0.2 Satellite-based models 57
5.0.3 Modelling Techniques 57
5.1 Conclusions 57

5.1.1 Models 58
5.1.2 Comparison of C and L SAR bands 58
5.3 Recommendations 58
LIST OF REFERENCES:.....60

LIST OF FIGURES

Figure 2. 1: The Fast carbon cycle (NASA, 2020) 8
Figure 2. 2**:** Illustration of C and L band with vegetation interactions (Kumar et al., 2019) ..14

Figure 3. 1**:** Study locality map showing AENP located in the province of Eastern Cape, the LiDAR footprints and the field sample plots. 19
Figure 3. 2: Workflow for working with Satellite SAR data 20
Figure 3. 3: (a) ALOS PALSAR-2 mosaics & Sentinel-1a over AENP, displaying HV and VH polarizations, respectively. 28
Figure 3. 4: Flow chart of steps used for forest above-ground biomass using SAR imagery .....
……..30
Figure 3. 5: Histograms of Sentinel-1 VH time series 39
Figure 3. 6: Boxplots of Sentinel-1 VH time series 39
Figure 3. 7: QQ plots of Sentinel-1 VH time series 40
Figure 4. 1: Field AGB vs CC*MH scatterplot 44
Figure 4. 2: Observed vs Predicted scatter plot 45
Figure 4. 3: LiDAR AGB map 46
Figure 4. 4: Observed vs Predicted scatter plot of Sentinel SLR 47
Figure 4. 5: Observed vs Predicted scatter plot of the PALSAR 48
Figure 4. 6: Predicted vs Actual scatter plot Sentinel -1 RF 49
Figure 4. 7: Predicted vs Actual scatter plot ALOS PALSAR-2 mosaics 50
Figure 4. 8: The scatter plots of sampled AGB and SAR backscatter from Sentinel 1 52
Figure 4. 9: Above Ground Biomass Map Over the Sunday River district 53
Figure 4. 10: Comparison of AGB map results to google earth image 55

LIST OF TABLES

Table 2. 1: Summary of SAR bands and their applications (Franz, 2019) 13
Table 3. 1: Species encountered during the survey and the available species-specific models and dry: wet ration used for each to estimate dry biomass (DM). 23
Table 3. 2: Basic statistics of the field measurements of the sample trees 25
Table 3. 3: Summary of satellite images to be used in this research 28
Table 3. 4: Plot level AGB in Tonnes per Hector 32
Table 3. 5: Model Regression statistics 34
Table 3. 6: Sentinel-1 and ALOS PALSAR-2 predictors 37
Table 3. 7: Times series VH bands nomenclature 38
Table 4. 1: Addo thickets Sample Field AGB 43
Table 4. 2: Addo thickets Sample Field AGB Summary Statistics 44
Table 4. 3: Summary of the goodness of fit Statistics values of the models 51

-

CHAPTER 1: INTRODUCTION

1.0 Introduction

Over the years, there has been an increase in the occurrence of catastrophic natural phenomena such as droughts, floods, heatwaves, and wildfires, which can be attributed to global climate change. One of the factors that contribute to these occurrences is the increase of greenhouse gases in the atmosphere of which carbon dioxide is one of the major. Therefore, it is important to quantify terrestrial carbon stocks at various spatial and temporal scales (Waring and Running, 2007) which can help in formulating measures and strategies for combating global climate change. One of the most important carbon sinks within the carbon cycle on land are vegetation (80% of carbon sequestrated on the terrestrial biosphere is found in vegetation). Therefore, plants play a very important role in mitigating climate change (Fuchs et al., 2009; Chen et al., 2018).

Based on the assessment of the Intergovernmental Panel on Climate Change (IPCC), deforestation and unsustainable use of energy which result in increased carbon dioxide emissions negatively impact the fight against global climate change. IPCC estimated that 11% of the world's anthropogenic emissions can be attributed to carbon dioxide emissions (Van derWerf et al., 2009; Berninger et al., 2018). On the contrary, some studies have shown that increase in atmospheric carbon dioxide has resulted in woody thickening (Zaloumis and Bond, 2016) whereby broad-leaved trees shade grass resulting in a complete biome shift from savanna to a bush encroached thicket-type state (Parr et al., 2012). Bond & Midgley (2016) found that there is rapid bush encroachment in southern Africa across various land use types (i.e. conservation, commercial and communal lands). Most importantly, Wigely et al (2010) proved the hypothesis that increased woody thickening is driven by an increase in atmospheric carbon. The consequences of this vegetation shift have negative effects on ecosystem functions, human livelihoods, and biodiversity (Parr et al., 2012). Therefore, tools that can help in understanding the quantities and distribution of vegetation Above Ground Biomass (AGB) are of paramount importance in quantifying carbon stocks, management of forest as well as monitoring the impacts of climate change (Fuchs et al., 2009; Maselli, 2011).

In recent years, it has been demonstrated that remote sensing techniques can provide a multi-scale contiguous estimate of AGB, which are ideally suited for modelling over broad spatial and temporal scales. Remote sensing can be easily used to acquire measurements in areas that

are difficult to reach (e.g. private forest, game reserves, difficult terrain, and remote areas) using traditional methods, thereby enabling us to efficiently estimate carbon stocks in such area (Palace et al., 2015; Asner and Mascaro, 2014).

The establishment of relationships between forest structure and remotely sensed variables, like spectral signatures and synthetic aperture radar (SAR) backscatter has enabled us to design satisfactory AGB estimation models. The ability of SAR to penetrate the forest canopy has enabled it to produce higher accuracies in retrieving forest-AGB, compared to optical sensors. The ability of SAR signal to penetrate different types of vegetation is wavelength dependent. Longer wavelengths tend to penetrate deeper into the vegetation while shorter wavelengths only interact with the vegetation canopy. Hence SAR backscatter has a problem of saturation, particularly for high forest-AGB levels ranging from 60 to 100 Mgha−1 (Li et al., 2018, Ni et al.,2014, Hamdan et al., 2014). To solve this problem, one needs to incorporate variables such as backscatter ratios and various textures, which help to maximise the woody structure related signal sensitivities while reducing signal noise (e.g., topographic bias) (Page et al., 2011; MacKinnon et al., 2013; & Li et al., 2018).

Li et al., (2018) highlighted that airborne LiDAR is another active remote sensing tool that is immune to the saturation problem prevalent in multispectral and SAR sensors and at the moment there is no spaceborne LiDAR available with only NASA's Global Ecosystem Dynamics Investigation (GEDI) LiDAR sensor still in the testing phase. Forest parameters such as canopy height (Naesset & Bjerknes, 2001), crown size, canopy cover (CC), leaf area index (LAI), (Morsdorf et al., 2006; Sasaki et al.,2008; Zhao & Popescu,2009) and AGB (Pang & Li,2013; E. Hansen et al.,2015) have been successfully estimated using airborne LiDAR.

However, it is expensive to acquire airborne LiDAR data over large areas, making this method of AGB estimation unachievable for many resource managers, especially for those in developing regions (Li et al., 2018). Therefore, there is an increase in the number of studies showing ways of utilising small-scale LiDAR-derived AGB to upscale satellite imagery variables to estimate AGB at much larger scales (Li et al., 2018; Berninger et al., 2018).

The basis of upscaling LiDAR Forest AGB models using Satellite data lies in establishing a correlation between LiDAR and satellite imagery variables (Chen et al., 2011; Tsui et al., 2013; Hudak et al., 2012; Korhonen et al., 2011; Berninger et al., 2018). Some studies have been carried out to show the existence of such relationships. For example, Korhonen et al., (2013) determined that LiDAR canopy cover metrics were closely related to SPOT vegetation index

and field measurements. Also, a correlation has been proven between LiDAR-based AGB and SAR backscatter (Joshi et al.,2015; Chen et al 2018, Berninger et al., 2018; Li et al., 2018). However, it should be noted that these relationships vary depending on factors such as tree species, terrain conditions, quality of data and environmental conditions (Li et al., 2018). Another important factor that should be considered when dealing with LiDAR metrics is point density which can be influenced by LiDAR data collection pattern and by forest structure heterogeneity (Li et al.,2015).

Various methods have been used to estimate the amount of above-ground forest biomass across land-scapes and to create biomass maps for specific stands or pixels across ownership or project areas. Without an accurate estimation method, land managers might end up with incorrect biomass estimate maps, which could lead to poor decisions in their future management plans (Donal et al 2014).

1.1 Problem Statement

Due to the density of thickets, it is difficult to collect field-based samples, hence thicket biomes are generally underrepresented in remote sensing-based research particularly on research where SAR is used. Recently van der Vyver and Cowling (2019) estimated aboveground biomass of this subtropical thicket using species-specific allometric models, a method which can be tedious and limited in spatial coverage. Hence, this research seeks to derive Addo thicket AGB by upscaling field-based AGB measurements using LiDAR metrics and then subsequently derive a regional model of the Addo thicket AGB model using satellite-based SAR backscatter values. This has the advantage of not only increasing the sample size considerably but also makes it possible to have samples in difficult to access landscapes.

Furthermore, this research will investigate the performance of temporal Sentinel-1 C band SAR data versus (one-off) ALOS PALSAR-2 L band data in retrieving AGB. This is especially relevant because even though ALOS PALSAR-2 data has been proven to perform well in forest biomes, the behaviour of both Sentinel-1 C band and ALOS PALSAR-2 L have not been extensively investigated in the Albany thicket biome that is prevalent in the Addo Elephant Nation Park (AENP). Should Sentinel-1 data be able to satisfactorily estimate AGB of the AENP thickets, this will assist in the monitoring and management of this endemic vegetation type as Sentinel-1 datasets are hyper-temporal and free whilst ALOS PALSAR-2 free data set is a global mosaic which is released once every year.

1.2 Aim of the study

To quantifiably estimate the AGB of the Albany thicket dominated AENP using up-scaled field measurements and remotely sensed SAR data, while also assessing the estimation errors associated with different modelling algorithms and SAR frequencies.

1.3 Objectives

I. Use field-collected AGB plots and airborne LiDAR data to develop and apply a LiDAR-based thicket AGB model to the sample LiDAR dataset.

II. Use the LiDAR-based AGB product as calibration and validation data in exploring different modelling techniques (e.g., SLR, RF each representing parametric and machine learning algorithms respectively) for estimating regional scale SAR-based AGB.

III. Investigate the accuracies associated with different SAR wavelengths in retrieving thicket AGB by comparing multi-temporal C-band Sentinel-1 data to once-off L-band ALOS PALSAR-2 global mosaic data.

The study of the vegetation type found in AENP which belongs to the broader Albany Thicket biome of South Africa is important due to the endemism of this biome to the Easter Cape, it is a highly fragmented biome nested in a mosaic of other biomes and is under threat of extensive degradation due to overgrazing and invasion by non-thicket species (e.g. Grassland and Nama-Karoo elements) (Pierce & Mader, 2006, Mucina & Rutherford 2006). Hence this biome is worth studying and conserving as its destruction may result in habitat loss, degradation, and establishment and spread of invasive vegetation, and increased poaching of rare and endangered fauna and flora in this area (Vlok et al., 2003; Eastern Cape Biodiversity Conservation Plan (ECBCP), 2017).

The novelty of this research lies in upscaling AGB from field measurements to LiDAR and then to satellite measured SAR data in mapping the Albany thicket biome AGB as well as investigating the accuracies associated with different SAR wavelengths in retrieving thicket AGB. This study attempted to contribute to the development of remote sensing-based predictive mapping techniques for thicket AGB using freely accessible multi-source remote sensing data with a relatively high spatial resolution.

CHAPTER 2: LITERATURE REVIEW

2.0 Global warming and Climate change

The Intergovernmental Panel on Climate Change (IPCC) defined Global warming as the rise in the earth's mean climate temperature owing to human activity. Climate change is an umbrella term that is used to describe rising average temperatures, extreme weather events, shifting wildlife populations and habitats, rising seas, and a range of other impacts of global warming (NASA, 2020). NASA has demonstrated the phenomena of climate change by direct global temperature measurements as well as measurements of various effects such as shrinking ice sheets, glacial retreat, decreased snow cover, sea-level rise, declining arctic sea ice, extreme events such as wildfires and intense rain fall, and ocean acidification (NASA, 2020).

All these changes are emerging as humans continue to add heat trapping greenhouse gases to the atmosphere. The Intergovernmental Panel on Climate Change (IPCC) concluded that "anthropogenic activities that have continuously added heat-trapping greenhouse gases such as carbon dioxide, methane, and nitrous oxide to the atmosphere since the mid-20th century are the major drivers of climate change" (Len et al., 2018). The by-products of fossil combustion are the major contributors to these gases, together with agricultural emissions and deforestation being other notable contributors (EPA 2019).

Climate change has adversely affected agricultural productivity resulting in food scarcity in some communities and an increase in mean sea levels also threatens the existence of coastal communities and their infrastructure (NASA, 2020). Extinction or migration of certain species due to change of ecosystems is one of the prominent consequences of climate change and loss of snow and permafrost increases water vapour which results in "feedback effects" on the carbon cycle which further accelerate global warming (EPA 2017).

Development of technologies with low carbon footprints, policies that are aimed at reducing carbon emissions, forest conservation and reforestation efforts as well as research in climate engineering are some of the strategies that are being pursued to mitigate climate change. On the other hand, research on more resilient crop types, fortification of coastal infrastructure, development of efficient disaster monitoring and management tools as well as adaptation to current and future impacts of climate change are some of the ways governments and societies are addressing the challenge of climate change.

Carbon dioxide as the main component of the earth's biogeochemical cycle has been monitored for the last 400 years (NASA, 2020). The records indicate that since the Industrial Revolution, atmospheric carbon dioxide concentration has been increasing from approximately 280±10 parts per million (ppm) to the 2020 record high of 413±10 ppm and these trends are expected to continue in that trajectory into the future (United Nations IPCC,2007; NASA 2020).

Even though about half of anthropogenic carbon dioxide is absorbed through natural processes (Jones et al., 2006), studies suggest that feedbacks (climate-carbon feedbacks) will further reduce the ability of the terrestrial and oceanic carbon cycles to absorb anthropogenic carbon dioxide. As a result, the carbon storage capacity of forest ecosystems and its potential to ameliorate climate change has become one of the focus areas of research (Stoffberg et al., 2010; IPCC 2007; McGranahan et al., 2005).

2.1 The Carbon Cycles

Carbon is the main component of carbon dioxide and is one of the prominent drivers of global warming. It is released to the atmosphere through deforestation, burning of organic matter as well as natural means like volcanic activity and respiration. Therefore, there is no way we can understand climate change processes without studying the carbon cycle.

The carbon cycle is composed of reservoirs or sinks namely:

- rocks
- oceans
- Atmosphere
- Plants
- soil and
- fossil fuels.

The carbon cycle is the process that describes the propagation of carbon through these sinks. The propagation process includes two distinct routes, namely:

- Slow cycle
- Fast cycle.

Any imbalance within these cycles results in the depletion of carbon from one reservoir and accumulation of carbon on the other. Accumulation of more carbon in the atmosphere results in a rise in temperature and vice versa.

2.2 The Slow Cycle

In the slow cycle, it takes more than 100 million years for carbon to complete this cycle. Major drivers of the slow cycle are tectonic activities and chemical reactions. Rocks constitute the biggest sink in this cycle sequestering 80% of carbon with the remainder being stored as organic carbon. The products of this organic carbon cycle are shale, a form of sedimentary rock formed under heat and pressure and in some special cases organic carbon is converted into natural gas, coal, and oil. The main vehicle by which carbon finds its way back to the atmosphere is by way of volcanoes.

2.3 The Fast Carbon Cycle

The completion of one cycle in the fast carbon life cycle is measured by way of a lifespan of the major components of this cycle namely:

- Plants and
- Phytoplankton (microscopic organisms found in the ocean)

The fast cycle can be described as primarily the propagation of carbon through life forms in the biosphere. In one year, 1 000 to 100 000 million metric tons of carbon moves through this cycle whilst anthropogenic activity alone moves at least 1015 grams of carbon over the same period

The primary method by which carbon is stored in this cycle is when plants and phytoplankton combine carbon dioxide, water, and solar energy in a chemical reaction to produce sugar and oxygen.

In turn, carbon is released back into the atmosphere when sugar and oxygen react in a process called combustion to release carbon dioxide water and energy.

The fast cycle is so dependent on plants that the concentration of carbon in the atmosphere fluctuates according to different plant growing seasons.

Figure 2.1 shows carbon propagation through the three major sinks in the fast carbon cycle in gigatons. Red numbers represent emissions attributed to humans, and white numbers represented stored carbon and yellow are natural movements of carbon through the cycle.

As can be seen from figure 2.1, plants contribute significant storage of carbon at 550 gigatons, hence forest AGB is a notable variable in the carbon cycle. Considering this fact, we can conclude that, accurate estimation of AGB is vital in understanding the carbon cycle (Solberg et al., 2010; Banskota et al., 2011). Initiatives such as Reduction of Emissions due to Deforestation and Forest Degradation (REDD) and other international efforts aimed at reducing global warming can benefit from spatially explicit measurement of global biomass (Englhart et al., 2011; Toan et al.,2011).

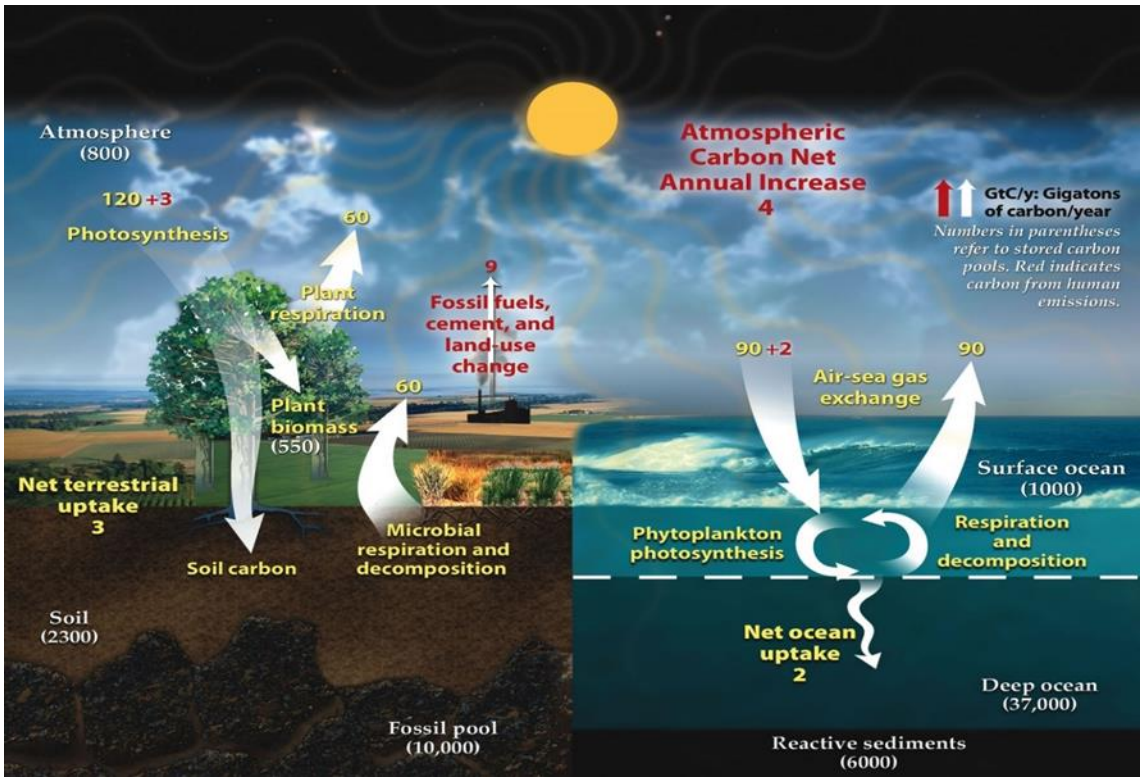


Figure 2. 1: The Fast carbon cycle (NASA, 2020)

2.4 Biomass

Biomass can be divided into two types i.e., above-ground and below-ground living mass. Above-ground biomass includes trees, shrubs, vines etc while below-ground biomass includes roots, tubers and dead mass of vegetation litter associated with soil. Biomass is always quantified in dry weight terms e.g., tree above-ground biomass (AGB) is the weight of that part

of a tree above the ground surface when dried until its weight is constant. Plot-level biomass result from the summation of individual trees within that plot and then standardized for the area covered by that plot, hence their unit is given as kilograms per square metre (kg m^{-2}).

When it comes to the global ecosystems, the impact of biomass cannot be overemphasised with regards to the distribution of nutrients in the soil, accrual of fuel, the carbon cycles as well as habitat environments. This is so because regional biomass changes directly influence climate change and the main functions of the ecosystem as biomass potentially controls the release of carbon into the atmosphere because of deforestation.

Biomass governs the potential carbon emission that could be released to the atmosphere due to deforestation, and regional biomass changes have been associated with important outcomes in ecosystem functional characteristics and climate change. The roles and impacts of biomass on carbon cycles, soil nutrient allocations, fuel accumulation, and habitat environments in terrestrial ecosystems have long been recognized. Hence Foody et al., (2003) rightly stated that "Accurate delineation of biomass distribution at scales ranging from local and regional to global becomes significant in reducing the uncertainty of carbon emission and sequestration, understanding their roles in influencing soil fertility and land degradation or restoration and recognizing their roles in environmental processes and sustainability".

Above-ground biomass has been widely researched compared to below-ground biomass partly due to the difficulty of collecting field samples associated with below-ground biomass. Traditional methods that have been used to estimates biomass include the harvest method where trees are manually harvested, oven-dried and then weighed. This method is destructive in nature, and it is impractical over large scales.

After the harvest method came in the use of allometric equations, these empirical equations are derived from the harvest method, and they use tree dimensions to calculate AGB. Limitations of this method include that they are usually built on locally available tree samples and do not encompass all species. Also, allometric equations are limited by a particular height and Diameter at Breast Height (DBH) range (based on what is sampled). Trees outside these ranges can yield unpredictable estimates.

Even though remote sensors do not directly measure biomass, remote sensing and GIS techniques are progressively becoming important in forest biomass estimation (Lu 2006). Besides the ease of collecting data, remote sensing has many advantages in AGB estimation over traditional methods with the main one being its ability to estimate AGB at different scales.

Methods of indirectly estimating AGB with the use of remotely sensed variables include parametric and non-parametric models as well as a plethora of machine learning algorithms like Random Forest and neural networks (Breiman 2001; Goetz et al., 2005; Hansen et al., 2008), and MaxEnt (maximum entropy) (Phillips et al., 2006; Saatchi et al., 2011). Data from various sensors can be used for regional and global biomass mapping e.g., biomass data from field measurements (e.g., FIA-Forest Inventory and Analysis plots; Blackard et al., 2008), LiDAR (GLAS-Geoscience Laser Altimeter System) (Blackard et al., 2008; Nelson et al., 2009), and image data from LANDSAT, MODIS have been used together to perform regional biomass mapping.

2.5 Methods of estimating forest above-ground biomass

2.5.1 Harvest method

The harvest or direct method consists of harvesting the actual tree, oven drying it and then weighing each of the tree components e.g., stem, roots, branches, and leaves (Brown et al., 1989, Parresol, 1999). This destructive method is not only environmentally unfriendly but laborious too, especially if there are large trees involved and the study includes Below Ground Biomass (BGB) as well. Plot size also influences the outcome of calculated biomass for example, using large plots with small trees can underestimate the average biomass density and vice versa.

In general, it is impractical to use this method in areas with high biomass density, over large scales or to use this method for biomass monitoring as it is not possible to repeat the measurements (Houghton et al., 2009).

2.5.2 Allometric Equations

The use of empirically based allometric equations is another method used to calculate AGB. This method allows us to use easy to measure tree dimensions like DBH and height as inputs to estimate tree biomass (Whittaker and Woodwell, 1968). DBH and height are some of the parameters used to build allometric equations. Some are linear while others are exponential or logarithmic. Wood density (species dependent) and broadleaf versus fine leaf categorisation are also considered in some equations.

2.5.3 Remote Sensing based methods

Remote sensing methods can easily acquire data on a large scale thereby saving time and money. Remotely sensed observations do not directly measure biomass, instead, the correlation between AGB and radiometrically sensed plant parameters, shadows and textures are used to deduce biomass (Baccini et al., 2008). Biomass estimation sensors in remote sensing can be classified as passive (e.g. optical multi-spectral) or active (e.g. SAR and LiDAR).

2.5.3.1 Passive optical sensors

Using optical sensors, biomass modelling is based on establishing a relationship between biomass vegetation indices or spectral variables derived from multi-spectral images and field measurements. Lu (2002), established that there is a strong relationship between vegetation biomass and the visible bands of the electromagnetic spectrum, whilst some tree parameters can have a positive (Spanner et al., 1990) or negative (Danson and Curran, 1993) correlation to near-infrared wavelength factors such as canopy shadowing and diminished understory brightness etc. In general, in all bands, shadowing plays a crucial role in AGB modelling.

Several AGB estimation methodologies have been developed to utilise the correlation between optical vegetation indices and forest biophysical parameters (such as leaf area index, the fraction of absorbed photosynthetically active radiation, and percent ground cover) in order to circumvent challenges caused by atmospheric effects, sun view angle, soil background and geometry of forest canopy.

Even though results can be dependent on the properties of the study area, studies have shown that the use of vegetation indices (such as the Perpendicular Vegetation Index (PVI); Soil-Adjusted Vegetation Index (SAVI); Modified Soil-Adjusted Vegetation Index (MSAVI); Global Environmental Monitoring Index (GEMI)) in estimating biomass can partially reduce the impact on reflectance (Richardson and Wiegand, (1977); Huete, (1988); Qi et al., (1994); Pinty and Verstraete, (1992)).

In general, Lu (2004) concludes that “vegetation indices can partially reduce the impacts on reflectance caused by environmental conditions, thus improve the correlation between AGB and vegetation indices, especially in those sites with complex vegetation stand structures” (Lu et al., 2004).

Although optical sensors have been used to estimate AGB, they will not be utilised in this study because using a 2D image product to ascertain a 3D environment variable is not suitable (Cai et al., (2019); Urbazaev et al., (2018). Furthermore Urbazaev et al., (2018), stated that optical sensors saturate at high biomass. The other disadvantages in using optical data for AGB estimation are that they are affected by cloud cover and are strongly dependent on environmental, seasonal and acquisition conditions (e.g., solar zenith angle).

2.5.3.2 Active Sensors

Two active remote sensing technologies, SAR and LiDAR have been used to add a 3D dimension to AGB models. SAR and LiDAR differ in that they use the energy of different wavelengths; SAR uses microwave energy and LiDAR uses visible and near-infrared wavelengths (Hudak et al., (2002)).

The use of SAR data to estimate AGB assumes that the forest floor, large branches, and tree trunks will scatter all SAR frequencies similarly, but that the foliar components of the vegetation will primarily scatter higher frequencies (Dobson et al., (1992)). The SAR sensor first transmits microwave signals and then receives back the signals that are returned, or backscattered, from the Earth's surface. Using its backscatter, radar data has been widely used to estimate AGB due to its ability to measure various layers of forest structure including woody and canopy components. Another advantage of the Radar systems is that they are active sensors, and they are not dependent on the sun but rather emit their own energy and hence can collect data day and night. Depending on the wavelength, e.g., X-band can display atmospheric effects in imagery while C- and L-band do not. Radar has been used successfully in AGB estimation on several occasions (Hussin et al., 1991; Saatchi and Moghaddam, 1995; Imhoff et al., 2000; Saatchi et al., 2007; Simard et al., 2008; Koch, 2010; Sexton et al., 2009; Ghasemi et al., 2011). SAR based biomass estimations are limited by canopy density because closed canopies often do not allow enough SAR penetration for biomass estimates to be accurate (Dobson et al., 1992).

LiDAR biomass estimations differ from optical estimations in that LiDAR takes measurements of physical canopy properties rather than spectral properties. Airborne Light Detection and Ranging (LiDAR) systems yield the most accurate estimates of aboveground biomass for

forested areas over a wide range of biomass values and it is not hindered by canopy density in measuring the forest's biophysical properties.

2.6 SAR signal wavelengths

Radar sensors make use of various microwaves, which are used for a variety of applications in remote sensing. Table 2.1 list some of the available SAR bands based on their wavelengths as well as their respective uses in environmental applications.

Table 2. 1: Summary of SAR bands and their applications (Franz, 2019)

BAND	FREQUENCY	WAVELENGTH	TYPICAL APPLICATION
X	8 - 12 GHz	3.8 - 2.4cm	High resolution SAR (urban monitoring, ice and snow, little penetration into vegetation cover, fast coherence decay in vegetated areas)
C	4 - 8 GHz	7.5 - 3.8cm	SAR (global mapping; change detection, monitoring of areas with low to moderate penetration, higher coherence); ice, ocean maritime navigation
L	1 - 2 GHz	30 - 15cm	Medium resolution SAR (geophysical monitoring, biomass, and vegetation mapping; high penetration, InSAR)
P	0.3 - 1 GHz	100 - 15cm	Biomass, vegetation mapping and assessment. Experimental SAR

- The X-band is scattered by leaves and canopy cover surface, so it is suitable to extract information about the surface layer of the trees.
- The C-band penetrates through leaves and scatters by small branches and under layer elements.
- The L-band penetrates through the surface layers and is scattered by the trunk and main branches.

- The P-band penetration into the canopy cover and the major parts. Hence P-band backscattering is caused by the trunk and the ground reflectance.

Therefore, P-band and L-band are generally regarded as the most suitable in retrieving tree biophysical variables (Ghasemi et al., 2011). Figure 2.2 illustrates typical interactions of C-band and L-band SAR signals with vegetation in a forest with broad leaves and thicker trunks and branches, the L-band signal penetrates deeper through the vegetation and can acquire more information about the forest structure than the C-band signal. However, in this study, C-band will be used to investigate its ability to retrieve biomass for the Albany thicket biome due to the unique characteristics of this biome composed of small shrubby leaves and small or twig-like branches which theoretically should allow shorter wavelength signals like that of C-band to penetrate further the canopies. Sentinel-1 was particularly proposed for this research as it is the only operational C-band sensor with freely available hyper-temporal data.

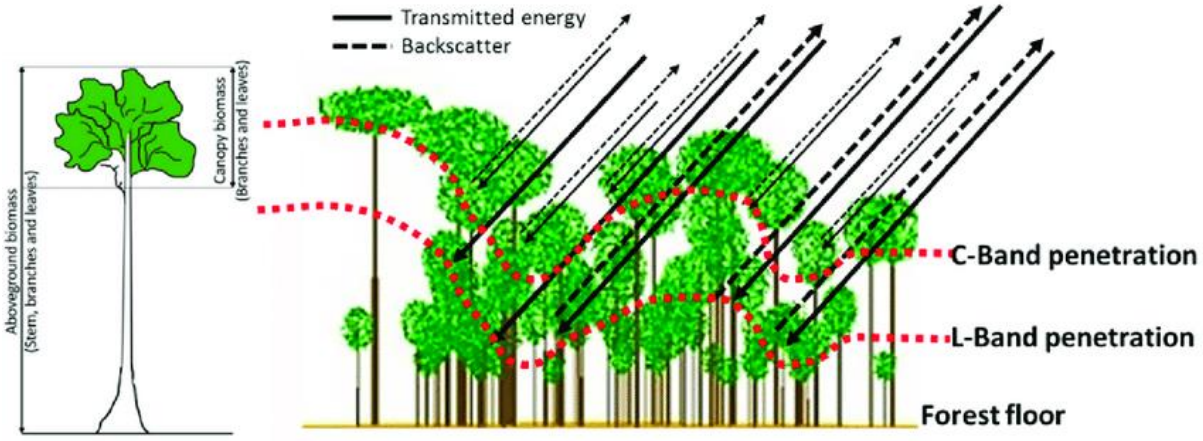


Figure 2. 2**:** Illustration of C and L band with vegetation interactions (Kumar et al., 2019)

2.7 SAR Polarizations

Signal polarization is another important property of SAR data. According to Ghasemi et al (2011), polarization is the direction of the electric field in the electromagnetic waves, and it is the main factor into the interaction between signals and the reflectors. Ordinarily, horizontal (H) and vertical (V) polarizations are the primary signals emitted by microwave instruments, however, SAR data may contain four polarizations i.e. (HH, HV, VH and VV).

Sun (2002) demonstrated that HV polarization and longer wavelengths (L and P bands) are better suited for retrieving AGB for dense forests and woodlands (Sun et al., 2002), because they have longer wavelengths which better penetrates foliage and because HV measures the volumetric response associated with vegetation. This improves AGB estimation due to better linear correlation with woody biomass and the problem of saturation at higher AGB is reduced (Milne and Dong 2002). Lu (2006) also highlighted that saturation levels for SAR data also depended on vegetation structure, topography and variations in canopy and soil moisture of the study area. High diversity of succulents may have a negative effect on the backscatter interactions due to high moisture content and subsequent changing dialectic constants hence in this study a time series of Sentinel-1 and multi-temporally filter will be used to minimise this effect.

2.8 LiDAR

A LiDAR system is composed of a laser, which emits laser pulses mounted on a platform, it illuminates the scene, then measures the lapse between the pulse emission and reception, hence it is an active sensor. LiDAR footprint refers to the laser pulse illumination scene which is governed by the laser divergence and the altitude of the LiDAR instrument (Rosette et al., 2012). There are two main classes of LiDAR i.e., full waveform and discrete return LiDAR with the former measuring entire returned laser signal which is related to the vertical distribution of canopy structure (Dubayah and Drake, 2000) and the latter captures the first, intermediate and the last laser pulses (Rosette et al., 2012).

The ability of LiDAR to derive tree volumes has enabled us to use the correlation of volume to measured tree biomass to estimate biomass on much larger scales. When using LiDAR discrete returns two surfaces are extracted i.e., Digital Surface Model (DSM) which comes from the first returns and Digital Terrain Models (DTM) with the former defining vegetation canopy and the latter defining the ground of the study area (Hyyppa et al., 2004). Then Canopy Height Model is calculated by subtracting DSM from DTM to get the vegetation height.

Advantages of using LiDAR systems lies in its ability to collect data quickly and very accurately, high sample density, its ability to penetrate dense forest to accurately collect elevation and terrain data, LiDAR data does not suffer any geometric distortions like a side-looking radar and that it can be integrated with other data sources. Its disadvantage is that it is a relatively new technology hence it is very expensive to use for data acquisition over large

areas and also acquiring sufficient point density is needed for detailed tree structure modelling, which depends on flight height and sensor fidelity.

2.9 Modelling Techniques

Regression analysis is simply a statistical technique for investigating the relationship among variables. What differs are methods of doing it, associated errors and how to interpret the results. Regression analysis is used when you want to predict a continuous dependent variable or response from several independent or input variables. If the dependent variable is dichotomous, then logistic regression should be used. In this study, we used two classes of regression techniques namely parametric and non-parametric machine learning algorithms

2.10 Simple linear regression model (SLR)

Linear regression is a linear approach to modelling the relationship between a response variable and one or more explanatory variables. In the case of one explanatory variable, the model is called "simple linear regression". If there is more than one explanatory variable, the process is called multiple linear regression (Cohen et al., 2003).

The general equation of a linear relation is as follows:

$$\hat{y} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 x} \quad \text{..... 2.1}$$

"Where $\hat{y}$ is the response variable and x is the explanatory variable and β_0 and β_1 is the intercept and the slope coefficient respectively".

To fit a linear regression, the following assumptions must be made about the predictor variables, the response variables, and their relationship:

Weak exogeneity: the predictor variables x can be treated as fixed values, rather than random variables.

Linearity: there should exist a linear relationship between response and explanatory variables.

Homoscedasticity: the variance of the residuals should be constant.

Independence of errors: the errors of the response variables are uncorrelated with each other.

2.11 Random Forest machine learning algorithm (RF)

In the past two decades, machine learning techniques, such as the artificial neural network (ANN) (Zeng et al., 2017), support vector machines (SVM) (Dusseux et al., 2014), and random forest (RF) (Fassnacht et al., 2014) have been increasingly used for the study of biomass. Random Forest in particular have the capacity to reduce bias and overfitting by creating as many trees as possible on the subset of the data and combining the output of all the trees (Mutanga et al., 2012; Karlson et al., 2015; Ramoelo et al., 2015). Moreover, studies have shown that Random Forest can automatically handle missing values, can efficiently deal with non-linear parameters, does not require feature scaling i.e., standardisation and normalisation of the data and it is usually robust in handling outliers (Wang et al. 2017, Anaya et al., 2009; Gleason and Im, 2012).

As a classification and regression tree (CART) technique proposed by Breiman (2001), RF combines bagging (Breiman 1996; Breiman 2001) with random variable selections at each node (Amit and Geman, 1997) to iteratively generate a large group of CARTs. The classification output represents a majority vote (classification), or an average (regression) from the whole ensemble, and hence achieves a more robust model than a single classification tree that is produced by a single model run (Breiman 2001; Dhanda et al.; 2017; Genuer et al., 2017).

Several decision trees in RF choose their best splitting attributes from a random subset of predictors at each internal node without pruning. Based on the bootstrap sampling procedure, RF ensures at the same time the smallest obtainable bias and very low data variance (Jovic and Bogunovic., 2011).

The performance of a Random Forest is dependent upon the values of two model parameters, namely ntree and mtry. The parameter ntree denotes the number of trees to be used in the model. Li et al (2016) used 500 trees as a starting point in their research, hence in this research, ntree=500 was used as a starting value and the number will be increased until the value where the model performs optimally is determined. As a rule of the thump, the number of trees should be large enough so that the model error no longer meaningfully declines as the number of trees increases and this was achieved through experimentation since this value is dependent on the data set being used.

The other parameter is mtry which denotes the random number of predictor variables chosen at each split in each tree in the forest. To determine the optimal value requires experimentation as above since mtry also depends on the dataset being used. In this research mtry ($\frac{1}{3}$) of the total number of predictor variables was used as the starting point and the number was increased progressively until an optimal value was determined.

CHAPTER 3: MATERIALS AND METHODS

3.0 Study area

Addo Elephant National Park (AENP) situated in the endemic-rich succulent thicket of the Albany Centre (Van Wyk & Smith, 2001; Vlok et al. 2003), is a diverse wildlife conservation park established in 1931 located 60km north-east of Ggeberha (Port Elizabeth) in the province of the Eastern Cape in South Africa (Landman et al. 2008).

The park covers over 1 640km^2 in a semi-arid region with annual rainfall ranging from 260mm to 530mm with the highest rainfall in June to July and lowest rainfall in December to February. According to the South African Weather Service (SAWS) temperatures are mild from May to October ranging from an average maximum of 21 – 25^{o}C and from November to April temperatures are hot averaging 30^{o}C.

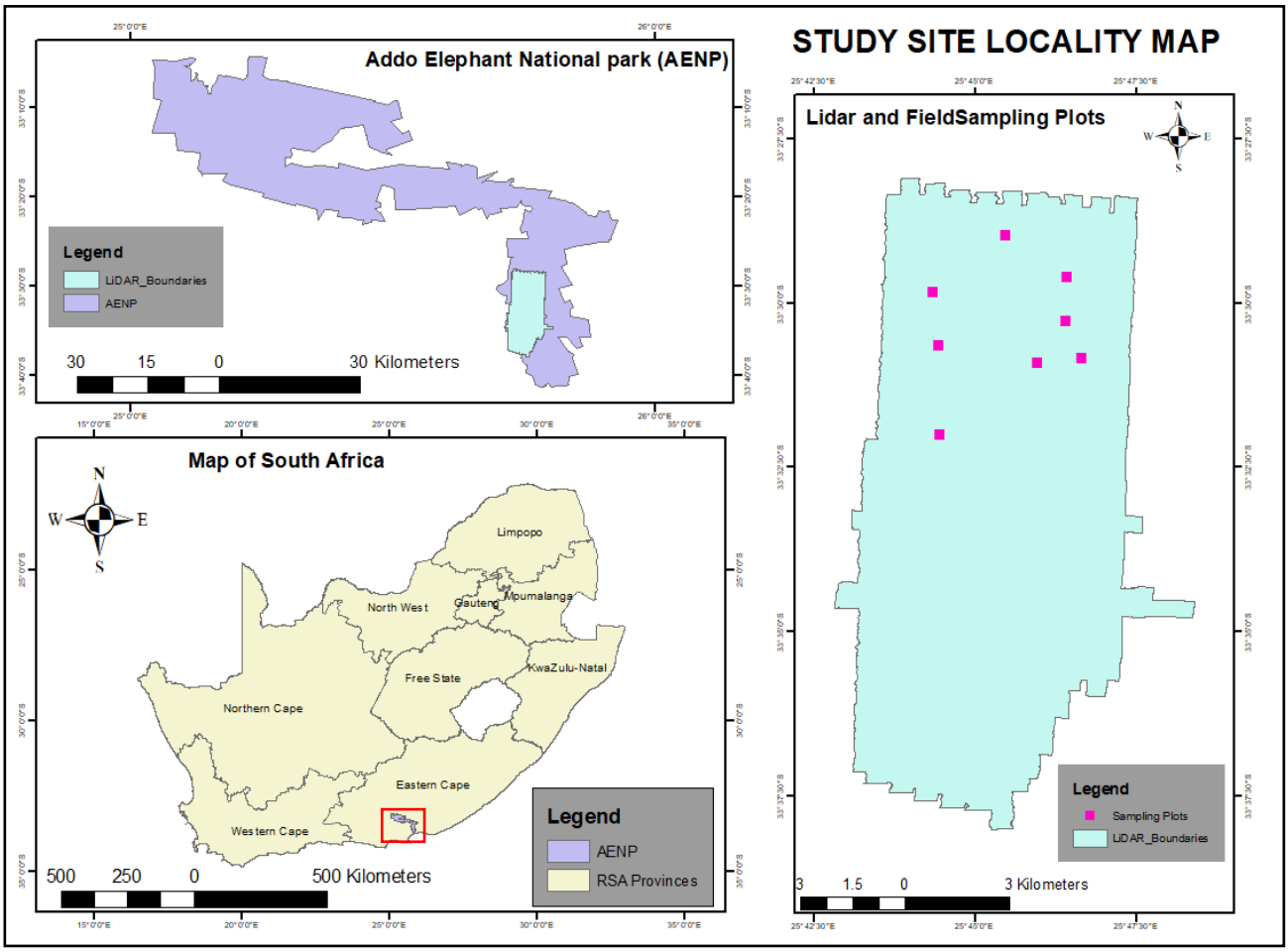


Figure 3. 1**:** Study locality map showing AENP located in the province of Eastern Cape, the LiDAR footprints and the field sample plots.

The vegetation is characterized by a high diversity of growth forms, including succulents, evergreen and deciduous shrubs, lianas, herbs, geophytes, and grasses (Vlok et al., 2003). There is not much literature as to how the Albany thicket biome interacts with SAR but based on the characteristics of the vegetation which according to van der Vyver and Cowling (2019), consist of a matrix of the large succulent shrub, within which other canopy species, both woody and succulent, often form clumps. It is worth investigating how this type of vegetation interact with different types of SAR wavelength signals. In this study, both C-band and L-bands were used to model the AGB of this biome.

3.1 Synthetic Aperture Radar (SAR) data pre-processing steps

This study focused on using SAR imagery to retrieve AGB of a subtropical Addo thicket biome. Hence it is important to zoom in and pay attention to the pre-processing steps to be followed such as downloading satellite SAR data from various web portals before meaningful analysis and inferences can be drawn from such data.

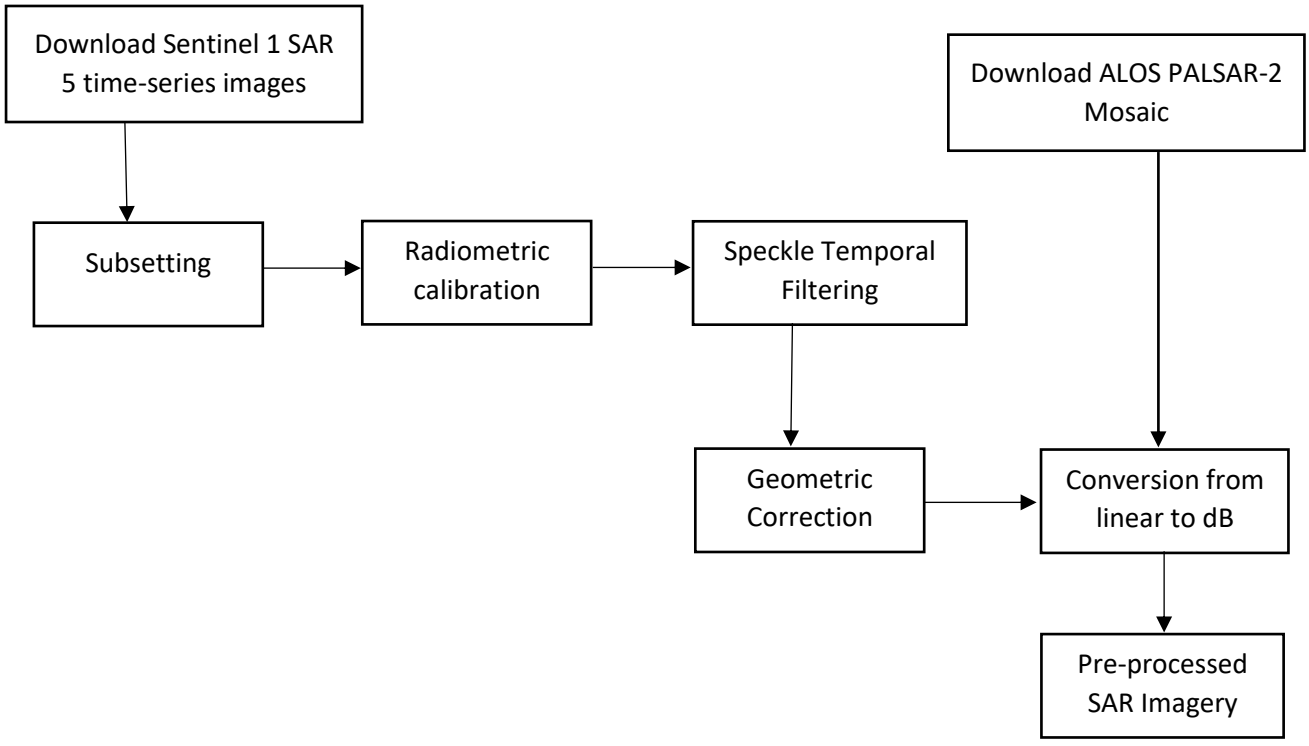


Figure 3. 2: Workflow for working with Satellite SAR data

Figure 3.2 shows a general approach one can follow in pre-processing satellite-based SAR data. The workflow consisted of subsetting, radiometric calibration, speckle reduction using temporal filtering, and geometric correction to acquire accurate radar intensity backscatter coefficient with a map projection (Veci, 2015; Liu, 2016) as well as converting the backscatter values from their default DN values to decibels (dB) values.

3.2 Subsetting

Satellite SAR imageries are usually big with a swath size of approximately 250km^2 hence, to preserve computing power, it is necessary to subset to a smaller and manageable size by using the raster subset function which can be found in various softwares such as in SNAP, Erdas Imagine, ArcMap and QGIS.

3.3 Radiometric calibration

All SAR data should be calibrated before they can be used in any quantitative analysis. Radiometric calibration corrects a SAR image so that the pixel values truly represent the radar backscatter of the reflected surface. The main radiometric distortions are due to signal loss as it propagates, non-uniform antenna pattern, the difference in gain, saturation, and speckle.

The corrections that are applied during calibration are mission-specific, therefore the software will automatically determine the required input product and corrections need based on the product's metadata. Radiometric calibration is important for comparing images from different sensors, modalities, processors, or images acquired at different times (NASA).

3.4 Speckle reduction

Speckle is caused by random constructive and destructive interference resulting in salt and pepper noise throughout the image and makes interpretation and analysis of the images difficult. Speckle filters and muiltilooking are some of the techniques that are used to extract radar information from images that have speckle. In this study, a series of SAR images were temporally filtered, using the Quegan filter (Quegan and You, 2001). This filter linearly combines images of given time series to create a new set of speckle reduced data, as well as a local mean backscattering coefficient that is estimated in a window around each pixel (Quegan

and You, 2001). The filter serves to reduce the variance within the image while preserving both radiometric and spatial resolution. The whole temporal filtering process was run in R-Studio.

3.5 Terrain Correction

Terrain corrections geocode the image by correcting SAR geometric distortions using a digital elevation model (DEM) to produce a map projected product. Geocoding converts an image from Slant Range or Ground Range Geometry into a Map Coordinate System. In this research, a Range Doppler Terrain Correction was applied. Using SNAP, the relevant SRTM30m DEM tile was automatically downloaded and used to minimise inherent SAR geometry effects such as foreshortening, layover and shadow. Bilinear interpolation was used to resample both the DEM and the SAR image to 10m resolution, and the product was reprojected to UTM zone 35 South on a WGS 1984 datum.

3.6 Convert from sigma0 to Direct Backscatter (dB)

The sigma0 values are converted to a physical quantity called the backscattering coefficient or normalised radar cross-section measured in decibel (dB) units with values ranging from +5 dB for very bright objects to -40 dB for very dark surfaces.

3.7 Data acquisition

3.7.1 Field data

In selecting sampling sites, the study area was first divided into six distinct strata based on the vegetation type (Vlok et al., 2003) and length of exposure to elephants and the stratum were named as follows:

- E31Y = Elephants present for 31 Years (Sundays Spekboom Thicket vegetation)
- E37Y = Elephants present for 37 Years (Sundays Spekboom Thicket vegetation)
- OEC = Original Elephant Camp, longest continuous exposure to elephants (Sundays Spekboom thicket vegetation)
- BRSST = Sundays Spekboom Thicket vegetation in the Botanical Reserve (elephants excluded)
- BRST = Sundays Thicket vegetation in the Botanical Reserve (Spekboom not major component)
- BV = Bontveld vegetation

Because of the difficulty of gaining access into the Addo thickets, eight $25m^2$ plots were selected using purposive sampling. While the bontveld thicket type was sampled as a single $25m^2$ plot, of the remaining seven plots a further three $5m^2$ subplots were selected within these large plots for sampling. Hence a total of 22 subplots were sampled during the field survey.

Fieldwork was carried out by Marius L. van der Vyver and his team from Nelson Mandela University in May 2018 and raw data collected in the field were, the height of vegetation (hgt), two perpendicular crown diameter measurements (Dai1) and (Dai2). For Aloe speciose, Stem length (SL) was also recorded since it is a predictor variable required by the Aloe speciose's species allometric model. The table below shows a list of 63 species that were sampled and their respective allometric models used to estimate green/wet biomass (Model) as well as a list of the species-specific dry: wet ration models (DWratio).

Table 3. 1: Species encountered during the survey and the available species-specific models and dry: wet ration used for each to estimate dry biomass (DM).

No.	Species	Model	DWratio
1	Azima tetracantha	Azima tetracantha	Azimatetracantha
2	Schotia afra	Schotia afra	Schotia afra
3	Euclea undulata	Euclea undulata	Euclea undulata
4	Portulacaria afra	Portulacaria afra	Portulacaria afra
5	Capparis sepiara	Capparis sepiara	Searsia longispina
6	Pappea capensis	Pappea capensis	Pappea capensis
7	Sensevieria hyacinthoides.elliptica	Aloe striata	Aloe striata
8	Eleadendron zeyheri	Ptaeroxylon obliquum	Pappea capensis
9	Rhoicissus digitata	Ehretia rigida	Searsia longispina
10	Searsia longispina.pterota	Gymnosporia polyacantha	Searsia longispina
11	Lauridia tetragona	Pappea capensis	Pappea capensis
12	Putterlickia pyracantha	Putterlickia pyracantha	Gymnosporia capitata
13	Asparagus crassicladus	Asparagus capensis	Asparagus capensis
14	Crassula pubescens	Crassula pubescens	Crassula pubescens
15	Scutia myrtina	Capparis sepiaria	Gymnosporia polyacantha
16	Asparagus suavolens	Asparagus capensis	Asparagus capensis
17	Grewia occidentalis	Grewia robusta	Grewia robusta

18	Hippobromus pauciflorus	Ptaeroxylon obliquum	Schotia afra
19	Sideroxylon inerme	Pappea capensis	Pappea capensis
20	Gymnosporia capitata	Gymnosporia polyacantha	Gymnosporia capitata
21	Diospyros dichrophylla	Euclea undulata	Euclea undulata
22	Vachellia karroo	Vachellia karroo	Vachellia karroo
23	Olea europaea	Boscia oleoides	Boscia oleoides
24	Clausena anisata	Ptaeroxylon obliquum	Schotia afra
25	Carissa haematocarpa	Carissa haematocarpa	Carissa haematocarpa
26	Euclea natalensis	Euclea undulata	Euclea undulata
27	Zanthoxylum capense	Ptaeroxylon obliquum	Schotia afra
28	Pittosporum viridiflorum	Ptaeroxylon obliquum	Schotia afra
29	Searsia pallens	Grewia robusta	Searsia longispina
30	Euphorbia mauritianica	Euphorbia mauritianica	Euphorbia coerulescens
31	Crassula perforata	Crassula perforata	Crassula perforata
32	Asparagus striatus	Asparagus capensis	Asparagus capensis
33	Drosanthemum sp.	Drosanthemum lique	Drosanthemum lique
34	Crassula muscosa	Crassula muscosa	Crassula muscosa
35	Crassula tetragona	Crassula perforata	Crassula perforata
36	Psilocaulon junceum	Psilocaulon junceum	Mesembryanthemum splendens
37	Pteronia incana	Galenia filiformis	Galenia filiformis
38	Lycium cinereum	Lycium cinereum	Lycium ferocissimum
39	Asparagus sp.	Asparagus capensis	Asparagus capensis
40	Cadaba aphylla	Euphorbia mauritanica	Euphorbia coerulescens
41	Asparagus africanus	Asparagus capensis	Asparagus capensis
42	Gymnosporia heterophylla	Gymnosporia polyacantha	Gymnosporia polyacantha
43	Plumbago auriculata	Plumbago auriculata	Plumbago auriculata
44	Chrysocoma ciliata	Pteronia incana	Pteronia incana
45	Justicia capensis	Plumbago auriculata	Plumbago auriculata
46	Kalanchoe rotundifolia	Crassula perforata	Crassula perforata

47	Cotyledon woodii	Crassula ovata	Crassula ovata
48	Crassula mesembryanthemoides	Crassula mesembryanthemoides	Crassula esembryanthemoides
49	Crassula ovata	Crassula ovata	Crassula ovata
50	Aloe speciosa	Aloe speciosa	Aloe striata
51	Cotyledon orbiculata	Cotyledon velutina	Crassula ovata
52	Asparagus multiflorus	Asparagus capensis	Asparagus capensis
53	Diospyros whyteana	Euclea undulata	Euclea undulata
54	Plectranthus madagascariensis	Plumbago auriculata	Plumbago auriculata
55	Helichrysum rosum	Pteronia incana	Pteronia incana
56	Jathropa capensis	Jathropa capensis	Jathropa capensis
57	Sarcostemma viminale	Euphorbia mauritanica	Euphorbia coerulescens
58	Ptaeroxylon obliquum	Ptaeroxylon obliquum	Schotia afra
59	Searsia crenata	Grewia robusta	Searsia longispina
60	Euphorbia sp.	Euphorbia coerulescens	Euphorbia coerulescens
61	Searsia lucida	Grewia robusta	Searsia longispina
62	Pelargonium peltatum	Jathropa capensis	Jathropa capensis
63	Brachylaena ilicifolia	Brachylaena ilicifolia	Boscia oleoides

Table 3.2 summaries the basic statistics of the sampled trees. The height of the Addo thicket biome ranges from 5cm to 520cm which means the vegetation type is characterised by bushes and shrubs. The mean crown diameter refers to the average measurement of the two perpendicular crown diameter measurements.

Table 3. 2: Basic statistics of the field measurements of the sample trees

Field measurements	**Min**	**Max**	**Mean**	**Median**	**Std**
Height (cm)	5.00	520	116.00	80.00	99.44
Mean Crown Diameter (cm)	3.50	733	105.43	64.00	113.19
Stem length (cm)	27.00	79.00	49.20	50.00	20.94

The mean crown diameter ranges from 3.5cm to 733cm and the height distribution of the sample vegetation is skewed to the right. The stem length was measured only for Aloe speciosa which required that predictor in its best available allometric model.

3.7.2 LiDAR data

LiDAR data were collected over 10 310 ha in May 2018, with an Optech ALTM M300 and Optech Gemini laser scanner mounted on Partenavia P68B aircraft operated by Southern Mapping Company.

The LiDAR system was flown at an altitude of approximately 700m and had a density of approximately 9 points per m^2. An opposing flight line side-lap of ≥25% was maintained to increase the point density.

Due to budget constraints, independent verification of the LiDAR accuracy was not carried out, however, the vertical and horizontal accuracy of ±0.10m is assumed. The procedure for deducing the two LiDAR metrics from the LiDAR point cloud namely: Mean vegetation height (CHM) and Canopy Cover (CC) was adopted from (Naidoo et al., 2015), as follows:

Raw LiDAR point cloud data were classified into ground and non-ground points, and then processed to deduce a 2m Digital Elevation Model (DEM) and top-of-canopy surface models (CSM). Then a Canopy Height Model (CHM) with 2m pixel size was created by subtracting the DEM from the CSM.

The Woody Canopy Cover (CC) which Jennings et al (1999) defined simply as the area which woody plant canopies vertically project on a horizontal surface was created by classifying the LiDAR CHM image into a binary image of two classes, namely zeros (no woody canopy) and 1's (presence of woody canopy). The percentage cover could then be estimated by summing the area of all the vegetation pixels (i.e. 1's) and dividing by the area of the chosen aggregated pixel resolution (i.e. $25m^2$).

3.7.3 Sentinel-1

Sentinel 1 SAR imagery data acquired in January and February 2018, which coincides with the regions dry season were downloaded free of charge from the Alaska Satellite Facility (ASF) website. Sentinel 1 C-band images were collected in the interferometric wide swath mode of the VH and VV polarization, with a temporal resolution of ±6 days, a pixel size of 30m and a swath width of 250km. SAR bands are sensitive to moisture hence the images downloaded coincided with the dryer period rather than the time field and LiDAR surveys were conducted. Taking advantage of the hyper-temporal characteristic of Sentinel 1 data, five Sentinel imagery

were downloaded so that a multi-temporal filter can be run over the dataset to further reduce the effect of speckle and hence increase the chance of getting a good model.

3.7.4 ALOS PALSAR-2 imagery

One image tile of the 2018 ALOS PALSAR global mosaic was downloaded covering the study area. The downloaded tile was approximately 100km x 100km in extent and has a spatial resolution of 25m and a temporal resolution of 1 year. The download product has HH + HV polarization only and was projected to a geographic (latitude and longitude) coordinate system. These data are freely available at https://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/data/index.htm.

According to Shimada et al (2014), this image was calibrated using published coefficients and output with 16-looks to reduce speckle. The image is also orthorectified and slope corrected using SRTM 90m DEM. Then a de-striping process was applied to equalize the intensity differences between neighbouring strips, which can be attributed largely to seasonal and daily differences in surface moisture conditions. Unfortunately, ALOS PALSAR 2 mosaics have a temporal resolution of 1 year and hence we did not have the luxury of running a multi-temporal filter on the imagery.

The ALOS PALSAR-2 mosaic was converted from DN to Gamma dB using the formula provided by the JAXA metadata file expressed as follows:

$$\gamma_0 = 10\log_{10}(DN^2) - 83.0 \text{ dB} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 3.1$$

In summary, Sentinel-1 carries a C band sensor and dual VV and HV polarization whilst ALOS PALSAR-2 carries an L band sensor and dual HH and HV polarizations figure 3.3 below show (a) ALOS PALSAR-2 mosaics HV polarization image band and (b) shows Sentinel-1 VH polarisation image band over AENP.

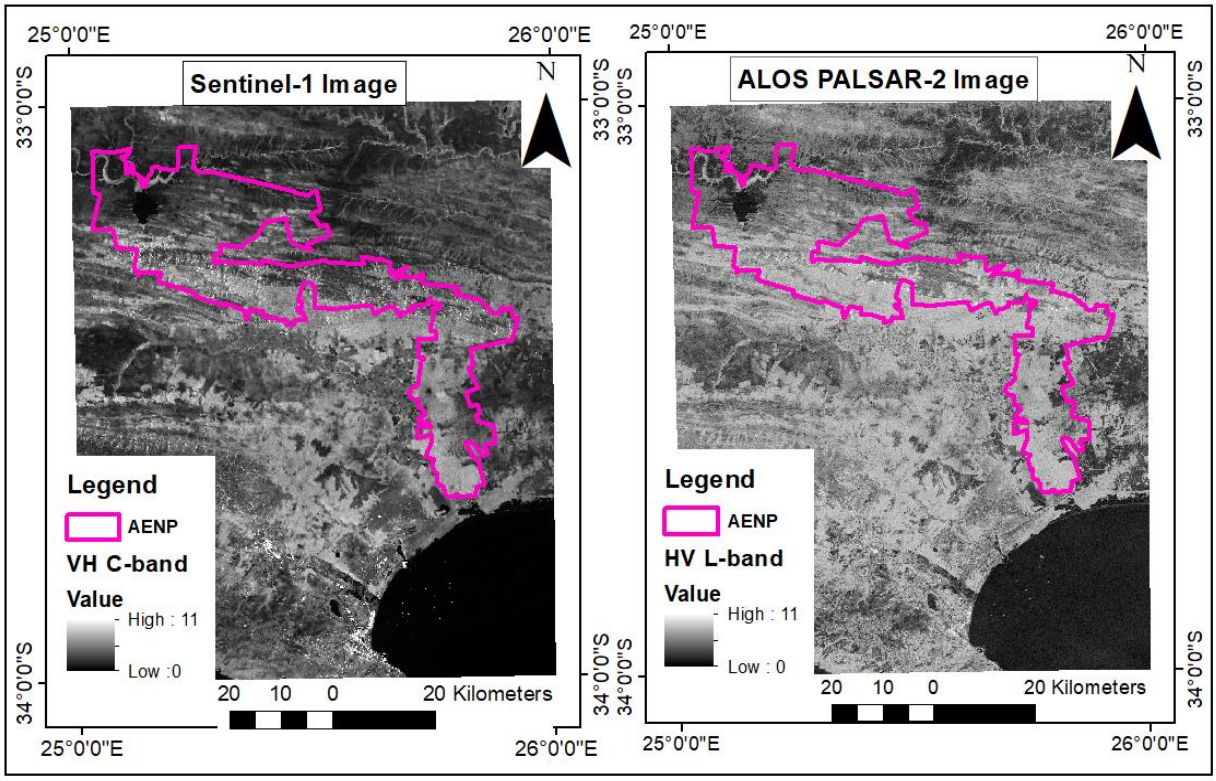


Figure 3. 3: Sentinel-1 & ALOS PALSAR-2 mosaics over AENP, displaying VH and HV polarizations, respectively.

Table 3.3 shows the information on the downloaded Sentinel-1 and ALOS PALSAR-2 dataset that were used in this study.

Table 3. 3: Summary of satellite images to be used in this research

Sensor	Product	Date of Acquisition	Polarization	Wavelength	Spatial Resolution	File Name
PALSAR-2	Global Mosaic	2018	HH, HV	L-band ±15cm	25m	S33E025_18_MOS_F02DAR.tar
Sentinel-1A	GRDH	Jan 18	VV, VH	C-band ±5cm	30m	S1A_IW_20180111T170110_DVP_RTC30_G_spufem_4BDB
Sentinel-1A	GRDH	Jan 18	VV, VH	C-band ±5cm	30m	S1A_IW_20180123T170109_DVP_RTC30_G_spufem_C379
Sentinel-1A	GRDH	Feb 18	VV, VH	C-band ±5cm	30m	S1A_IW_20180204T170109_DVP_RTC30_G_spufem_3C97
Sentinel-1A	GRDH	Feb 18	VV, VH	C-band ±5cm	30m	S1A_IW_20180216T170109_DVP_RTC30_G_spufem_351D
Sentinel-1A	GRDH	Feb 18	VV, VH	C-band ±5cm	30m	S1A_IW_20180228T170109_DVP_RTC30_G_spufem_0B21

3.8 Methodology

In this research, two different models will be explored in estimating AGB namely:

Parametric models, like Least Squares regression (LSR): are the most commonly used regression techniques (Ni-Meister et al.,2015). Parametric models assume some finite set of parameters θ. Given the parameters, future predictions, x, are independent of the observation data, D:

P (x/θ, D) = P(x/θ) ……..**3.2**

Therefore, θ captures everything there is to know about the data.

As a result, the complexity of the model is bounded even if the amount of data is unbounded. This makes these models very rigid.

Machine Learning algorithms are computer algorithms that improve automatically through experience using data and they are part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms have gained popularity in recent years due to advancements in computer technology and that they can make reliable predictions even where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. In this research, Random Forest (RF) was used to represents the category of non-parametric machine learning algorithms whereby no assumptions of normality as in the case of SLR are made about the data that one is modelling.

Another important factor that will be explored in this research is the ability and efficacy of L and C band SAR sensors to retrieve AGB. This is important as different wavelengths interact with different parts of the vegetation structure. For instance, the L band with its ±15cm wavelength penetrates through the canopy hence interact with the main branches as well the tree trunk while ±5cm C band wavelength interacts primarily with the leaves and twigs. However, for the Addo thicket biome, the vegetation type is comprised of mainly smaller woody structure and leaves, hence it is worthwhile to test and see which one of the two wavelengths better captures the structural properties of this biome.

Figure 3.4 shows the main method used in this study illustrated as a flow chart. Field data and LiDAR data were combined to create more accurate biomass predictions for the area within the LiDAR image. This upscaling from field measurements to LiDAR sample enables us to

considerably increase the sample for calibration and validation of the SAR derived AGB model (Englhart et al., 2013).

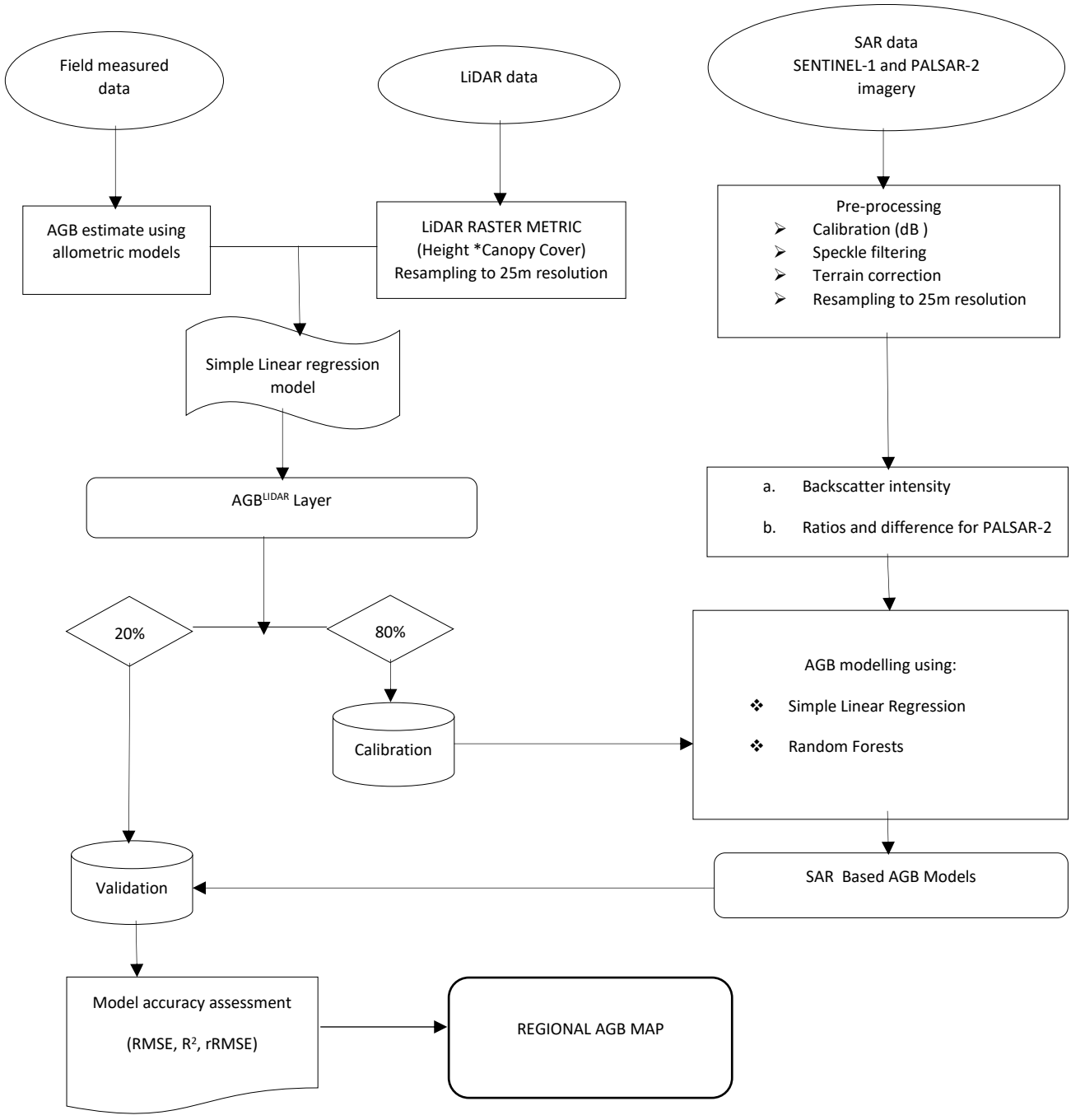


Figure 3. 4: Flow chart of steps used for forest above-ground biomass using SAR imagery

Where:

R^2 = coefficient of determination

rRMSE = relative Root Mean Square Error

RMSE = Root Mean Square Error

3.9 Field plot data and LiDAR data

In this study, the following field survey data and LIDAR data were supplied by CSIR as secondary data:

- LiDAR sample boundary in UMT35S coordinate system and .shp file format.
- A shape (.shp) showing the location of eight field sampling plots.
- A 25m resolution Canopy Cover (CC) LiDAR metric in UMT35S coordinate system
- A 25m resolution Vegetation Height (CHM) LiDAR metric in UMT35S coordinate system
- An excel spreadsheet showing calculated AGB per subplot.

3.10 Calculation of Field Plot AGB

Since only the location of the $25m^2$ plot and not the $5m^2$ subplot was available, there was a need to extrapolate these subplots AGB values (in kg) to the $25m^2$ plot level and then convert the resultant biomass to tonnes per hectare. For example for BRSTT_1 plot:

1. Add up all the Subplots AGB for a particular plot

   A = (BRSTT_SP1 + BRSTT_SP2 + BRSTT_SP3)
   = (90.58 + 92.88 + 141.7)Kg
   =325.16Kg
2. Upscale to 25m X 25m plot level
   BRSTT_1= A(Kg) * 8.33 ((i.e $625m^2$ area of 25X25m plot divided by $75m^2$ area of 5X5m subplot) =8.33)

   BRSTT_1 = 325.16*8.33

=2708.583Kg

3. Then convert to tons per hector

   = (BRSTT_1/1000(to convert to tons)) *16 (10000m^2 area of 1ha divide by 625m^2 area of 25X25m plot)

   = (2708.583/1000) *16

   = 43.337 t/ha

Therefore, for a 25mX25m BRSTT_1 plot, AGB is equal to 43.337t/ha.

The same procedure is applied on BRST_1; E31Y_1; E37Y_1; E37Y_2; OEC_2 and OEC_3. BV1 requires the conversion to t/ha only. This will result in eight field plot samples (see table below).

Table 3. 4: Plot level AGB in Tonnes per Hector

ID	AGB (Kg)	AGB(t/ha)	Plot No.
BRSTT_1_SP1	90.58	43.337	1
BRSTT_1_SP2	92.88		
BRSTT_1_SP3	141.7		
BRST_1_SP1	205.11	39.107	2
BRST_1_SP2	39.29		
BRST_1_SP3	49.02		
BV1	571.93	9.151	3
E31Y_1_SP1	115.72	37.973	4
E31Y_1_SP2	53.52		
E31Y_1_SP3	115.67		
E37Y_1_SP1	1.6	1.935	5
E37Y_1_SP2	6.47		
E37Y_1_SP3	6.44		
E37Y_2_SP1	33.08	16.759	6
E37Y_2_SP2	68.45		
E37Y_2_SP3	24.21		
OEC_2_SP1	75.14	27.186	7
OEC_2_SP2	58.9		
OEC_2_SP3	69.94		
OEC_3_SP1	22.99	5.199	8
OEC_3_SP2	15.01		
OEC_3_SP3	1.01		

3.11 LiDAR Metrics and AGB model

The LiDAR metric Canopy Cover (CC) and Vegetation height (CHM) were analysed using ArcGIS 10.6.1 software. First, the LiDAR metrics were cleaned of all the "no data value" by converting them to zero. Then using ArcGIS 10.6.1, a third raster metric which is the interaction term of CHM and CC was created by multiplying CHM*CC.

A spatially explicit AGB map over the entire area covered by the LiDAR survey was produced using plot-level LiDAR metrics and field AGB. LiDAR data was used to produce the AGB reference map due to its high level of structural detail and resolution, which can ensure the precision and reliability of reference maps (Li et al. 2015). The metrics derived from LiDAR, and therefore used to upscale field AGB plots to LiDAR, were the vegetation height (CHM), canopy cover (CC) and the product of the two metrics (CHM*CC).

Simple linear regression (SLR) modelling using the backward elimination approach was used to determine the suitable predictor variable(s) (Li et al. 2016), by regressing the field AGB and LiDAR metrics. Then the significant predictor variable(s) were determined using the P-value. The P-value is used to prove or disprove a null hypothesis, by comparing the P-value to a given confidence level (in this case 5%). If the P-value is less than the confidence level, then the null hypothesis is accepted otherwise the null hypothesis is rejected. The R-squared value of the final model indicates the percentage of the variation in the data that is explained by the model, the so-called "model fit".

The 8 field AGB plots within the LiDAR survey area were used as training and validation data. Due to the small sample size, Leave-one-out cross-validation (LOOCV) was used to validate the accuracy of the final SLR model.

3.12 Validation method for small samples

The leave-one-out cross-validation method (LOOCV) was used in this study to assess the prediction accuracy of the model due to the small sample size of (n=8) (Li et al, 2016). The LOOCV approach made the field data from each plot act simultaneously as training and validation samples.

Then finally using the raster calculator, the model was applied over the entire LiDAR raster to retrieve the AGB of the entire LiDAR survey area.

3.13 Calculation of the prediction model

Due to the size of the data sample, the LOOCV method was used to calculate the LiDAR AGB model as shown below. The equation for this model can be generally stated as follows:

Predicted AGB = β_0 + β_1(CC*MH) …….. **3.3**

Where: β_0 is the intercept

β_1 is the coefficient CC*MH

Table 3. 5: Model Regression statistics

Plot_ID	AGBf	CC*HM	Model	β_0	β_1	P(t)$_0$	P(t)$_1$	P(F)	R^2
BRST1	39.107	127.318	M1	0	0.251	0.184	0.0000	0.0000	0.98
BRSTT 1	43.337	185.960	M2	0	0.282	0.409	0.0000	0.0000	0.81
BV1	9.151	18.082	M3	0	0.261	0.440	0.0000	0.0000	0.82
E31Y-1	37.973	149.721	M4	0	0.265	0.288	0.0000	0.0000	0.81
E37Y-1	1.935	21.672	M5	5.428	0.222	0.042	0.0001	0.0001	0.96
E37Y-2	16.759	53.800	M6	0	0.260	0.334	0.0000	0.0000	0.98
OEC-2	27.186	91.946	M7	0	0.258	0.311	0.0000	0.0000	0.81
OEC-3	5.199	5.869	M8	0	0.262	0.406	0.0000	0.0000	0.98
Total				**0.678**	**0.258**	**0.302**	**0.0000**	**0.0000**	**0.89**

Where ID denotes each sampling plot name; AGBf denotes field Above Ground Biomass; M_{n-8}; β_0 is the intercept; β_1 is the coefficient of CC*HM; P(t)$_0$ is the value of β_0 ; P(t)$_1$ is the value of β_1 ; P(F) is the p-value of the model and R^2 is the coefficient of determination of the model.

Table 3.5 shows a summary of individual model regression statistic. P(F) and P(t) values are used to test the significance of each model and its coefficients, respectively.

3.14 Hypothesis Test of each model

Using the F-test to test the significance at α= 0.05 of the individual models, the following hypothesis was made about the models:

H_0: the linear model is not significant

H_a: the linear model is significant

P(F) = 0.0001<0.05

From table 3.5, for all the models the P(F) values are less than 0.05. Therefore, the null hypothesis is rejected, hence there is sufficient evidence to conclude in favour of the alternative hypothesis which states that the linear models are all significant.

3.15 Hypothesis Test of the Intercept B_0

From table 3.5 all models have p(t) values greater than 0.05 except model number 5 (M5) which has a p(t) = 0.042. Hence the hypothesis tests of models M1, M2, M3, M4, M6, M7, M8 were set as follows:

H_0: $\beta_0 = 0$

H_a: $\beta_0 \neq 0$

The p(t) values for these models are all greater than 0.05 hence we fail to reject the null hypothesis meaning that for all 7 models the intercept is not significant hence the intercepts of all these models are equal to zero.

However, for model number 5 (M5), we set up the hypothesis as follows:

H_0: $\beta_0 = 0$

H_a: $\beta_0 \neq 0$

$p(t) = 0.042 < 0.05$

Therefore, we reject the null hypothesis. Hence there is sufficient evidence to conclude in favour of the alternative hypothesis. Which states that β_0 is not equal to zero. Hence the intercept for model 5 is equal to 5.428

3.16 Hypothesis Test of the coefficient of CC*MH for each model

Using the t-test to test the significance at $\alpha = 0.05$ of the individual coefficients of CC*MH, the following hypothesis was made about the coefficients of CC*MH for each model:

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

$P(t) = 0.0001 < 0.05$

From table 3.5, all the models have p(t) values less than 0.05. Therefore, we reject the null hypothesis. Hence there is sufficient evidence to conclude in favour of the alternative hypothesis which states that the coefficients of CC*MH on all the models are significant.

3.17 Final LiDAR AGB model

Using the sum of each of the test statistics in Table 3.5, the following hypothesis test about the overall linear model and its coefficients were determined.

3.18 Hypothesis Test of Final LiDAR AGB model

hypothesis about the final LiDAR AGB models is that,

H_0: the linear model is not significant

H_a: the linear model is significant

$P(F) = 0.0001<0.05$

From Table 1 above, the total $P(F)<0.05$. Therefore, we reject the null hypothesis. Hence there is sufficient evidence to conclude in favour of the alternative hypothesis which states that the linear model is significant.

3.19 Hypothesis of the Final LiDAR AGB model intercept

Using the total regression statistic in Table 3.5, the following hypothesis test about the model intercept β_0 was investigated using the t-test

H_0: $\beta_0 = 0$

H_a: $\beta_0 \neq 0$

p(t)=0.302 which is greater than 0.05 hence we fail to reject the null hypothesis meaning the intercept of the final regression model is not significant. Hence the $\beta_0 = 0$

3.20 Hypothesis of the Final LiDAR AGB model β_1

Using the total regression statistic in Table 3.5, the following hypothesis test about the model coefficient of CC*MH was investigated using the t-test

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

p(t) =0.000 which is less than 0.05. Therefore, we reject the null hypothesis and conclude in favour of the alternative hypothesis which states that the coefficient of CC*MH is not zero. Therefore, the model was expressed as follows:

Predicted AGB =0.258(CC*MH) …….. **3.4**

3.21 Sar-based predictors variable selection

Based on previous studies (Pan and Sun, 2018; Chen et al., 2018; Berninger et al., 2018; Omar et al., 2017; Haralick et al., 1973), the following SAR predictors were found to be relevant in above-ground biomass modelling and will be used in this research.

Table 3. 6: Sentinel-1 and ALOS PALSAR-2 predictors

LIST OF PREDICTOR VARIABLES	
Sentinel-1 C band	**ALOS PALSAR-2 L band**
L1_VH	HH
L2_VH	HV
L3-VH	Ratio = HH/HV
L4_VH	Diff = HH-HV

3.22 Data Cleaning

3000 points were randomly generated over the entire area covered by the LiDAR-derived AGB and these points were used to extract LiDAR AGB and corresponding VH backscatter values of those locations for all the 5 Sentinel-1 time series. To avoid confusion, Table 3.7 describes the different names used during regression analysis in SLR and RF.

Table 3. 7: Times series VH bands nomenclature

Date of Imagery Acquisition	Name of the band in SLR	Name of the band in RF
11-01-2018 VH	tb$b1_s1	Layer_1
23-01-2018 VH	tb$b2_s1	Layer_2
04-02-2018 VH	tb$b3_s1	Layer_3
16-02-2018 VH	tb$b4_s1	Layer_4
28-02-2018 VH	tb$b5_s1	Layer_5

To fit a linear regression model to a data set, an assumption that the residuals of the predictor values must follow a normal distribution must be satisfied. Therefore, extracted dataset values were cleaned of all zero values in ArcMap. Then the outliers were removed in R-Studio using the upper and lower limits defined as follows:

Uplim = 3rd Quatile +1.5*IQR(X) ……. **3.5**

Lolim = 1st Quatile +1.5*IQR(X) …….**3.6**

Then all values which are lower than the lower limit and greater than the upper limit were classified as outliers and hence were removed from the sample dataset. Then boxplots, histograms and QQ plots were used to test for the assumption of normality of the dataset.

3.23 Histogram

Figure 3.5 Shows five histograms of the sentinel 1 C band backscatter values, all five histograms appear to be leptokurtic, meaning that more values are concentrated around the mean and therefore less variance.

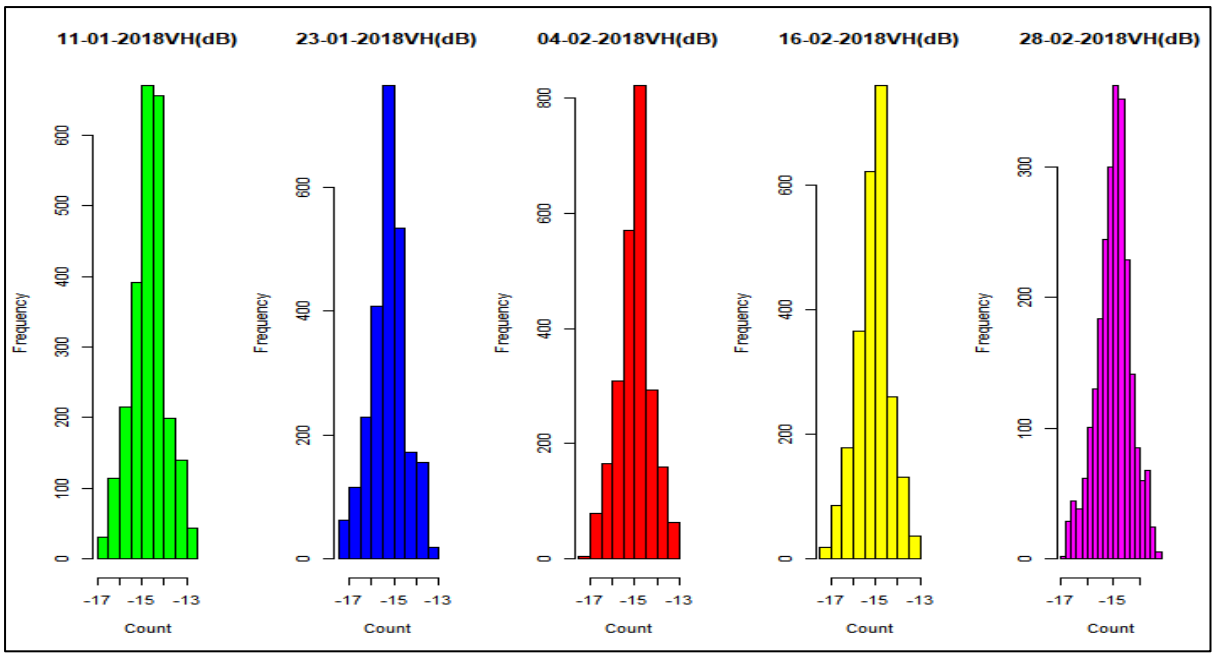


Figure 3. 5: Histograms of Sentinel-1 VH time series

3.23 Boxplots

When using SLR, outliers can greatly influence the performance of the model, hence boxplots were used to check for outliers in the data set.

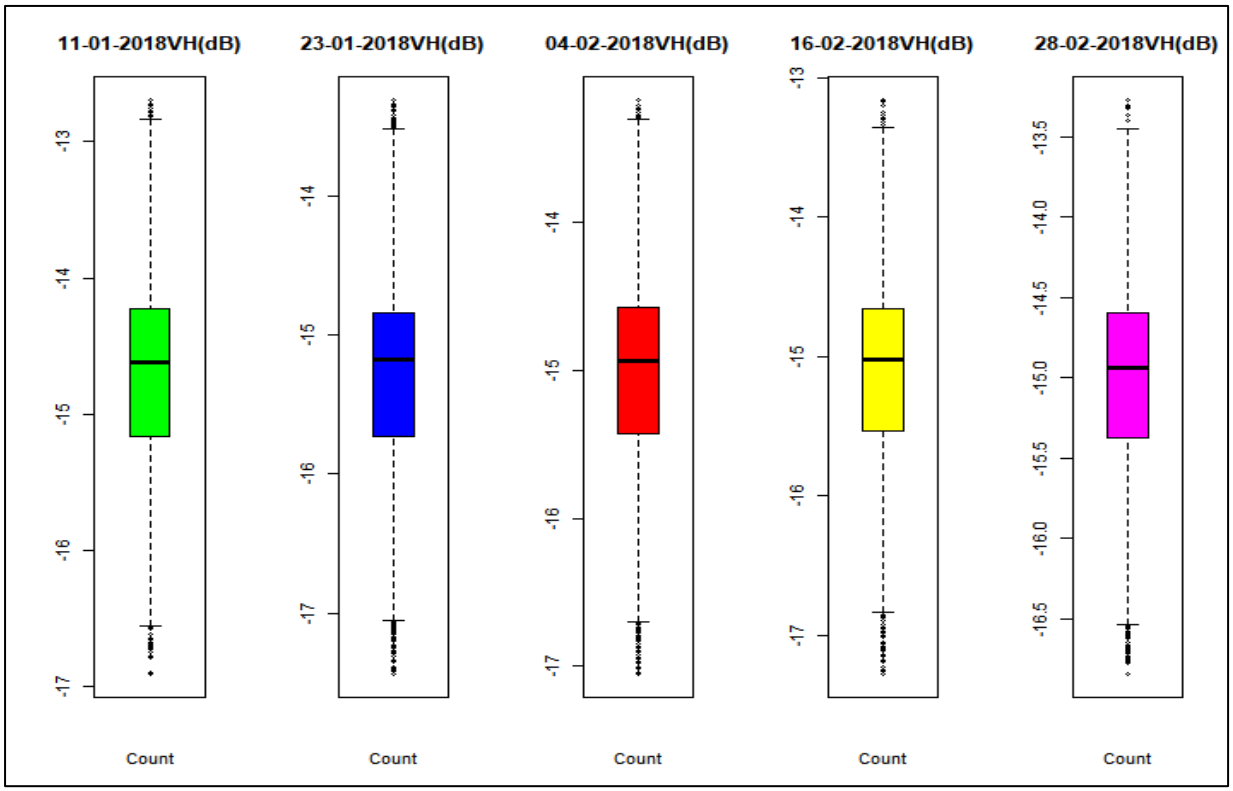


Figure 3. 6: Boxplots of Sentinel-1 VH time series

As can be seen from figure 3.6 there are observations beyond both the upper and the lower limit of all the time series boxplots. However, these outliers are not severe hence normality of the data set can be assumed.

3.24 QQ plots

QQ plots were also used to test for normality, as can be seen from figure 3.7 all the observations lay along the diagonal line which means that the data set follows a normal distribution.

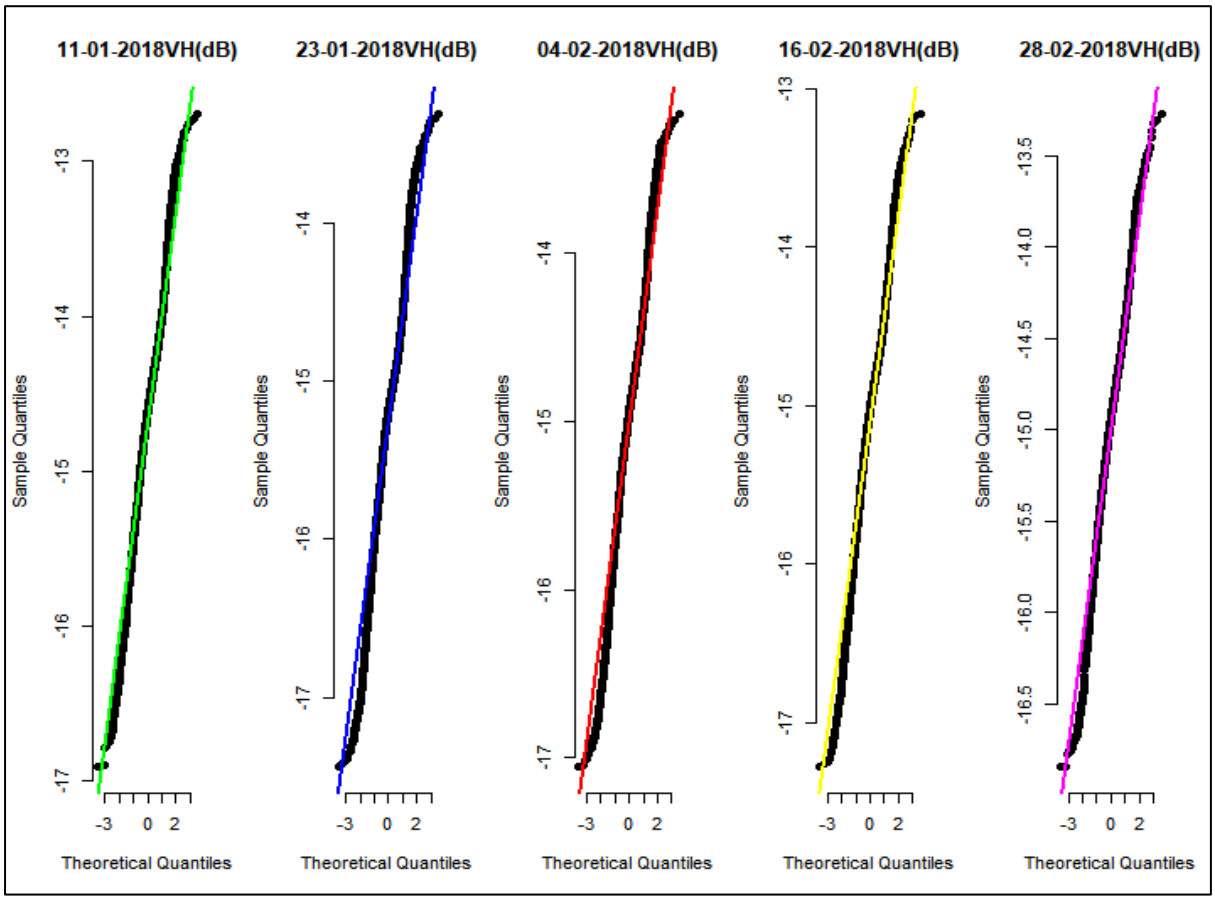


Figure 3. 7: QQ plots of Sentinel-1 VH time series

3.25 Analysis of Predictor variables and SAR AGB Model

The predictor variables were analysed for their significance using the p(t) value statistic test and the overall significance of the models were tested using the p(F) value statistical test at a 5% level of confidence.

After that, SLR and RF were used to model AGB based on Sentinel-1 and ALOS PALSAR-2 predictors, respectively.

3.26 Mapping procedure and interpretation

The general mapping procedure of the AGB was at 25m pixel resolution, this is partly informed by field plots which were sampled using $25m^2$ plots. The AGB mapping procedure involves mainly two modelling stages i.e:

1. modelling the relationship between field plot AGB and LiDAR metrics
2. Modelling the relationship between LiDAR AGB and SAR variables

3.27 Modelling LiDAR AGB and SAR variables

A total of 3000 points covering the entire LiDAR AGB raster were randomly generated in Arcmap using a distance threshold of 25m and both the LiDAR AGB and the SAR metrics were extracted. Then the relationship between AGB^{LIDAR} and the SAR variables from both Sentinel 1 and ALOS PALSAR-2 modelled. Hence two models were used namely Simple Linear Regression and Random Forest.

Simple Linear Regression general equation to be used is as follows:

$$\mathbf{AGB^{LIDAR} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots .+ \beta_{n-1} X_{n-1} + \beta_n X_n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{3.7}$$

Random Forest regression general equation to be used is as follows:

$$\mathbf{AGB^{LIDAR} = \frac{1}{B}\,(CART_1 + CART_2 + CART_3 + \ldots\ldots\ldots .. CART_B)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{3.8}$$

Where: $CART_1$, $CART_2$, $CART_3$, ... $CART_B$ denotes Classification and Regression Trees (CART) constructed on bootstrap samples of the original data.

Using the two models above, 80% of the AGB LIDAR sample produced in step 1 were used to train the models and the remain 20% were used to validate the final model of each of the modelling techniques.

3.28 Model Selection

The selection of the most suitable model was based on the following accuracy indicators: The coefficient of determination (R^2), Root Mean Square Error (RMSE) and the Mean Square Error (MSE), (Chen et al., 2018; Berninger et al., 2018, Huang et al., 2018).

- The coefficient of determination(**R^2**) is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions perfectly fit the data.
- Root Mean Square Error (**RMSE**) is a standard way to measure the error of a model in predicting quantitative data. It serves as a heuristic for training models and to evaluate trained models for usefulness or accuracy. The smaller the RMSE the better the model
- Relative Root Mean Square Error **(rRMSE)** is a statistical indicator which highlights the percentage of uncertainty in the predicted values by a model. rRMSE is calculated by dividing RMSE with the average value of measured data. According to (Despotovic et al., 2016), model accuracy is considered excellent when rRMSE < 10% ; good if 10% < rRMSE <20%; fair if 20%< rRMSE <30% and poor if rRMSE > 30%.

3.29 Comparison of Sentinel-1 C band and ALOS PALSAR-2 L band

Then R^2, RMSE and rRMSE for both the best model using both Sentinel 1 and ALOS PALSAR-2 data were compared and the model with the highest R^2, and the smallest RMSE and rRMSE values was considered to be the best model for estimating Addo thicket AGB. After the final model selection, a regional map of AGB was generated and presented in figure 4.9.

CHAPTER 4: ANALYSIS AND RESULTS

4.0 LiDAR AGB model

Table 4.1 below list the sample field plot AGB data as provided by CSIR in tonnes per hectare (t/ha).

Table 4. 1: Addo thickets Sample Field AGB

Plot_ID	Field AGB (t/ha)
BV	9.151
BRST	39.107
BRSTT	43.337
E37Y-2	16.759
E37Y-1	1.935
OEC-3	5.199
E31Y	37.973
OEC-2	27.186

Table 4.2 summarises the distribution of the sample field AGB. The BRSTT has the highest AGB while E37Y (i.e.) the thicket type most exposed to elephants has the least AGB. The sampled plots have a range of 41.40t/ha and a standard deviation of 16.50 t/ha and a mean of 22.6 t/ha. Hence Addo thicket biome can be classified as a low AGB.

Table 4. 2: Addo thickets Sample Field AGB Summary Statistics

Field AGB Summary Statistics	
Mean	22.58
Median	21.97
Standard Deviation	16.50
Range	41.40
Minimum	1.93
Maximum	43.33
Sum	180.64
Count	8

4.1 Exploratory data analysis

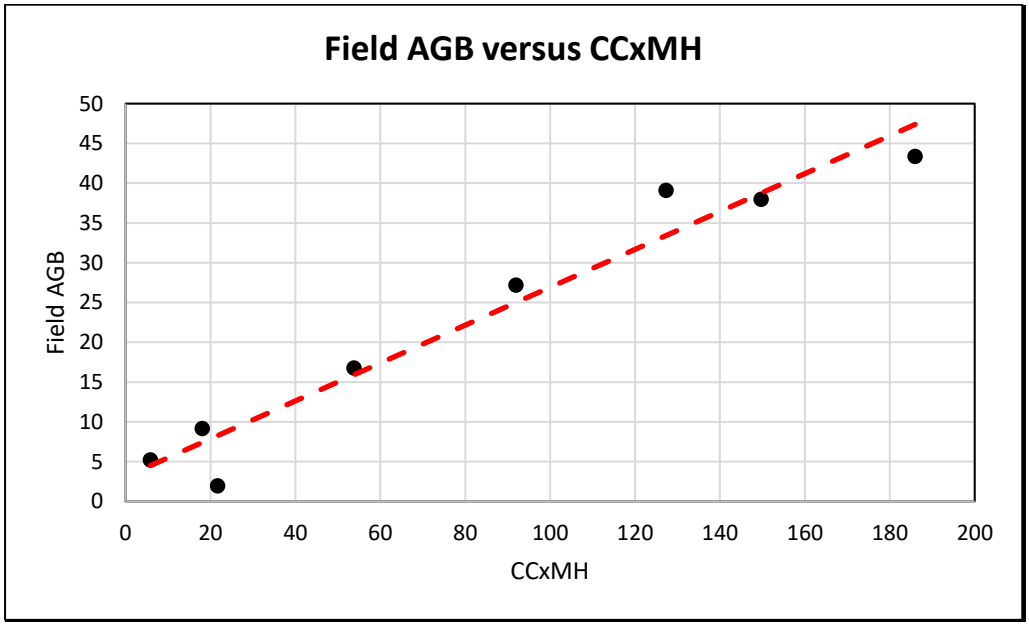


Figure 4. 1: Field AGB versus CC*MH scatterplot

A scatter plot in figure 4.1 was used to explore the trends in the data by plotting field AGB against the predictor LiDAR metric CCxMH. Based on the trend exhibited by the scatter plots, it is evident that there is a strong linear relationship between the field AGB and CC*MH. Therefore, a linear model was fitted to the data to predicted LiDAR-based AGB.

LiDAR AGB SLR model

The SLR model was calculated in chapter 3 and the equation can be stated as follow:

Predicted AGB =0.258(CC*MH) …… **4.1**

This equation was applied to the sample data using the LOOCV and a summary of the test LiDAR-AGB model calculated

4.2 LiDAR AGB Summary statistics

The model achieved R^2 of 89% which means that 89% of the variation in the data is explained by the model. Hence this is a good model fit. The overall RMSE is 3.779 t/ha. However, rRMSE is 35.6% meaning the residual variance is moderate.

4.3 LiDAR AGB Model Validation

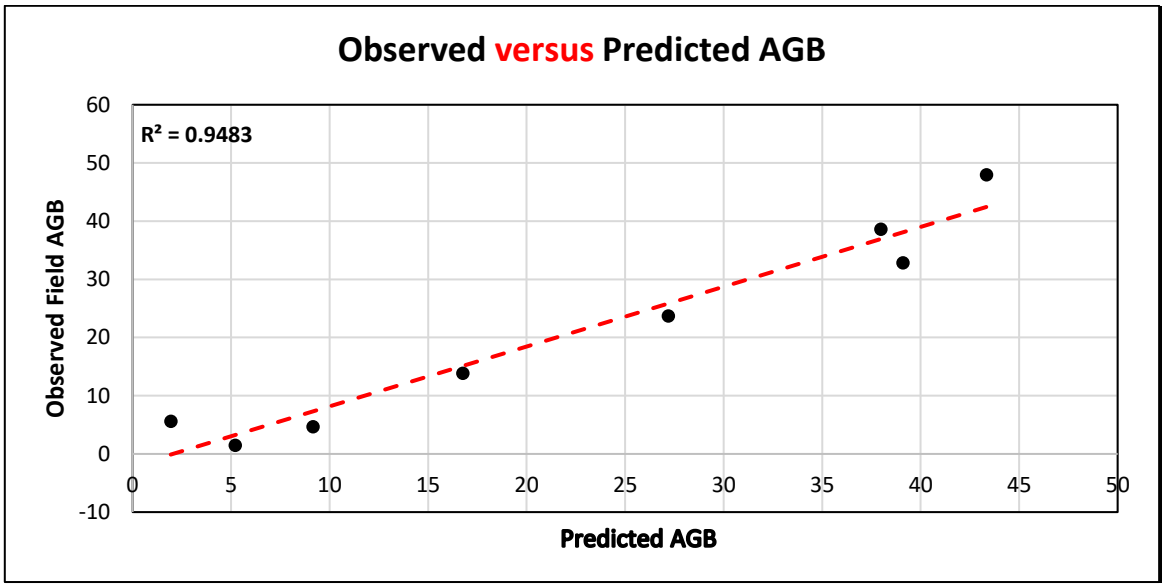


Figure 4. 2: Observed versus Predicted scatter plot

Using LOOCV, each sample was predicted using equation 4.6 and the final predictions plotted against the AGB deduced from field observations. As can be seen from Figure 4.2, the LiDAR derived AGB model could predict 95% of the field AGB. Even though the prediction of this model is excellent, it is worth mentioning that the field sample size of n=8 is very small hence

it is prone to overfitting. However, despite this limitation in the data, the derived model is accepted and used as is to upscale the field AGB using LiDAR metric data (CC*MH).

4.4 LiDAR-derived Addo Thicket AGB map

Figure 4.3 shows the distribution of field sampled plots over the LiDAR AGB raster, the AGB raster was trimmed such that it covers only that area which have mostly vegetation species that were sampled in the field.

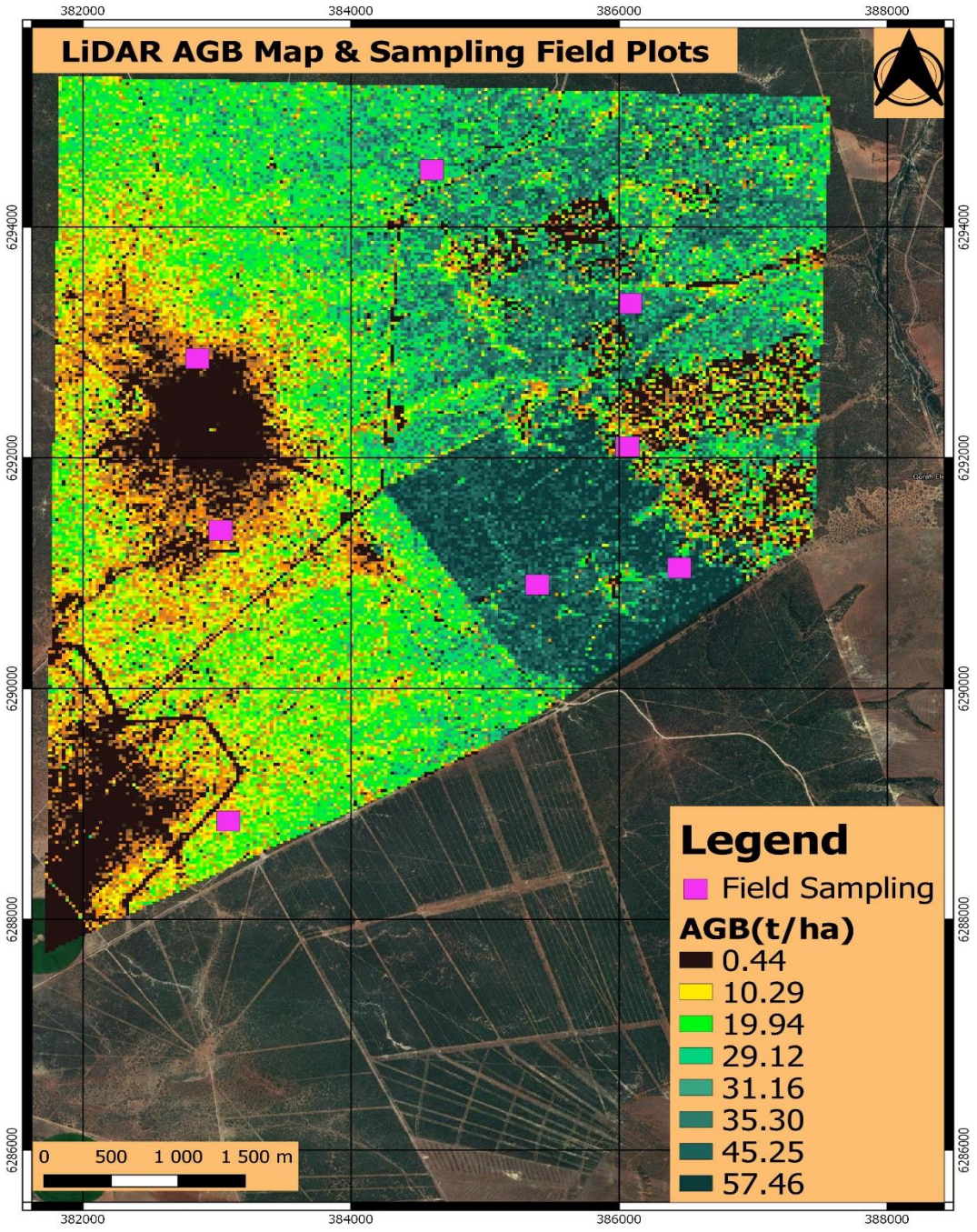


Figure 4. 3: LiDAR AGB map

The LiDAR AGB raster covers 34 km^2 and AGB ranges from 0.44 to 57.46 t/ha. The raster AGB was calculated in ArcMap using equation 4.6 in ArcMap.

4.5 Goodness of fit Sentinel-1 C and ALOS PALSAR-2 L Models

4.5.1 Sentinel Simple Linear Regression model

The summary statistics for this model are R^2=0.52, RMSE = 10.03t/ha and rRMSE = 44.2%. Therefore, the model has an R^2 of 0.52 which means that 52% of the variation in the data is explained by the model.

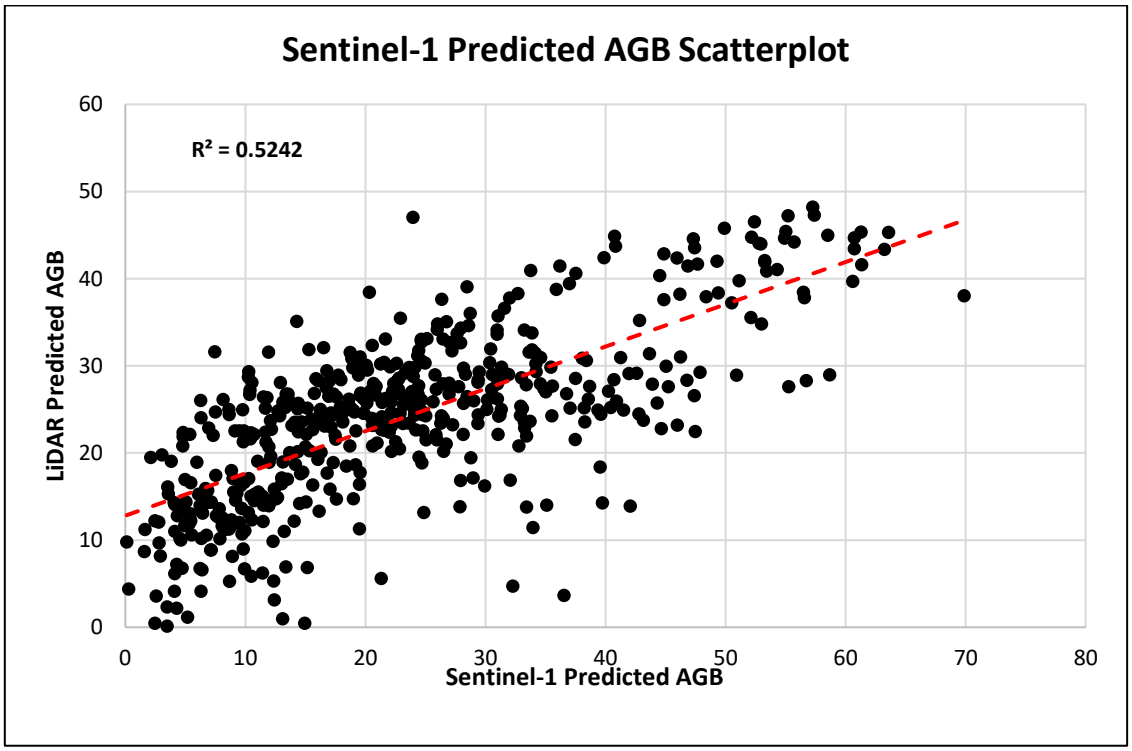


Figure 4. 4: Observed versus Predicted scatter plot Sentinel-1 using SLR

The scatter plot in figure 4.4 shows the relationship of the LiDAR AGB against the predicted Sentinel-1 VH predicted AGB. Based on the scatter plot, the SLR model was able to predict 52% of the LiDAR-derived AGB.

4.5.2 ALOS-PALSAR Simple linear Regression model

The summary statistics for this model are R^2=0.41, RMSE = 11.20t/ha and rRMSE = 49.1%. Therefore, the model has R^2 of 0.41 which means that 41% of the variation in the data is explained by the model.

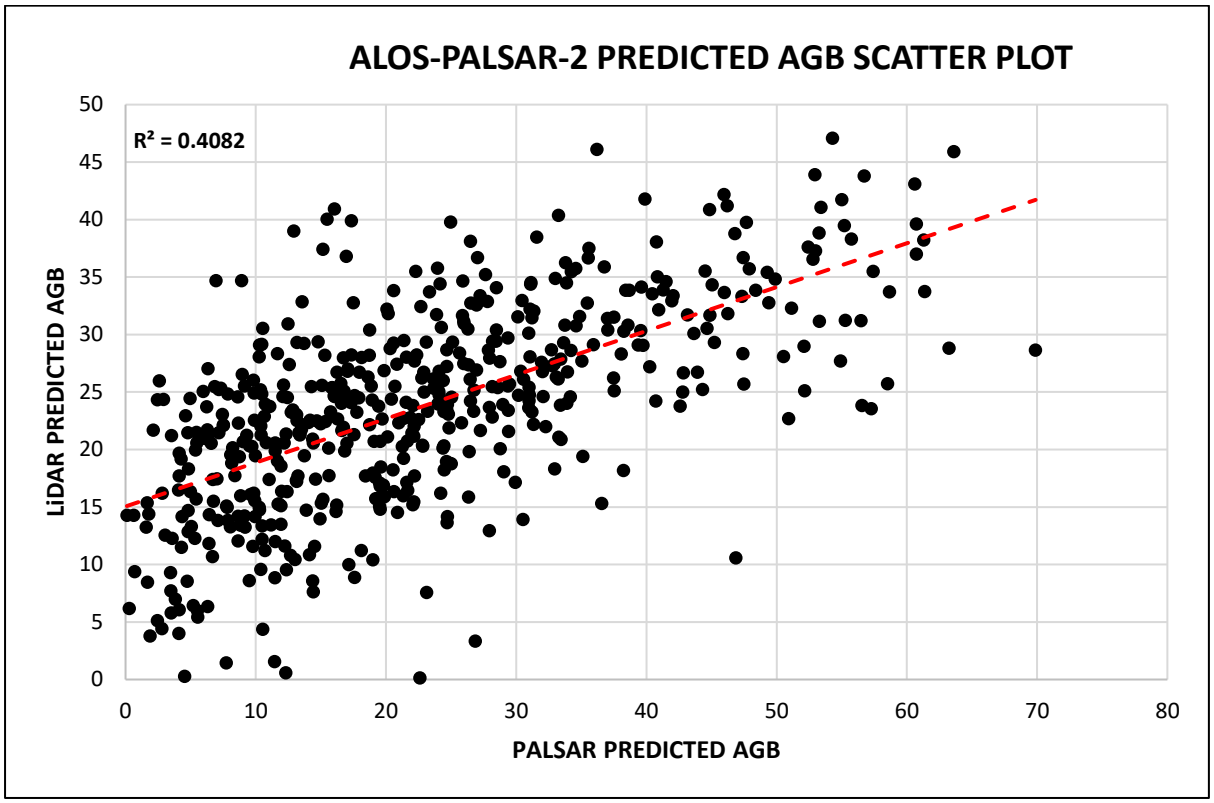


Figure 4. 5: Observed versus Predicted scatter plot ALOS-PALSAR-2 using SLR

4.5.3 Random Forest Regression Model

The summary statistics for this model are R^2=0.67, RMSE = 8.96t/ha and rRMSE = 38.4%. Therefore, the model has R^2 of 0.67 which means that 67% of the variation in the data is explained by the model.

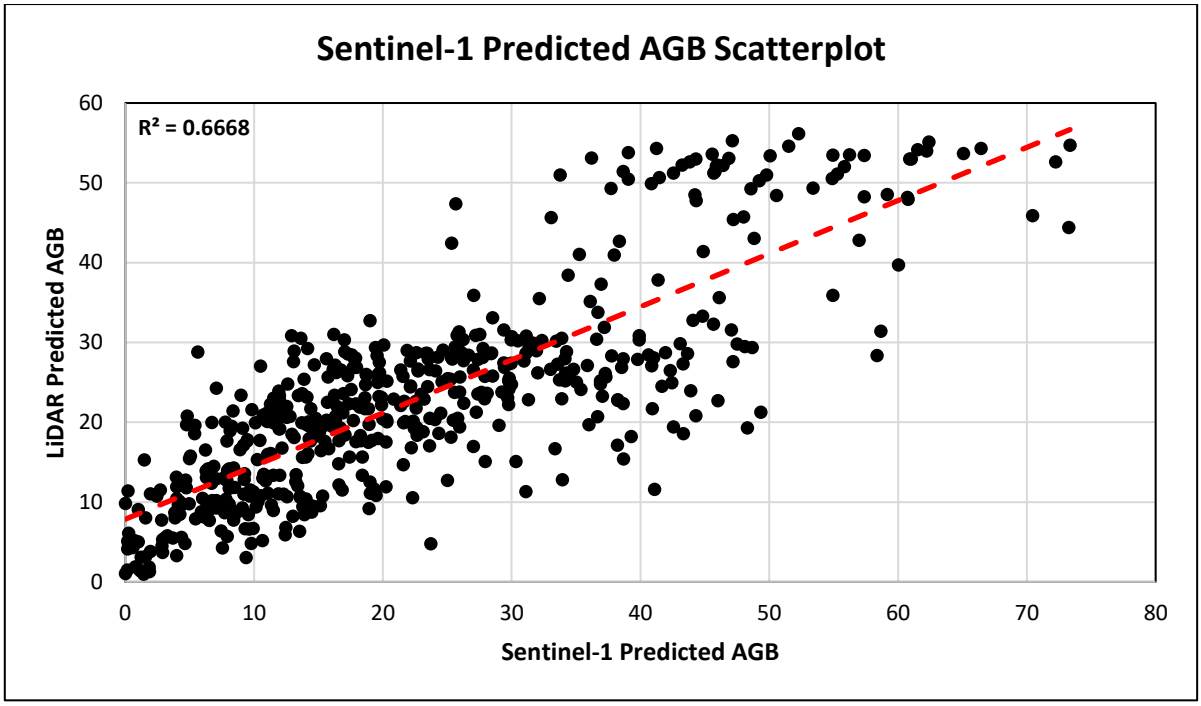


Figure 4. 6: Observed versus Predicted scatter plot Sentinel-1 using RF

The scatter plot in Figure 4.6 above shows that the Random forest regression model for predicting AGB using sentinel VH bands has an R^2 = 0.67 mean that this model can explain 67% of the variability in the data set.

4.5.4 ALOS PALSAR-2 Random Forest

The summary statistics for this model are R^2=0.51, RMSE = 9.93t/ha and rRMSE = 42.5%. Therefore, the model has R^2 of 0.51 which means that 50% of the variation in the data is explained by the model.

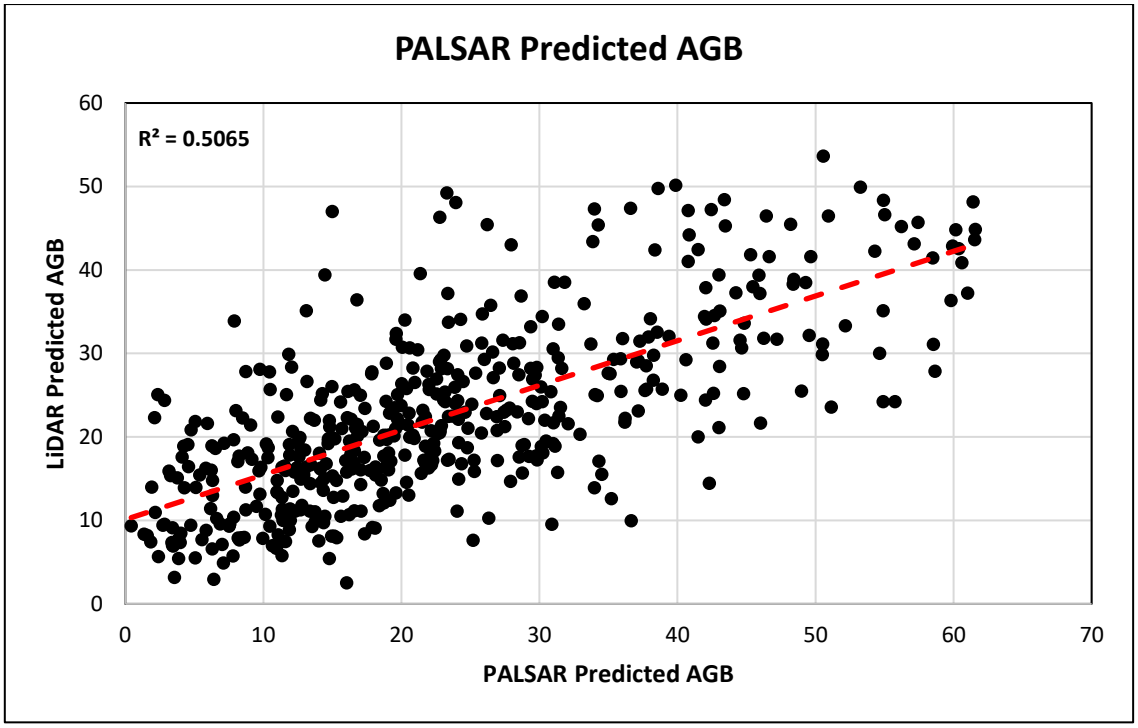


Figure 4. 7: Observed versus Predicted scatter plot ALOS PALSAR-2 mosaics using RF

The scatter plot in Figure 4.7 shows that the Random forest regression model for predicting AGB using ALOS PALSAR-2 mosaics L- band has $R^2 = 0.51$ mean that this model can explain 50% of the variability in the data set.

4.6 Final Best Model Selection

Table 4.8 shows a summary of the two types of regression techniques and using the R^2, the RMSE and the rRMSE, the ALOS PALSAR derived SLR is the least performing model and Sentinel-1 derived RF is the best model to predict AGB using Sentinel 1 C VH bands as it has the highest R^2, and the lowest RMSE and rRMSE values. Then the saturation of the best performing data set that is Sentinel-1 was tested for signal saturation.

Table 4. 3: Summary of the goodness of fit Statistics values of the models

Model	R^2	RMSE(t/ha)	rRMSE (t/ha)
SLR (Sentinel-1)	0.52	10.03	44.2%
RF (Sentinel-1)	0.67	8.96	38.4%
SLR (ALOS PALSAR-2)	0.41	11.20	49.1%
RF (ALOS PALSAR-2)	0.51	9.93	42.5%

4.7 Signal Saturation

SAR backscatter signals usually suffer from saturation especially at medium (50t/ha<AGB<100t/ha) to high AGB (AGB>100t/ha) values (Li et al, 2008). Therefore, after identifying Sentinel 1 C-band VH time series as the best data set in this study. Hence the problem of saturation was only test on the Sentinel data set and the results are shown in figure 4.8. Signal saturation was tested by generating a scatterplot of the VH backscatter values against the predicted LiDAR AGB observations for each of the images on the Sentinel-1 time series. SAR signal saturation is present if, an increase in LiDAR AGB does not result in a proportionate increase in the SAR backscatter values. As a result, the scatter plot flattens out as the AGB values increase. Based on the scatterplot in figure 4.8, there was no problem of SAR signal saturation as the scatterplot does not level up as AGB increases.

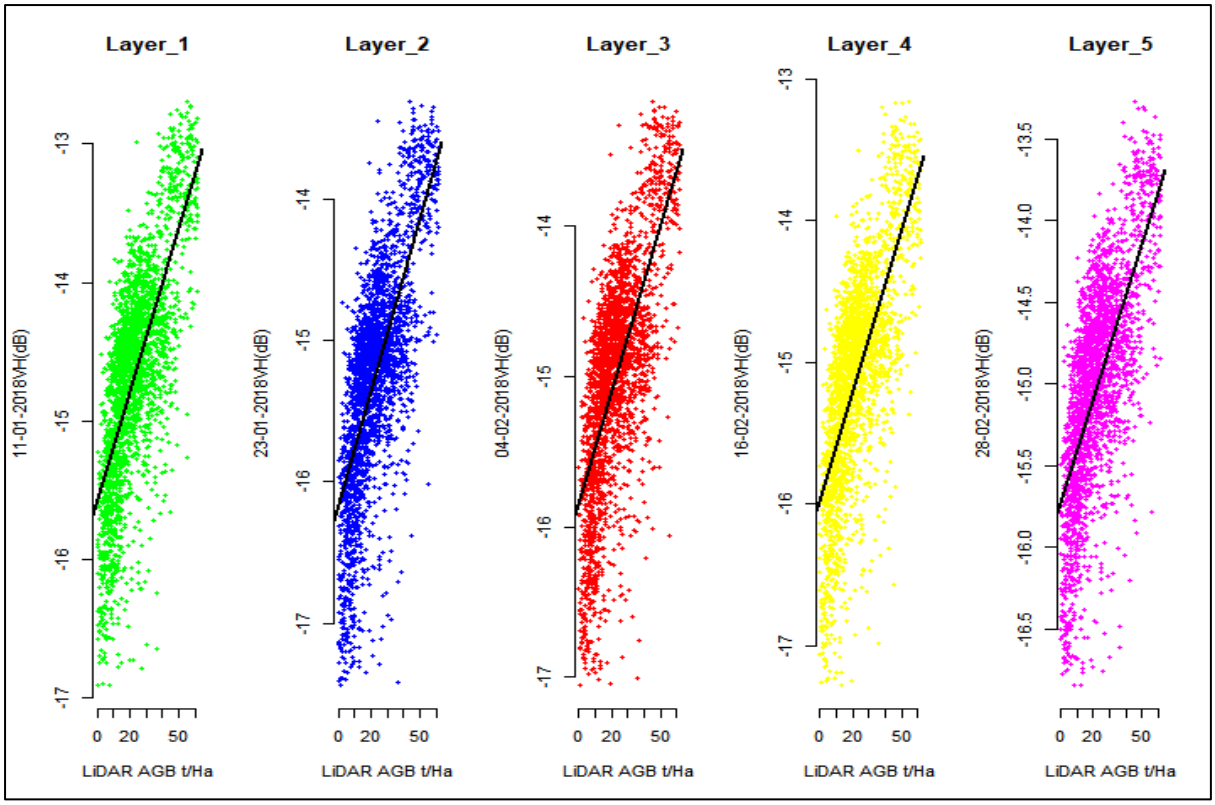


Figure 4. 8: The scatter plots of sampled AGB and SAR backscatter from Sentinel 1

Therefore, as a result, Sentinel data set was used to map the Addo Thicket biome AGB of the entire Sunday River Valley district which encompasses the sampled area, and the resultant map is shown in figure 4.9.

4.8 Results: Sentinel-1 VH C-band Addo Thicket AGB map

Figure 4.9 shows the AGB map covering the Sunday River Valley district of the Eastern Cape in South Africa. This map was produced by using the best model in this research using the free data set from Sentinel 1 SAR VH cross polarisation C band five-day time series and Random Forest algorithm. The model has a goodness of fit with R^2 = 0.67, RMSE =8.96% and rRMSE=38.5%. The biomass ranged from 0.44 t/ha to 57.46 t/ha and the AGB was upscaled from a LiDAR sample measuring 34 km^2 to cover the entire region measuring 5993km^2.

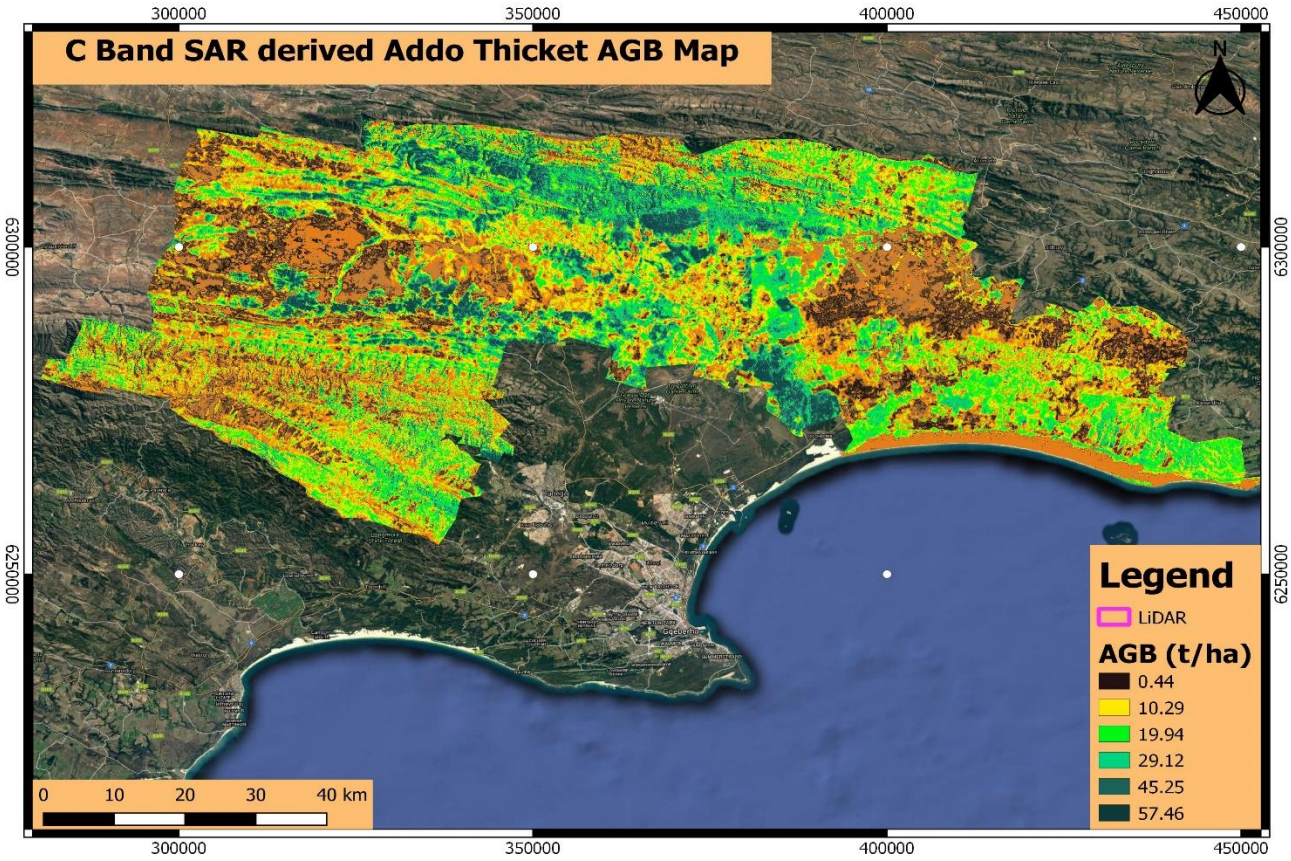


Figure 4. 9: Above Ground Biomass Map Over the Sunday River district

4.9 Sentinel 1 SAR Results

The field AGB and LiDAR surveys were conducted in May of 2018, however, the Sentinel images used in this research were acquired in January and February of 2018. The dates of acquisition for the Sentinel time-series images were chosen specifically to coincide with the dry season in the Eastern Cape region. This is important because SAR backscatter is sensitive to moisture. In this study, a time series of five Sentinel 1 SAR images were used, and (Quegan and Jiong Jiong Yu, 2001) a multi-temporal filter was applied to mitigate the effects of speckle noise that is inherent in all SAR data and that temporal filter was essential to counter the effects of water in the succulent leaves of most of the Addo thickest vegetation species.

Sentinel 1 SAR data comes in two polarizations i.e., VH and VV, however, in this study only the VH bands were used to model AGB as the VH bands are established as being more sensitive to multiple scattering that typically occurs in vegetation. The VV polarization was not used as it is more sensitive to surfaces like water and bare soil.

Figure 4.10 shows scenes of the SAR AGB map compared with the google earth image, these scenes were chosen to highlight some of the success, limitations, and pitfalls of this model.

4.10 Fence line effect

Scene A1 and B1 highlights some of the successfully mapped areas by the RF model, by comparison to image B1, image A1 clearly shows the delineation of bare soil and vegetation.

4.11 Unsampled land cover

Scene A2 and B2 shows the limitation of this model, B2 shows clearly agricultural plots but because this land cover class was not sampled during the field survey, the RF failed to effectively map AGB in this area resulting in a fuzzy image as shown in scene A2. Hence caution should be exercised in applying the model over a large area as the model can incorrectly predict AGB over unsampled vegetation types.

4.12 Topographical influence

According to (Paloscia et al., 1999 and Li et al 2008), topography influences the backscattering of SAR, as can be seen in scene A3 and B3. The model seems to have overestimated AGB because of the presence of bare rock and steep surfaces which resulted in a higher-than-normal return of backscatter.

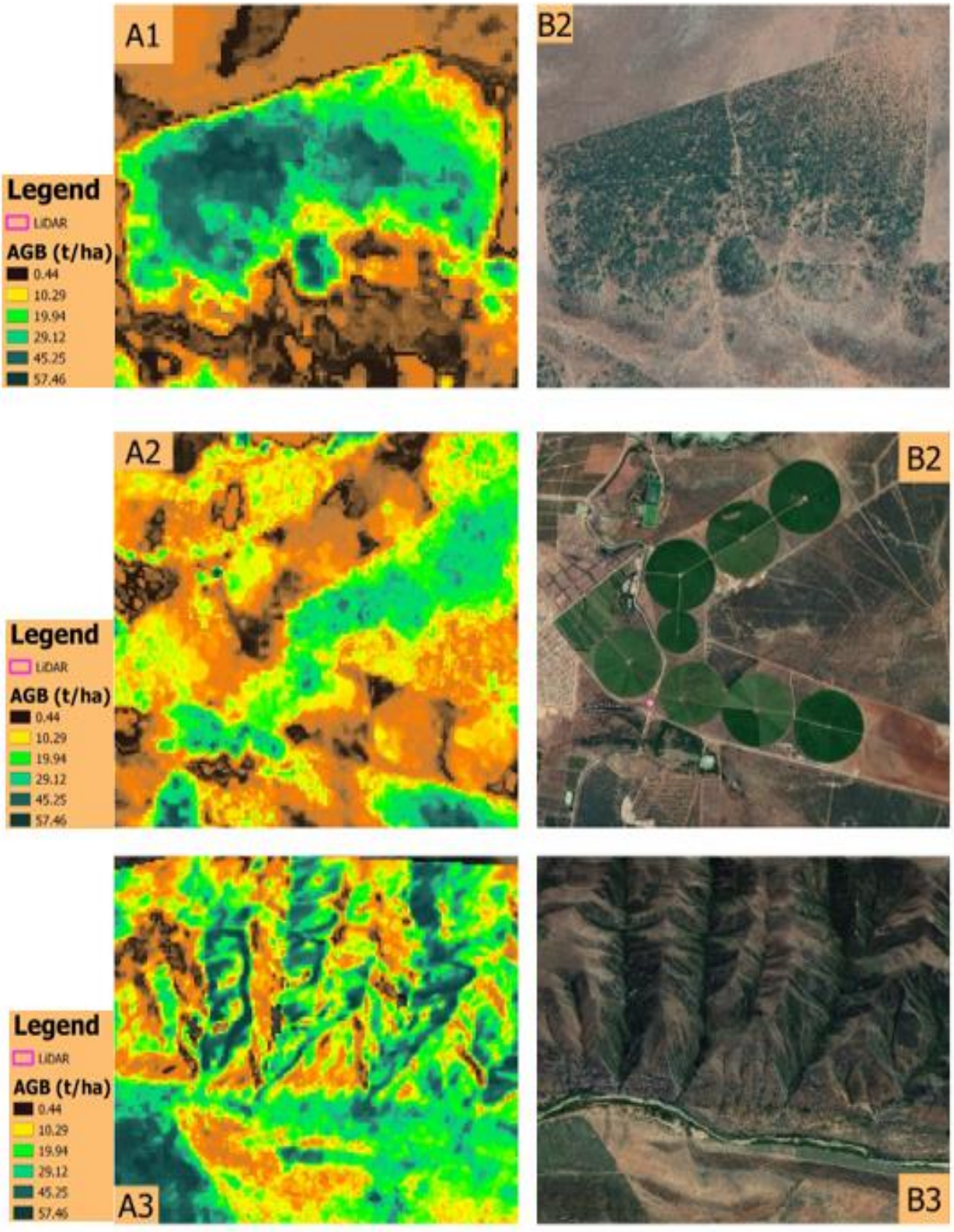


Figure 4. 10: Comparison of AGB map results to google earth image

CHAPTER 5: DISCUSSIONS CONCLUSIONS AND RECOMMENDATIONS

In this study, we set out to explore different modelling techniques to upscale field sample AGB data to LiDAR level AGB products, which then acted as calibration data for regional-scale satellite-based thicket AGB estimates. The Addo Elephant National Park was chosen as the study area since government conservation efforts are already underway and that some of the other parts of the Eastern Cape have experienced urbanisation, overgrazing, as well as general environmental degradation which have resulted in a considerable change in the indigenous tree species that this study aimed to investigate.

The other important part of this research was to investigate the ability of the C band Sentinel 1 SAR data in modelling thicket AGB and to compare it with the well-proven ALOS PALSAR 2 L band mosaic data. This is important due to the hyper temporal nature of Sentinel 1 data as well as that this data set is available to the public free of charge.

5.0 Discussions

5.0.1 LiDAR model

Field determined AGB of 8 plots which were carefully chosen in the Addo Elephant National Park were successfully upscaled using LiDAR mean vegetation height (MH) and percentage canopy cover (CC) metrics using a simple linear regression. Due to the small size of the field samples, the Leave One Out Validation technique was used to model the relationship between field AGB and the LiDAR metrics. Even though a model with a very good R^2 of 89% was derived, this model could have been improved by avoiding the possibility of overfitting as the training data was also used as validation data. However, there is enough precedence in the literature that supports the techniques used in upscaling field AGB using LiDAR metric to give this researcher confidence in the derived model. Therefore, this model was applied to the LiDAR survey covering 34km^2 and this LiDAR AGB significantly increased the sample size which was subsequently used to train and validate satellite-based models. This technique was used before by Li et al, (2016) and Palace et al., (2015) with comparable results.

5.0.2 Satellite-based models

As has already been alluded to in the sections above, satellite data was chosen primarily due to its wide spatial coverage. Sentinel 1 C-band SAR data was an ideal candidate dataset, due to its wide swath (250km^2), hyper temporal resolution and open data distribution policy. The ALOS PALSAR mosaic was used as a check due to its well-document track record of estimating AGB with its L band SAR sensor. However even though PALSAR's global mosaics can be accessed free of charge, this data has a poor temporal resolution of 1 year which can be a limiting factor. However, Addo thicket consists of mostly evergreens hence its AGB is not significantly affected by changes in seasons.

5.0.3 Modelling Techniques

The broad modelling techniques covered in this research included parametric and machine learning with each of these broad techniques represented by a particular modelling technique (i.e.) SLR and RF. Each of these techniques has their strengths and weaknesses in handling different data types and scenarios. The SLR model assumed that the variability in the model can be entirely interpreted by the parameters of the model (i.e.) the variability in AGB across the area of study can be fully explained by the SAR image backscatter metrics. RF tapped into advancements in computer technology and high dimensionality and correlation of spatial data to explain the variation in AGB across the study area. The strength of each of these was then seen in their ability to correctly predict AGB using the validation data.

5.1 Conclusions

The AGB presented here add to the limited knowledgebase and rare spatial data products available for the Albany thicket biome (van der Vyver and Cowling, 2019; Mills et al., 2005a, 2005b; Powell, 2009; Penzhorn et al., 1974; Mills and Cowling, 2006) and are potentially some of the only thicket AGB estimates derived via an upscaling process that included field data, airborne LiDAR, and satellite-based SAR imagery data. This research successfully explored different modelling techniques in estimating the Albany thicket AGB and furthermore, it investigated the ability of the Sentinel-1 C-band imagery in modelling and estimating Albany thicket AGB. Therefore, the conclusions drawn from this research are as follows:

5.1.1 Models

Firstly, a SLR model was derived by regressing field derived AGB and LiDAR metrics and the model was able to explain 89% of the variation in the data set. Then the LiDAR model was used to upscale satellite SAR backscatter values using SLR and RF models. The SLR model was the worst performing and the RF was the best performing model using Sentinel 1 C SAR data. SLR had a coefficient of determination of 0.51, RF had R^2=0.67. This can be attributed to the ability of RF modelling to not overfit the data and its ability to take advantage of correlation and multidimensionality of the data. Therefore, according to this research RF was the best performing method used to model and estimate AGB of the Addo thicket biome.

5.1.2 Comparison of C and L SAR bands

After determining that RF is the best model that can be used to model Sentinel 1 C-band data, the same method was used to model the Addo Thicket AGB using ALOS PALSAR 2 L-band data. According to this study, the C-band performed better than the L-band. The C-band best model has an R^2 of 0.67 while the L-band has a best R^2 of 0.51 which is 16% more.

Hence, in summary, we can broadly conclude that multi-temporal C-band is better to use in thickets than a single L-band image. Furthermore, this research has once more proven the superiority of machine learning algorithms over parametric modelling techniques in modelling Addo thicket AGB. Overall, the performance of the models in this study has contributed to the knowledge of predictive techniques that are suitable for estimating thicket AGB, while the map that was generated can be used in sustainable land management strategies across the district.

5.3 Recommendations

The most important data that influenced the research outcome is the LiDAR AGB that was used to upscale field AGB. One uncertainty in this process is whether there was overfitting due to the use of the SLR model due to the small sample size of field AGB. Hence, in future studies, it is recommended that field sampling should be increased. The other improvement that can be made in this research is that of investigating the significance of including (Chen et al 2018 and Li et al 2016) grey level textures as independent variables in the models.

Another aspect that is worth investigating is that of combining both the L and C bands to model Addo thicket AGB. This has been shown to improve the models in other studies investigating different biomes (Li et al., 2016; Fuchs et al., 2009; Pham et al., 2019; Stelmaszczuk-Górska et al., 2015) but this technique has not been applied in the Addo thicket biome as far as this researcher is aware. The SLR might have performed worse because of the existence of spatial dependence therefore is recommended that non-parametric models that consider spatial heterogeneity such as Geographical Weighted Means (GWR) be test too.

In the future, the model performance should be improved by including more vegetation groups in the sampled data as well as the stochastic component of AGB. The study demonstrated promising results in thicket AGB mapping of the Albany thicket biome using remote sensing-based (i.e. LiDAR and SAR) upscaling of field data; thus, it is helpful and valuable for vital information mining from the Sentinel series when it is applied to the greater Albany thicket AGB estimation endemic in the Eastern Cape province of South Africa. The novelty of this study included the successful upscaling of AGB using field measurements, LiDAR metric as well as SAR satellite data, with the performance of the Sentinel-1 data and RF models being other notable successes.

LIST OF REFERENCES:

Anaya, J.A., Chuvieco, E., Palaciosorueta, A., 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. Forest Ecology and Management. 1237-1246. 10.1016/j.foreco.2008.11.016.

Berninger, A., Lohberger, S., Stängel, M., Siegert, F., 2018. SAR-Based Estimation of Above-Ground Biomass and Its Changes in Tropical Forests of Kalimantan Using L- and C-Band. Remote Sensing 10, 831. https://doi.org/10.3390/rs10060831

Bond, W., Zaloumis, N.P., 2016. The deforestation story: testing for anthropogenic origins of Africa's flammable grassy biomes. Phil. Trans. R. Soc. B 371, 20150170. https://doi.org/10.1098/rstb.2015.0170

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., Asner, G.P., 2018. An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR. Remote Sensing of Environment 206, 156–173. https://doi.org/10.1016/j.rse.2017.12.030

Cai, Y., Lin, H., Zhang, M., 2019. Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data. Advances in Space Research 64, 2233–2244. https://doi.org/10.1016/j.asr.2019.08.042

Chen, L., Ren, C., Zhang, B., Wang, Z., Xi, Y., 2018a. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. Forests 9, 582. https://doi.org/10.3390/f9100582

Chen, L., Ren, C., Zhang, B., Wang, Z., Xi, Y., 2018b. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. Forests 9, 582. https://doi.org/10.3390/f9100582
Colombia: a remote sensing approach. For. Ecol. Manage. 257, 1237–1246.

Dusseux, P., Corpetti, T., Hubert-Moy, L., Corgne, S., 2014. Combined use of multi-temporal E., 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality Earth Observ. Geoinf. 43, 43–54.

ECBCP (Eastern Cape Biodiversity Conservation Plan). 2017. Eastern Cape Biodiversity Conservation Plan Handbook. 9 Department of Economic Development and Environmental Affairs (King Williams Town). Compiled by G. 10 Hawley, P. Desmet and D. Berliner. Draft version, December 2017.

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

Fotheringham, A.S., 2009. "The Problem of Spatial Autocorrelation" and Local Spatial Statistics: Spatial Autocorrelation and Local Spatial Statistics. Geographical Analysis 41, 398–403. https://doi.org/10.1111/j.1538-4632.2009.00767.

Ghasemi, N., Sahebi, M.R., Mohammadzadeh, A., (2011). A review on biomass estimation methods using synthetic aperture radar data. INTERNATIONAL JOURNAL OF GEOMATICS AND GEOSCIENCES. 1. 776-778.

Gleason, C.J., Im, J., 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches": https://www.sciencedirect.com/science/article/pii/S0034425712002787

Guo, Y., Li, Z., Zhang, X., Chen, E., Bai, L., Tian, X., He, Q., Feng, Q., Li, W., 2012. Optimal Support Vector Machines for forest above-ground biomass estimation from multisource remote sensing data, in: 2012 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2012 - 2012 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Munich, Germany, pp. 6388–6391. https://doi.org/10.1109/IGARSS.2012.6352721.

Kumar, A., Kishore, B.S.P.C., Saikia, P., Deka, J., Bharali, S., Singha, L.B., Tripathi, O.P., Khan, M.L., 2019. Tree diversity assessment and above ground forests biomass estimation using SAR remote sensing: A case study of higher altitude vegetation of North-East Himalayas, India. Physics and Chemistry of the Earth, Parts A/B/C 111, 53–64. https://doi.org/10.1016/j.pce.2019.03.007.

Li, W., Niu, Z., Li, Z., Wang, C., Wu, M., Muhammad, S., 2016. Upscaling coniferous forest above-ground biomass based on airborne LiDAR and satellite ALOS PALSAR data. J. Appl. Remote Sens 10, 046003. https://doi.org/10.1117/1.JRS.10.046003.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest 2, 5.
machine learning approaches. Remote Sens. Environ. 125, 80–91.

Mucina, L. & Rutherford, M. C. 2006. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia.

Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. International Journal of

Applied Earth Observation and Geoinformation 18, 399–406. https://doi.org/10.1016/j.jag.2012.03.012.

Naidoo, L., Mathieu, R., Main, R., Kleynhans, W., Wessels, K., Asner, G., Leblon, B., n.d. Savannah woody structure modelling and mapping using multi- frequency (X-, C- and L-band) Synthetic Aperture Radar (SAR) data 49.
of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. Sci. Rep. 7, 6940.

Omar, H., Misman, M.A., Kassim, A.R., 2017. Synergetic of PALSAR-2 and Sentinel-1A SAR Polarimetry for Retrieving Aboveground Biomass in Dipterocarp Forest of Malaysia 21. optical and radar satellite images for grassland monitoring. Remote Sens. 6, 6163–6182. landsat 8 and random forest. Remote Sens. 7, 10017–10041.

Paloscia, S., Macelloni, G., Pampaloni, P., Sigismondi, S., 1999. The potential of C- and L-band SAR in estimating vegetation biomass: the ERS-1 and JERS-1 experiments. IEEE Trans. Geosci. Remote Sensing 37, 2107–2110. https://doi.org/10.1109/36.774723.

Parr, C.L., Gray, E.F., Bond, W.J., 2012. Cascading biodiversity and functional consequences of a global change-induced biome switch: Savanna biome switch. Diversity and Distributions 18, 493–503. https://doi.org/10.1111/j.1472-4642.2012.00882.

Pham, T., Yokoya, N., Bui, D., Yoshino, K., Friess, D., 2019. Remote Sensing Approaches for Monitoring Mangrove Species, Structure, and Biomass: Opportunities and Challenges. Remote Sensing 11, 230. https://doi.org/10.3390/rs11030230.

Pierce S.M. & Mader, A.D. 2006. The STEP Handbook. Integrating the natural environment into land use decisions at 30 the municipal level: towards sustainable development. Centre for African Conservation Ecology Report 31 Number 47 (2nd Edition). Nelson Mandela Metropolitan University, South Africa.

Quegan, S., Jiong Jiong Yu, 2001. Filtering of multichannel SAR images. IEEE Trans. Geosci. Remote Sensing 39, 2373–2379. https://doi.org/10.1109/36.964973.

Shimada, M., Isoguchi, O., Rosenqvist, A., Tadono, T., 2008. PALSAR CALVAL and generation of the continent scale mosaic products for Kyoto and Carbon Project, in: Meynart, R., Neeck, S.P., Shimoda, H., Habib, S. (Eds.), Presented at the SPIE Remote Sensing, Cardiff, Wales, United Kingdom, p. 71060I. https://doi.org/10.1117/12.800897.

Stelmaszczuk-Górska, M., Rodriguez-Veiga, P., Ackermann, N., Thiel, C., Balzter, H.,Schmullius, C., 2015. Non-Parametric Retrieval of Aboveground Biomass in Siberian Boreal Forests with

ALOS PALSAR Interferometric Coherence and Backscatter Intensity. J. Imaging 2, 1. https://doi.org/10.3390/jimaging2010001

Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M., Schmullius, C., 2018. Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. Carbon Balance Manage 13, 5. https://doi.org/10.1186/s13021-018-0093-5

van der Vyver, M.L., Cowling, R.M., 2019. Aboveground biomass and carbon pool estimates of Portulacaria afra (spekboom)-rich subtropical thicket with species-specific allometric models. Forest Ecology and Management 448, 11–21. https://doi.org/10.1016/j.foreco.2019.05.048 vegetation using WorldView-2 imagery and random forest regression algorithm. Int.J. Appl. Earth Observ. Geoinf. 18, 399–406.

Vlok, J.H.J., Euston-Brown, D.I.W. & Cowling, R.M., 2003. Acocks' Valley Bushveld 50 years on: new perspectives on the 1 delimitation, characterization and origin of thicket vegetation. South African. Journal of Botany, 69: 27–51.

Yun Guo, Xin Tian, Zengyuan Li, Feilong Ling, Erxue Chen, Min Yan, Chunmei Li, 2014. Comparison of estimating forest above-ground biomass over montane area by two non-parametric methods, in: 2014 IEEE Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2014 - 2014 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Quebec City, QC, pp. 741–744. https://doi.org/10.1109/IGARSS.2014.6946530

Zaloumis, N.P., Bond, W.J., 2016. Reforestation or conservation? The attributes of old growth grasslands in South Africa. Phil. Trans. R. Soc. B 371, 20150310. https://doi.org/10.1098/rstb.2015.0310.

Zeng, N., Ren, X., He, H., Zhang, L., Li, P., Li, Z., Zhang, L., 2017. Aboveground biomass of grassland in the three-river headwaters region based on neural network. Res. Environ. Sci. 30, 59–66.