

https://dspace.org/

Modelling and mapping canopy cover in African savanna using C-band Synthetic Aperture Radar (SAR); the case study of Bushbuckridge Local Municipality (BLM)

Khoza, Gladness Mikateko

2021

https://hdl.handle.net/10539/33038

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

2021



UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Modelling and Mapping Canopy Cover in African Savanna using C-band Synthetic Aperture Radar (SAR); The Case Study of Bushbuckridge Local Municipality (BLM)

BY:

Gladness Mikateko Khoza Student No: 1109534

A research report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in GIS and Remote Sensing

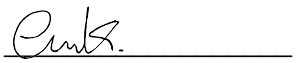
Supervisor: Prof Elhadi Adam

Co-Supervisor: Dr Russell Main

01 August 2021

Declaration

I ***Gladness Mikateko Khoza*** student no. ***1109534*** declare that this research report being submitted for the Degree of Master of Science (GIS and Remote Sensing) at the University of the Witwatersrand, Johannesburg is my own work and has never been submitted before for any degree or examination at any other University.



Gladness Mikateko Khoza

Date: 01 August 2021

Abstract

The savanna biome covers approximately 20% of the earth's land surface and more than half of the African continent. It is home to millions of people and different fauna and flora species but is under pressure due to natural and anthropogenic factors threating the existence of the biome. Remote sensing techniques have the capabilities of assessing savanna vegetation by mapping and monitoring savanna woody structure. Remote sensing systems are divided into passive (e.g., Landsat) and active sensors (e.g., Light Detection and Ranging- LiDAR and Synthetic Aperture Radar- SAR), with active sensors better suited for assessing woody structure than passive sensors. This is because their signals can penetrate through dense vegetation cover. LiDAR sensors are more accurate than SAR sensors, hence often used to train SAR-based models. However, LiDAR data is expensive to collect over large regions. Therefore, spaceborne SAR datasets are more useful for vegetation studies as they have a large spatial coverage and can capture large images at once.

Due to the nature of savanna vegetation, monitoring vegetation within this biome is a challenging endeavour. This is because savannas are characterised by a mixture of grass and sparsely distributed trees mostly in the form of shrubs and woody vegetation. The success of monitoring savannas is dependent on the availability of up-to-date, accurate and well-validated spatial information. Accessing the information requires users to go through several pre-processing steps to turn the data into easily analysed formats. However, the pre-processing steps are challenging for users to implement and computationally expensive. Analysis Ready Data (ARD) products promise a future where remote sensing data will no longer include computationally expensive or have challenging pre-processing steps but instead provides easy-to-use satellite data mapping and modelling savanna vegetation.

This study assesses the utility of multi-temporal 'Analysis Ready' C-band SAR data to estimate woody canopy cover. It seeks to test different regression methods; Linear (LR), Support Vector Machine (SVM) and Random Forest (RF) regression models and their effectiveness in modelling woody canopy cover. Identify the optimal season for modelling canopy cover using different temporal combinations (dry & wet seasons of 2017 & 2018) and polarisations (VV & VH) of C-band SAR data, and subsequently mapping canopy cover using the most effective regression method and optimal season. The results show that although SVM produced higher results regardless of the season, noise or number of images that were analysed, it predicted for a smaller range while RF predicted for a larger range and performed better the more data or

noise was added to the analysis. LR was not sensitive, or robust, enough to find good relationships between the variables as its performance dropped (compared to RF) as more images were added to the analysis. A combination of all seasons (dry & wet), years (2017 & 2018), and all polarisation bands (VV & VH) yielded higher results than when the images were analysed individually. Although the ARD products used in this study are still in the experimental phase, the results produced here are comparable to other savanna vegetation studies that used the operational C-band SAR data.

Keywords: Synthetic Aperture Radar, Analysis Ready Data, C-band, Canopy Cover, Savanna, Linear Regression, Support Vector Machine, Random Forest

Acknowledgements

I would like to thank the almighty God who made it possible for me to get accepted to the University of the Witwatersrand, for being the source of my strength, for giving me wisdom and guiding me throughout this journey. I can never give myself credit but give it all to Him (Philippians 4:13). My Parents Gloria and Glyden Khoza, and my siblings Courage M. and Khensani T. Khoza, I can never thank you enough for all that you have done for me throughout my academic journey. Thank you for allowing me to be the best that I could ever be. Your everyday prayers have sustained me throughout this journey.

Mr Russell Main, my Co-Supervisor. I thank you for your support, supervision, patience, and guidance through this academic journey. I am entirely grateful for your contribution to the completion of this study. I am grateful to you and the Council for Scientific and Industrial Research (CSIR) for providing me with data for my research and assisting me in choosing a research topic. If it were not for your help, I do not know how my research journey would have been like. My main supervisor, Professor Elhadi Adam thank you for taking your time to help me where you could, for giving me the opportunity to select a topic from the CSIR, little did I know, I would learn so much from them and my research project journey.

My proof-readers, Thembelani H. Dlamini and Mandisa S. Mntambo, I can never thank you enough, words fail me at this point, your support throughout this journey, literally dropping everything you are doing just to help me with my academics, checking my errors and proofreading my work was your responsibility, thank you for always giving me your honest opinions and always helping me wherever you could. Liberty Mlambo, my Zimbabwean friend who helped me with everything, for your invaluable advice and wisdom. I will always remember how you encouraged me to always do better. Tshepo Matsana, my best friend, thank you for being there for me throughout my academic journey, for your unwavering support and willingness to help even when you did not understand anything about my research. Thank you for taking care of me when I was admitted in the hospital and could not do my work. You are appreciated.

People often get carried away thanking others forgetting to thank themselves for without their own effort they would have not achieved what they did. So, I Gladness Mikateko Khoza, would like to give a special thanks to myself for I could have not done this without the motivation and determination I had to finish my master's degree. It was not easy, but I managed. I made it.

This research contains modified Copernicus Sentinel-1 data [2017 & 2018], processed by NORCE (Tromsø, Norway) under ESA contract No. 4000125675/18/INB".

TABLE OF CONTENTS

Declaration i
Abstract ii
Acknowledgements iv
List of Figures viii
List of Tables ix
List of Abbreviations x
CHAPTER 1: Introduction 1
1.1. General Introduction 1
1.2. Problem Statement 2
1.3. Aim 3
1.4. Objectives 3
1.5. Outline of the Research Report 3
CHAPTER 2: Literature Review 5
2.1. The Savanna Biome 5
2.2. Remote Sensing Background 7
2.3. Passive Sensors 8
2.4. Active Sensors 8
2.4.1. LiDAR 8
2.4.2. RADAR 9
2.5. SAR Sensors 9
2.6. SAR Interaction 11
2.7. SAR Polarisation 12
2.8. Modelling Algorithms 12
2.8.1. Linear Regression (LR) 13
2.8.2. Support Vector Machine (SVM) 13
2.8.3. Random Forest (RF) 14
2.9. Application of Modelling Algorithms on Savanna Vegetation Studies 15
2.10. Summary of Literature 16
CHAPTER 3: Methods and Materials 16
3.1. Study Area 16
3.2. Remote Sensing Data Acquisition and Pre-processing 18
3.2.1. LiDAR Data 18
3.2.2. Sentinel-1 Images 20
3.2.3. SAR Pre-processing 21
3.3. Calibration and Validation Dataset 23

3.4. Statistical Analysis 23
3.4.1. Linear Regression 23
3.4.2. Support Vector Machine 23
3.4.3. Random Forest 24
3.4.4. Accuracy Assessment and Modelling 25
CHAPTER 4: Results 26
4.1. Optimal Season and Most Effective Regression Method 26
4.1.1. Individual Scenes, per polarisation 26
4.1.2. Individual Scenes with both Polarisations combined 28
4.1.3. Seasonal Combinations of Scenes by Polarisation and Year 31
4.1.4. Seasonal Combinations of scenes of all Polarisations by Year 32
4.1.5. Multi-year Combination of Scenes per Season and Polarisation 33
4.1.6. Multi-Year combinations of scenes for all polarisation bands by season or both 34
4.2. Mapping Canopy Cover 36
4.2.1. Canopy Cover Distribution 36
4.2.2. LiDAR CC Frequency Plot 39
4.3. Summary of Results 39
CHAPTER 5: Discussion and Conclusion 40
5.1. Discussion 40
5.1.1. Optimal Season and Most Effective Regression Method 40
5.1.2. Mapping Canopy Cover 43
5.2. Conclusion 43
REFERENCES 46

List of Figures

Figure 1: Location and Geographical Distribution of the Savanna Biome around the World (Source: ESRI, 2019) 6
Figure 2: Study area map of the Bushbuckridge Local Municipality in Mpumalanga, South Africa, with the spatial coverage of the LiDAR boundaries (red rectangles; Source: CSIR; S.A Municipal Demarcation Board) 17
Figure 3: Total Monthly Rainfall for Bushbuckridge Local municipality, Mpumalanga 2017-2018. (Source: Global Precipitation Measurement (GPM)) 18
Figure 4: ARD processing chain from GRD (Level 1) to ARD products (Level 3; Source: Haarpaintner, 2019). 22
Figure 5 (a-f): Observed vs predicted canopy cover scatter plots of LR, SVM and RF individual scenes, all polarisation March 2017 scenario (b, d, f), and LR, SVM and RF multi-year combinations of scenes for all polarisation bands and all seasons combined (a, c, e). The black dashed line represents the 1:1 line, while the red solid line is the regression line..... 35
Figure 6: Multi-Temporal C-band SAR predictive canopy cover (CC) map using Random Forest (RF). Letters A, I, J & W represent areas covered by the LiDAR CC. 37
Figure 7: Multi-Temporal C-band SAR predictive canopy cover (CC) map using Support Vector Machine (SVM). Letters A, I, J & W represent areas covered by the LiDAR CC..... 37
Figure 8: Observation of the representation of Random Forest vs SVM maps 38
Figure 9: Frequency plot showing the distribution of the RF (red) vs SVM (blue) predicted canopy cover 39

List of Tables

Table 1: Average density and spacing of LiDAR returns for the study areas in Bushbuckridge Local Municipality, Mpumalanga. (Source: CSIR) 19
Table 2: Summary of the Sentinel-1 Data Acquisition. (Source: Haarpaintner, 2019) 20
Table 3: Validation results from the scenario whereby 2017 VH polarised scenes were analysed..... 26
Table 4: Validation results from the scenario whereby 2018 VH polarised scenes were analysed 27
Table 5: Validation results from the scenario whereby both polarisations (VV & VH) for 2017 were analysed..... 29
Table 6: Validation results from the scenario whereby both polarisations (VV & VH) 2018 polarised scenes were analysed..... 30
Table 7: Validation results from the scenario whereby VH polarised scenes from both 2017 and 2018 were analysed seasonally 31
Table 8: Validation results from the scenario whereby all polarised scenes from both 2017 and 2018 were analysed seasonally. 32
Table 9: Validation results from the scenario whereby VH polarised scenes from both 2017 and 2018 were combined 33
Table 10: Validation results from the scenario whereby all polarised scenes from both 2017 and 2018 were combined 34

List of Abbreviations

ANN	Artificial Neural Network
ARD	Analysis Ready Data
BLM	Bushbuckridge Local Municipality
CART	Classification and Regression Trees
CC	Canopy Cover
CEOS	Committee on Earth Observation Satellites
CHM	Canopy Height Model
CSIR	Council for Scientific and Industrial Research
CSM	Top-of-canopy Surface Model
DEM	Digital Elevation Model
DT	Decision Tree
DTM	Digital Terrain Model
ESRI	Environmental Systems Research Institute
GRD	Ground Range Detected
GSAR	Generic Synthetic Aperture Radar
LiDAR	Light Detection And Ranging
LR	Linear Regression
NDI	Normalized Difference Index
NORCE	Norwegian Research Centre

RADAR	Radio Detection and Ranging
RF	Random Forest
SAR	Synthetic Aperture Radar
SRTM	Shuttle Radar Technology Mission
SVM	Support Vector Machine
UTM	Universal Transverse Mercator
WGS	World Geodetic System

CHAPTER 1: Introduction

1.1. General Introduction

The savanna biome is a mixture of tree and grass species (Naidoo et al., 2015; Higginbottom et al., 2018; Wessels et al., 2019) characterised by essential ecosystem services such as the availability of ecological and socioeconomic benefits, natural resources, as well as providing habitats for the human population. The biome supports communities by providing crucial ecological benefits and socioeconomic opportunities (e.g., tourism, agriculture, conservation; Fisher et al., 2015). Natural resources such as fuelwood, medicinal and timber products often act as a safety net for alleviating poverty in the surrounding communities (Madonsela et al., 2018). Due to the extensive use of natural resources, the landscape of savannas has increasingly been altered (Mograbi et al., 2015). The alterations may lead to a change in the diversity and distribution of wildlife species (Adjorlolo, 2008) through the loss of biodiversity (Fisher et al., 2015) and thus contribute to fuelwood scarcity (Mograbi et al., 2015). Therefore, knowledge about the spatial information of woody vegetation density and canopy cover in the savanna biomes is crucial for timely biome assessments and management.

According to Horning et al. (2010) and Jansen (2014), remote sensing is a powerful tool used to map and monitor ecosystems such as the savanna vegetation and document the changes that take place over time. Remote sensing tools are divided into passive sensors and active sensors. The passive sensors monitor the earth's resources, using the sun as their source of energy, to provide information about the surface of the canopy on vegetation (Horning et al., 2010). However, their operational capabilities are limited as they are unable to provide structural information beyond the top-of-canopy or when atmospheric conditions are less than ideal (e.g., cloud cover, dust, smoke; Srivastava et al., 2008). To acquire subcanopy information and to gain a greater understanding of the vegetation structure, active sensors; Synthetic Aperture Radar (SAR) and Light Detection and Ranging (LiDAR) are recommended (Fisher et al., 2015). According to Kumar and Mutanga (2017), active sensors are better at modelling the vertical distribution of elements that make up the canopy, by penetrating through the canopy's upper layers.

Therefore, SAR and LiDAR are, the most effective tools for assessing savanna woody vegetation due to their capability to acquire information about the subcanopy or vertical structure of the vegetation. Furthermore, a SAR-based approach can map large extents of woody vegetation in all types of weather conditions. In contrast, capabilities of mapping large

extents of woody vegetation are not easily achieved in a LiDAR-based approached (Naidoo et al., 2015). For this reason, this study uses the SAR-based approach (multi-temporal Sentinel-1 C-band SAR) together with modelling algorithms: Linear Regression (LR), Support Vector Machine (SVM) and Random Forest (RF) to model and map woody canopy cover (CC) in African savannas.

Apart from the C-band, the SAR also has the L-band that, according to Mathieu et al. (2013), Urbazaev et al. (2015), Naidoo et al. (2015) and Main et al. (2016), has predominantly been the most common band of use in the savanna because it penetrates deeper into the canopy and produces good results. However, the L-band is not open-source and not routinely available. Thus, the C-band was the choice of use in this study. According to Main et al. (2016), the C-band continues to be the only freely available SAR dataset. Furthermore, the savanna has sparsely distributed trees that prevent closed-canopy covers (ESRI, 2019), which means that the C-band can perform relatively well under multi-temporal data regardless of its short wavelength. This is because the multi-temporal SAR images are used to improve the influence of vegetation-related backscatter to reduce the noise coming from other environmental sources (Main et al., 2016).

1.2. Problem Statement

Approximately 20% of the earth's land surface is covered by the savanna biome (Shorrocks, 2007; Jensen, 2014) primarily located at the tropical and subtropical zones of continents such as Africa (Graw et al., 2016). The savanna is an essential ecosystem characterized by a balance of tree-grass species covering more than half of Africa's land surface and supports many communities within the continent (Main et al., 2016). It provides a habitat to millions of people, plants, and animals, and offers ecosystem services such as food, water, and clean air (Fisher et al., 2015; Stafford et al., 2017). However, there have been predictions that savannas will become severely impacted by climate and land use changes (Sankaran et al., 2005).

Areas such as Bushbuckridge Local Municipality (BLM), Mpumalanga, for instance, might be characterised by unsustainable use of natural resources (Fisher et al., 2015), with Wessels et al. (2013) predicting that at the rate which wood is being harvested, it could be depleted by the year 2024. Although majority of South African households have access to electricity (Wessels et al., 2011), wood is still a primary source of energy for many areas within the BLM (Shackleton and Shackleton, 2000; Twine et al., 2003; Wessels et al., 2011; 2013). Over 90% of households are dependent on fuelwood for cooking, heating, and other means of thermal

energy purposes despite having access to electricity (Wessels et al., 2011; Twine and Holdo, 2016). Due to this unsustainable harvesting of wood, the savanna woody vegetation structure has become severely affected particularly causing a decline in biomass, vegetation height, canopy cover, and density of larger stems (Twine and Holdo, 2016).

Therefore, the savannas need to be monitored to avoid major losses. The success of monitoring them is dependent on the availability of up-to-date, accurate and well-validated spatial information. However, there is a shortage of detailed and validated woody structure related (e.g., woody canopy cover) spatial products in the (southern) African savannas (Naidoo et al., 2016), and there are no up-to-date spatial products concerning the condition of (southern) African savannas (Main et al., 2016). Therefore, this study seeks to assess the utility of multi-temporal 'Analysis Ready' data to model and predict woody canopy cover of an African savanna, using the case study of BLM to contribute to the timely assessment and monitoring of the savanna vegetation.

1.3. Aim

This study aims to assess the utility of multi-temporal 'Analysis Ready' C-band SAR data to estimate savanna woody canopy cover.

1.4. Objectives

- To test different regression methods (LR, SVM and RF) and their effectiveness in modelling woody canopy cover.
- To identify the optimal season for modelling canopy cover using different temporal combinations (dry & wet seasons of 2017 and 2018) and polarisation bands (VV & VH) of C-band SAR data.
- To map canopy cover using the most effective regression method and optimal season

1.5. Outline of the Research Report

This research consists of five chapters which include:

- **Chapter 1:** This chapter provides the background of the research and describes the problem statement, aim and objectives.
- **Chapter 2:** This chapter reviews the relevant literature related to the savanna biome, background on remote sensing, and information on the different remote sensors and their capabilities. This chapter also discusses the modelling algorithms used in this research and their applications in other savanna vegetation studies.

- **Chapter 3:** This chapter describes the research methodology. This includes the study area, the research methods and materials that entail the data acquisition and processing and the statistical analysis used in this research.
- **Chapter 4:** This chapter presents the study results by answering the research objectives using tables, graphs, and maps.
- **Chapter 5:** This chapter discusses the findings by linking them to the aim and objectives and determining whether the study's aim and objective were achieved. This chapter is also the final chapter which concludes the study and present recommendations.

CHAPTER 2: Literature Review

2.1. The Savanna Biome

Due to the co-dominance of tree and grass species, savannas are generally water limited and mostly influenced by natural and anthropogenic processes (e.g., wood-harvesting, overgrazing, and fire regimes; Sankaran et al., 2008). The biome has noticeable seasonal variations that results in arid, semi-arid and humid savanna types (Ward et al., 2013). During the dry season, savannas often have a negative water balance that affects flora and fauna present in the ecosystem. During the wet season, the water balance changes (Shorrocks, 2007). If not properly monitored, the changes between the dry and wet season may alter the tree-grass balance and result in the savanna being converted to a forest or grassland biome (Sankaran et al., 2004).

There are intensifying concerns all over the world over their health and sustainability (Higginbottom et al., 2018). In 1993, Scholes and Walker (1993) documented that savannas occupied a fifth (20%) of the earth's land surface, while in 1997, Scholes and Archer (1997) recorded that they covered an eighth (12.5%) of the land surface. Sankaran et al. (2005) and Jensen (2014) noted that the savannas covered a fifth (20%) of the earth's land surface. Although savannas are poorly understood (Sankaran et al., 2005; Urbazaev et al., 2015). It is certain that they are changing, and these changes have been influenced by water and nutrient availability (Frost et al., 1986), livestock production (Sankaran et al., 2005), increasing human population (Fischer et al., 2015), changes in air temperature and precipitation (Urbazaev et al., 2015), amongst others. These change variables need to be thoroughly understood to understand the savanna ecosystem thereby providing appropriate management practices.

The African continent's land cover is remarkably diverse due to varying climatic and ecological conditions (Graw et al., 2016). The continent's land cover ranges from the dry non-vegetated areas of the Sahel desert to the humid area covered by the rainforests of Central Africa. The transitional zones between the deserts and rainforests consist of tropical and subtropical zones primarily dominated by the savanna ecosystem. The savannas cover more than half of the African continent (Scholes and Walker, 1993; Scholes and Archer, 1997; Twine et al., 2003; Shackleton and Shackleton, 2004). They are known for their dynamic tree-grass interaction and co-dominance (Sankaran et al., 2005; Ward et al., 2013; Urbazeav et al., 2015) that supports a large number of communities and contributes greatly to reducing the world's carbon footprint (Naidoo et al., 2015; Naidoo et al., 2016; Main et al., 2016).

Apart from Africa, savannas are also found in South America, India (Asia) and Australia as shown in the figure below (Figure 1). They cover more than half the area of Australia, 45% of South America, and 10% of India (Scholes and Archer, 1997). They mainly occur close to the equator where the climate is relatively warm and dry (Shorrocks, 2007; ESRI, 2019). As savannas have a dry climate, they tend to experience seasonal drought, with the vegetation often affected by rainfall, fire, grazing and soil types. As a result, the biome has adopted a wide variety of fauna and flora which has adapted to the ecosystem and its climate (ESRI, 2019). However, savanna ecosystems are likely to differ due to the different conditions of the areas they are found. For instance, Africa's savanna trees tend to be deciduous while those of South America and Australia tend to be ever green (Shorrocks, 2007).

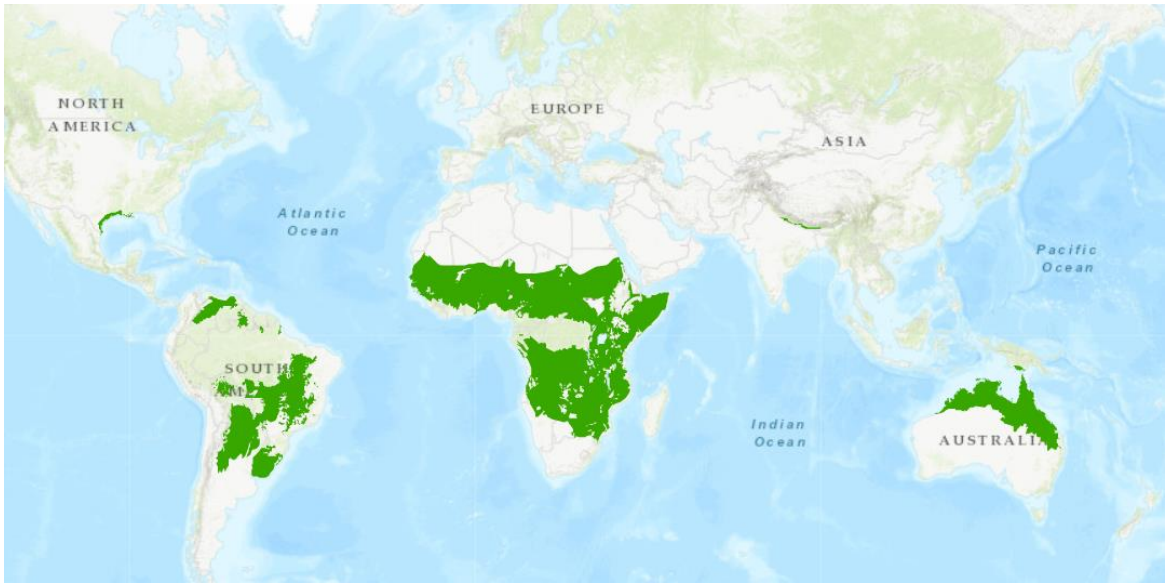


Figure 1*:* Location and Geographical Distribution of the Savanna Biome around the World (Source: ESRI, 2019)

Wigley et al. (2020) described African savannas as unique due to their diversity of herbivore mammals, which differ in sizes ranging from a few kgs to over 6 tonnes. The African savanna is vast as shown in Figure 1. It covers more than half the continent ranging from the Northwest to the East and down to the South of the continent. As one moves from the wet Congo basin to the desert of the Sahara, the savanna changes from densely covered areas to open grassland. South of the Congo basin towards the areas of Angola, Namibia and South Africa, the savanna becomes semi-arid. The savannas of Kenya and Tanzania join the North and South regions of the savanna (Shorrocks, 2007). The African savanna's vegetation differs substantially due to varying climatic conditions (Graw et al, 2016). It is comprised of species that range from the Acacia trees to the savanna shrubs, the Terminalia woodland and many more (Shorrocks, 2007).

In South America, the savanna is spread out mostly across the south of the equator and includes plant species such as the cerrado, llanos de Moxos, Gran Sabana, savannas of the Rio Branco-Rupununi and many more (Shorrocks, 2007). They are the 2nd largest type of vegetation after the tropical forests of the continent and are a great example of adaptive shift as they have contributed significantly to the origination of the diversity of tropical species (Borghetti et al., 2019). In Australia, the savannas are widespread over the north and include species such as the Monsoon tallgrass, Tropical tallgrass, Tussoc and many more. In India, the savanna covers the Northeast region of the country as illustrated in Figure 1. The savannas of India are believed to be from woodland systems brought by deforestation, abandoned cultivation as well as burning and they comprise of species such as the Zizyphus, rubicaulis and Acacia catechu (Shorrocks, 2007).

Over the years, there have been inadequate explorations of woody canopy cover mapping using SAR images in savannas across the world, mainly in southern Africa (Urbazaev et al., 2015). Furthermore, savannas have been susceptible to many negative effects, hence often affected by phenomena such as bush encroachment (Kgosikoma and Mogotsi, 2013). Bush encroachment may be because of over-grazing, over-harvesting, or an unsustainable fire regime (Fisher et al., 2015) which may lead to food insecurity since as bush encroachment increases, agricultural potential decreases (Kgosikoma and Mogotsi, 2013). Moreover, savannas are important ecosystems as the availability of natural resources and disturbance regimes such as fire and grazing in the biome play a role in regulating woody cover (Sankaran et al., 2005). Therefore, it is of utmost importance that savanna ecosystems always be monitored and managed in order to ensure continued provision of ecosystem services not only for human beings but also for the wildlife that exists within the biome.

2.2. Remote Sensing Background

Remote sensing is the process of acquiring information about the state or condition of an object through sensors without physical contact (Chuvieco and Huete (2010). It is the most suitable tool for assessing woody structure across large regions (Naidoo et al., 2016). There are a variety of sensors available now a days, with some sensors (e.g., LiDAR and RADAR) being better suited for assessing woody structure than others. Remote sensing systems are divided into passive and active sensors. Passive sensors collect radiation derived from external sources such as the sun while active sensors produce their own source of energy to detect reflection of the energy from the observed surfaces (Chuvieco and Huete, 2010).

2.3. Passive Sensors

Passive sensors also known as optical sensors are limited to measuring the electromagnetic radiation derived from external sources. Their energy is either reflected from solar radiation or emitted by the earth's surface. Passive sensors have been extensively used in remote sensing for many decades to map vegetation (Chuvieco and Huete, 2010). These passive sensors include systems such as the Landsat satellites which have been used to collect land observation satellite data since the year 1972 and have thus been deemed significant for large scale monitoring applications and many other land observation applications (Staben et al., 2018). The monitoring applications include those such as rangeland monitoring while the other applications include tropical forest mapping and natural hazard assessment.

However, due to the sensitivity of solar illumination differences, the use of Landsat and many other passive sensors tends to be limited by weather conditions and time of the day as they are sun dependent and cannot acquire data at night (Campbell and Wynne, 2011). Furthermore, using passive sensors for mapping woody cover is challenging due to their inability to differentiate between woody and non-woody vegetation layers. Additionally, vegetation in the savanna biome shed their leaves during the dry season to preserve water. Due to this season often being short-lived, it is usually difficult to target for a reliable woody cover mapping using images acquired by passive sensors as there is often a short window to acquire the phenological differences in the vegetation which then produces poor results (Urbazaev et al., 2015).

2.4. Active Sensors

Active sensors can produce energy pulses and collect them after the surface target reflects them back to the sensor and can also penetrate through clouds to ensure continuous data acquisition even during harsh weather conditions (Srivastava et al., 2008). They can generate, transmit, and acquire electromagnetic energy which can penetrate through thick vegetation canopy covers and able to provide data on the 3D distribution of vegetation structures (Wessels et al., 2019). The major sensors for active remote sensing are LiDAR which stands for Light Detection and Ranging, and RADAR which stands for Radio Detection and Ranging.

2.4.1. LiDAR

LiDAR is a system that emits pulses of polarised light to the observed surface. The pulses are detected after reflecting off the target surface (Chuvieco and Huete, 2010). LiDAR was built to transmit energy in a narrow range of frequencies and receive the reflected energy to produce an image of the earth's surface. It can compare the characteristics of the emitted and the

returned energy, the timing pulses, the wavelengths, and the angles in order to measure its angular position, changes in frequency, the timing of reflected pulses as well as brightness of the backscatter (Naidoo et al., 2015). Therefore, such knowledge shows that LiDAR data can be used to obtain information that illustrates the structure of objects that cannot be portrayed by conventional passive sensors.

Although LiDAR data is more accurate than SAR data and provides high-resolution geo-located measurements of vegetation vertical structure and ground elevations beneath thick canopy covers, LiDAR data is not often available on a large or regional scale (Naidoo et al., 2015; Staben et al., 2018). This is because there has been limited availability of space-borne satellite LiDAR and because LiDAR is frequently operated from an aerial platform which usually result in financial constraints due to the high costs of the LiDAR systems. Consequently, its data acquisition is restricted to small spatial coverage and as a result, usually restricted to local scale studies. Hence, LiDAR data is utilised to provide reference data for SAR-based modelling and upscaling (Urbazaev et al., 2015).

2.4.2. RADAR

Radars are the most well-known active sensor systems that work between the 0.1cm to 1m range in the electromagnetic spectrum (Chuvieco and Huete, 2010). They emit a microwave signal, then receive its reflection as the foundation for creating images of the earth's surface. Their signals can penetrate through dense vegetation cover and hard soil surface. Their penetration is associated with frequency, surface roughness and incidence angle (Campbell and Wynne, 2011). They receive a backscatter coefficient which increases in value in relation to the power of the signal received (Campbell and Wynne, 2011). Due to their ability of acquiring images under any weather condition and independent of solar illumination, Radars have been shown to be successful at differentiating between land cover features according to their surface roughness and their canopy penetrating capabilities (Lechner et al., 2020). Synthetic Aperture Radar (SAR) uses a doppler effect (the change in radiation frequency resulting from the relative movement between the sensor and the object) to avoid spatial resolution issues by using a virtual antenna that is artificially synthesized (Chuvieco and Huete, 2010).

2.5. SAR Sensors

SAR sensors function on the notion that features within a scene are illuminated by the radar over a time interval, as the sensor platform moves along its flight path. The SAR sensor receives backscatter from the land surface during this interval and stores the history of the reflections

from each feature (Campbell, 2006). The backscatter is a key parameter of SAR data that computes the returned energy from a feature on the ground and is controlled by the geometry of the features and dielectric properties, which are often determined by water content. The backscatter is dependent on the wavelength, polarisation, and angle of incidence of the emitted wave from the SAR system (Srivastava et al., 2008; Mathieu et al., 2013; Urbazaev et al., 2015). Hence, careful selection of sensor properties is required as they can play a role in defining the sensitivity of the signal to the variable being investigated (Srivastava et al., 2008).

Using SAR data to map savanna woody cover has been demonstrated by authors such as Mathieu et al., 2013, Naidoo et al., 2015, and Main et al., 2016. SAR sensors have been widely used for mapping woody structure due in part to the signal's ability to penetrate the tree canopy and their capability of acquiring data in all weather conditions (Naidoo et al., 2016). SAR has become a powerful tool for mapping woody vegetation using either airborne or spaceborne platforms. The airborne systems allow for local-scale imaging, while spaceborne platforms enable regional to national-scale monitoring of woody resources (Martorella et al., 2012). This makes the understanding and interpretation of SAR images vital in order to be able to detect and map vegetation as well as other features of the environment (Chen and Tzeng, 2012).

SAR sensors, such as Sentinel-1 (A & B) have become popular due to their ability of cloud penetration and being routinely and freely available (Trunckenbrodt et al., 2019). However, accessing the information stored within these sensors requires users to go through several pre-processing steps to turn the data into formats that can be easily analysed. Although the extent of pre-processing required depends on the application, the pre-processing steps can entail applying a precise orbit of acquisition, removing thermal and image boarder noise, performing radiometric calibration, and applying a range doppler and terrain correction (Filipponi, 2019). These steps are in a set of pre-processing levels that can range from Level 0 to Level 3 with Level 0 being raw instrument data and Level 3 being data that can be analysed without extensive pre-processing (Piwowar, 2001). However, the pre-processing steps are challenging and computationally expensive hence there has been a growing need for easy-to-use satellite data (Potapov et al., 2020).

Level 3 data has become the popular easy to use format often referred to as Analysis Ready Data. The Committee on Earth Observation Satellites (CEOS) has defined Analysis Ready Data as "satellite data that has been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort, and

interoperability both through time and with other datasets" (http://ceos.org/ard/). Analysis Ready Data was introduced to deal with the challenges of scalability, increased data volume size, and to allow instant analysis without extensive pre-processing (Trunckenbrodt et al., 2019). Although there are currently few SAR Analysis Ready Data products available for analysis (Potapov et al., 2020). SAR Analysis Ready Data has been used for savanna biomass mapping in South Africa (Haarpaintner et al., 2020) and forest land cover mapping in the Democratic Republic of Congo (Haartpainter and Hindberg, 2019).

2.6. SAR Interaction

SAR sensors operate on a microwave region of 0.1cm to 1m within the electromagnetic spectrum which has been divided into numerous sub-regions that may or may not be useful for woody vegetation studies. The sub-regions have been assigned a band according to their wavelength. There is the K-band which ranges from 0.8cm to 2.4cm, the X-band from 2.4cm to 3.8cm, the C-band from 3.8cm to 7.5cm, the S-band from 7.5cm to 15cm, the L-band from 15cm to 30cm and P-band from 30cm to 100cm (Srivastava et al., 2008). However, the C and L bands are the most common for vegetation studies with the L-band often used for vegetation with thick canopy covers. The preference of the L-band over the C-band is supported by Wessels et al. (2019) who argues that it performs better at retrieving woody cover in savannas and woodlands regions.

According to Urbazaev et al. (2015), the interaction of SAR wavelengths and vegetation elements is influenced by their size and shape. For instance, when detecting vegetation, the short SAR waves (X- and C- bands), interact mostly with small vegetation parts such as leaves and twigs while long waves (L- and P-bands) mostly interact with large vegetation parts such as branches and trunks. The short wavelengths then result in canopy level backscattering with limited signal penetration while longer wavelengths result in a backscattering that can penetrate deeper into the vegetation (Naidoo et al., 2015). As such, studies have shown that short wavelengths are more suitable for retrieving canopy biophysical parameters but not effective for predicting CC in places such as the savannas (Wu et al., 2011).

Long SAR wavelengths (L-band) have a stronger relationship with woody structure than short wavelengths (C-band). This is because the backscatter of longer wavelengths is mainly due to interactions within-canopy (branches, trunks etc.) rather than the top-of-canopy (twigs, leaves etc.) which changes through time (Wessels et al., 2019). However, even though C-band is purportedly not as sensitive as the L-band, the C-band has less limitations when it comes to

mapping woody cover due its regular and operationalised acquisition strategy. The use of L-band images is affected by data continuity issues, sensors failures, and high data costs of SAR systems. Although the different SAR wavelengths may have limitations on their own, merging them has also been shown to be beneficial in woody cover studies (Higginbottom et al., 2018). Naidoo et al. (2015) documented the benefits by comparing the accuracy of modelling woody above ground biomass, canopy cover and total canopy volume using the X, C and L-bands in South African Savannas.

2.7. SAR Polarisation

Remote sensing systems record polarisation characteristics which are crucial for many earth observations studies. Polarisation is the orientation of the emitted and received signal (Naidoo et al., 2015) and it is feasible to selectively emit and receive polarised energy using SAR systems (Jensen, 2015). Polarisation and frequency of SAR data play a crucial role in detecting vegetation structure. Polarised SAR sensors either emit or receive horizontal wave orientation or vertical wave orientation which can be referred to as HH, HV, VH or VV (Liu et al., 2005; Campbell, 2006; Wu et al., 2011; Urbazaev et al., 2015). The polarisations allow for a more comprehensive categorization of the scattering properties of ground targets which then permit the extraction of structural information from vegetation.

For example, HV or VH are more suitable for mapping canopy structures due to the water content in canopies, which results in volumetric scattering within the canopy. This often changes the polarisation of the produced wave from H to V or from V to H (Naidoo et al., 2015). Furthermore, the polarisation also affects the backscatter of vegetative features like the interaction of V orientated waves which are greater with vertical features such as branches and trunks while the interaction of H orientated waves is greater with horizontal features such as leaves and twigs (Urbazaev et al., 2015). Therefore, the polarisation of a SAR signal represents the field direction of the electromagnetic energy that is produced or received by the antenna (Liu et al., 2005; Campbell, 2006).

2.8. Modelling Algorithms

Modelling algorithms are data analysis methods that are chosen based on the nature of the questions and/or desired outputs in a project (McCue, 2007). The modelling algorithms can either be parametric or non-parametric. Parametric modelling algorithms make assumptions about the distribution of their input data while non-parametric algorithms do not make assumptions about their input data but construct their distribution based on information

acquired from its dataset (Garcia et al., 2010). The parametric approaches include LR while nom-parametric approaches include the machine learning algorithms of RF, SVM, REPTree, Artificial Neural Network (ANN), Decision Tree (DT) and many more. The Machine learning algorithms have been used in different remote sensing studies and have produced results higher than parametric approaches (Naidoo et al., 2014, 2015, 2016; Forkuor et al., 2017; Garcia et al., 2018).

According to Naidoo et al. (2014), LR is the simplest algorithm to carry out but is sensitive to outliers and not suitable for data that does not have a linear distribution. RF and SVM are powerful algorithms for facilitating data analysis and use advanced strategies to enhance the probability of reaching the desired goals (Alijani et al., 2018). ANN is a powerful algorithm suitable for analysing images due to its ability to extract spatial features (Bayr & Puschmann, 2019). However, because of their required level of difficulty and customization, ANN and SVM are time consuming and computationally intensive (Naidoo et al., 2015). The modelling accuracies of LR, SVM and RF were compared in this study due to their simplicity, robustness, and powerfulness.

2.8.1. Linear Regression (LR)

LR is a common statistical tool for modelling the relationship between dependent and independent variables (Shai and Shai, 2014). It is the simplest form of regression analysis used when there is only one independent variable. LR is based on the linear relationship between two or more variable and can be represented by the equation $y = a + bx$ (Pietersen and Maree, 2007). The LR model has been used in savanna studies by authors such as Main et al. (2016) to describe the relationship between C-band backscatter and LiDAR metrics. Urbazaev et al. (2015) also used the LR model to analyse the effects of spatial resolution and seasonality between SAR parameters and LiDAR based woody canopy cover.

2.8.2. Support Vector Machine (SVM)

According to Ge et al. (2018), SVM is a supervised non-parametric machine learning technique for regression and classification analyses. It was presented first by Vapnik (1995) with principles based on statistical training theory (Alijani et al., 2018). It is an advanced machine learning method that offers exceptional advantages when dealing with small samples, non-linear dataset and a high dimensional pattern recognition which presents a high correlation with predictor variables (Ge et al., 2018). The SVM is without prior assumption to the original distribution of data and appealing to remotely sensed data due to its ability to handle small

training sample datasets while still producing accurate results (Mountrakis et al., 2011). SVM attempts to achieve a balance between overfitting and prediction accuracy which is different from traditional linear and non-linear regression methods, such as Maximum Likelihood Estimation (Ge et al., 2018).

Furthermore, SVMs are insensitive to the sample size because they can produce good results with a small training sample dataset, unlike DT and ANN which are unable to produce comparable accuracies with small sample sizes. This insensitivity in SVM is overcome by transmitting data into a higher dimension space which uses non-linear kernels, thereby creating a hyper-plane (Alijani et al., 2018). The hyper-plane is created for optimization purpose to increase the boundary between two different variables (Mountrakis et al., 2011). Garcia et al. (2018) used SVM to estimate forest canopy height by demonstrating the potential of using a multi-sensor method for extrapolating LiDAR-based canopy height measures on different ecosystems of diverse forest biomes. They discovered that SAR data alone was successful in estimating canopy height over temperate broadleaf and mixed forest biomes while in areas situated within the conifer biome, SAR data was limited, and multispectral data produced higher accuracies. Overall, the combination of SAR and multispectral data yielded higher results and improved the retrieval of forest canopy height

2.8.3. Random Forest (RF)

RF is an ensemble machine learning algorithm by Breiman (2001) that builds on the Classification and Regression Trees (CART) algorithm. It is a collaborative machine learning method that integrates a multitude of DTs where each DT is implemented using a training set that was constructed by a bootstrap sampling method (Hu et al., 2020). RF generates many DTs and combines them using bootstrapping and bagging. Unlike CART, RF is not sensitive to small alterations in the training data and is not susceptible to overfitting (Naidoo et al., 2015). This is because for RF to avoid overfitting, trees are grown from a random selection of predictors at every point and with a random subset of samples (Breiman, 2001). Therefore, the RF model is regarded as an improvement of CART due to its ability to use many DTs, bagging, and bootstrapping (Naidoo et al., 2016).

RF is a robust model that can be used either for categorical or continuous data in order perform classification or regression analysis (Cutler et al., 2012). In instances where the number of available training samples are limited, the RF still performs well due to its dependence on two user defined values which are the number of decision trees in the forest and the number of

possible features for each node (Waske et al., 2012; Naidoo et al., 2014). Moreover, the RF has proven to be a robust model for denoting the complexity of relationships that may not be linearly distributed (Naidoo et al., 2014). As a result of its robustness, the model has been successfully used to map vegetation structural metrices due to its ability of producing high accuracies (Urbazaev et al., 2015). Higginbottom et al. (2018) used RF to map fractional woody cover in semi-arid savannas by comparing the accuracies of single season Landsat metrics multi-seasonal data, or multi-sensor (Landsat and SAR) data at different spatial resolutions. They found that although Landsat can produce accurate maps of woody cover on its own, the combination of Landsat and SAR yielded higher results.

2.9. Application of Modelling Algorithms and SAR Datasets on Savanna Vegetation Studies

Naidoo et al. (2014) compared the modelling accuracies of woody canopy cover in South African savannas by using techniques such as LR, SVM, REPTree, Decision Tree (DT), ANN and RF using the X-band, C-band, and L-band datasets. Their study used LiDAR data to train its models and evaluate the modelling accuracies of the SAR models. The results showed that the ANN, REPTree and RF models produced high canopy cover prediction accuracies using different combinations of SAR wavelengths. The results also showed that the L-band data should be prioritised because a single L-band image produced higher accuracies compared to the single images of shorter wavelengths such as the X-band and C-band. Urbazaev et al. (2015) demonstrated that the L-band produced promising sensitivity and therefore suggests that the band can be used to map and monitor woody cover changes in South African savannas.

Naidoo et al. (2015) used RF to model and map the accuracies of the structural metrices of aboveground biomass (AGB), woody canopy cover (CC) and total canopy volume (TCV) using the SAR datasets of X-band, C-band, L-band, and their combinations in South African savannas. Although the combination of all bands (X, C, and L-band) produced the best results for all three metrices, the L-band was still more effective in modelling all three metrices than the X and C-bands whether individually or combined. However, the modelling of C-band alone showed promising results which shows that in the absence of L-band datasets, the C-band can produce comparable results in similar environments. Hence, Mathieu et al. (2013) concluded that the C-band SAR datasets show encouraging results for open, semi-arid savannas for use when L-band is not available and makes the implementation of the Sentinel-1 C-band SAR sensor viable.

Main et al. (2016) assessed the strength of the relationship between savanna woody vegetation cover and volume and a time-series of C-band ASAR data using RF and LR models. They used LiDAR data to calibrate and validate the predictive SAR models for total woody canopy cover and total canopy volume. Their results showed that with enough C-band data, the RF model produces accuracies equivalent to studies done using the high-resolution expensive L-band SAR data. The study further produced high prediction accuracies for the models even though the SAR wavelength used was the C-band with a spatial resolution of 75m and a single HH polarisation to estimate woody vegetation structure. This is because according to Haarpaintner et al. (2020), when there is enough data (multi-temporal combinations of C-band data), C-band images reduce signal noise and improve model sensitivity to woody structural related parameters thereby being able to produce results comparable to the L-band.

2.10. Summary of Literature

The savanna biome is home to millions of people and different fauna and flora species but is under pressure due to natural and anthropogenic factors threating the existence of the biome. Remote sensing techniques can assess savanna vegetation by mapping and monitoring savanna woody structure. Remote sensing systems are divided into passive and active sensors. Active sensors, especially SAR are better suited for assessing woody structure than LiDAR or passive sensors due to their capability of penetrating through dense vegetation cover. However, the nature of savanna vegetation makes it a challenging endeavour for monitoring vegetation within the biome. This is because savannas are characterised by a mixture of grass and sparsely distributed trees mostly in the form of shrubs. Analysis Ready Data products promise a future where remote sensing data will no longer include computationally expensive or challenging pre-processing steps but instead provide easy-to-use satellite data for mapping and modelling savanna vegetation.

CHAPTER 3: Methods and Materials

3.1. Study Area

The Bushbuckridge Local Municipality (BLM) is a semi-arid savanna area situated in the Lowveld region of South Africa (Mograbi et al., 2015). It is situated in the north-eastern part

of Mpumalanga (Figure 2), and as of 2016, the Kruger National Park (KNP) forms part of the municipality (Bushbuckridge Local Municipality, 2018). The municipality is made up of two former apartheid homelands, namely Gazankulu and Lebowa which were established with the Native Land Act (No. 27) of 1913 (Fisher et al., 2015). Although BLM falls under state control, in 1994 after the homelands were dismantled, the municipality was divided into communal rangelands administered by tribal authorities who zoned the land into residential, arable, and communal areas for grazing and collection of timber and non-timber products (Mograbi et al., 2015; Fisher et al., 2015). Due to its history, the BLM is densely populated with the years between 1972 and 2012 seeing an increase in human population in the area to 209 people/km^2 which resulted in an increased land utilization intensity and economic impoverishment (Fisher, 2013).

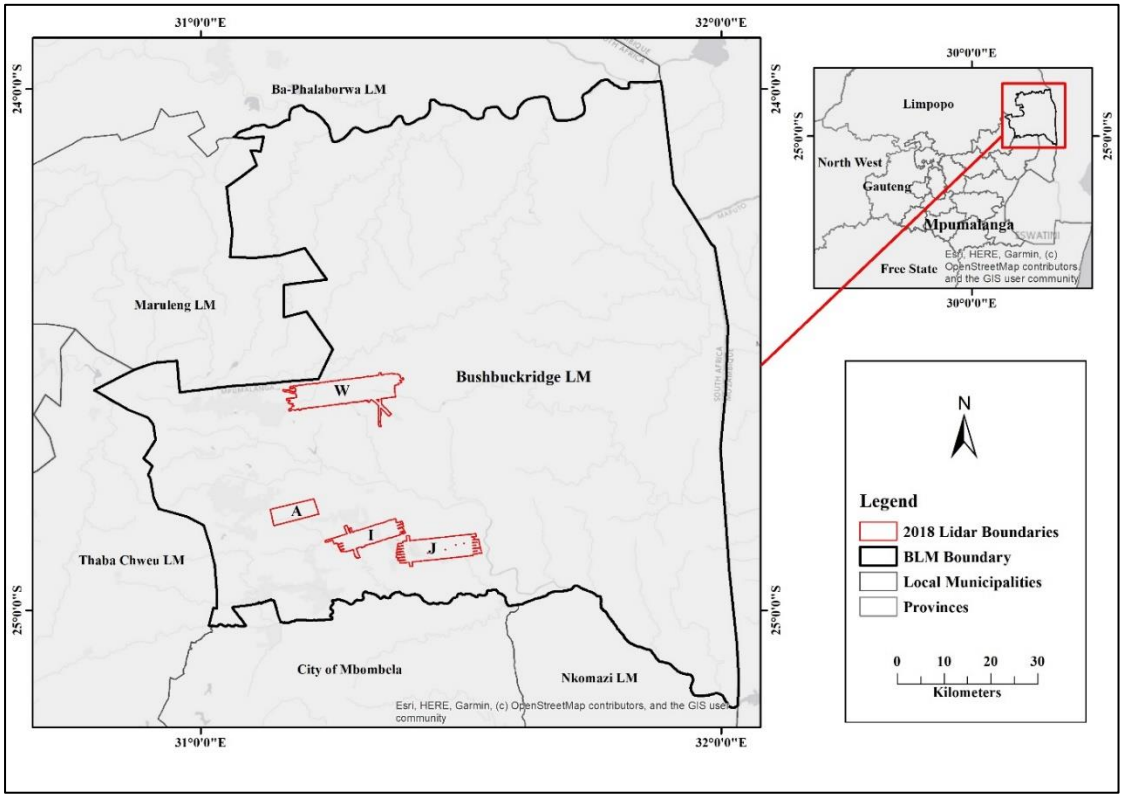


Figure 2: Study area map of the Bushbuckridge Local Municipality in Mpumalanga, South Africa, with the spatial coverage of the LiDAR boundaries (red rectangles; Source: CSIR; S.A Municipal Demarcation Board)

The study area (BLM) has varying natural vegetation types, comprised of mixed bushveld, sweet and sour lowveld, bushveld and clay thorn bush (Mucina and Rutherford, 2006). Woody vegetation is prominent in the study area with canopy cover ranging between 5% in the savanna area to 60% in woodlands and 80% in riparian zones (Venter et al., 2003; Main et al., 2016), and tree height ranging from 2-5m (Naidoo et al., 2015). In combination with the climatic

conditions, the vegetation distribution and density are influenced by the geology, fire regimes, and grazing and harvesting practices of woody vegetation (Main et al., 2016).

The study area experiences a hot and humid summer with temperatures averaging 30° C (Mograbi et al., 2015), and rainfall that primarily occurs between the months of October and May with an average of 250mm/annum in the east and 1000mm/annum in the west (Main et al., 2016). Winter is short and dry with an average temperature of 23° C (Mograbi et al., 2015), and no frost (Naidoo et al., 2016). The dry season occurs from late May to late September (Urbazaev et al., 2015). Figure 3 below shows the total monthly rainfall for BLM, Mpumalanga for the years 2017 to 2018. It shows the variation in rainfall between the months and shows that 2018 had a relatively lower total rainfall than 2017. This variation is evident on the results in section (Chapter 4) where different months responded differently due to variations on the amounts of rainfall received.

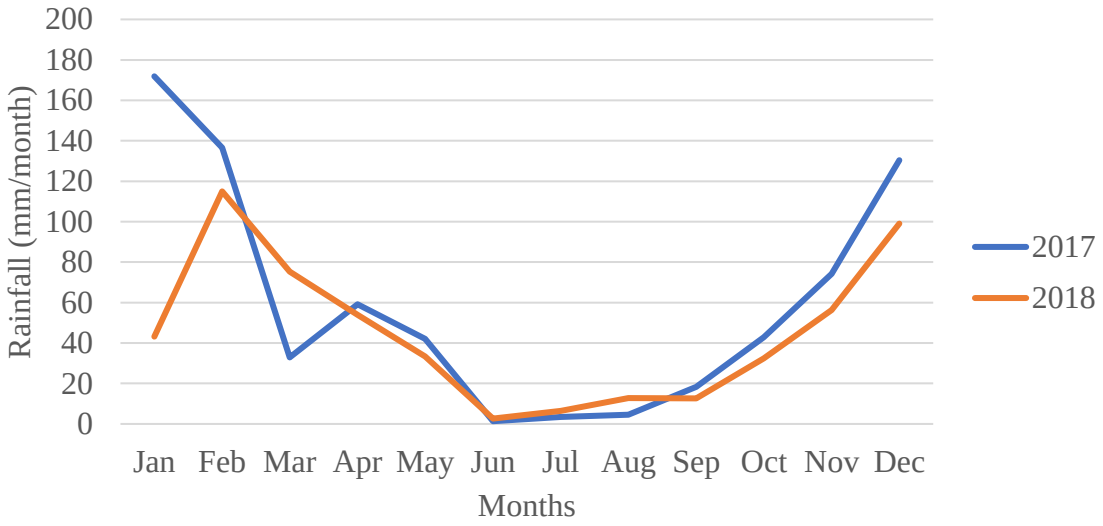


Figure 3: Total Monthly Rainfall for Bushbuckridge Local municipality, Mpumalanga 2017-2018. (Source: Global Precipitation Measurement (GPM))

3.2. Remote Sensing Data Acquisition and Pre-processing

3.2.1. LiDAR Data

This study used LiDAR-based products as reference data. The products were collected by the Council for Scientific and Industrial Research (CSIR) as part of a project titled CSIR TP45 "National woody vegetation monitoring system for ecosystem and value-added services". Southern Mapping undertook the topographical survey to produce rectified colour images and a Digital Terrain Model (DTM) of the surveyed project area. The topographical survey was done using an aircraft mounted LiDAR system that scanned the ground below at a high laser

frequency rate (150 kHz), resulting in a dense DTM of the ground surface and objects above the ground. Digital colour images were also taken from the aircraft and rectified to produce colour orthophotos of the surveyed area. The survey was flown at a height of approximately 700m and ortho images with a 7cm pixel resolution were produced.

The LiDAR products were acquired in June 2018 using the Patenavia P68B (ZS-OWLand ZS-BRD) aircraft with Optech ALTM M300(13SEN327) and Optech Gemini (09SEN258) laser scanner and a Phase One iXA180 camera. The LiDAR data covered the communal lands of Agincourt (A), Ireagh (I), Justicia (J) and Welverdient (W) in BLM, Mpumalanga (Figure 2) and are mainly used by residents for farming practices, harvesting of fuelwood, and ranching of livestock (Shackleton, 2000). The data had the following acquisition attributes (Table 1):

Table 1: Average density and spacing of LiDAR returns for the study areas in Bushbuckridge Local Municipality, Mpumalanga. (Source: CSIR)

	Avg Density (All Returns)	Avg Density (Last Return)	Avg Spacing (All Returns)	Avg Spacing (Last Returns)
Angincourt	5.1	4.6	0.4	0.5
Ireagh	9.2	8.2	0.3	0.4
Justicia	8.1	6.7	0.4	0.4
Welverdient	10.8	9.8	0.3	0.3

The pre-processing of LiDAR data is demonstrated by Naidoo et al. (2015, 2016), where they showed that to derive CC from LiDAR data, a LiDAR canopy height model (CHM) must be generated first. To do this, the digital elevation model (DEM) and top-of-canopy surface model (CSM) were produced first. The DEM was subtracted from the CSM to calculate the CHM (pixel size of 2m). To create the LiDAR woody CC, a data mask was applied to the CHM to generate a spatial arrangement of "no woody canopy cover" vs "woody canopy cover". In total, three woody structural metrics (CHM, CSM & CC) were generated from processing the LiDAR data and were all resampled to a 25m spatial resolution. For this study, the acquired LiDAR data was resampled to a 100m spatial resolution using the nearest neighbour resampling technique in R software. This was done to spatially average the dataset in order to increase the chances of obtaining higher accuracies.

3.2.2. Sentinel-1 Images

Sentinel-1 is a constellation of satellite missions launched by the European Space Agency as part of the European's Copernicus Program (Trunckenbrodt et al., 2019). It comprises of two satellites namely Sentinel-1A launched in 2014 and Sentinel-1B launched in 2016 (S1A and S1B; Haarpaintner and Hindberg, 2019). The satellites provide C-band SAR images of the whole world on a 6 to 12 days revisit cycle and are somewhat independent from adverse weather (cloud) and solar illumination conditions (Haarpaintner et al., 2020). They carry the C-band sensor at a 5.405 GHz frequency, with an incidence angle between 0° and 90° (Zhang et al., 2019). The Sentinel-1 products are freely and systematically available for all users for public to commercial and scientific uses. The products are distributed in the Sentinel-Standard Archive Format for Europe (SAFE) and can be found on https://sentinel.esa.int/web/sentinel/sentinel-data-access (Haarpaintner, 2019). Each scene of the Sentinel-1 products either has 1 or 2 of 4 polarisation bands: single band VV and HH, or dual-band VV+VH and HH + HV (Zhang et al., 2019).

Table 2 below shows a summary of the Sentinel-1 data acquisitions for the years 2017 and 2018 by specifying the satellite (S1A or S1B), the path number, number of orbits and the flight direction stating whether the satellite was ascending (ASC) or descending (DES). These Sentinel-1 products are from the Norwegian Research Centre (NORCE) and were prepared by Haarpaintner (2019). They cover a period spanning from 1 January 2017 to 31 December 2018. Individual scenes from these products were averaged for each month to produce mean monthly composites. The composites comprised of the co-polarised VV, cross-polarised VH and the Normalized Difference Index (NDI) SAR bands. For this study, the products were divided into two different seasons: the dry and wet seasons. The dry season covered the months from May to October while the wet season covered from November to April for both 2017 and 2018.

Table 2: Summary of the Sentinel-1 Data Acquisition. (Source: Haarpaintner, 2019)

Site	2017	2018

South Africa	Satellite	ASC or DES	Path	Nr of orbits	Satellite	ASC or DES	Path	Nr of orbits
	S1A	ASC	043	31	S1A	ASC	043	30
	S1A/S1B	DES	079	27A+6B	S1A/S1B	DES	079	13A+27B
	S1A	ASC	145	30	S1A	ASC	145	31
	S1B	DES	152	7	S1B	DES	152	26
Total	S1A/S1B	ASC/DES		101	S1A/S1B	ASC/DES		127

3.2.3. SAR Pre-processing

Norut's (now NORCE) Generic SAR (GSAR) processing system was used for pre-processing because it can process big data. The SAR data was geocoded, radiometrically calibrated, and radiometric slope correction was applied once all other processing procedures were implemented. A 1Arcs SRTM DEM interpolated on a 20m spatial resolution 36s UTM grid using a cubic convolution resampling was applied to pre-process the data using a geocoding software (Larsen et al., 2005). The 20m spatial resolution of the DEM was defined to ensure that all pre-processed single orbits are processed on the same UTM grid (Haarpaintner, 2019). Figure 4 below shows the processing done from Level-1 Sentinel-1 GRD data to pre-processed georeferenced and slope-corrected single orbit images (Level 2), and lastly, to the complete set of ARD products (Level 3).

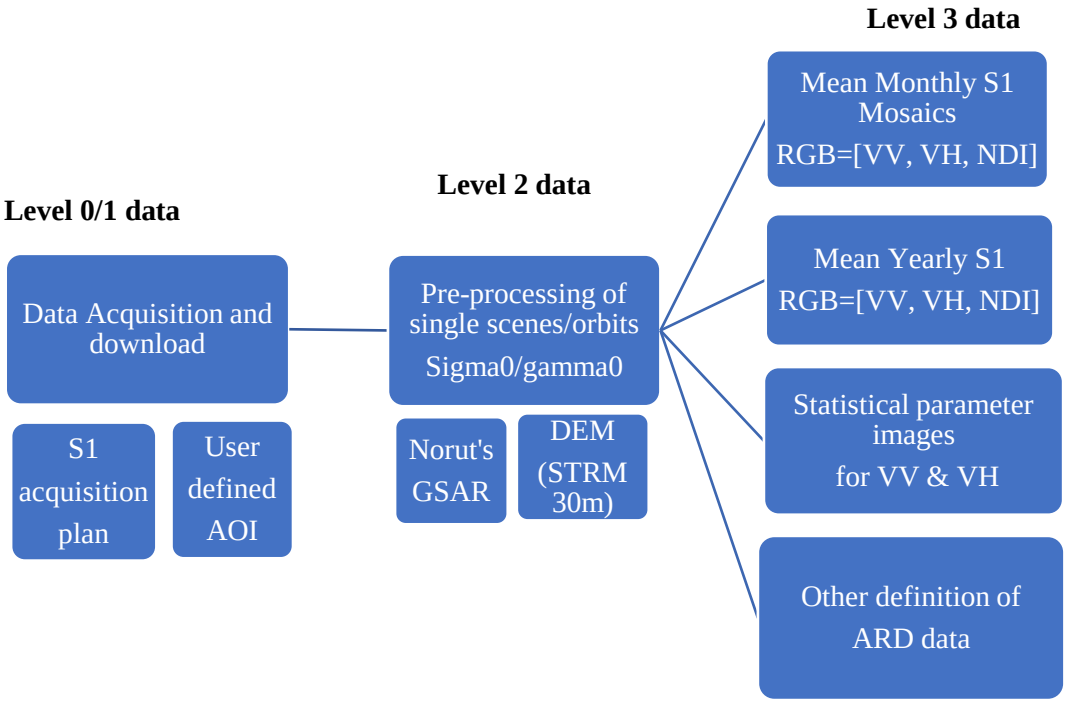


Figure 4: ARD processing chain from GRD (Level 1) to ARD products (Level 3; Source: Haarpaintner, 2019).

The nearest neighbour resampling technique was used to resample the ARD products to a 100m spatial resolution. This was done to ensure that the SAR and LiDAR data products have the same resolution and align with each other so values would be extracted within the extent of the defined study areas. The 'rasterToPoint' function in R was used to convert the LiDAR data to sampling points. The points were used to randomly extract SAR data using the 'raster extract' function in R. This was done to predict canopy cover models using LR, SVM and RF algorithms. Literature has established that the dry season is the best period to acquire SAR data because SAR sensors are negatively affected by moisture/rainfall during the wet season (Archibald and Scholes, 2007; Urbazaev et al., 2015; Naidoo et al., 2016). Therefore, Merged satellite-gauge precipitation Final Run (recommended for general use) GPM_3IMERGM_v06 data was acquired from the Giovanni website accessed through https://giovanni.gsfc.nasa.gov. The data was monthly rainfall in mm/month units with a Spatial Resolution of 0.1° (11.1km) for the BLM region for 2017-2018. The acquired data was exported to a CSV format to create

a monthly rainfall graph (shown in Figure 3) in excel to help understand the results of the SAR predictive models.

3.3. Calibration and Validation Dataset

The dataset (n=26269) was randomly split into 70% and 30% for calibration and validation. 18388 (70%) points were used for calibration while 7881 (30%) were used for validation. The calibration dataset was used to train the LR, SVM and RF predictive models while the validation was used to test the validity and how well the predictive models performed.

3.4. Statistical Analysis

This study adopted a regression analysis approach to assess the utility of multi-temporal 'Analysis Ready' C-band SAR data for woody canopy cover estimation. LR, SVM and RF were used in open-source R software (R Core Team, 2018) to model the statistical relationships between the LiDAR CC dependent variable and C-band ARD independent variable.

3.4.1. Linear Regression

The "lm" function was used to fit the LR (Sugiura, 1978) model in R software (R Core Development Team, 2018). LR is the simplest and most used regression model. It is easy to calculate and interpret but not suitable for data not linearly distributed (Naidoo et al., 2014). LR assumes that there is a linear relationship between the dependent variable and the independent variable (Pietersen and Maree, 2007).

The LR model is expressed by the equation:

$$Y = a + bx$$

Where Y is the dependent variables, a is the intercept, b is the regression coefficient and x is the independent variable.

3.4.2. Support Vector Machine

SVM (Vapnik, 1995) implemented in the "e1071" package (Meyer et al., 2019) was carried out in R statistical software (R Core Development, 2018) to model canopy cover. SVM was originally used for classification but was later extended to include regression. As a result, the e1071 package can be used for both classification and regression studies (Meyer et al., 2019). The aim for SVM regression is to estimate a function $f(x)$ that is very close to the target values and at the same time as flat as possible to allow for generalizations (Shevade, 2000).

A linear function in the feature space is used to express the function f represented by the equation:

$$f(x) = \omega \cdot \emptyset(x) + b$$

Where x is the input vector of the SVM, ω denotes the weight vector,"·" represents the inner product in the feature space, $\emptyset(x)$ is a kernel to transfer the data to a vector feature space and b denotes the bias vector.

SVM relies on the use of kernel functions to plot input data into a new hyperplane where separations are performed and aims to reach an optimal hyperspace for fitting and predicting data using the ε-insensitive loss function (Forkuor et al., 2017). Vapnik (1995) suggested that the ε-insensitive loss function be used if the error produced is less than ε (Shevade (2000). As such, the "caret" package (Kuhn, 2020) in R statistical software was carried out to tune the SVM model to determine the best parameters by ignoring errors less than ε.

3.4.3. Random Forest

RF is a robust and computationally efficient model capable of denoting the difficulty of relationships that may not be linearly distributed and is hardly affected by issues of overfitting (Naidoo et al., 2014) even though overfitting could occur when the dataset being modelled is noisy (Statnikov et al., 2008). To avoid overfitting, RF grows trees by randomly selecting predictors at each node with random subset of samples for each tree (Breiman, 2001). RF (Breiman, 2001) was applied using the "randomForest" package (Liaw and Wienner, 2002) in R statistical software (R Core Development Team, 2018) as one of the algorithms for modelling canopy cover. To model RF, two main user defined parameters are required to tune the model (Ismail et al., 2010). These parameters are the number of trees built in the forest (ntree) and the number of possible splitting variables for each node (mtry). The default 500 trees were used for ntree and the default for mtry which is the square root of the number of predictor variables used was 1 for the individual scenes and 2 for the multi-temporal scenes. These are the parameters that affect the modelling ability of RF which were set using the "caret" (Kuhn, 2020) package in R statistical software (R Core Development, 2018).

During the modelling process, RF combines bootstrapping and bagging in order to grow trees. To construct the trees, 2/3 of its samples are selected randomly for the training dataset and a matching DT is created based on the random samples (Breiman, 2001). The 1/3 samples that were not selected from the Out-of-Bag (OOB) test dataset. The OOB is then used as a sample

for determining errors when processing the RF model (Zhang et al., 2017). For the ntree and mtry parameters to be at their most optimal, the value of the OOB error needs to be at its lowest and can thus be able to calculate the mean square error (MSE_{OOB}; Liaw and Wienar, 2002)

The mean square error MSE_{OOB} is expressed by the equation:

$$MSE_{OOB} = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_{iOOB})^2$$

Where n is the number of observations, O_i is the average prediction of the ith observation, P_{iOOB} is the average of all OOB predictions from all trees, MSE_{OOB} is normalized as it depends on the unit of the response variable and the percentage of explained variance

The explained variance is expressed by the equation:

$$Var = 1 - \frac{MSE_{OOB}}{Var_{resp}}$$

3.4.4. Accuracy Assessment and Modelling

The accuracy of the models was measured using regression statistics such as the Coefficient of Determination (R^2), Root-mean-Square Error (RMSE), relative RMSE (rRMSE). R^2 assesses the strength of the relationship between the SAR and LiDAR canopy cover while RMSE and rRMSE assist in evaluating the predictive performance and error in the models (Robison, 2005). The co-polarised VV bands is not as sensitive to volumetric scattering as the VH bands (Watanabe et al., 2016), making it not as beneficial on its own for modelling savanna woody structure. Cross-polarised VH band is sensitive to volumetric scattering and has been documented as the more effective and superior band for modelling savanna woody structure (Haarpaintner et al., 2020). As a result, the analysis of this study was simplified by only presenting VH results and where VV was combined with VH.

To model canopy cover, the datasets were analysed using both individual scenes and multi-temporal combinations of scenes, which created several modelling scenarios: individual scenes, per polarisation, individual scenes with both polarisations combined, seasonal combinations of scenes by polarisation and year, multi-year combinations of scenes per season and polarisation, and lastly, multi-year combinations of scenes where all seasons and polarisations were analysed at the same time. The modelling process of the datasets was repeated ten times to ensure robustness and cross-validation to allow different combinations of

training and validation datasets to determine the regression statistics of R^2, RMSE and rRMSE (Naidoo et al., 2014). Predicted vs observed scatterplots as well as canopy cover maps were produced to illustrate and show the distribution of the predictive models. It should be noted that spatial autocorrelation was not considered in this study. As such, the results presented in chapter 4 might be affected by it.

CHAPTER 4: Results

4.1. Optimal Season and Most Effective Regression Method

4.1.1. Individual Scenes, per polarisation

Validation results for the VH polarised individual scenes for 2017 are shown in Table 3, where only the VH polarisation band was analysed per image using SVM, RF and LR. The results show that the VH polarisation band is most effective when the environment is drier during the late dry season (i.e., September) and early wet season (i.e., November). The trends show that the transition into the wet season and the first few wet season months (i.e., October-November), show higher accuracy than the late wet season months (i.e., March-April). Furthermore, the results also show higher accuracy later in the dry season (i.e., September) than in the early dry season (May). SVM shows the highest accuracy, followed by LR and RF having the lowest accuracy. SVM and RF have their highest accuracies in November while LR is in October. This shows that regardless of which model is used, the drier the environment, the better the modelling accuracies.

Table 3: Validation results from the scenario whereby 2017 VH polarised scenes were analysed.

		Random Forest			***SVM***			***Linear Regression***		
	Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
Wet	***January***	0.31	16.56	63.43	0.38	15.8	60.1	0.35	15.97	60.96
	February	0.29	16.77	63.9	0.35	16.09	61.41	0.33	16.23	61.94
	March	**0.23**	**17.65**	**67.1**	**0.28**	**17.03**	**64.99**	**0.27**	**17**	**64.89**
	April	0.3	16.64	64.15	0.37	15.85	60.6	0.35	15.94	60.83
Dry	***May***	0.34	16.25	61.64	0.4	15.45	58.97	0.39	15.55	59.33
	June	0.44	14.94	57.71	0.49	14.29	54.54	0.46	14.54	55.47
	July	0.47	14.4	55.45	0.53	13.66	52.14	0.49	14.17	54.09

	August	0.49	14.1	53.8	0.54	13.48	51.43	0.51	13.87	52.93
	September	0.51	14.01	53.7	0.56	13.33	50.86	0.52	13.71	52.31
	October	0.52	13.85	53.02	0.57	13.15	50.18	**0.53**	**13.65**	**52.11**
Wet	***November***	**0.52**	**13.8**	**52.66**	**0.57**	**13.07**	**49.8**	0.51	13.92	53.13
	December	0.44	14.94	57.56	0.49	14.23	54.29	0.44	14.81	56.51
	Mean Wet	0.35	16.06	61.47	0.41	15.35	58.53	0.38	15.65	59.71
	Mean Dry	0.46	14.5	55.89	0.52	13.88	53.02	0.48	14.25	54.37

** Dry season results on the tables are shaded using a grey colour while the best results are highlighted using red and the poorest results on individual scenes per polarisation and all polarisations combined are highlighted using a blue colour (here and hereafter).*

Validation results for the VH polarised individual scenes for 2018 are shown below in Table 4, where only the VH polarisation band was analysed per image using SVM, RF and LR. The results (Table 4) show similar trends as the 2017 validation results shown in Table 3 where the transition into the wet season and the first few wet season months (i.e., October-November), show higher accuracies than the late wet season months (i.e., March-April). However, unlike the 2017 validation results (Table 3) where the highest accuracies were recorded in November for SVM and RF, and October for LR, the 2018 validation results (Table 4) show that the highest accuracies for all models (i.e., SVM, RF and LR) were recorded in August. Although these results also show that the drier the environment becomes, the more the modelling accuracies improve, they also show that the highest accuracy in 2018 (Table 4) was in the mid-dry season rather than the early wet season. Furthermore, when comparing the 2018 validation results (Table 4) with the 2017 validation results (Table 3), they show higher accuracies were recorded in 2018 rather than 2017. This suggests that the year 2018 was drier than the year 2017 (Figure 3).

Table 4: Validation results from the scenario whereby 2018 VH polarised scenes were analysed

		Random Forest			***SVM***			***Linear Regression***		
	Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
Wet	***January***	0.43	14.81	57.16	0.5	14.14	53.96	0.42	15.11	57.65
	February	0.35	16.05	61.5	0.41	15.33	58.51	**0.35**	**16.02**	**61.16**
	March	0.42	15.2	57.84	0.48	14.47	55.21	0.43	14.98	57.15
	April	**0.34**	**16.27**	**62.03**	**0.41**	**15.38**	**58.69**	0.38	15.68	59.84

Dry	***May***	0.44	14.92	57.27	0.48	14.35	54.96	0.45	14.72	54.19
	June	0.46	14.64	55.96	0.52	13.74	52.43	0.49	14.22	54.26
	July	0.5	14.11	54.04	0.56	13.27	50.65	0.51	13.82	52.75
	August	**0.53**	**13.69**	**52.07**	**0.58**	**13**	**49.56**	**0.54**	**13.51**	**51.57**
	September	0.49	14.08	54.36	0.55	13.34	50.91	0.51	13.86	52.91
	October	0.51	13.92	53.19	0.56	13.23	50.49	0.52	13.77	52.54
Wet	***November***	0.51	13.95	53.29	0.56	13.25	50.56	0.52	13.76	52.51
	December	0.47	14.44	55.34	0.51	13.91	53.1	0.47	13.76	54.87
	Mean Wet	0.42	19.58	57.86	0.48	14.08	55.01	0.43	14.89	57.2
	Mean Dry	0.49	14.23	54.48	0.54	13.49	51.5	0.5	13.9	53.04

The results for the individual scenes, per polarisation presented in Table 3 and 4 show that the season and model which produced the highest accuracies was the same in both 2017 and 2018. Amongst all the individual scenes, per polarisation datasets, August 2018 (dry season; Table 4) produced the highest model accuracy while March 2017 (wet season; Table 3) produced the lowest results. Overall, the late dry and early wet seasons (i.e., Sept-Nov) produced the highest results in all models while the late wet seasons (i.e., Feb- Apr) produced the lowest results with their rRMSE ranging between 58.69% and 67.1%. The mean results for all seasons in both 2017 (Table 3) and 2018 (Table 4) show that on average 2018 model accuracies were higher than those of 2017 in all models and seasons.

4.1.2. Individual Scenes with both Polarisations combined

Validation results for the scenario where both VV and VH polarisation bands were combined and analysed together for 2017 are shown below in Table 5. The validation results in Table 5 show that the trends are comparable to those of the single polarisation band analysis above in Tables 3 and 4 where SVM shows the highest accuracy followed by LR and RF showing the lowest accuracy. Compared to the 2017 and 2018 individual scenes validation results shown in Tables 3 and 4, with modelling accuracies for both polarisations (VV and VH; Table 5) improve the drier the environment. The transition into the wet season and the first few wet season months (i.e., October-November), show higher accuracies compared to the late wet season months (i.e., March-April), with November showing the highest accuracies for SVM and RF and October showing the highest accuracies for LR.

Table 5: Validation results from the scenario whereby both polarisations (VV & VH) for 2017 were analysed

		Random Forest			***SVM***			***Linear Regression***		
	Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
Wet	***January***	0.32	16.4	63	0.38	15.9	60.9	0.36	15.9	60.5
	February	0.3	16.7	63.4	0.35	16.1	61.6	0.33	16.3	62.2
	March	**0.2**	**17.8**	**68.6**	**0.27**	**17**	**65.2**	**0.26**	**17**	**65.2**
	April	0.3	16.7	63.8	0.37	15.9	60.4	0.35	15.9	61
Dry	***May***	0.35	16.2	61.7	0.4	15.4	58.8	0.39	15.5	59.5
	June	0.44	15.1	57.8	0.49	14.3	54.6	0.46	14.5	55.6
	July	0.48	14.4	55.7	0.53	13.7	52.3	0.49	14.2	54.3
	August	0.48	14.4	55.2	0.54	13.5	51.6	0.51	13.9	52.8
	September	0.51	14	53.4	0.56	13.3	50.8	0.52	13.7	52.3
	October	0.52	13.9	53.2	0.57	13.2	50.2	**0.53**	**13.7**	**52**
Wet	***November***	**0.52**	**13.7**	**52.7**	**0.57**	**13.1**	**50**	0.51	13.9	52.9
	December	0.43	14.9	56.8	0.49	14.2	54.2	0.44	14.8	56.6
	Mean Wet	0.35	16	61.4	0.41	15.4	58.7	0.38	15.6	59.7
	Mean Dry	0.46	14.6	56.2	0.52	13.9	53.1	0.48	14.3	54.4

Validation results where both VV and VH polarisation bands were analysed together per image for the year 2018 are shown below in Table 6. Compared to the previous individual scenes modelling scenarios shown in Table 3, 4 and 5, results presented in Table 6 also show increasing accuracies with the transition into the wet season. The first few wet season months (i.e., October-November) show higher accuracies compared to the late wet season months (i.e., March-April). SVM outperformed both LR and RF with the highest accuracy for all being in August 2018 mid-dry season. These results show similar trends and accuracies with those of Tables 3, 4 and 5 and suggest that there are marginal gains when the VH polarisation band is analysed individually or when VV and VH are combined and analysed together when dealing with a single scene.

Table 6: Validation results from the scenario whereby both polarisations (VV & VH) 2018 polarised scenes were analysed

		Random Forest			***SVM***			***Linear Regression***		
	Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
Wet	***January***	0.44	14.9	57.5	0.5	14.1	54	0.42	15.1	57.8
	February	0.36	15.8	60.7	0.41	15.3	58.5	**0.35**	**16**	**61.3**
	March	0.42	15.1	58.2	0.48	14.4	55.1	0.43	14.9	57.2
	April	**0.35**	**15.9**	**61.3**	**0.41**	**15.4**	**58.7**	0.38	15.7	60
Dry	***May***	0.43	15.1	57.2	0.48	14.4	54.9	0.45	14.7	56.4
	June	0.47	14.7	55.9	0.53	13.7	52.4	0.49	14.2	54.3
	July	0.5	14	53.6	0.56	13.2	50.7	0.52	13.8	52.8
	August	**0.54**	**13.5**	**51.2**	**0.58**	**13**	**49.6**	**0.54**	**13.5**	**51.7**
	September	0.5	14.1	54.7	0.55	13.3	50.9	0.51	13.8	52.7
	October	0.51	13.8	53.2	0.56	13.2	50.6	0.52	13.8	52.4
Wet	***November***	0.5	14.1	53.5	0.56	13.3	50.7	0.52	13.8	52.4
	December	0.48	14.4	54.9	0.52	13.9	53.1	0.48	14.4	55
	Mean Wet	0.43	15	57.7	0.48	14.4	55	0.43	14.9	57.3
	Mean Dry	0.49	14.2	54.3	0.54	13.5	51.5	0.51	13.9	53.4

The individual scenes, both polarisations (VH & VH) combined (Table 5 and 6), show similar trends as the single polarisation individual scenes (Table 3 and 4) where the season and model with the highest accuracy was the same in both 2017 and 2018. Compared to the VH only validation results in Table 4, August 2018 (dry season) produced the highest model accuracies while March 2017 produced the lowest results. Overall, the dry season outperformed the wet season regardless of the model while SVM returned the highest accuracies and RF produced the lowest. The mean dry and wet seasons results in both Tables 5 and 6 show that the average performance for the dry season is higher than that of the wet season with ~4% rRMSE difference in 2018 and ~5% rRMSE difference in 2017.

4.1.3. Seasonal Combinations of Scenes by Polarisation and Year

Although some of the results are comparable to those of the individual scenes for 2017 and 2018 in the tables above (Table 3 to Table 6), combining the individual scenes into seasonal and/or yearly composites resulted in improved model accuracies with the rRMSE improving by ~2.5% for 2017 and by ~2% for 2018 when compared to the highest individual scene within each year of the previous scenarios (Table 3 to Table 6). SVM remains the highest model by showing high accuracies in all seasons. The performance of RF improved compared to the individual scenes while the performance of LR showed insensitivity for modelling large datasets.

Validation results for seasonal single-polarisation band analysis where individual VH scenes were combined into single season or yearly composites for both 2017 and 2018 are shown in Table 7. The 2017 VH dry & wet seasons combination produced the highest results in all models compared to the 2018 VH dry & wet seasons combination. The validation results for 2018 wet season returned higher accuracies compared to the validation results for 2017 wet season in all models. In the dry season, RF and SVM produced their highest accuracies in the 2017 dry season while LR produced its highest accuracy in the 2018 dry season. Overall, the 2017 SVM dry & wet season combination showed the highest accuracies with an rRMSE of 47.53% while the 2017 LR wet season showed the lowest results with an rRMSE of 52.35%.

Table 7: Validation results from the scenario whereby VH polarised scenes from both 2017 and 2018 were analysed seasonally. Here and hereafter, the red colours in the tables represents the best results for the period being analysed.

	Random Forest			***SVM***			***Linear Regression***		
Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
2017 VH Dry Season	**0.58**	**12.71**	**48.63**	**0.59**	**12.71**	**48.65**	0.54	13.44	51.42
2018 VH Dry Season	0.57	12.96	49.6	0.59	12.74	48.75	**0.54**	**13.38**	**51.2**
2017 VH Wet Season	0.58	12.81	49.17	0.59	12.72	48.65	0.52	13.68	52.35
2018 VH Wet Season	**0.59**	**12.78**	**49.6**	**0.6**	**12.6**	**48.21**	**0.53**	**13.55**	**51.87**

2017 VH Dry & Wet	**0.6**	**12.41**	**47.63**	**0.61**	**12.42**	**47.53**	**0.55**	**13.33**	**50.99**
2018 VH Dry & Wet	0.6	12.52	47.77	0.61	12.44	47.58	0.55	13.33	51

4.1.4. Seasonal Combinations of scenes of all Polarisations by Year

Validation results for seasonal combinations of both VV and VH polarisations for both 2017 and 2018 are shown below in Table 8. SVM generally shows high accuracies compared to the RF and LR validation results in Table 8. The validation results for 2017 dry season (Table 8) show that RF has higher accuracies than both the SVM and LR. The validation results for the 2018 multi-seasonal combination produced the highest results for RF and LR while the 2017 multi-seasonal combination yielded the highest result for SVM. Compare to the validation results in Table 7 (where individual VH scenes were combined into single season or yearly composites), the validation results for the 2018 wet season in Table 8 yielded the highest accuracies for all models compared to the 2017 wet season. In the dry season, RF and SVM produced their highest accuracies in the 2017 dry season and LR produced its highest accuracies in the 2018 dry season.

Table 8: Validation results from the scenario whereby all polarised scenes from both 2017 and 2018 were analysed seasonally.

	Random Forest			***SVM***			***Linear Regression***		
Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
2017 Dry Season	**0.59**	**12.69**	**48.42**	**0.59**	**12.71**	**48.65**	0.54	13.44	51.42
2018 Dry Season	0.57	12.97	49.49	0.59	12.74	48.75	**0.54**	**13.38**	**51.2**
2017 Wet Season	0.57	12.98	49.13	0.59	12.72	48.65	0.52	13.68	52.35
2018 Wet Season	**0.6**	**12.72**	**48.13**	**0.6**	**12.6**	**48.21**	**0.53**	**13.56**	**51.87**
2017 Dry & Wet	0.6	12.7	48.89	**0.61**	**12.42**	**47.53**	0.55	13.38	50.99
2018 Dry & Wet	**0.6**	**12.3**	**47.78**	0.61	12.43	47.58	**0.55**	**13.33**	**51**

RF results improved the more images were added to its analysis compared to the validation results of the individual scenes shown in Tables 3 and 4. The highest RF accuracy from seasonal combinations of scenes of all polarisations improved by ~3.4% rRMSE when compared to RF validation results of the individual scene (August 2018) shown in Table 5. The RF seasonal combination results presented in Table 7 are ~0.15% (rRMSE) lower than those presented in Table 8. This suggests that there was a marginal increase in the results when VV and VH were analysed together. Although the models of SVM and LR show similar validation results in both Tables 7 and 8, SVM produced the highest accuracies with the 2017 dry and wet season combinations being the overall highest accuracies by an rRMSE of 47.53%.

4.1.5. Multi-year Combination of Scenes per Season and Polarisation

Validation results for seasonal and single-polarisation analysis where images of 2017 and 2018 were combined and analysed using only the VH polarisation band according to season are shown in Table 9. The validation results show that all models presented in the table below (Table 9) produced their highest accuracies during the dry and wet season combination for both 2017 and 2018. Amongst all the models, SVM produced the highest results with an rRMSE of 46.83%, followed by RF with an rRMSE of 47.62% and the lowest being LR with an rRMSE of 50.31%. RF improved (compared to the individual scenes from Table 3 to Table 6) with more images added to the analysis. Its highest results were comparable to SVM's highest results where the rRMSE of RF is only 0.8% lower than that of SVM (Table 9).

Table 9: Validation results from the scenario whereby VH polarised scenes from both 2017 and 2018 were combined

	Random Forest			***SVM***			***Linear Regression***		
Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
Combined VH Dry Season	0.59	12.86	48.77	0.61	12.52	47.91	0.55	13.24	50.66
Combined VH Wet Season	0.6	12.54	48.15	0.61	12.41	47.51	0.55	13.32	50.94
Combined VH Multi-Seasonal	**0.6**	**12.55**	**47.62**	**0.62**	**12.24**	**46.83**	**0.56**	**13.15**	**50.31**

4.1.6. Multi-Year combinations of scenes for all polarisation bands by season or both

Validation results for all polarisation bands multi-year combination analysis are shown below in Table 10. The polarisation bands and images for different seasons and years were analysed together at the same time. The combined multi-seasonal showed the highest accuracies in all models with SVM showing the highest accuracies followed by RF and LR showing the lowest results. Overall, SVM showed higher accuracies than RF and LR. LR showed higher accuracies than RF on individual scene analysis (Tables 3 and 4). However, the performance of RF improved the more images were added to the analysis (Table 5 to Table 10.). The RF (and SVM) models were able to better find meaningful relationships between the backscatter and the woody canopy cover when more data is added to the model. LR was not sensitive, or robust, enough to find good relationships between the variables.

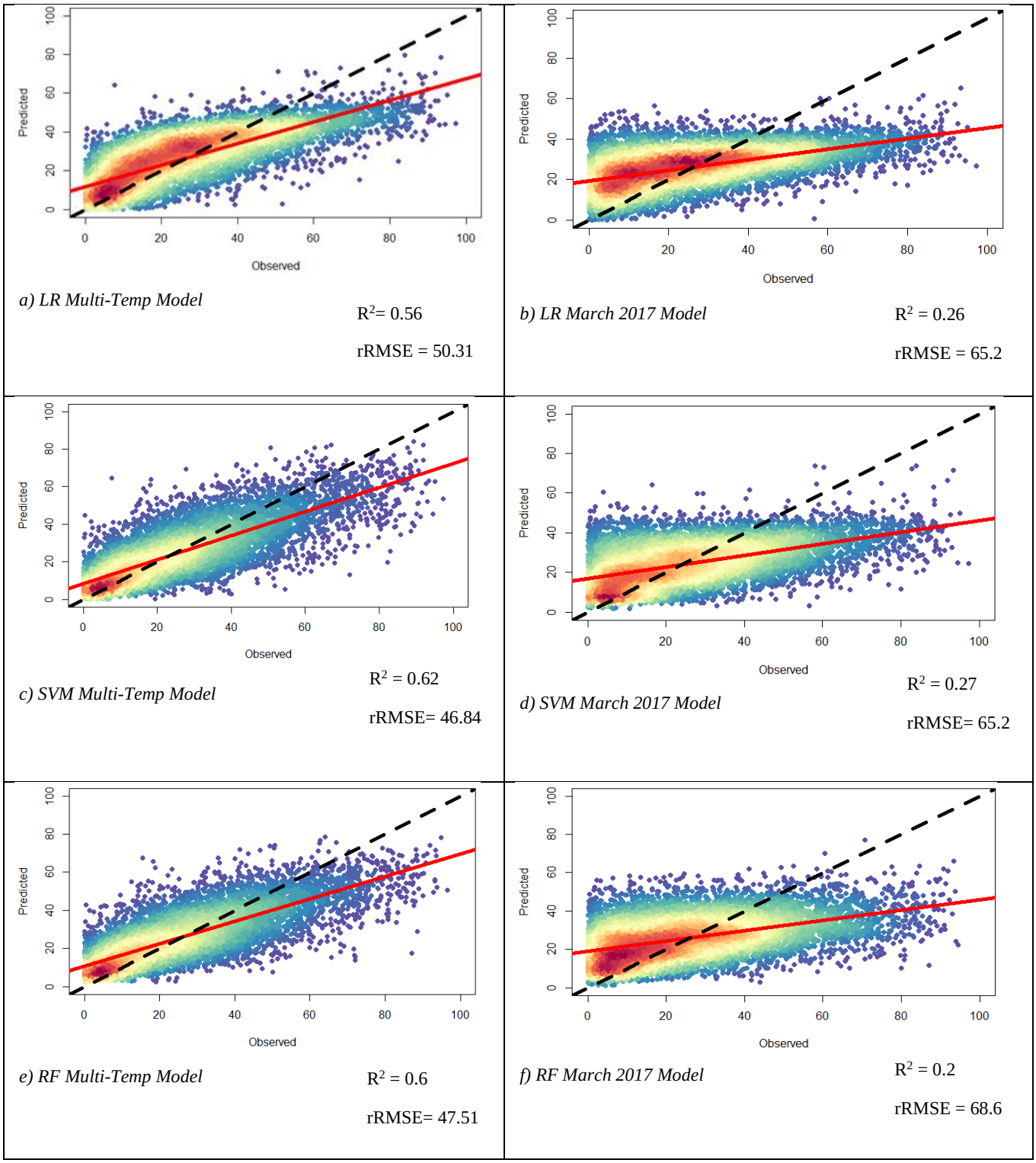
Table 10: Validation results from the scenario whereby all polarised scenes from both 2017 and 2018 were combined

	Random Forest			***SVM***			***Linear Regression***		
Dataset	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*	*Rsq*	*RMSE (%)*	*rRMSE (%)*
Combined Dry Season	0.6	12.67	48.62	0.61	12.52	47.91	0.55	13.24	50.66
Combined Wet Season	0.59	12.57	48.51	0.61	12.42	47.51	0.55	13.15	50.94
Combined Multi-Seasonal	**0.61**	**12.34**	**47.42**	**0.62**	**12.24**	**46.84**	**0.56**	**13.15**	**50.31**

4.1.7. Observed vs predicted canopy cover scatter plots of LR, SVM and RF

The observed vs predicted scatter plots are shown below in Figure 5a-f for LR, SVM and RF individual scenes, all polarisations for the March 2017 (Table 5) scenario and the multi-year combinations of scenes for all polarisation bands and all seasons combined. The March 2017 (Table 5) scenario showed the lowest model accuracies compared to all models. The multi-year combinations of scenes for all polarisation bands and all seasons combined (combined multi-seasonal; Table 10) showed the highest model accuracies compared to all models. The observed vs predicted scatter plots (Figure 5a-f) support the findings from Table 3 to Table 10. The deviation of the regression line from the 1:1 line (Figure 5a-f) suggests that all models generally

show the same trends by being overestimated but reduced when it comes to the multi-year RF and SVM which show underestimation at the top end of the range (Figure 5c, e).



a) LR Multi-Temp Model R^2= 0.56 rRMSE = 50.31

b) LR March 2017 Model R^2 = 0.26 rRMSE = 65.2

c) SVM Multi-Temp Model R^2 = 0.62 rRMSE= 46.84

d) SVM March 2017 Model R^2 = 0.27 rRMSE= 65.2

e) RF Multi-Temp Model R^2 = 0.6 rRMSE= 47.51

f) RF March 2017 Model R^2 = 0.2 rRMSE = 68.6

Figure 5 (a-f): Observed vs predicted canopy cover scatter plots of LR, SVM and RF individual scenes, all polarisation March 2017 scenario (b, d, f), and LR, SVM and RF multi-year combinations of scenes for all polarisation bands and all seasons combined (a, c, e). The black dashed line represents the 1:1 line, while the red solid line is the regression line.

4.2. Mapping Canopy Cover

4.2.1. Canopy Cover Distribution

Based on the results shown in Section 4.1: optimal season and most effective regression method, the multi-year combinations of all polarisation bands and season scenarios of RF and SVM were used to map canopy cover. The maps are shown in Figures 6 and 7. They show the spatial distribution of canopy cover in BLM. Figure 6 is an RF canopy cover map while Figure 7 is an SVM canopy cover map. When assessing the performance or accuracies of the SAR models presented in Section 4.1. (Chapter 4), SVM outperformed RF and LR. However, RF showed the most improvement between all the models thus both models were chosen as the effective methods and were used to map canopy cover.

The maps shown in Figures 6 and 7 range from brown to green with green showing high canopy cover and brown showing little to no canopy cover. The West region is dominated by canopy cover, while the Eastern region depicts little to no vegetation cover in both maps. This may be due to a rainfall gradient present in the area where the amount of rainfall received decreases from the West to the East. Furthermore, the canopy cover of RF is shown to be higher than that of SVM. RF canopy cover is at a maximum of 86% while that of SVM has a maximum of 77%. This shows that although SVM shows higher results (Table 3 to Table 10), it is predicting for a smaller range of canopy cover than the RF model.

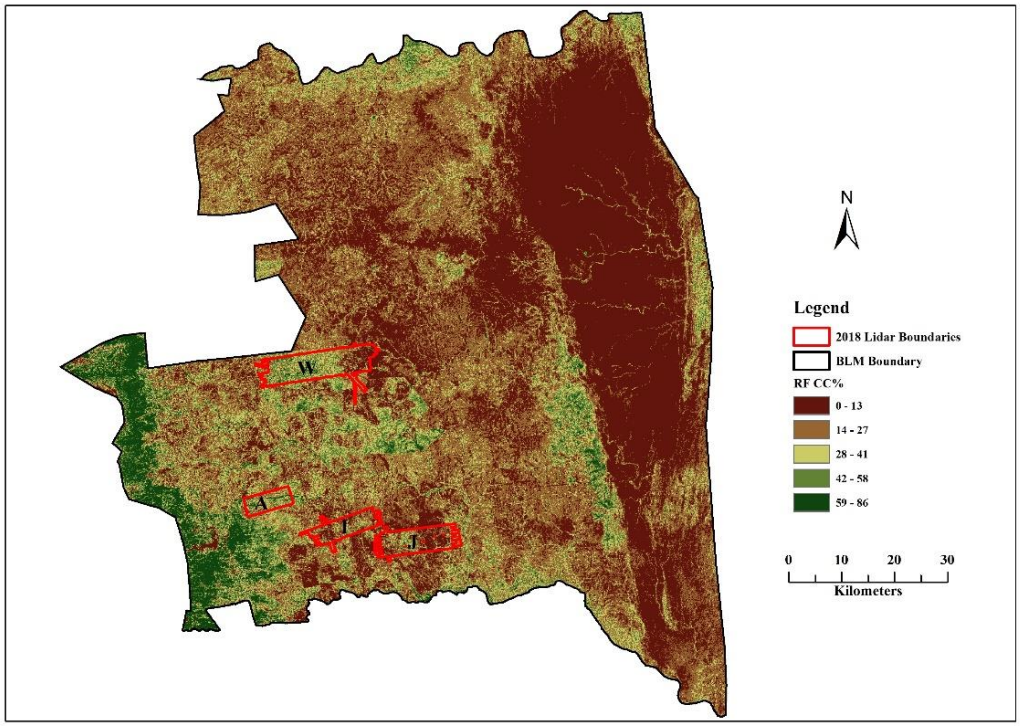


Figure 6: Multi-Temporal C-band SAR predictive canopy cover (CC) map using Random Forest (RF). Letters A, I, J & W represent areas covered by the LiDAR CC.

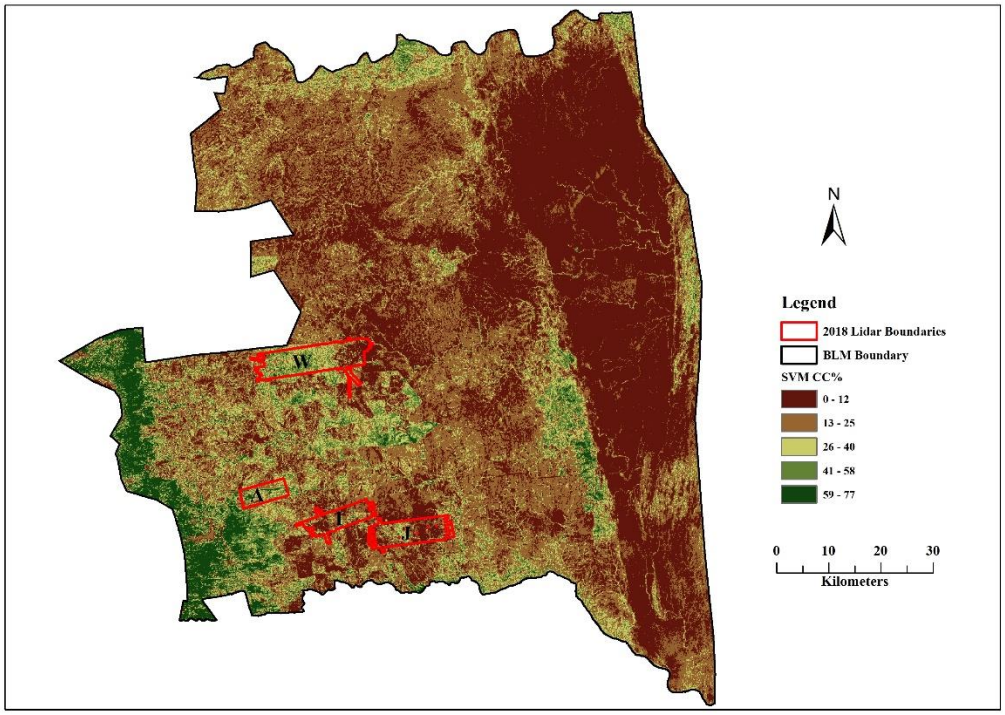


Figure 7: Multi-Temporal C-band SAR predictive canopy cover (CC) map using Support Vector Machine (SVM). Letters A, I, J & W represent areas covered by the LiDAR CC.

A zoomed-in area for RF and SVM to determine which of the two models produced a more accurate predictive map is shown in Figure 8. The area is in the West where the highest percentage of canopy cover in BLM is found. However, the area chosen is without canopy cover as it is a dam. In RF, this dam has clearly defined boundaries falling under the 0-13% category of canopy cover and ranges from a dark brown to a light brown colour showing that there is little to no canopy cover in the dam. Whereas on the SVM, the dam ranges from a light brown to a dark brown colour and falls under the 13-25% category of canopy cover which shows a much higher presence of canopy cover. The canopy cover categories and clear boundary definition of the dam show that RF produced a more accurate predictive map than SVM as the categories or percentage of canopy cover increase further away from the dam.

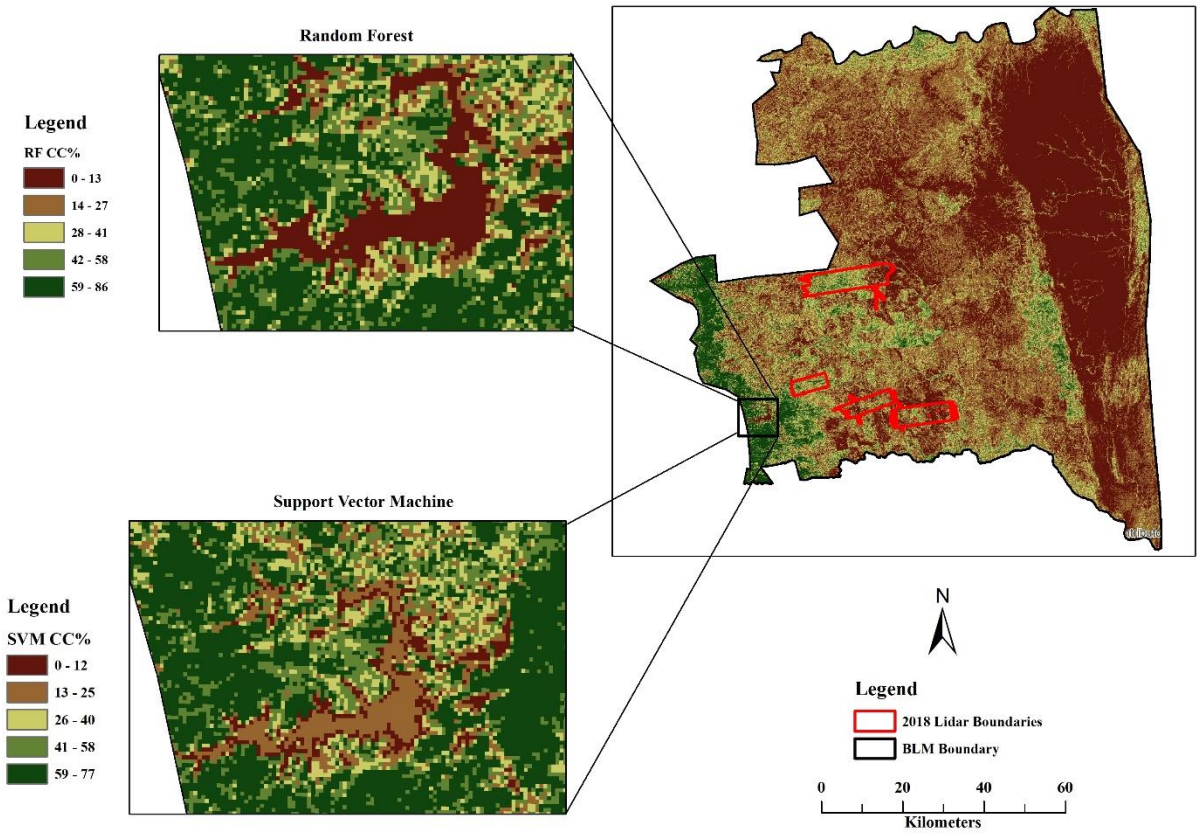


Figure 8: Observation of the representation of Random Forest vs SVM maps

4.2.2. LiDAR CC Frequency Plot

The distribution of the predicted LiDAR canopy cover is shown below in Figure 9. This distribution plot supports the canopy cover findings of Figure 6 to Figure 8 where canopy cover is mapped using SVM and RF. The plot is skewed towards lower values of canopy cover and shows that although the canopy cover predictive values are positive, the SVM (blue) model has a lower prediction range than the RF (red) model.

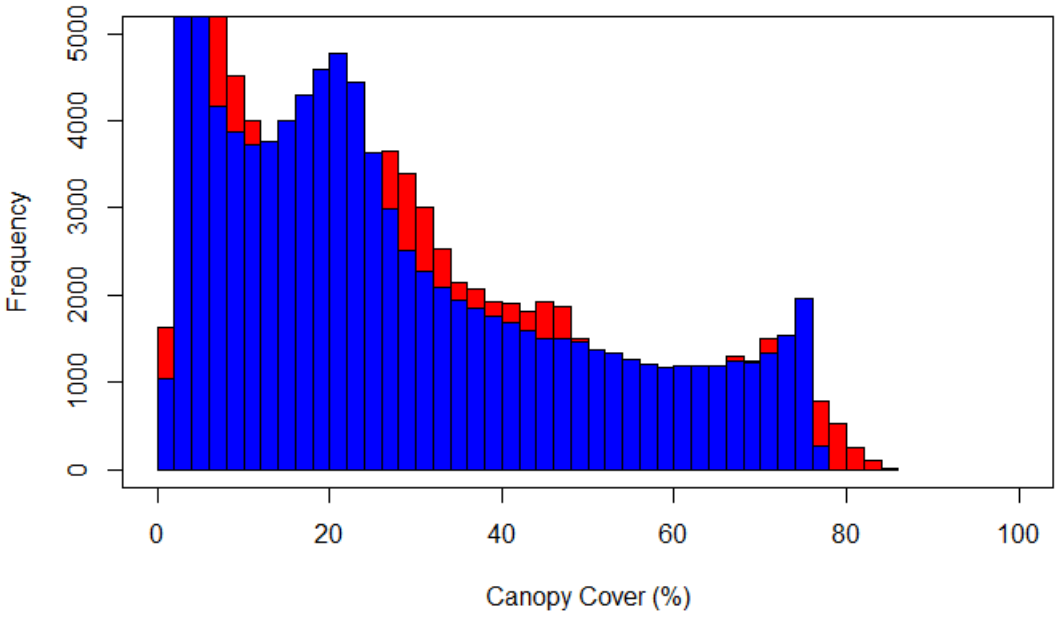


Figure 9: Frequency plot showing the distribution of the RF (red) vs SVM (blue) predicted canopy cover

4.3. Summary of Results

The analysis showed that although SVM showed higher accuracies regardless of the season, noise or number of images that were analysed, it predicted for a smaller range while RF predicted for a larger range and improved its performance the more data or noise was added to the analysis. LR was not sensitive, or robust, enough to find good relationships between the variables as its performance dropped (compared to RF) the more images were added to the analysis.

CHAPTER 5: Discussion and Conclusion

5.1. Discussion

5.1.1. Optimal Season and Most Effective Regression Method

The analysis presented in chapter 4 assessed the utility of multi-temporal 'Analysis Ready' C-band SAR data for woody canopy cover estimation using different regression methods and different combinations of C-band SAR temporal data (dry & wet seasons of 2017 and 2018) to find the best results for modelling canopy cover in the savanna areas of BLM. Table 3 to Table 10 show model accuracies of different regression methods (RF, SVM & LR) and different combinations of C-band SAR temporal data to predict canopy cover using LiDAR data as the reference datasets.

The results of this study agree with other studies (Matheiu et al., 2013, Urbazaev et al., 2015) in this environment which also showed that the drier the environment became, the more the accuracies increase. The transition into wet season and the first few wet season months (i.e., Oct-Nov) generally show higher accuracies than the wet season (i.e., Mar-Apr) images. This is likely due to trees greening up in early spring, leading to more volumetric scattering, while background soil and grass are still dry and do not negatively affect the SAR signal (Archibald and Scholes, 2007). These results are similar to those of Urbazaev et al. (2015), that observed that the highest woody canopy cover correlations were obtained during the dry season where all grass layers are completely dry, and most trees are without leaves which then allows radar signal to penetrate further into the tree canopy and interact with the larger twigs/branches.

Rainfall in the Bushbuckridge region mostly occurs between late October and May with highest rainfall amount expected between the months of January and February (Urbazaev, 2015). However, unexpected results were produced in 2018, where some wet season months (e.g., January) performed similarly to the early dry season months (e.g., May-June) with an rRMSE difference of ~2%. This might be because 2018 was relatively drier than 2017 with total rainfall figures being 129mm (decrease from 171,86mm in January 2017 to 43,23mm in January 2018) lower in January (Figure 3). Generally, the results produced in the 2018 wet season were closely related to those of the 2018 dry season. Similarly, unexpected results were found by Main et al. (2016), where some wet season accuracies were similar and/or higher than dry season image combinations.

Apart from environmental conditions influencing the 2018 validation results, the LiDAR reference data used in this study was acquired in 2018. Literature suggests that model

performance decreases as the time difference increases between the SAR and LiDAR datasets (Urbazaev et al., 2015; Haarpaintner et al., 2020). Higher model performance is expected in SAR images when the year of acquisition corresponds with the acquisition of LiDAR datasets (Urbazaev et al., 2015; Main et al., 2016). This may well have been the same situation with the dataset of this study. Images for 2018 could be showing higher accuracies than 2017 images due to the LiDAR datasets being of the same year as the 2018 SAR images.

Literature shows that the cross-polarised VH band produces higher accuracies than co-polarised VV band (Haarpaintner et al., 2020). As such, this study presented the results of VH only and where both VV & VH were combined and analysed together. This is because VH is more associated with volumetric scattering and is more sensitive to canopy structure variations, due to differences in canopy volumetric water content (Naidoo et al., 2015). Furthermore, Cross-polarisation (i.e., VH) occurs when the microwave energy is emitted in vertical orientation and then received in the horizontal orientation, after interacting with a target and becoming de-polarised. Co-polarised signals (i.e., VV) emit and receive microwave energy in the same orientation. VV polarisations are said to interact with vertical features (tree branches and trunks), while VH polarisations would interact with both vertical and horizontal features within the canopy and therefore are generally better representations of volume scattering and subsequently produce higher accuracies for modelled woody structure variables (Urbazaev et al., 2015). Hence when VV & VH were combined and analysed together, the validation results showed improvements.

A study by Urbazaev et al. (2015) showed that the comparisons of polarisation bands were only meaningful for imageries of the same season. During the dry season, the HH and HV results showed a high correlation with the LiDAR reference data while during the wet season, the HH and VV results showed a lower correlation than the HV and VH. This is similar to the results of this study where months of the dry season for VH showed higher accuracies than those of the wet season. Furthermore, when months are combined according to their seasons for each year, VH showed higher sensitivity and accuracies, with 2018 showing the highest accuracies. However, when the months and years were combined according to their seasons, there was no apparent difference between the seasons. For example, the validation results presented in Table 7 showed marginal differences regardless of the regression method used.

Naidoo et al. (2014) compared the modelling accuracies of canopy cover using LR, SVM and RF, amongst others. In the study, RF obtained high accuracies while SVM and LR showed

poor results. This is somewhat in contrast to this study as the individual scenes' analysis (e.g., Table 3) showed RF to have lower accuracies than SVM and LR. The differences in these studies may be because Naidoo et al. (2014) used L-band SAR, which is inherently more sensitive to woody structure while C-band is less sensitive and noisy. Therefore, it needs more data/images for the models to build accurate relationships to woody structure variables. However, as more images were added to the analysis, RF showed higher accuracies while SVM remained constant at high accuracies, but with lower predictive ranges (i.e., SVM CC Max = 77%, RF CC Max= 86%; Figures 6 & 7).

The improvement of RF is comparable to the observations of Naidoo et al. (2014), where they found that when more datasets were combined, the RF results returned higher accuracies compared to other regression methods used in their study. Hastie et al. (2009) noted that when the number of data being analysed is small, RF is likely to return poor results, but performance becomes robust when large datasets are analysed. Furthermore, Hastie et al. (2009) stated that in situations where a small dataset is used, LR can sometimes perform better than robust models such as the RF. Hence in this study when more images were added, the RF model showed higher accuracies than the LR model. This is because the LR model is not sensitive enough to find good relationships/correlations between noisy variables while the SVM attempts to achieve a balance between overfitting and prediction accuracy (Ge et al., 2018).

Although the use of ARD products is not common and the products used in this study were distributed in 2019, the usefulness of ARD products has been demonstrated by Ruetchi et al. (2017) for vegetation mapping and dynamics. Furthermore, even though the results show deviations from the regression line (Figure 5.), the results produced are expected for the C-band and can be comparable to Matheieu et al. (2013) who used multi-temporal full polarimetric RADARSAT-2 C-band imagery to map canopy cover in the savannas of the southern African Lowveld. Their models showed that the most appropriate season for modelling woody structural metrics is in the dry season when trees have very few leaves left within their canopy. Urbazaev et al. (2015) also produced comparable results using RADARSAT-2 C-band to model woody canopy cover using individual scenes, multi-seasonal and multi-temporal scenarios. Their results showed that the cross-polarised datasets (HV & VH) produced higher results compared to the co-polarised bands (HH & VV). Like Matheieu et al. (2013), their results also showed that woody canopy cover models are more robust during the dry season.

5.1.2. Mapping Canopy Cover

Canopy cover ranges between 5% in the savanna area to 60% in woodlands and 80% in riparian areas of the BLM (Venter et al., 2003; Main et al., 2016; Janecke, 2020). The canopy cover maps presented in Figure 6 and 7 show that canopy cover is most dominant in the West region, while the Eastern region of the municipality shows little to no vegetation cover. This is due to the topography of the area whereby the West is on a high altitude thus more moisture while the rest of the area has an elevation less than 400m (Urbazeav et al., 2015). Furthermore, the area is characterized by the East to West increasing rainfall gradient (Mathieu et al., 2013), whereby the mean annual rainfall for the West is 1200mm while for the East is 550mm.

The Western region of the maps in Figures 6 and 7 shows a heavily vegetated area which is evidence of commercial forestry. The Eastern region shows a river drainage network while the middle (Close to Welverdient (W)) shows fence line contrasts. The fence line contrast in the region could be delineating the different land use or management practices present in the BLM. Furthermore, the decrease in canopy cover in the East region could be due to the increased land utilization and high population density (Mograbi et al., 2015; Fisher et al., 2015) living in areas with the legacy of intensive over-grazing left behind by white-owned cattle farms from the early 20th century (Pollard et al., 2013). The municipality also comprises communal rangelands, private game reserves (Sabi Sands) and a National Park (the Kruger National Park) which influence land utilization and management practices, which can result in variable distributions of vegetation cover.

Although previous studies have shown that L-band SAR performs well when retrieving woody canopy cover in southern African savannas (Naidoo et al., 2016), and has a potential for accurately mapping canopy cover (Naidoo et al., 2015), Figure 6. and Figure 7. show a good representation of the spatial distribution of canopy cover in BLM even though C-band SAR images were used to make the predictions and SVM showed to have a lower prediction range than RF. This may be in line with Main et al. (2016) who stated that C-band SAR images have the possibility of being an alternative to L-band SAR images in producing woody resources maps, and Mathieu et al. (2013) who in their study found that C-band SAR data provide encouraging results for savanna vegetation studies.

5.2. Conclusion

Canopy cover is the simplest system of measurement for assessing the presence of woody elements in the savanna biome (Naidoo et al., 2016). However, well-validated and detailed

maps are not widely available for southern African savannas. Those that are available often use the optical sensors and/or those that can be prohibitively expensive (e.g., SAR L-band). Thus far, SAR C-band imagery continues to be the only freely available and systematically collected SAR dataset that has the possibility of being an alternative to optical or expensive high-resolution SAR images for mapping canopy cover (Main et al., 2016). Even so, Analysis Ready Data is not commonly available and is still considered a future goal for being able to select the level/readiness of analysis for one to conveniently process data without going through the steps of pre-processing data. This study aimed to assess the utility of multi-temporal 'Analysis Ready' composites of C-band SAR data, for savanna woody canopy cover estimation using different regression methods. Through the objectives; determining the most effective regression method (RF, SVM and LR) for modelling canopy cover, identifying the optimal season for modelling canopy cover (dry & wet seasons of 20017 and 2018) and mapping canopy cover using the most effective regression method and optimal season.

The results of this study showed that although a single season or model is not conclusively better than the others, there are results that can or cannot be recommended for use to model canopy cover. RF showed the most improvement by the number of images used in the analysis. Its performance increased the more images were added while SVM showed high accuracies from the beginning and was not affected by the number of images used. LR performed better than RF when analysing a single image, but when more images were added the results of RF performance surpassed it. This shows that the SVM can be recommended for use regardless of the number of images while the use of LR and RF depends on the number of images being used. The results also showed that the drier the environment, the higher the model accuracies, with a combination of all seasons showing the highest accuracy compared to when each scene was analysed individually.

The utility of the models used in this study differ considerably. SVM is the most effective model when only looking at highest R^2 and lowest rRMSE because it does not get affected by the number of datasets being processed but has a smaller predictive range. RF has a higher computational cost as the more images added to the analysis, the better the results. On the other hand, adding more images to the analysis of LR makes it to not be sensitive or robust enough to find good relationships between the variables being used. On seasonal analysis, dry season was the most optimal for modelling canopy cover, but a combination of both dry and wet seasons yielded better results. This is however dependent on the conditions of the year as seen in the results where 2018 was outperforming 2017 because it (2018) was the driest year.

Furthermore, seasonal variations differ per year, the beginning and end of seasons are not constant and as such early wet season months tend to appear dry while early dry season months appear wet.

The ARD C-band results are comparable to studies that did not use Analysis Ready datasets but instead the authors pre-processed their dataset. The goal of ARD products is to present users the choice of not having to pre-process data by doing computationally expensive processes such as geometric rectification. The use of ARD products also presents an option of easily undertaking large area or long time series analyses studies as much time will no longer be spent on pre-processing the data at hand. This study used unique imagery compared to others especially those in savanna. The ARD products are an experiment by NORCE which means they are not readily available for download on the Sentinel website even though they are Sentinel-1 images. They are composites (i.e., averaged individual scenes over a whole month, and produced a single/monthly composite). This could be good in a noisy environment like the savanna, as it removes a bit of the noise and improves backscatter sensitivity to savanna woody structure but in doing so, it could also remove sensitivity to the woody structure component too. Although other authors have demonstrated the usefulness of ARD C-band products, there is still an ongoing research on its performance, and it can be recommended that more ARD studies be undertaken especially in the, often under-represented, savannas to determine how well these products are able to model and map savanna woody vegetation before they can become the mainstream products of analysis.

REFERENCES

Adjorlolo, C. (2008) *Estimating Woody Vegetation Cover in an African Savanna using Remote Sensing and Geostatistics*. MSc. University of KwaZulu-Natal. Available at: https://pdfs.semanticscholar.org/0d9e/11ae765831fb341f2794fd30b0a01addd5cf.pdf?_ga=2.105013123.1655515847.1570097495-2027520203.1569908946 (Accessed: 2 October 2019).

Alijani, G., Mahdi, H. and Zahra, A. (2018) 'Classifying UAVSAR Polarimetric Synthetic Aperture Radar (PolSAR) Imagery Using Target Decomposition Features', *Proceedings*, 2(333), pp. 1–8. doi: 10.3390/ecrs-2-05146.

Archibald, S. and Scholes, R. J. (2007) 'Leaf green-up in a semi-arid African savanna - separating tree and grass responses to environmental cues', *Journal of Vegetation Science*, 18(4), pp. 583–594. doi: 10.1111/j.1654-1103. 2007.tb02572.x.

Bayr, U. and Puschmann, O. (2019) 'Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network', 50, pp. 220–233. doi: 10.1016/j.ecoinf.2019.01.012.

Borghetti, F. *et al.* (2019) 'South American Savannas', in *Savanna Woody Plants and Large Herbivores*. John Wiley & Sons, Ltd, pp. 77–122. doi: 10.1002/9781119081111.ch4.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.

Bushbuckridge Local Municipality, (2019). Integrated Development Plan. Available at: http://bushbuckridge.gov.za/wp-content/uploads/2019/03/FINAL-IDP-BLM-2018-19.pdf (Accessed: 3 November 2020).

Campbell, J. B. (2006) *Introductio to Remote Sensing*. New York: Taylor & Francis Group.

Campbell, J. B. and Wynne, Randolph. H. (2011) *Introduction to Remote Sensing*. fifth. New York: The Guilford Press.

Chen, K.-S. and Tzeng, Y.-C. (2012) 'On SAR Image Processing: From Focusing to Target Recognition', in *Signal and Image Processing for Remote Sensing*. Second. Boca Raton: CRC Press.

Chuvieco, E. and Huete, A. (2010) *Fundementals of Satellite Remote Sensing*. United States of America: CRC Press.

Cutler, A., Cutler, D. R. and Stevens, J. R. (2012) 'Random Forests', in Zhang, C. and Ma, Y. (eds) *Ensemble Machine Learning: Methods and Applications*. Boston, MA: Springer US, pp. 157–175. doi: 10.1007/978-1-4419-9326-7_5.

Dwyer, J. L. *et al.* (2018) 'Analysis Ready Data: Enabling Analysis of the Landsat Archive', *Remote Sensing*, 10(9). doi: 10.3390/rs10091363.

ESRI (2019) *The Savanna Biome, ESRI's Story Map Journal*. Available at: https://arcg.is/09LHG50 (Accessed: 6 March 2020).

Filipponi, F. (2019) 'Sentinel-1 GRD Preprocessing Workflow', *Proceedings*, 18(1). doi: 10.3390/ECRS-3-06201.

Fisher, J. T. (2013) *People, Parks and Rangelands: An Analysis of three-dimensional woody vegetation structure in a semi-arid savanna*. PhD. University of the Witwatersrand.

Fisher, J. T. *et al.* (2015) 'What lies beneath: detecting sub-canopy changes in savanna woodlands using a three-dimensional classification method', *Applied Vegetation Science*. Edited by D. Rocchini, 18(3), pp. 528–540. doi: 10.1111/avsc.12160.

Forkuor, G. *et al.* (2017) 'High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models', *PLOS ONE*, 12(1), p. e0170478. doi: 10.1371/journal.pone.0170478.

García, M. *et al.* (2018) 'Modellingforest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery', *International Journal of Applied Earth Observation and Geoinformation*, 66, pp. 159–173. doi: 10.1016/j.jag.2017.11.017.

García, S. *et al.* (2010) 'Advanced Nonparametric Tests for Multiple Comparisons in the Design of Experiments in Computational Intelligence and Data Mining: Experimental Analysis of Power', *Inf. Sci.*, 180(10), pp. 2044–2064. doi: 10.1016/j.ins.2009.12.010.

Graw, V., Oldenburg, C. and Dubovyk, O. (2016) 'Bush Encroachment Mapping for Africa: Multi-Scale analysis with remote sensing and GIS'. Center for Development Research, University of Bonn. Available at: http://ssrn.com/abstract=2807811 (Accessed: 6 March 2010).

Haarpaintner, J. (2019) *D3.1 - Initial ARD Product Report, ESA EO4SD project 'SAR-4-Africa'*. NORCE Report N809. Norwegian Research Centre AS, p. 39.

Haarpaintner, J. and Hindberg, H. (2019) 'Multi-Temporal and Multi-Frequency SAR Analysis for Forest Land Cover Mapping of the Mai-Ndombe District (Democratic Republic of Congo)', *Remote Sensing*, 11(24). doi: 10.3390/rs11242999.

Haarpaintner, J. *et al.* (2020) *Advanced Sentinel-1 Analysis Ready Data for Africa (ESA EO4SD: SAR-4-Africa), D5 – Final Report*, *78*. NORCE Klima. Available at: https://norceresearch.brage.unit.no/norceresearch-xmlui/handle/11250/2651389 (Accessed: 31 March 2021).

Hastie, T., Tibshirani, R. and Friedman, J. H. (2009) *The elements of statistical learning: data mining, inference, and prediction*.

Higginbottom, T. P. *et al.* (2018) 'Mapping fractional woody cover in semi-arid savannahs using multi-seasonal composites from Landsat data', *ISPRS Journal of Photogrammetry and Remote Sensing*, 139, pp. 88–102. doi: 10.1016/j.isprsjprs.2018.02.010.

Horning, N. *et al.* (2010) *Remote Sensing for Ecology and Conservation: A Handbook of Techniques*. New York: Oxford University Press.

Hu, X. *et al.* (2020) 'Improving surface roughness lengths estimation using machine learning algorithms', *Agricultural and Forest Meteorology*, 287, p. 107956. doi: 10.1016/j.agrformet.2020.107956.

Ismail, R., Mutanga, O. and Kumar, L. (2010) 'Modeling the Potential Distribution of Pine Forests Susceptible to Sirex Noctilio Infestations in Mpumalanga, South Africa', *Transactions in GIS*, 14(5), pp. 709–726. doi: 10.1111/j.1467-9671.2010. 01229.x.

Janecke, B. B. (2020) 'Vegetation structure and spatial heterogeneity in the Granite Supersite, Kruger National Park', *Koedoe; Vol 62, No 2 (2020)*. doi: 10.4102/koedoe. v62i2.1591.

Jensen, J. R. (2014) *Remote Sensing of the Environment*. 2nd edn. Essex: Pearson.

Jensen, J. R. (2015) *Introductory Digital Image Processing: A Remote Sensing Perspective*. 4th edition. United States of America: Pearson.

Kuhn, M. (2020) 'caret: Classification and Regression Training. R package version 6.0-85.' Available at: https://CRAN.R-project.org/package=caret.

Kumar, L. and Mutanga, O. (2017) 'Remote Sensing of Above-Ground Biomass', *Remote Sensing*, 9(9), pp. 1–8. doi: 10.3390/rs9090935.

Lechner, A. M., Foody, G. M. and Boyd, D. S. (2020) 'Applications in Remote Sensing to Forest Ecology and Management', *One Earth*, 2(5), pp. 405–412. doi: 10.1016/j.oneear.2020.05.001.

Liaw, A. and Wiener, M. (2002) 'Classification and Regression by RandomForest', *R News*, 2(3), pp. 18–22.

Liu, C., Vachon, P. W. and Geling, G. W. (2005) 'Improved ship detection with airborne polarimetric SAR data', *Canadian Journal of Remote Sensing*, 31(1), pp. 122–131. doi: 10.5589/m04-056.

Madonsela, S. *et al.* (2018) 'Estimating tree species diversity in the savannah using NDVI and woody canopy cover', *International Journal of Applied Earth Observation and Geoinformation*, 66, pp. 106–115. doi: 10.1016/j.jag.2017.11.005.

Main, R. *et al.* (2016) 'Hyper-Temporal C-Band SAR for Baseline Woody Structural Assessments in Deciduous Savannas.', *Remote Sensing*, 8(1), pp. 1–19. doi: 10.3390.

Martorella, M. *et al.* (2012) 'An ISAR Technique for Refocusing Moving Targets in SAR Images', in *Singal and Image Processing for Remote Sensing*. Second. Boca Raton: CRC Press.

Mathieu, R. *et al.* (2013) 'Toward structural assessment of semi-arid African savannahs and woodlands: The potential of multitemporal polarimetric RADARSAT-2 fine beam images', *Remote Sensing of Environment*, 138, pp. 215–231. doi: 10.1016/j.rse.2013.07.011.

McCue, C. (2007) '4 - Process Models for Data Mining and Analysis', in McCue, C. (ed.) *Data Mining and Predictive Analysis*. Burlington: Butterworth-Heinemann, pp. 45–66. doi: 10.1016/B978-075067796-7/50026-X.

Meyer, D. *et al.* (2019) 'e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.', *R package version 1.7-3*. Available at: https://CRAN.R-project.org/package=e1071.

Mograbi, P. J. *et al.* (2015) 'Biomass Increases Go under Cover: Woody Vegetation Dynamics in South African Rangelands', *PLoS ONE*, 10(5), pp. 1–21. doi: 10.1371.

Mountrakis, G., Im, J. and Ogole, C. (2011) 'Support vector machines in remote sensing: A review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), pp. 247–259. doi: 10.1016/j.isprsjprs.2010.11.001.

Mucina, L. and Rutherford, M. C. (2006) *The Vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute.

Mutanga, O., Adam, E. and 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, pp. 399–406. doi: 10.1016/j.jag.2012.03.012.

Naidoo, L. *et al.* (2014) 'The assessment of data mining algorithms for modelling Savannah Woody cover using multi-frequency (X-, C- and L-band) synthetic aperture radar (SAR) datasets', in *2014 IEEE Geoscience and Remote Sensing Symposium*. *2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 1049–1052. doi: 10.1109/IGARSS.2014.6946608.

Naidoo, L. *et al.* (2015) 'Savannah woody structure modelling and mapping using multi-frequency (X-, C- and L-band) Synthetic Aperture Radar (SAR) data', *Journal of Photogrammetry and Remote Sensing*, 105(1), pp. 234–250. doi: 10.1016/j.isprsjprs.2015.04.007.

Pietersen, J. and Maree, K. (2007) 'Overview of Statistical Techniques', in *First Steps in Research*. Revised. Pretoria: Van Schaik (Maree K.), pp. 225–252.

Piwowar, J. M. (2001) *Remote Sensing Imagery Processing Levels*. Available at: http://uregina.ca/piwowarj/Think/ProcessingLevels.html (Accessed: 5 February 2021).

Pollard, S., Toit, D. du and Biggs, H. (2013) 'River management under transformation: The emergence of strategic adaptive management of river systems in the Kruger National Park', *Koedoe*, 53(2), p. 14. doi: 10.4102/koedoe. v53i2.1011.

Potapov, P. *et al.* (2020) 'Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping', *Remote Sensing*, 12(3). doi: 10.3390/rs12030426.

R Core Team (2018) *R: The R Project for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.r-project.org/ (Accessed: 9 November 2020).

Robinson, D. L. (2005) 'Accounting for bias in regression coefficients with example from feed efficiency', *Livestock Production Science*, 95(1), pp. 155–161. doi: 10.1016/j.livprodsci.2004.12.017.

Rüetschi, M., Schaepman, M. E. and Small, D. (2018) 'Using Multitemporal Sentinel-1 C-band Backscatter to Monitor Phenology and Classify Deciduous and Coniferous Forests in Northern Switzerland', *Remote Sensing*, 10(1). doi: 10.3390/rs10010055.

Sankaran, M. (2019) 'Droughts and the ecological future of tropical savanna vegetation', *Journal of Ecology*, 107(4), pp. 1531–1549. doi: 10.1111/1365-2745.13195.

Sankaran, M. *et al.* (2005) 'Determinants of woody cover in African savannas', *Nature*, 438(7069), pp. 846–849. doi: 10.1038/nature04070.

Sankaran, M., Ratnam, J. and Hanan, N. (2008) 'Woody cover in African savannas: the role of resources, fire and herbivory', *Global Ecology and Biogeography*, 17(2), pp. 236–245. doi: 10.1111/j.1466-8238.2007. 00360.x.

Sankaran, M., Ratnam, J. and Hanan, N. P. (2004) 'Tree–grass coexistence in savannas revisited – insights from an examination of assumptions and mechanisms invoked in existing models', *Ecology Letters*, 7(6), pp. 480–490. doi: 10.1111/j.1461-0248.2004. 00596.x.

Scholes, R. J. and Archer, S. R. (1997) 'Tree-Grass Interactions in Savannas', *Annual Review of Ecology and Systematics*, 28(1), pp. 517–544. doi: 10.1146/annurev.ecolsys.28.1.517.

Scholes, R. J. and Walker, B. H. (1993) *An African Savanna: Synthesis of the Nylsvley Study*. Cambridge: Cambridge University Press (Cambridge Studies in Applied Ecology and Resource Management). doi: 10.1017/CBO9780511565472.

Shackleton Charlie and Shackleton Sheona (2004) 'The importance of non-timber forest products in rural livelihood security and as safety nets: a review of evidence from South Africa', *South African Journal of Science*, 100(11), pp. 658–664. doi: 10.10520/EJC96169.

Shackleton, C. M. (2000) 'Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa', *Biological Conservation*, 94(3), pp. 273–285. doi: 10.1016/S0006-3207(00)00001-X.

Shai, S.-S. and Shai, B.-D. (2014) *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press.

Shevade, S. K. *et al.* (2000) 'Improvements to the SMO algorithm for SVM regression', *IEEE Transactions on Neural Networks*, 11(5), pp. 1188–1193. doi: 10.1109/72.870050.

Shorrocks, B. (2007) *The Biology of African Savannahs*. New York: Oxford University Press.

Srivastava, H. S. *et al.* (2008) 'Potential Applications of Multi-Parametric Synthetic Aperture Radar (Sar) Data in Wetland Inventory: A Case Study of Keoladeo National Park (A World Heritage and Ramsar Site), Bharatpur, India', in *Proceedings of Taal 2007. The 12th World Lake Conference*, Bharatpur, India: Space Applications Centre, pp. 1862–1879. Available at: https://www.academia.edu.

Staben, G., Lucieer, A. and Scarth, P. (2018) 'Modelling LiDAR derived tree canopy height from Landsat TM, ETM+ and OLI satellite imagery—A machine learning approach', *International Journal of Applied Earth Observation and Geoinformation*, 73, pp. 666–681. doi: 10.1016/j.jag.2018.08.013.

Stafford, W. *et al.* (2017) 'The economics of landscape restoration: Benefits of controlling bush encroachment and invasive plant species in South Africa and Namibia', *Ecosystem Services*, 27, pp. 193–202. doi: 10.1016/j.ecoser.2016.11.021.

Statnikov, A., Wang, L. and Aliferis, C. F. (2008) 'A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification', *BMC Bioinformatics*, 9(1), p. 319. doi: 10.1186/1471-2105-9-319.

Sugiura, N. (1978) 'Further analysts of the data by akaike' s information criterion and the finite corrections', *Communications in Statistics - Theory and Methods*, 7(1), pp. 13–26. doi: 10.1080/03610927808827599.

Truckenbrodt, J. *et al.* (2019) 'Towards Sentinel-1 SAR Analysis-Ready Data: A Best Practices Assessment on Preparing Backscatter Data for the Cube', *Data*, 4(3). doi: 10.3390/data4030093.

Twine, W. C. and Holdo, R. M. (2016) 'Fuelwood sustainability revisited: integrating size structure and resprouting into a spatially realistic fuelshed model', Journal of Applied Ecology, 53(6), pp. 1766–1776. doi: 10.1111/1365-2664.12713.

Twine, W. *et al.* (2003) 'Consumption and direct-use values of savanna bio-resources used by rural households in Mametja, a semi-arid area of Limpopo province, South Africa'. Available at: http://hdl.handle.net/10204/1903.

Urbazaev, M. *et al.* (2015) 'Assessment of the mapping of fractional woody cover in southern African savannas using multi-temporal and polarimetric ALOS PALSAR L-band images', *Journal of Remote Sensing Environment*, 166(1), pp. 138–153. doi: 10.1016/j.rse.2015.06.013.

Venter, F. J., Scholes, R. J. and Eckhardt, H. C. (2003) The abiotic and template its associated vegetation pattern. In *the Kruger experience. Ecology and Management of Savanna Heterogeneity;* J. T. Du. Toit, K. H. Rogers, and H. C. Biggs., Eds. London: London Island Pres

Ward, D., Wiegand, K. and Getzin, S. (2013) 'Walter's two-layer hypothesis revisited: back to the roots!', *Oecologia*, 172(3), pp. 617–630. doi: 10.1007/s00442-012-2538-y.

Waske, B., Benediktsson, J., Atli and Sveinsson, J., R. (2012) 'Random Forest Classification of Remote Sensing Data', in *Signal and Image Processing for Remote Sensing*. Second. Boca Raton: CRC Press.

Watanabe, M. *et al.* (2006) 'Forest Structure Dependency of the Relation Between L-Band and Biophysical Parameters', *IEEE Transactions on Geoscience and Remote Sensing*, 44, pp. 3154–3165. doi: 10.1109/TGRS.2006.880632.

Wessels, K. *et al.* (2019) 'Mapping and Monitoring Fractional Woody Vegetation Cover in the Arid Savannas of Namibia Using LiDAR Training Data, Machine Learning, and ALOS PALSAR Data', *Remote Sensing*, 11(22). doi: 10.3390/rs11222633.

Wessels, K. J. et al. (2013) 'Unsustainable fuelwood extraction from South African savannas', Environmental Research Letters, 8(1), p. 014007. doi: 10.1088/1748-9326/8/1/014007.

Wessels, K.J., Mathieu, R., Erasmus, B.F.N., Asner, G.P., Smit, I.P.J., van Aardt, J.A.N., Main, R., Fisher, J., Marais, W., Kennedy-Bowdoin, T., Knapp, D.E., Emerson, R., Jacobson, J., 2011. Impact of communal land use and conservation on woody vegetation structure in the Lowveld savannas of South Africa. Forest Ecology and Management 261, 19–29. https://doi.org/10.1016/j.foreco.2010.09.012

Wiegand, K., Ward, D. and Saltz, D. (2005) 'Multi-Scale Patterns and Bush Encroachment in an Arid Savanna with a Shallow Soil Layer', *Journal of Vegetation Science*, 16(3), pp. 311–320.

Wigley, B. J. *et al.* (2020) 'Grasses continue to trump trees at soil carbon sequestration following herbivore exclusion in a semiarid African savanna', *Ecology*, n/a(n/a), p. e03008. doi: 10.1002/ecy.3008.

Wu, F. *et al.* (2011) 'Rice Crop Monitoring in South China With RADASAT-2 Quad-Polarisation SAR Data.', *IEEE Geoscience and Remote Sensing Letters*, 8(2), pp. 196–200. doi: 10.1109/LGRS.2010.2055830.

Zhang, H. *et al.* (2017) 'Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model', *Science of The Total Environment*, 592, pp. 704–713. doi: 10.1016/j.scitotenv.2017.02.146.

Zhang, W. *et al.* (2019) 'From woody cover to woody canopies: How Sentinel-1 and Sentinel-2 data advance the mapping of woody plants in savannas', *Remote Sensing of Environment*, 234, p. 111465. doi: 10.1016/j.rse.2019.111465.