Informing Reforestation Practices

Quantifying Live Forest Above Ground Biomass of a Randomly Mixed Natural Forest Plantation using GIS and Remote Sensing Models

By

Sizwe Thamsanqa Hlatshwayo

Submitted in fulfilment of the academic requirements for the degree of Master of Science in the School of Agricultural, Earth and Environmental Sciences, Discipline of Geography, University of KwaZulu-Natal, South Africa

February 2017

Abstract

Restoration of natural forests is viewed as one of the effective and viable approaches for mitigating and adapting to climate change. However, maximising the carbon capture and storage of naturally mixed forest plantations is currently a challenge for forest managers, due to the complex nature of species interaction and environmental controls that inhibit the distribution and growth rates of certain species. Monitoring the amount of carbon captured and stored in natural forest ecosystem is vital in verifying their productivity and detecting areas of concern that could be unproductive. In this study the productivity of the Buffelsdraai reforestation site was monitored using above ground biomass (AGB) of planted trees. While there are traditional approaches for monitoring forest AGB with high accuracy, these approaches are unfavourable because they are timeous and spatially restricted. Fortunately, the inception of remote sensing has provided viable approaches for estimating forest AGB at a synoptic scale and with low cost. The purpose of this study was to apply remote sensing and GIS models to quantify the ecological benefits of the Buffelsdraai reforestation project on AGB productivity. The study investigated the potential of the spatially optimised three band texture combinations in predicting and mapping forest AGB and structural diversity. This research study has potential to contribute to the importance of spatial planning and design of naturally mixed forest plantations to improve their diversity and AGB productivity. The first part of the study focused on mapping the temporal and spatial distribution of forest AGB using spatially optimised three band texture combinations computed from SPOT-6 imagery and random forest regression algorithm. The results indicated that the three band texture combinations were superior in predicting forest AGB compared to raw texture bands and two band texture combinations. The second part of the thesis focussed on assessing the effects of forest structural diversity and topographic variables on forest AGB productivity using GIS and remotely sensed data. The forest structural diversity measures were predicted using three band texture combinations modelled using random forest and stochastic gradient boosting algorithms. The topographic variables were derived using the digital elevation model in ArcMap 10.3. Results indicated that random forest yielded overall higher accuracies in predicting the forest structural diversity measures compared to stochastic gradient boosting. More importantly, the study showed that forest diversity and topographic variables have significant influences on forest AGB variability. Overall the study provided insight into the management of natural forests and to the importance of spatial planning and design of these mixed forests.

Preface

The research undertaken in this thesis was conducted in the School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, from March 2015 to February 2017, under the supervision of Professor Onisimo Mutanga.

I, Sizwe Thamsanqa Hlatshwayo declare that the work submitted in this thesis represents my own original work and has never been submitted for examination at any other university. Where the work of others has been used, I have duly acknowledged it in the text and reference sections of this thesis.

Sizwe Thamsanqa Hlatshwayo Signed: _____ Date: _____

As the candidate's supervisor, I certify the above statements and have approved this thesis for submission.

Professor Onisimo Mutanga Signed: _____ Date: _____

As the candidate's co-supervisor, I certify the above statements and have approved this thesis for submission.

Dr. Romano Lottering Signed: _____ Date: _____

Declaration 1: Plagiarism

I, Sizwe Thamsanqa Hlatshwayo declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written, but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics, or tables copied and pasted from the Internet, unless specifically acknowledged and the source being detailed in the thesis and in the References section.

Signed: _____

Declaration 2: Publication and manuscripts

1. Hlatshwayo, S. T., Mutanga, O., Lottering, R., Kiala, Z., & Ismail, R. (Under second round revision). An Innovative Technique for Mapping Temporal and Spatial Variation of Forest Aboveground Biomass in the Reforested Buffelsdraai Landfill Site using Texture Combinations Computed from SPOT-6 Pan-sharpened Imagery. *Remote Sensing of Environment*.

2. Hlatshwayo, S. T., Mutanga, O., Lottering, R.T, and Peerbhay, K., (In preparation). Evaluating the Effects of Forest Structural Diversity and Topography on Forest Above Ground Biomass using Three Band Texture Combinations Computed from SPOT-6 imagery and Advanced Machine Learning Algorithms: A case study of the Buffelsdraai Community Reforestation Project in KwaZulu-Natal, South Africa.

Dedication

Dedicated to my mother who is my pillar of strength

ACKNOWLEDGEMENTS

I would like to first extend my gratitude to the almighty God for giving me strength to complete this project.

To my supervisor, Professor Onismo Mutanga, my warm gratitude goes out to you for all your guidance and invaluable ideas throughout my research. Having you as a supervisor has been such a great honour.

To Dr. Romano Lottering, I thank you for your continuous support and guidance throughout my MSc. You have guided me through the challenging times of my research, though I gave you problems and complaints you never runout of patience with me, I am truly grateful to you.

Dr. Kabir Peerbhay, I thank you for your continuous support and assistance with some of the technical aspects of my MSc. Your skill and capacity within the remote sensing domain has been a very stimulating experience for me.

My deepest gratitude goes to the geography department staff Mr. Donavin DeVos, Mrs. Ramroop, Dr. Desai and Mr. Victor Bangamwabo. I thank you for providing me the best support structure and working conditions to be at my best all the time.

My dear best friend Shenelle Sewell I thank you for your support throughout the year, through your undying love and encouragement, you practically was with me all the way through the project I am most grateful to you.

My colleagues and friends who supported me and encouraged me to complete my project I thank you all.

I have to express my gratitude to The EThekwini municipality for providing me with funds to complete this project, it would not have been possible without their support, and they played a significant role in my research.

Last but never least, I would like to thank my family for their support throughout the year, through their encouragement and love I am deeply grateful especially my mother who has been my pillar of strength and my motivation to complete this MSc degree.

Table of Contents

Abstract ii

Preface iii

Declaration 1: Plagiarism iv

Declaration 2: Publication and manuscripts v

Dedication vi

ACKNOWLEDGEMENTS vii

CHAPTER 1 1

General Introduction 1

1.1 Background Information 2

1.2 Benefits of Reforestation 3

1.3 Remote sensing of Forest above Ground Biomass 4

1.4 Aims and Objectives 6

1.5 Area Description 7

1.6 Thesis Structure 8

CHAPTER 2 9

Predicting and Mapping Temporal and Spatial Variation of Forest Aboveground Biomass using Image Texture Combinations Computed from SPOT-6 Pan-sharpened Imagery. 9

Abstract 10

2.1 Introduction 11

2.2 Methods and Material 13

2.2.1 Study area 13

2.2.2 Image Acquisition and Pre-processing 14

2.2.3 Field Measurements 15

2.2.4 Wood density measurements 16

2.2.5 Field Biomass Calculations 18

2.2.6 Optimum Window Selection 18

2.2.7 Characterising forest AGB representative classes 18

2.2.8 Resampling the SPOT-6 pan-sharpened image 19

2.2.9 Minimum variance calculation 19

2.2.10 Image Texture analysis 20

2.2.11 Extracting texture parameters 23

2.2.12 Statistical Analysis 24

2.2.13 Relationship between texture indices and AGB 24

2.2.14 Model Validation 25
2.3 Results 26
2.3.1 Descriptive statistics 26
2.3.2 Window Size Selection 26
2.3.3 Correlation Analysis 29
2.3.4 Predicting AGB of the 2009-2011 and 2011-2012 plantation phases 30
2.3.5 Variable Importance Measures 31
2.3.6 Predictive performance of SPOT6 Texture Combination Models 35
2.3.7 Frequency Analysis 37
2.4 Discussion 41
2.5 Conclusion 45
CHAPTER 3 46
Evaluating the Effects of Forest Structural Diversity and Topography on Forest Above Ground Biomass using Three Band Texture Combinations and Advanced Machine Learning Algorithms 46
Abstract 47
3.1 Introduction 48
3.2 Methods and materials 53
3.2.1 Study area 53
3.2.2 Image acquisition 54
3.2.3 Forest Inventory Data 55
3.2.4 Structural and Species Diversity Indices 55
3.2.5 Topographic Variables 57
3.2.6 Remotely Sensed Biomass Model 61
3.2.7 Optimum Window Size Selection 62
3.2.8 Minimum variance 62
3.2.9 Image texture computation 63
3.2.10 Extracting image texture 63
3.2.11 Statistical Analysis 64
3.2.12 Model validation 65
3.3 Results 65
3.3.1 Descriptive statistics 65
3.3.2 Window Size Selection 66
3.3.3 Correlation Analysis 67
3.3.4 Forest structural diversity predictions: RF and SGB Regression Model Performance 69

3.3.5 Best Performing Models' Variable Importance measurements 72
3.3.6 Impact of forest structural diversity and topographic variables of forest AGB. 78
3.3.7 Principal Component Analysis 83
3.4 Discussion 86
3.4.1 Predictive performance of RF and SGB Algorithm 86
3.4.2 Ecological Interactions 88
3.4.3 The effects of FSD measures on Forest AGB 88
3.4.4 The Effects of Topographic variables 89
3.4.5 Implications for Spatial Planning of Reforestation sites 90
3.5 Conclusions 91
CHAPTER 4 93
Quantifying the Ecological Benefits of Above Ground Biomass using GIS and Remote Sensing Models: A synthesis 93
4.1 Introduction 94
4.1.1 Estimating forest AGB of trees planted at different phases using texture band combinations and random forest regression. 95
4.1.2 Determining the effects of forest structural diversity and topographic variables on forest AGB productivity using remote sensing and GIS models. 97
4.2 Conclusions 100
REFERENCES 102

List of Figures:

Figure 2.1: Location of study site and the SPOT-6 pan-sharpened image of the Buffelsdraai Reforestation site covering the study area 14

Figure 2.2: The relationship between wood density values obtained from field measurements and wood density values obtained from online wood density database of African dry forests, (Glenday, 2007). 17

Figure 2.3: Optimum window sizes for the 2009-2011 plantation phase selected based on minimal variance of pixel values, under varying forest AGB classes. Herein, a, b, and c represent 0-220 (kg m^{-2}), 221-440 (kg m^{-2}) and 441-680 (kg m^{-2}) forest AGB, respectively. 27

Figure 2.4: Optimum window sizes for the 2011-2012 plantation phase selected based on minimal variance of pixel values, under varying canopy cover percentage classes. Herein a, b and c represent 0-115 (kg m^{-2}), 116-230 (kg m^{-2}), and 231-360 (kg m^{-2}) percentage canopy cover, respectively. 28

Figure 2.5: Selection optimum number of variables (texture) for predicting forest AGB using backward elimination search function, a, b and c, represent raw band texture, two band texture ratios and three band texture ratios, respectively.....34

Figure 2.6: Shows that the three texture band models produced the overall highest predicted performance with a R^2 of 0.88 compared to both the two texture band ratios (R^2 = 0.85) and raw band texture (R^2 = 64) based on test dataset of the pan-sharpened image. Herein a, b, and c represent the raw texture bands, two band texture and three band texture combinations, respectively and i and ii represent test dataset and cross-validation dataset, respectively. 36

Figure 2.7: Presents the frequencies of a: SPOT6 bands, b: window size, and c: texture measure in the selected models of single texture bands, 2 texture band ratios and 3 texture band ratios for the pan-sharpened image. 37

Figure 2.8: Above ground biomass map derived from the best performing three texture band combinations computed from the pan-sharpened image for the 2009-2013 plantation period. 39

Figure 2.9: Mean AGB for the five year successional dates from 2009 to 2013 computed from the predicted above-ground biomass map shown in Figure 2.8.....40

Figure 3.1: Location of study site and the plantation phases in the Buffelsdraai Reforestation site covering the study area.....54

Figure 3.2: The digital elevation map of the study area computed from 5 m interval contour lines. 57

Figure 3.3: Slope steepness in angles derived from the DEM 58

Figure 3.4: The topographic wetness of the study area based on the levitation and runoff accumulation areas. 59
Figure 3.5: Annul shortwave radiation map (MJ m^{-2} $year^{-1}$) derived using Geographic Information System (GIS) and digital elevation model. 60
Figure 3.6: Above ground biomass map derived from the best performing three texture band combinations computed from the pan-sharpened image for the 2009-2013 plantation period. 61
Figure 3.7: Shows one to one relationship between measured and predicted forest structural diversity attributes derived from three texture band models. Only best performing models are displayed here, a, b, c, d and e represent SD, DD, SR, TR and GC, respectively. 71
Figure 3.8: Variable importance measurements of texture models in predicting forest AGB using RF. Higher OOB error indicates high importance, a, b, c, d and e represent SD, DD, SR, TR and GC, respectively. 73
Figure 3.9: Map showing predicted forest structural diversity measures derived from the best performing three texture bands and regression algorithms. 75
Figure 3.10: Reclassified map showing categorical groups of forest structural diversity measures derived from the predicted maps in figure 3.9. 77
Figure 3.11: Mean forest AGB for the forest structural diversity categorical groups in figure, computed from the predicted above-ground biomass map shown in Figure 3.6. 79
Figure 3.12: Mean forest AGB for the topographic variables' categorical groups extracted from the predicted above-ground biomass map shown in Figure 3.6, a, b, c and d represent altitude, slope, solar radiation and topographic wetness, respectively. 80
Figure 3.13: Principal component analysis plots for forest AGB categorical groups (dotted lines) and forest structural diversity attributes (solid lines) that had a significant influence ($p < 0.01$) on spatial distribution of forest AGB. 84
Figure 3.14: Principal component analysis plots for forest AGB categorical groups (dotted lines) and topographic variables (solid lines) that had a significant influence ($P < 0.01$) on spatial distribution of forest AGB. 85
Figure 4.1: Above ground biomass map derived from the best performing three texture band combinations computed from the pan-sharpened image for the 2009-2013 plantation period. 97
Figure 4.2: Principal component analysis plots for forest AGB categorical groups (dotted lines) with forest structural diversity (in A) and topographic variables (in B) that had a significant influence ($p < 0.01$) on spatial distribution of forest AGB. 99

List of Tables

Table 2.1: The spectral bands and spatial resolutions of the SPOT-6 pan-sharpened image .15
Table 2.2: Descriptive statistics of measured wood densities for the selected dominant tree species and published wood density values 16
Table 2.3: The selected representative classes for optimum window size selection.....19
Table 2.4: Definition and equations of GLOM texture measures 21
Table 2.5: Definitions and equations of GLCM texture measures.....22
Table 2.6: Explanatory statistics of the observed above ground biomass (kg m^{-2}).....26
Table 2.7: Canopy cover and wood density classes used for optimum window size selection for AGB estimation and their corresponding window sizes. 29
Table 2.8: Significant three band texture combinations computed from the pan-sharpened image ($p < 0.05$). 30
Table 2.9: Accuracy of raw band texture in predicting forest AGB at different plantation phases 31
Table 2.10: Variable importance measurements of texture models in predicting forest AGB using RF. Higher OOB error signifies higher variable importance 33
Table 2.11: Predictive Performance of the texture models.....35
Table 2.12: Variables that were selected for constructing texture models using forward selection in RF regression. 38
Table 3.1: Descriptive statistics of the observed forest structural diversity measures. 66
Table 3.2: Forest structural diversity attributes' classes used for optimum window size selection and their corresponding window sizes. 67
Table 3.3: Correlation between forest structural diversity attributes with three band texture combinations significant ($p < 0.05$). 69
Table 3.4: Random forest and stochastic gradient boosting regression results illustrating the correlation between forest structure and band texture combinations with p-value<0.05. 70
Table 3.5: One-way Analysis of Variance results for the forest structural diversity attributes and the topographic variables 81
Table 3.6: Games Howell post-hoc analysis tests for the forest structural diversity and topographic variables' categorical groups 82
Table 3.7: Principal component analysis fitting forest structural diversity attributes as explanatory variables and forest AGB as the target categorical variable. 84

Table 3.8: Principal component analysis topographic variables as explanatory variables and forest AGB as the target categorical variable. 85
Table 4.1: Canopy cover and wood density classes used for optimum window size selection for AGB estimation and their corresponding window sizes. 95
Table 4.2: Predictive Performance of the texture models 96
Table 4.3: Random forest and stochastic gradient boosting regression results illustrating the correlation between forest structure and band texture combinations with p-value<0.05. 98

CHAPTER 1

General Introduction

1.1 Background Information

Forests represent one of the severely degraded biomes worldwide (FAO, 2010). In this case, forests can be broadly divided into three categories, namely; natural, plantations and woodlands. In South Africa, natural forests cover approximately 0.5 million hectares of the country's total land surface (Mucina & Rutherford, 2006a). On the other hand, woodlands cover the largest land area in South Africa, with the area covered approximated to be between 39-42 million hectares (DAFF, 2012). Conversely, commercial forests cover 1.3 million hectares of land area, which is approximately 1.1% of the country's land area (Lottering & Mutanga, 2012). While these ecosystems cover a small portion of the South African land area, they are capable of providing various economic, social and environmental benefits.

Benefits derived from forests include, their ability to sequestrate carbon from the atmosphere, moderate the hydrological cycle, control the rate of erosion, and provide various resources that are critical for the survival of rural livelihood (Nasi & Wunder, 2002; Díaz et al., 2009; FAO, 2010). In addition, the aesthetic nature of forest settings provides potential for ecotourism and non-consumptive source of income for the poor. Furthermore, commercial forests play a significant role in South Africa's economy, as it contributes approximately 2% to the gross domestic product of the country (DAFF, 2012). However, South African forests, more especially indigenous forests face a wide variety of threats that have caused degradation of this valuable biome.

Indigenous forests face a wide variety of threats that have resulted in the loss or fragmentation of the forest habitat and consequently loss of biodiversity (Leblois et al., 2017). Threats to the forest biome result from agricultural expansion, increase development and other activities such as harvesting of timber for commercial or subsistence usage (Noriko et al., 2012). Degradation of forests habitat has resulted in an increase in carbon concentration in the atmosphere, and subsequently reduction of carbon sequestration and storage capacity, both above and below ground (Díaz et al., 2009). Such an increase in carbon concentration in the atmosphere has resulted in global climate change.

1.2 Benefits of Reforestation

The threat and awareness of anthropogenic driven climate change has reached critical stage over the past decade, and has become a global concern. According to the UN report the year 2015 was the hottest year recorded in history (UN, 2015) and thus raises the alarm on the prevalence of climate change, despite efforts being made to fight against it. This has led to cooperative and integrated measures to manage climate change that are being undertaken at a global scale. Numerous studies have emphasised the significance of maximising ecosystem resilience through restoration of natural forest as an approach to control the causes and impacts of climate change (Turner et al., 2009). The emphasis on reforestation stems from the fact that forests play a crucial role in the biogeochemical cycle of carbon. Forests act as carbon sinks, through uptake and storage of carbon within its compartments (Pan et al., 2011; Bastin et al., 2014; Sousa et al., 2015). Currently there are no direct methods of measuring carbon stored in trees (Brown, 2002), and researchers have suggested the use of biomass as an index of carbon stored in trees (Nowak & Crane, 2002; Peichl & Arain, 2006).

Monitoring forest above ground biomass (AGB) is critical in climate change modelling and carbon accounting as they affect the concentration of greenhouse gases in the atmosphere (Dube et al., 2014). However, optimising AGB of natural forests is currently a challenge for forest managers due to the complex nature of species interaction (Shirima et al., 2015) and constraints of topographic variables on species growth and distribution (Gracia et al., 2007). Understanding the nature of tree species interaction is crucial in determining their compatibility and ensuring their survival (Forrester, 2004). Competitive exclusion is a common case in natural forests and can result in low species diversity and high tree mortality (Huston, 1979). Therefore, there is a need to manipulate of species functional traits to ensure higher biodiversity, AGB productivity and low interspecies competition. Species identity alone does not account for forest AGB productivity. Topographic variables such as slope, elevation, radiation and topographic wetness are critical in determining the rates of tree growth (Gracia et al., 2007). These variables affect the distribution of tree species and their growth rate across the landscape (Woollen et al., 2012). For example, high altitudes are characterised by low vegetation density and low growth rates, due to low temperatures and high pressure (Moser et al., 2011).

An in-depth understanding of the effects of forest structural diversity and topographic factors on forest AGB productivity is critical for spatial planning and design of natural forests. Traditionally forest structural attributes are measured directly by harvesting trees and measuring tree height, diameter and weighing their wet and dry mass (Chave et al., 2005). Although these methods are highly accurate, they however, have negative impacts on the environment including ecosystem disturbance, modification of ecological processes and loss of wildlife (Dube & Mutanga, 2015b). Furthermore, these methods are timeous, labour intensive, costly and only limited to a plot size, therefore lacking spatial continuity and repetitive coverage. Therefore, there is a need for accurate biomass estimation methods that are efficient and allow for spatial continuity and repetitive coverage.

1.3 Remote sensing of Forest above Ground Biomass

The advent of optical remote sensing has made it possible to provide AGB estimates that are spatially continuous, efficient and allow for repetitive coverage (Mather & Koch, 2011a; Lottering & Mutanga, 2012; Dube et al., 2014). The mapping of forest structural attributes exhorts the principle that trees with varying AGB content reflect and absorb radiation differently within the electromagnetic spectrum. According to Mather and Koch (2011b), vegetation that is vigorously growing reflects high in the NIR portion of the electromagnetic spectrum, due to multiple scattering effects and absorbs highly in the red portion of the electromagnetic spectrum, due to the presence of chlorophyll pigments. However, studies have shown that in instances where there is high variation in chlorophyll content (due to presence of stressed vegetation) the green wave band is more sensitive to chlorophyll variation compared to the red band (Carter, 1993a; Gitelson et al., 1996). Carter (1993a), discovered that the green spectral region (535-640) was more sensitive to vegetation stress. For example, Gitelson et al. (1999) found high sensitivity of the green and the red portion of the spectrum to vegetation with high variability in chlorophyll content ranging from 3 µg/cm^2 to 44 µg/cm^2. In another study, Daughtry et al. (2000) discovered that the green band was able to predict chlorophyll in corn leaf with varying nitrogen applications.

Attempts to remotely sense forest structural attributes have achieved plausible results over the past decades using multispectral, Synthetic Aperture Radar (SAR) and Lidar images (Cho et al., 2009). Higher accuracies were obtained using satellite images with high spatial resolution

and moderate spectral resolution with strategic bands in the red-edge region. Dube et al. (2014), found that the Rapid-Eye (5m resolution) was able to predict AGB of forest plantation with high accuracies (R^2 as high as 0.80 and RMSE as low as 16.93 t ha^{-1}). Eckert (2012), predicted AGB and carbon in degraded forest using red band derived from Wordview-2 with an R^2 of 0.82. Sarker and Nichol (2011), discovered that the AVNIR-2 image was able to explain 88% of thevariability in forest biomass (RMSE = 32 t ha^{-1}). As plausible as their results maybe, the use of these optical images, especially with high spatial resolution is limited by their cost, availability and accessibility (Barbosa et al., 2014b).

Due to the limitations in accessing satellite sensors with high spatial and spectral resolution, researchers have resorted to evaluating freely available images, such as; Landsat-8, and SPOT-6 imagery in estimating AGB (Dube et al., 2014; Kelsey & Neff, 2014; Dube et al., 2016). A major limitation of using freely available satellite images, is their low spatial and spectral resolutions. However, the use of advanced image transformation techniques have made it possible to increase the accuracy of these freely available images in predicting forest structural attributes (Mather & Koch, 2011a). Spectral vegetation indices have been extensively used as the main image transformation technique when estimating forest structural attributes (Shamsoddini et al., 2013). These indices have been widely adopted due to their capability of ironing out certain characteristics of spectral reflectance curves of various earth surfaces (Mutanga & Skidmore, 2004b). In addition, they are capable of eradicating errors associated with sun angle and variations in topography that affects the recorded upwelling radiance (Mather & Koch, 2011a). However, VIs have achieved inadequate results in tropical and sub-tropical areas, where there are closed forest canopies, with multiple layering, high density AGB and high species diversity (Dube & Mutanga, 2015b). Alternatively, some studies have been geared towards the use of image texture analysis techniques to address the issues associated with vegetation indices in fast growing vegetation, with high density vegetation (Mutanga & Skidmore, 2004b; Mather & Koch, 2011b; Bastin et al., 2014).

Texture analysis attempts to compute the degree of image tone variation in high resolution imagery for use in biomass estimation (Wulder et al., 1998). Image tone variation in forested areas is a function of changes in species type, level of crown closure, and stem density (Franklin, Wulder, & Gerylo, 2001). Texture analyses due to their capability to simplify canopy structure are capable of improving the estimation of forest structural attributes when compared to spectral VI's (Mather & Koch, 2011b). For example, Sarker and Nichol (2011), found a high performance of raw texture bands in estimating forest biomass compared to raw spectral bands.

The high performance of image texture in predicting AGB has been attributed to its ability to simplify forest canopy structure by smoothing the canopy surfaces using statistical texture variables (Franklin et al., 2001). Recent studies, however, have made new avenues in an attempt to improve AGB estimation using texture analysis, by introducing texture band combinations (Nichol & Sarker, 2011; Sarker & Nichol, 2011; Sarker et al., 2013; Dube & Mutanga, 2015b).

The use of texture band combinations have achieved a high degree of success in estimating forest structural attributes, especially when compared to raw texture bands and spectral vegetation indices. Sarkar & Nichol, (2011) obtained higher accuracies using texture ratios computed from AVNIR-2 image (R^2 = 0.88 and RMSE = 32 t ha^{-1}) compared to raw texture bands (R^2 = 0.76 and RMSE = 46 t ha^{-1}). Using image texture computed from AVNIR-data and SPOT-5 images Nichol and Sarker (2011) were able to improve AGB estimation from an R^2 of 0.91, using raw texture bands, to an R^2 of 0.94, using texture combinations. This study, extends from the work of Nichol & Sarker, (2011); Sarker & Nichol, (2011); Sarker et al., (2013); Dube & Mutanga, (2015b), by introducing the three texture band combination method. The motivation for using such a complex image processing level is that texture is capable of simplifying the canopy structure, while the textured band combinations are able to reduce errors that result from sun angle effect, topographic and illumination effects on upwelling radiance (Nichol & Sarker, 2011). In essence, this technique is a hybrid of both spatial and spectral information of remotely sensed data and needs to be tested in detecting and mapping AGB. According to our knowledge, no study has used the 3 band texture combination in detecting and mapping AGB within an indigenous forest.

1.4 Aims and Objectives

Based on the above discussion, the aim of this study was to quantify the benefits of the Buffelsdraai reforestation program on ecosystem productivity using above ground biomass derived from remotely sensed data as an indicator. The objectives of the study were set as follows:

1. To investigate the performance of image texture in various processing combinations to detect AGB.

2. To establish the temporal (i.e. depending on the plantation phase) and spatial distribution of forest AGB using the best performing texture combinations derived from SPOT-6 imagery and random forest regression.
3. To evaluate the utility of random forest (RF) with stochastic gradient boosting in predicting and mapping forest structural diversity using three band texture combinations.
4. To assess the effects of forest species and structural diversity on AGB productivity using remotely sensed data and multivariate statistical analyses.
5. To determine the contribution of topographic variables on the spatial distribution of forest AGB using remote sensing and GIS models.

1.5 Area Description

The Buffelsdraai community Reforestation project is situated along the east-coast of Durban KwaZulu-Natal, South Africa and is approximately 25 km North of Durban. The trees are planted in the buffer zone of the Buffelsdraai Landfill site, which consists of 50 indigenous tree species covering approximately 520.6 km^2. The landscape of the area is characterised by undulating steep heels that range from 200 m to 325 m above sea level. The Dwyaki Tillite and the Ecca shale rock group predominantly underlie the area. The Glenrosa, Shortlands and Swartland soils occupy the majority of the study area. Prior to the reforestation, the buffer zone of the landfill site was used for sugarcane farming and as a rangeland for communal cattle grazing. The area was planted at intervals commencing in 2009 to offset the 2010 world cup carbon footprint. The majority of its precipitation falls in summer with an annual rainfall of 766 mm $year^{-1}$on average. Temperatures are highest in summer and lowest in winter with averages of 27.5ºC and 22.2ºC, respectively. The area falls under the KwaZulu-Natal Coastal Belt that is listed as endangered under the South African vegetation map (Mucina and Rutherford 2006a).

1.6 Thesis Structure

The research objectives outlined in the introduction sectioned were achieved by breaking down the thesis into two papers, one of the papers is under revision and the second paper is under preparations for a journal submission. The paper under revision for publication is assigned a title and the name of the journal which the paper is being published in. Both paper one and paper two are written separately, however, their findings and conclusions are linked to at-least one of the objectives stipulated in the introduction. The two papers were constructed using the same dataset and image processing techniques with replications of certain sections, however, the application contexts of the papers are different, paper two is merely the continuation of paper one. The papers are separate and can be read independent of each other, thus the overlaps are insignificant. This thesis consists of four chapters.

Chapter 1: This chapter covers the general introduction and motivation of the study. **Chapter 2:** This chapter assesses the capability of texture band combinations derived from the SPOT-6 pan-sharpened image to predict forest AGB at various plantation phases. More specifically, the capability of three band texture combinations to predict forest AGB was compared to raw band textures and two band texture combinations. The predicted map was used to assess the differences in forest AGB across the plantation phases. **Chapter 3:** This chapter quantified the effects of forest structural diversity and topographic variables on forest AGB productivity using remotely sensed data. Three band texture combinations derived from the SPOT-6 image were used to predict forest structural diversity attributes and the digital elevation model was used to derive the topographic variables. The relationship between three band texture combinations was modelled using random forest and stochastic gradient boosting. The interaction between forests' AGB, forest structural diversity, topographic variables and forest AGB were assessed using univariate statistics and multivariate statistical measures. **Chapter 4:** This chapter contextualises the findings of the study for both chapter two and chapter three and recommendations are proposed.

CHAPTER 2

Predicting and Mapping Temporal and Spatial Variation of Forest Aboveground Biomass using Image Texture Combinations Computed from SPOT-6 Pan-sharpened Imagery.

This chapter is based on: Hlatshwayo, S. T., Mutanga, O., Lottering, R., Kiala, Z., & Ismail, R. (Under second round revision). An Innovative Technique for Mapping Temporal and Spatial Variation of Forest Aboveground Biomass in the Reforested Buffelsdraai Landfill Site using Texture Combinations Computed from SPOT-6 Pan-sharpened Imagery. *Remote Sensing of Environment*.

Abstract

Developing models for estimating aboveground biomass (AGB) in naturally growing forests is critical for climate change modelling. AGB models developed using satellite imagery vary with study area, depending on the complexity of vegetation and landscape structure, which affects the upwelling radiance. We assessed the potential of SPOT-6 imagery in predicting AGB of trees planted at different time periods, using image texture combinations. Image texture variables were computed from the SPOT6 pan-sharpened image data, which is characterised by a 1.5 m spatial resolution. In addition, we incorporated the minimal variance technique to select the optimum window sizes that best captures AGB variation in our study area. The results showed that image texture was able to detect AGB for both mature and young trees, however, models detecting mature trees were more superior, with accuracies of R^2 = 0.70 and 0.25 for 2009-2011 and 2011-2013 plantation phases, respectively. In addition, our results showed that the three band texture ratios yielded the highest accuracy (R^2 = 0.88 and RMSE = 54.54 kg m^{-2}) compared to two texture (R^2 = 0.85 and RMSE = 60.65 kg m^{-2}) and single texture band combinations (R^2 = 0.64 and RMSE = 94.13 kg m^{-2}). A frequency analysis was also run to determine which bands appeared more frequently in the selected texture band models. The frequency analysis revealed that both the red and green bands appeared more frequently on the selected texture band variables, indicating that they were more sensitive to the variation of AGB in our study area. The results showed high variation in AGB within the Buffelsdraai reforestation site, especially due to varying tree plantation phases as well as topography. In essence, the study demonstrated the possibility of image texture combinations computed from the SPOT-6 image in estimating AGB.

Keywords: SPOT-6 imagery, Reforestation, Biomass estimation, Image processing Techniques, Texture combinations

2.1 Introduction

Estimating forest above-ground biomass (AGB) distribution across a landscape provides invaluable information for ecological modelling (Dube & Mutanga, 2015a). This enables ecologists to understand the primary productive capacity of a particular landscape (Gower et al., 1996). Forests play a significant role in carbon cycling as they are responsible for carbon sequestration (Pan et al., 2011; Bastin et al., 2014; Sousa et al., 2015). Therefore, an understanding of forest AGB plays a critical role in climate change modelling and carbon accounting. The utility of forest AGB models for carbon estimations stems from the fact that half of the overall forest AGB is made up of carbon (Nowak & Crane, 2002). Hence, it is important to establish accurate and timely methods of computing forest AGB in natural forests. Conventional methods for estimating AGB have proven to be subjective, time-consuming, costly or spatially restrictive. Conversely, remote sensing approaches for estimating AGB have proven to be objective, inexpensive and spatially explicit (Sarker & Nichol, 2011; Lottering & Mutanga, 2012; Barbosa et al., 2014a; Bastin et al., 2014; Dube et al., 2014; Dube & Mutanga, 2015a).

The advent of optical remote sensing has made it possible to provide estimates of forest AGB that are timely, reasonably accurate and allow for repetitive coverage (Mather & Koch, 2011a; Lottering & Mutanga, 2012; Dube et al., 2014). Previous studies focusing on remote sensing of forest AGB have primarily used spectral reflectance of individual bands and vegetation indices with a reasonable degree of accuracy, however with some limitations (Anderson et al., 1993; Dube et al., 2014; Dube & Mutanga, 2015a). For example, studies have indicated that the relationship between forest AGB and spectral vegetation indices is negatively affected by rapid vegetation growth rates and canopy shadow (Lu et al., 2002; Lu et al., 2005; Mather & Koch, 2011a). As a result, several studies have geared towards the use of image texture analysis, which focuses more on the spatial distribution of grey tone levels within an image (Lu, 2005; Castillo-Santiago et al., 2010; Eckert, 2012).

Texture analysis attempts to compute the degree of image tone variation in high spatial resolution imagery, using statistical parameters to measure the spatial distribution of grey-tones within pixels falling in a particular window size (Wulder et al., 1998). In essence, image texture measures the local variance of grey-tone and is therefore largely dependent on the scale (Haralick et al., 1973). The local variance of grey tone in forested areas is a function of changes

in species type, the level of crown closure, and stem density (Franklin et al., 2001). Studies have confirmed that texture analyses are capable of improving forest AGB detection when compared to spectral vegetation indices (Lu & Batistella, 2005; Eckert, 2012). For example, Lu (2005) concluded that raw texture bands outcompeted raw spectral bands in estimating forest biomass. Similarly, Eckert (2012) improved the overall forest AGB estimation using texture bands when compared to raw spectral bands.

The high performance of image texture in predicting forest AGB is attributed to its ability to simplify the forest canopy structure by smoothing the canopy surfaces using statistical texture variables (Wulder et al., 1998; Franklin et al., 2001). However, research has proven that raw texture bands are incapable of eliminating topographic effects on reflected radiance and errors associated with sensor angle and radiance from sunlight (Sarker & Nichol, 2011). Therefore, studies have recently made new avenues in forest AGB estimation by introducing texture band combinations. The use of texture band combinations has achieved a high degree of success in estimating forest AGB, especially when compared to raw texture bands and vegetation indices (Nichol & Sarker, 2011; Sarker & Nichol, 2011; Dube & Mutanga, 2015b). The current study extends the work of Nichol and Sarker (2011), Sarker and Nichol (2011), and Dube and Mutanga (2015b), by introducing a new three band texture combination method. This novel method combines the strengths of image texture and band ratios, thus could potentially improve their ability to detect forest AGB. Therefore, in this study we compared the accuracy of the three band texture ratios against the commonly used two band texture ratios and raw texture bands computed from SPOT-6 pan-sharpened imagery, using random forest and multiple linear regressions (MLR). Since texture is largely dependent on scale, we used the minimum variance technique to establish optimum moving windows to detect forest AGB at various stages of succession. Finally, we compared the capability of the SPOT-6 pan-sharpened image to predict AGB of trees planted at different stages of succession.

2.2 Methods and Material

2.2.1 Study area

This study was conducted in the Buffelsdraai Landfill Site Community Reforestation Project, which is located roughly 25 km North of Durban in KwaZulu-Natal, South Africa (see Figure 2.1). The reforestation project is approximately 520.60 km^2 in size and is situated along the buffer zone of the Buffelsdraai landfill site, which is owned by the eThekwini Municipality. The topography of the area is highly undulating with steep hills ranging from 200 m to 325 m above sea level. According to Mucina and Rutherford (2006a), the area is located along the KwaZulu-Natal Coastal Belt, which is listed as endangered in the recent vegetation map of South Africa due to fragmentation. The area is predominantly utilised as a landfill site, and the buffer-zone was previously used for sugarcane and communal cattle grazing before it was reforested. Over 50 tree species have been planted in this area from 2009 to 2015 at various stages (i.e. 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014 and 2014-2015). Summer rainfall dominates this area as it falls along the Indian Ocean, which experiences high annual rainfall of approximately 766 mm/year. The midday average temperature is 22.2^oC and 27.5^oC in winter and summer, respectively. The geology of the area is dominated by the Dwyka Tillite. The poorly drained Glenrosa soil form dominates the study area with Shortlands and Swartland soil forms occupying the majority of the remaining area.

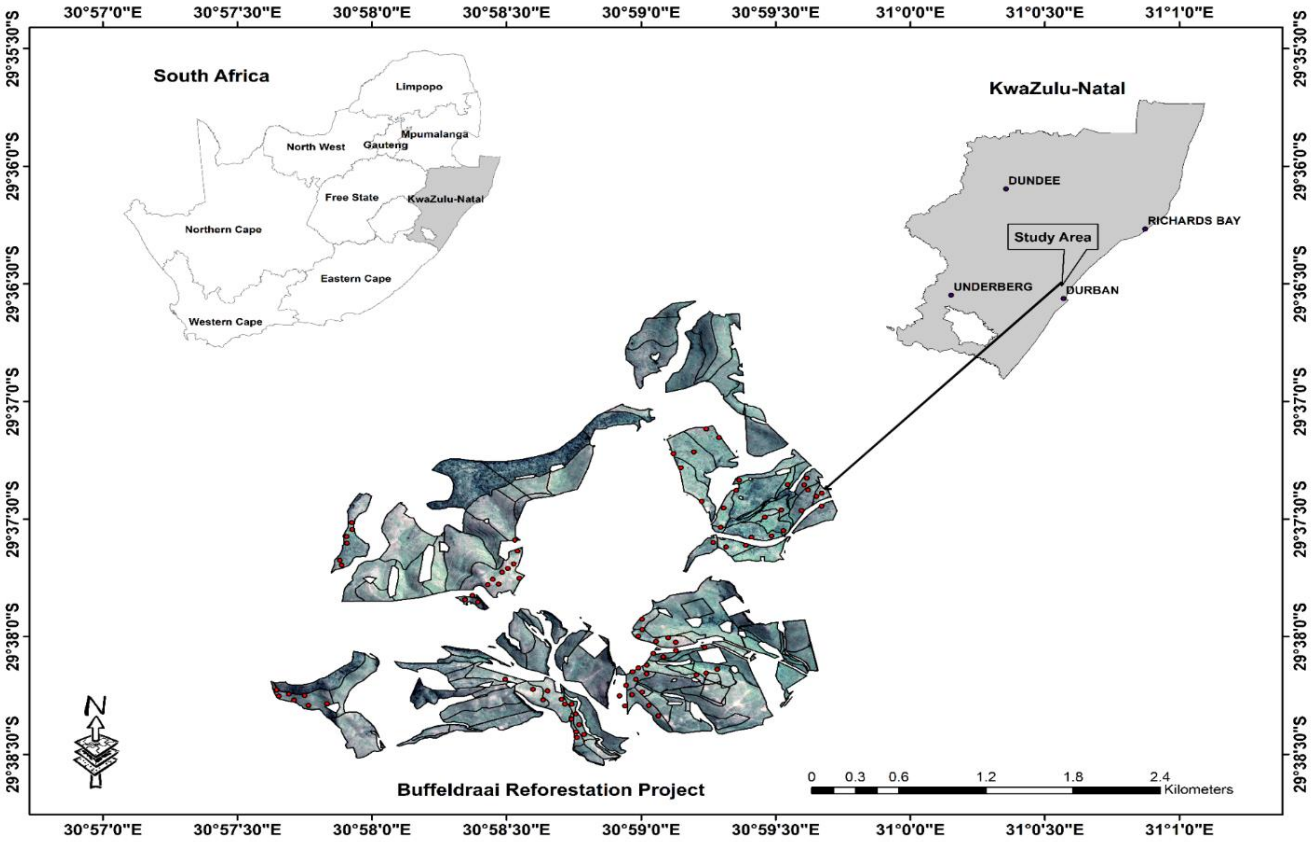

Figure 2.1: Location of study site and the SPOT-6 pan-sharpened image of the Buffelsdraai Reforestation site covering the study area

2.2.2 Image Acquisition and Pre-processing

The SPOT-6 pan-sharpened image was used due to its high spatial resolution of 1.5 m. This image consists of four bands ranging from 0.450-0.890 µm (see Table 2.1), covering the visible and near infrared region of the spectrum. The image was freely acquired from the South African Space Agency (SANSA) on the 11th of April 2015, at 07:44:38 am. The images were Ortho-projected courtesy of SANSA.

Table 2.1: The spectral bands and spatial resolutions of the SPOT-6 pan-sharpened image

Band Number	Spectral Color	Range (µm)	Spatial Resolution (m)
Band 1	Blue	0.450-0.520	1.5
Band 2	Green	0.530-0.590	1.5
Band 3	Red	0.625-0.695	1.5
Band 4	Near Infrared	0.760-0.890	1.5

The images were mosaicked using ERDAS imagine mosaic-pro tool for further processing. The SPOT-6 image was then atmospherically corrected to the top of the canopy reflectance using the algorithm Fast Line-of-Sight Atmospheric Spectral Hypercubes (FLAASH) in ENVI 4.7 software. In order to assess the full potential of the 1.5 m spatial resolution of the SPOT-6 panchromatic image, the image was pan-sharpened using the Intensity-Hue-Saturation (IHS) method. The IHS method was chosen based on the findings of Strait et al. (2008) who discovered that this method preserves the spectral resolution of remotely sensed imagery and improves the spatial resolution of the image. This makes the pan-sharpened image suitable for spatial analysis such as texture indices, which are in essence scale dependent.

2.2.3 Field Measurements

The purpose of this study was to establish a monitoring system for AGB in re-afforested ecosystems at different temporal scales. Therefore, the site was divided into different plantation phases. It is important to take into consideration the complex vegetation structure that exists within the study site, which is composed of trees and grasses that grow at the same heights. For the purpose of this study, only the plantation phases that were 3 years and older were considered, ranging from 2009-2010, 2010-2011 to 2011-2012. These age groups had enough canopy cover to minimise soil background reflectance. Field data collection was conducted in September 2015; the same year the image was acquired.

A set of ninety 35 X 35 m random plots were generated in ArcGIS 10.3, covering the Buffelsdraai planted sites. These random plots were then located infield using a Trimble Geo

7x GPS with sub-metre accuracy. In each plot, all the trees were measured for diameter at ankle height and total tree height using a caliper and ranging rod, respectively.

2.2.4 Wood density measurements

Estimating AGB in natural forest stands can be complex, due to finding allometric equations for all the tree species found in that forest and developed in the right bioclimatic zones. It is for this reason, Chave et al. (2006) emphasised the use of general equations that rely on wood density measurements. Wood density has been found to vary with species type and thus can give reasonable estimates of AGB for various tree species where allometric equations are unavailable. By definition, wood density is obtained by dividing the oven-dry mass of a sample with its green-volume. The study site has over 50 tree species planted, thus obtaining estimates of wood density for all these species can be tedious and impractical. Therefore, a field survey was conducted and 7 dominant tree species in the study area were selected for wood density measurements. Table 2.2 illustrates the dominant tree species selected for forest AGB calculation.

Table 2.2: Descriptive statistics of measured wood densities for the selected dominant tree species and published wood density values (g/cm^3)

Species Name	Min.	Max.	Mean	Std dev	Published	Reference
Bridelia micrantha	0.36	0.6	0.50	0.08	0.47	(Brower et al., 1997)
Vachelia natalitia	0.55	0.64	0.58	0.03	0.65	(Simpson, 1996)
Erythrina caffra	0.2	0.32	0.24	0.03	0.32	(Van Vuuren et al., 1978)
Acacia caffra	0.68	0.81	0.74	0.07	0.71	(Simpson, 1996)
Trema orientalis	0.31	0.37	0.36	0.04	0.37	(Simpson, 1996)
Syzygium cordatum	0.5	0.58	0.52	0.03	0.59	(Brower et al., 1997)
Trichilia dregeana	0.43	0.54	0.50	0.05	0.51	(Simpson, 1996)

For the purpose of this study, core-samples from adult specimens of the dominant tree species were collected in the field using an incremental borer of known height and diameter. The core samples were then stored in an airtight plastic bag and sealed to maintain constant humidity. The green mass of the tree core was determined using the water displacement method and was

carried out in the laboratory (Chave et al., 2006). The green volume was then recorded as the mass of displaced water (as the density of water is 1g/cm^3). The oven-dry mass of the sample was determined by placing the sample in a well-ventilated oven at 100^{o}C until the mass reaches a constant level (approximately 48-72 hours). A total of 10 core samples for each tree species were collected and their wood densities were calculated. The average wood density of the core samples was then calculated for each tree species to obtain a representative value and were recorded (Table 2.2). An accuracy assessment of measured wood density was performed using wood density values from published literature (Glenday, 2007). Figure 2.2 shows a scatterplot of measured and published wood density values.

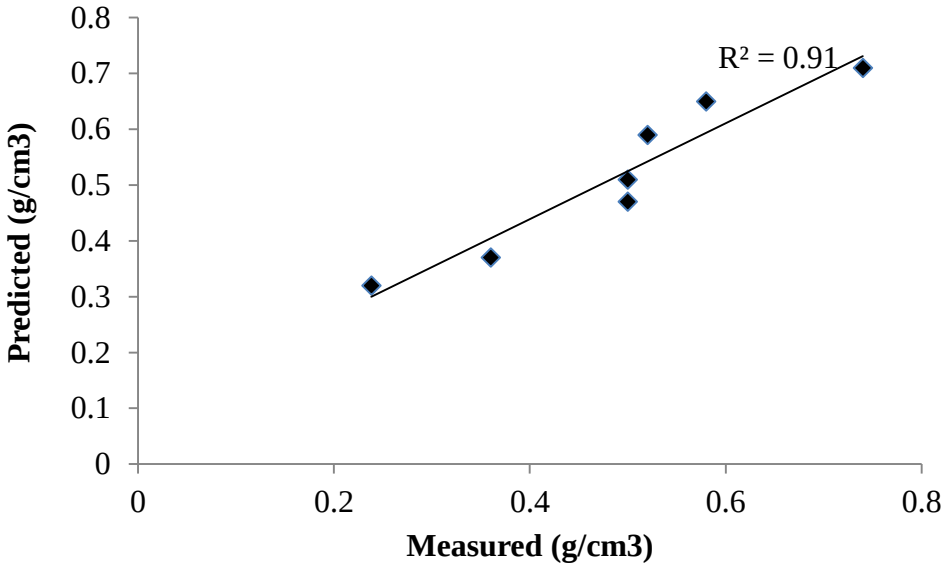

Figure 2.2: The relationship between wood density values obtained from field measurements and wood density values obtained from online wood density database of African dry forests, (Glenday, 2007).

The results indicated a high correlation between published and measured wood density (R^2 = 0.91). This finding affirmed the reliability of published wood density values for use in this study. As a result, measured and published wood density values were subsequently used in our forest AGB calculations.

2.2.5 Field Biomass Calculations

Dry biomass was calculated using the following general allometric equation developed by Chave et al. (2005), for tropical dry forest stands:

$$AGB = 0.112 \times (\rho D^2 H)^{0.916} \qquad (2.1)$$

Where AGB is total above-ground biomass (kg m^{-2}), ρ is wood density (g cm^{-3}), D is diameter at ground level (m) and H is total tree height (m).

The AGB for all tree species within a plot were determined and subsequently averaged to obtain a representative value per plot.

2.2.6 Optimum Window Selection

Since image texture is a function of spatial resolution of an image, we applied the method proposed by Marceau et al. (1994) to select the optimum window size that best captures forest AGB at various stages of development (i.e. 2009-2011 and 2011-2012). The variance is calculated for each window size to determine the level of pixel value variation within that particular window (Peerbhay et al., 2016). This method involves firstly, the selection of classes that best represent the geographic entity under investigation, secondly the resampling of the SPOT-6 pan-sharpened image to produce the images with varying window size, and thirdly the calculation of variance for each window size using the selected representative classes.

2.2.7 Characterising forest AGB representative classes

This study used, age of trees (i.e. 2009-2011 and 2011-2012) and forest AGB as a criteria for determining the optimum window size for forest AGB estimation. Herein, sampled plots that best represent varying forest AGB were identified. Table 2.3 shows the results for the representative classes used to characterise varying levels of forest AGB. Based on field measurements the old trees had higher forest AGB cover compared to young trees.

Table 2.3: The selected representative classes for optimum window size selection.

Succession Period	Biomass (kg m^{-2})	Number of Plots
2009-2011	0-220	14
	221-440	12
	441-680	21
2011-2013	0-115	20
	116-230	18
	231-360	4

2.2.8 Resampling the SPOT-6 pan-sharpened image

This step involves the resampling of the SPOT-6 pan-sharpened image prior to atmospheric correction using odd size windows (*n*x*n*). Varying window sizes were used in this study (i.e. 3x3, 5x5, 7x7, 9x9, 11x11, 13x13, 15x15, and 17x17). The image resampling procedure was carried out using the nearest neighbor resampling technique in ArcMap 10.3. This resulted in 8 images for each spectral band with varying spatial resolutions. Subsequently, the DN values of the pixels were extracted using the zonal statistic tool in ArcMap 10.3. The utility of this method to spatially resample remotely sensed images was recommended by Franklin et al. (1995), with an assumption that pixel values represent an average DN value of a particular area on the ground.

2.2.9 Minimum variance calculation

This step involves the calculation of the minimum variance of all the window sizes for each individual band. This method assumes that a smaller window size only captures a small component of the geographic entity (Marceau et al., 1994). In this case, variance increases as a result of intra forest AGB variation. Therefore, if the window size is larger than the geographic entity under study, more geographic entities are captured by the window size, as a result variance increases. The optimum window size is therefore attained when the geographic entity is homogenous, this is indicated by minimum variance (Lottering & Mutanga, 2016). Studies have confirmed that the selection of optimum window size is vital for improving

vegetation detection and the predictive performance of the models (Ismail et al., 2008; Lottering & Mutanga, 2016; Peerbhay et al., 2016). The variance of pixel values for all the bands of the pan-sharpened image (n = 4) was determined using equation (2.2) and (2.3). Since the variance for band 4 among the different window sizes was similar, only band 1, 2 and 3 were used to select the optimum window size.

$$Variance = \frac{\sum(x_{ij}-M)^2}{n-1} \qquad (2.2)$$

$$Mean = \frac{\sum x_{ij}}{n} \qquad (2.3)$$

Where M is the mean of the digital numbers (DN) within a moving window, x_{ij} denotes the DN values for the pixels and n is the number of pixels within a moving window.

The variance of each window size was plotted as a function of window size, wherein the initial trough of the variance for each band was regarded as the optimum window size.

2.2.10 Image Texture analysis

Texture parameters are commonly used to measure the spatial distribution of image tone variance (Moskal & Franklin, 2001), which can be used to estimate forest AGB. Herein image tone refers to the variation of grey scales of resolution cells in an image (Mather & Koch, 2011a). Variation in image tone can result from changes in stem density, species type, or crown closure (Franklin et al., 2001). In this study, two sets of texture measures were utilised namely: co-occurrence and occurrence measures. With the use of grey-level occurrence measures (GLOM), texture is calculated using the pixel intensities of the histogram within a processing window. This method does not consider the spatial dependency of pixel (St-Louis et al., 2006). The GLOM consists of five filters used to calculate texture, namely; mean, data range, variance, skewness and entropy. For the description of the GLOM filters, refer to Table 2.4.

Table 2.4: Definition and equations of GLOM texture measures

Parameter	Formula	Description
Mean	$Mean = \frac{\sum_k X_k}{k}$	Computes the average values of spectral reflectance at each window (Lottering & Mutanga, 2012).
Data range	$max\{X\} - min\{X\}$	Calculates the difference between the lowest and highest pixel values (St-Louis et al., 2006)
Entropy	$\sum_{i=0}^{M-1} p(i) log_2[p(i)]$	This is a measure of the degree of histogram uniformity (Materka & Strzelecki, 1998).
Skewness	$\mu_{3=\sigma^{-3}} \sum_{i=0}^{M-1} (i - \mu)^3 p(i)$	Measures the level of histogram asymmetry around the mean (Materka & Strzelecki, 1998).
Variance	$\frac{\sum(xij - M)^2}{m - 1}$	Variance determines the variability of the pixel values within a moving window (Materka & Strzelecki, 1998)

On the other hand, the grey-level co-occurrence matrix (GLCM) uses a spatial dependent grey tone matrix to compute texture (Haralick et al., 1973). The GLCM filters include, variance, mean, contrast, homogeneity, correlation, entropy, dissimilarity and second moment. Table 2.5 provides a brief description of GLCM filters.

Table 2.5: Definitions and equations of GLCM texture measures.

Parameter	Formula	Description
Contrast	$\sum_{i,j=0}^{M-1} P_{i,j}(i-j)^2$	Calculates the level of local variation within a window (Yuan et al., 1991)
Correlation	$\sum_{i,j=0}^{M-1} P_{i,j}\left[\frac{(i-\mu_i)(i-\mu_j}{(\sigma_i^2)(\sigma_j^2)}\right]$	Measures the grey-level linear-dependency within an image (Kayitakire et al., 2006)
Dissimilarity	$\sum_{i,j=0}^{M-1} P_{i,j}\lvert i-j\rvert$	Is a measure of the local variation (Rubner et al., 2001).
Homogeneity	$\sum_{i,j=0}^{M-1} \frac{P_{i,j}}{1+(i-j)^2}$	Measures the smoothness of image texture (Tuttle et al., 2006)
Mean	$\mu_i = \sum_{i,j=0}^{M-1} i(P_{i,j})$ $\mu_j = \sum_{i,j=0}^{M-1} j(P_{i,j)}$	Average grey-level in the small neighborhood (Materka & Strzelecki, 1998)
Second Moment	$\sum_{i,j=0}^{M-1} P_{i,j^2}$	Second moment is an indicator of local homogeneity (Yuan et al., 1991)
Variance	$\sigma_i^2 = \sum_{i,j=0}^{M-1} P_{i,j}(i-\mu)^2$ $\sigma_j^2 = \sum_{i,j=0}^{M-1} P_{i,j}(i-\mu_j)^2$	Variability of the spectral response of pixels (Materka & Strzelecki, 1998).
Entropy	$\sum_{i,j=0}^{M-1} P_{i,j}(-\ln P_{i,j})$	A statistical measure of uncertainty (Yuan et al., 1991)

Where *P(i,j)* is the normalised co-occurrence matrix where the sum of (*i,j=0, M-1*)(*P(i,j)*)=1

Both the GLOM and GLCM filters utilise a specified angle and direction to compute texture measures. However, the purpose of this study is to establish the appropriate filter and window size to extract texture for forest AGB estimation, thus one angle of 45º was used. The basis of choosing the 45º angle was founded upon the fact that it has minimal effect on coefficient of

determination (R^2) (Lottering & Mutanga, 2012). The window sizes used to compute texture images were selected based on the method explained above proposed by Marceau et al. (1994). Texture indices were derived from the SPOT-6 pan-sharpened image. The SPOT-6 pan-sharpened texture indices were processed in three steps:

Step1: The single texture parameters of the SPOT-6 pan-sharpened image were tested in predicting forest AGB using a random forest and multiple linear regression.

Step2: The pan-sharpened image was then processed further using two band texture parameters and their accuracy in predicting forest AGB was assessed in random forest and multiple linear regression. All possible two band texture combinations were computed using equation (2.4) to (2.6):

$$\frac{B1}{B2} \quad (2.4)$$

$$B1 - B2 \quad (2.5)$$

$$\frac{B1-B2}{B1+B2} \quad (2.6)$$

Where B1 and B2 are texture parameters.

Step3: The pan-sharpened image was then processed further by combining the bands using three texture parameters and their accuracy in predicting forest AGB was assessed using random forest and multiple linear regression. All possible three band texture combinations were computed using equation (2.7):

$$\frac{B1}{B2 \times B3} \quad (2.7)$$

Where B1, B2 and B3 are texture parameter.

2.2.11 Extracting texture parameters

The field data containing forest AGB plots and their GPS coordinates were used to establish a point map. The point map was superimposed upon the texture index images to establish a region of interest (ROI) map using the central points of the GPS coordinates per plot. The texture values were then extracted using the zonal statistics tool in ArcMap 10.3. The collected texture

dataset was then input into random forest and multiple linear regression to generate prediction biomass models.

2.2.12 Statistical Analysis

The relationship between natural forests aboveground biomass and image texture variables was modelled using RF algorithm and MLR. The advantage of using the RF algorithm is that, it is able to optimise the classification and regression trees (*ntree*) method by combining a large set of decision trees (Breiman, 2001). Each of these trees is constructed using a deterministic algorithm, whereby a random set of variables are selected and a random sample from the training dataset is selected (James et al., 2013). In addition, the random forest algorithm is able to improve model performance by optimising the *mtry* which refers to the number of different predictors tested at node (the default is $1/3$) and ntree which denotes the number of trees grown based on bootstrapped of observation (Mutanga et al., 2012). The machine learning technique was implemented using Python language statistical interface. In Python the random search function was used to optimise the *ntree* and *mtry* parameters, the function selected the best combination of parameters (i.e. *ntree* and *mtry*) based on the lowest root mean square error (RMSE) of the calibration dataset. The *ntree* values were tested in increments of 500 to 2500 and the *mtry* values were tested in increments of 1 to 5, both based on single value intervals (Dube et al., 2014). The results from the RF algorithm were then compared to the MLR to assess its performance.

2.2.13 Relationship between texture indices and AGB

The field data was tested for normality of distribution using the Shapiro-Wilk test to assess whether there were any normality violations. Normality tests are a prerequisite prior to running parametric statistical tests as they assume that the data follows a normal distribution (Mutanga & Skidmore, 2004a). The relationship between forest AGB dataset and image texture parameters were tested using Pearson's correlation. The input texture parameters undergone sequential forward selection prior to Pearson's correlation test, where only significant texture

indices were selected. The texture indices selected were then input into the RF and MLR algorithms to develop forest AGB models.

2.2.14 Model Validation

To evaluate model performance, the dataset (n = 90) was split into 70% training (n = 63) and 30% test (n = 27) dataset and the 5-fold cross validation technique was implemented to robustly test the performance of the algorithms. The training dataset was used to optimise and train the model and the test dataset was used to verify the predictive ability of the model. The performance of each model in estimating biomass was tested using the root mean square error (RMSE), % RMSE and the coefficient of determination (R^2).

$$RMSE = \sqrt{\frac{SSE^2}{n}} \qquad (2.8)$$

$$\%RMSE = \frac{RMSE}{Mean} \times 100 \qquad (2.9)$$

The SSE notation represents the sum of errors of (measured biomass-predicted biomass), n represents the number of predictors involved in the model construction and mean represent the average forest AGB measured in the field.

The model with the lowest RMSE and highest R^2 value was selected and used to predict forest AGB. The model that yielded the highest performance was then used to construct a forest AGB map for the study site in ArcMap 10.3.

2.3 Results

2.3.1 Descriptive statistics

Descriptive statistics of biomass measured in the field for both the plantation phases (i.e. 2009-2011 and 2011-2013) and combined dataset are shown in Table 2.6. The Shapiro-Wilk test revealed that the dataset was normally distributed ($p < 0.05$); subsequently the data was further analysed using parametric tests. The highest average tree biomass of 335.07 kg m^{-2} was obtained from the 2009-2011 plantation phase and the lowest was observed for the 2011-2013 plantation phase which was 193.03 kg m^{-2}. Furthermore, the 2009-2011 plantation phase contained the highest maximum AGB of 670.43 kg m^{-2} compared to the 2011-2013 plantation phase with a maximum AGB of 351.83 kg m^{-2}. These results show a directly proportional relationship between the time of plantation establishment and forest AGB.

Table 2.6: Explanatory statistics of the observed above ground biomass (kg m^{-2}).

Period	*n*	Mean	Std. dev	Min.	Max.	Range
2009-2011	48	335.07	144.65	98.33	670.43	572.09
2011-2013	42	193.03	73.62	39.73	351.83	312.11
Total Data	90	268.79	136.47	39.73	670.43	630.69

2.3.2 Window Size Selection

Figure 2.3 illustrate the variance of pixel values of older trees (i.e. 2009-2011) at various window sizes under varying forest AGB content. The optimum window size was reached when the geographic entity and the window size are equal, which is indicated by minimal variance. As stated, the optimum window size was indicated by the first trough in the variance. The results in Figure 2.3 indicate that for mature trees (i.e. 2009-2011) with moderate to high forest AGB, their optimum window size was 7x7 and for mature trees with lower forest AGB, their window size was 5x5.

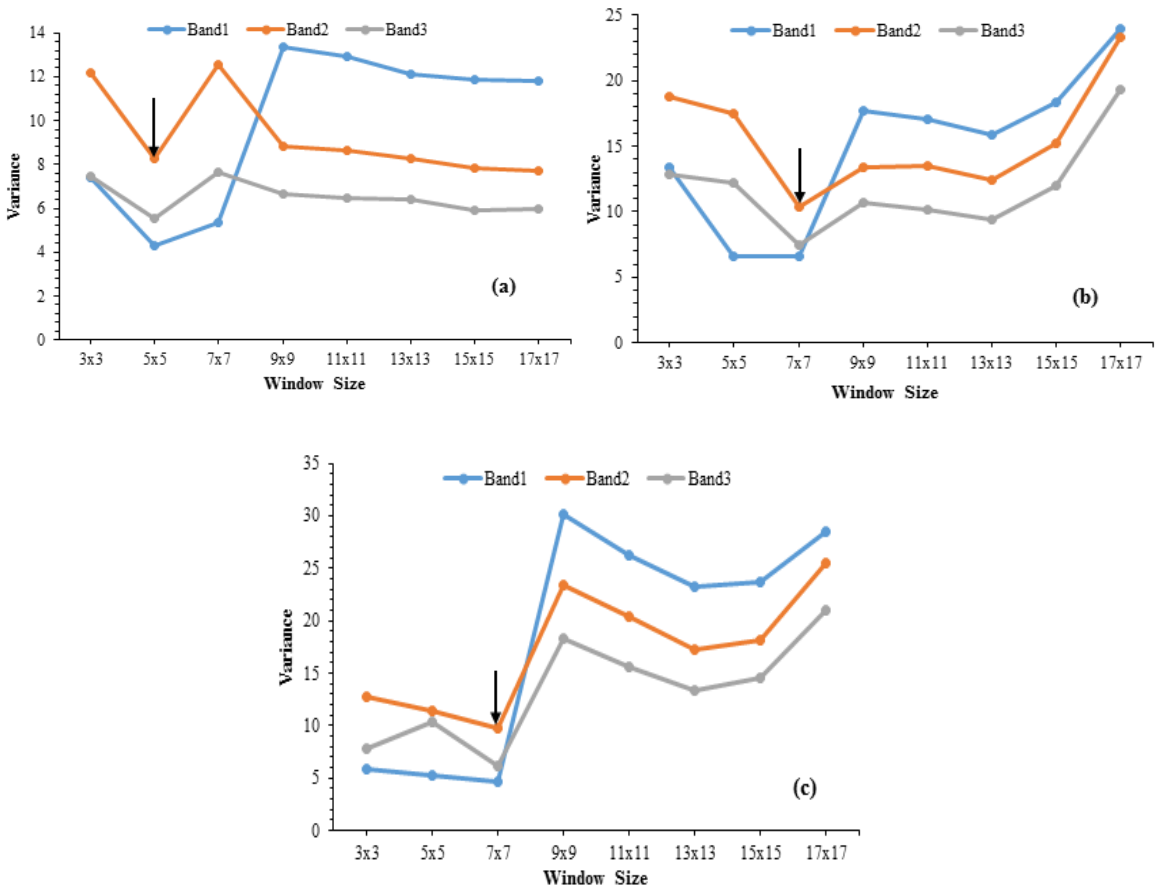

Figure 2.3: Selection of optimum window sizes for the 2009-2011 plantation phase based on minimal variance of pixel values, under varying forest AGB classes. Herein, a, b, and c represent 0-220 (kg m^{-2}), 221-440 (kg m^{-2}) and 441-680 (kg m^{-2}) forest AGB, respectively.

Figure 2.4 illustrates results of variance for the young trees (i.e. 2011-2012) at various window sizes and under varying forest AGB content. The results in Figure 2.4 indicate that the optimum window size for young

trees (i.e. 2011-2012) with moderate to high forest AGB was 5x5 and for young trees with low forest AGB, their window size was 3x3.

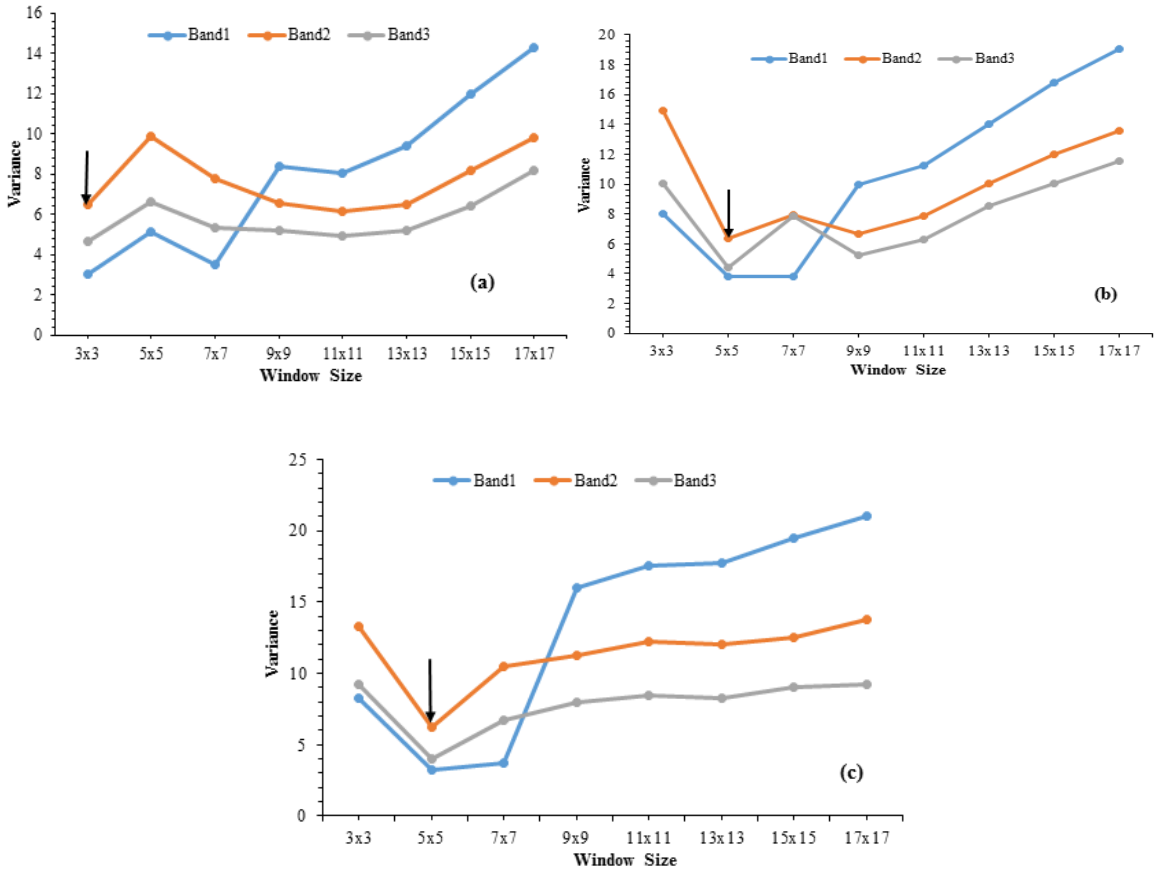

Figure 2.4: Selection optimum window sizes for the 2011-2012 plantation phase based on minimal variance of pixel values, under varying canopy cover percentage classes. Herein a, b and c represent 0-115 (kg m^{-2}), 116-230 (kg m^{-2}), and 231-360 (kg m^{-2}) percentage canopy cover, respectively.

The results in Table 2.7 show a summary of the selected plots at various successional periods and under varying forest AGB. Furthermore, the table shows the selected optimum window sizes for each forest AGB class under investigation. These window sizes were used to compute texture models for forest AGB in our study. Overall, the results indicate that young trees were best detected using small window due to the small amount of their AGB, whereas old trees exhibited homogeneity in spectral variance at slightly higher window sizes in this case 5x5 and 7x7.

Table 2.7: Canopy cover and wood density classes used for optimum window size selection for AGB estimation and their corresponding window sizes.

Succession Period	Biomass (kg m^{-2})	Number of Plots	Window Size
2009-2011	0-220	14	5x5
	221-440	12	7x7
	441-680	21	7x7
2011-2013	0-115	20	3x3
	116-230	18	5x5
	231-360	4	5x5

2.3.3 Correlation Analysis

The Pearson's correlation test was conducted to assess the significance of the relationship between band texture ratios and forest AGB. Table 2.8 shows the relationship between forest AGB and the texture variables that yielded the highest correlation scores. These texture variables were subsequently used in the MLR and RF regression to construct models for predicting forest AGB. The Pearson correlation's test demonstrated that there is a high agreement between tree AGB and the three band texture combinations. This was followed by the two texture band combinations with the Pearson's correlation as high as 0.76. The raw band texture variables yielded the lowest Pearson's correlations with the highest *r* score of 0.55. Moreover, it is also evident that the selected texture variables were developed mostly using the co-occurrence texture parameters computed predominantly from the red band (B3) and near infrared bands (B4).

Table 2.8: Significant three band texture combinations computed from the pan-sharpened image ($p < 0.05$).

Image Processing level	Image texture variable	*r*
Raw band textures	CN_C_5_B4	0.55
	MN_O_7_B4	0.54
	SM_C_3_B2	0.53
Two band texture ratios	$\mathrm{MN\_C\_3\_B4} - \mathrm{MN\_C\_5\_B2}$	0.72
	$\frac{\mathrm{DR\_O\_7\_B3} - \mathrm{SM\_C\_7\_B3}}{\mathrm{DR\_O\_7\_B3} + \mathrm{SM\_C\_7\_B3}}$	0.74
	$\mathrm{EN\_O\_7\_B3}/\mathrm{SM\_C\_7\_B3}$	0.76
Three band texture ratios	$\frac{\mathrm{MN\_C\_3\_B4}}{\mathrm{HM\_C\_7\_B2} \times \mathrm{MN\_C\_3\_B3}}$	0.82
	$\frac{\mathrm{EN\_O\_7\_B3}}{\mathrm{CR\_C\_7\_B3} \times \mathrm{HM\_C\_7\_B4}}$	0.80
	$\frac{\mathrm{DR\_O\_7\_B4}}{\mathrm{SM\_C\_7\_B3} \times \mathrm{VR\_O\_7\_B2}}$	0.79

B1, B2, B3, B4: Band 1, Band 2, Band 3 and Band 4; HM: Homogeneity, EN: Entropy, SM, Second moment, MN: Mean, DR: Data range, CR: Correlation, VR: Variance, CN: Contrast; O: Occurrence, C: Co-occurrence; 3, 5, 7: 3x3, 5x5 and 7x7.

2.3.4 Predicting AGB of the 2009-2011 and 2011-2012 plantation phases

A comparative analysis was conducted to assess the accuracy of the raw image texture bands computed from the SPOT6 pan-sharpened image in predicting biomass of the 2009-2011 and 2011-2013 forest plantation phases. The results in Table 2.9 show the predictive performance of the raw band texture for the divided dataset according to the plantation phases mentioned above and the pooled dataset.

Table 2.9: Accuracy of raw band texture in predicting forest AGB at different plantation phases

		2009-2011		2011-2013		Combined dataset	
Texture variable	**Dataset**	R^2	**RMSE (kgm^{-2})**	R^2	**RMSE (kgm^{-2})**	R^2	**RMSE (kgm^{-2})**
Multiple Linear Regression	Train	0.83	56.37 (16.82%)	0.48	50.93(26.38%)	0.51	91.39(34.00%)
	Test	0.86	46.19 (13.79%)	0.23	55.26(28.63%)	0.67	89.92(33.45%)
	5-FoldCV	0.76	70.04(20.90%)	0.22	80.303(41.60%)	0.29	114.72(42.68%)
Random Forest Regression	Train	0.86	36.68(10.95%)	0.70	27.85(14.43%)	0.79	42.70(15.89%)
	Test	0.70	81.55(24.34%)	0.25	54.17(28.06%)	0.63	94.13(35.02%)
	5-FoldCV	0.73	75.05(22.40%)	0.36	58.09(30.09%)	0.39	105.24(39.15%)

The results in Table 2.9 depicts that the raw texture variables computed from the SPOT6 pan-sharpened image were able to predict forest AGB for the 2009-2011 plantation phase better than the 2011-2013 plantation phase. For instance, an R^2 of 0.86 and 0.70 was obtained for the 2009-2011 plantation phase using the MLR and RF regression, respectively. Compared to the 2011-2013 plantation phase which produced an R^2 of 0.23 and 0.25, using the MLR and RF regression respectively, all the results are based on an independent test dataset. In addition, the models for 2011-2013 plantation phases produced the highest RMSE values when compared to the models for the 2009-2011 plantation phases. For example, the RMSE values for the 2011-2013 plantation phase were 55.26 kgm^{-2} (28.63%) and 54.17 kg m^{-2} (28.06%) using MLR and RF regression, respectively. Whilst for the 2009-2011 plantation phase RMSE values were 46.19 kgm-2 (13.79%) and 81.55 kgm-2 (24.34%), using MLR and RF regression, respectively. The combined dataset produced moderately accurate results, indicating the presence of the low accuracy bearing young trees combined with the high accuracy producing old trees with the highest R^2 of 0.67 and 0.63 for the MLR and RF regression respectively, based on an independent test dataset.

2.3.5 Variable Importance Measures

The importance of derived texture variables in predicting forest AGB was measured using the OOB error rate in RF. The RF algorithm explored the contribution of each texture variable in predicting forest AGB and ranked them according to their importance. Table 2.10 shows the top 20 important texture variables ranked according to the decreasing OOB error rate, which

indicates the deterioration of the model performance when each predictor is permutated. The results in Table 2.10 indicate that the number of texture variables that contributed significantly towards predicting forest AGB was high in raw band textures and decreased when using two and three band texture combinations. Notably, the co-occurrence texture measures appeared more frequently on the high ranking texture variables and band 4 (NIR-band) was frequently selected by the highly important variables (see Table 2.10).

.

Table 2.10: Variable importance measurements of texture models in predicting forest AGB using RF. Higher OOB error signifies higher variable importance

	Raw band textures		Two band textures		Three band texture	
Rank	**Variable**	**OOB Error**	**Variable**	**OOB Error**	**Variable**	**OOB Error**
1	SM_C_3_B2	0.128	MN_C_3_B4 − MN_C_5_B2	0.216	$\frac{\text{MN\_C\_3\_B4}}{\text{CR\_C\_7\_B2} \times \text{MN\_O\_5\_B2}}$	0.222
2	CN_C_5_B4	0.100	MN_C_3_B4 − MN_O_5_B2	0.175	$\frac{\text{MN\_C\_3\_B4}}{\text{MN\_O\_5\_B2} \times \text{CR\_C\_7\_B2}}$	0.209
3	MN_C_3_B4	0.0869	$\text{MN\_O\_5\_B4}/_{\text{CR\_C\_7\_B3}}$	0.083	$\frac{\text{MN\_C\_3\_B4}}{\text{HM\_C\_7\_B2} \times \text{MN\_C\_3\_B3}}$	0.139
4	EN_O_5_B1	0.0761	MN_C_5_B2 − MN_C_3_B4	0.081	$\frac{\text{MN\_C\_3\_B4}}{\text{MN\_C\_3\_B3} \times \text{HM\_C\_7\_B2}}$	0.115
5	MN_C_7_B4	0.0610	MN_C_5_B2 − MN_O_3_B4	0.054	$\frac{\text{MN\_O\_5\_B4}}{\text{MN\_C\_3\_B3} \times \text{HM\_C\_7\_B2}}$	0.0625
6	EN_C_7_B4	0.0527	MN_C_3_B4 − MN_O_3_B2	0.049	$\frac{\text{MN\_C\_3\_B4}}{\text{MN\_O\_5\_B1} \times \text{CR\_C\_7\_B3}}$	0.0615
7	CN_C_7_B4	0.0443	$\text{MN\_C\_3\_B4}/_{\text{CR\_C\_7\_B3}}$	0.0363	$\frac{\text{EN\_O\_7\_B3}}{\text{CR\_C\_7\_B3} \times \text{HM\_C\_7\_B4}}$	0.0523
8	DR_O_5_B4	0.0292	$\text{EN\_O\_7\_B3}/_{\text{SM\_C\_7\_B3}}$	0.0279	$\frac{\text{MN\_C\_7\_B4}}{\text{MN\_C\_3\_B3} \times \text{HM\_C\_7\_B2}}$	0.0238
9	HM_C_5_B4	0.0283	MN_O_3_B2 − MN_O_3_B4	0.0230	$\frac{\text{EN\_O\_7\_B3}}{\text{HM\_C\_7\_B4CR\_C\_7\_B3}}$	0.0199
10	DS_C_7_B4	0.0273	MN_O_3_B2 − MN_C_5_B4	0.0225	$\frac{\text{MN\_O\_5\_B4}}{\text{HM\_C\_7\_B2} \times \text{MN\_C\_3\_B3}}$	0.0140
11	MN_C_5_B3	0.0267	$\frac{\text{MN\_O\_3\_B1} - \text{DR\_O\_7\_B2}}{\text{MN\_O\_3\_B1} + \text{DR\_O\_7\_B2}}$	0.0208	$\frac{\text{MN\_C\_3\_B4}}{\text{CR\_C\_7\_B3} \times \text{MN\_O\_5\_B1}}$	0.0123
12	VR_C_5_B4	0.0242	$\text{EN\_O\_7\_B3}/_{\text{CR\_C\_7\_B2}}$	0.0197	$\frac{\text{MN\_C\_7\_B4}}{\text{HM\_C\_7\_B2} \times \text{MN\_C\_3\_B3}}$	0.0111
13	EN_O_7_B2	0.0206	MN_O_5_B3 − MN_C_3_B4	0.0159	$\frac{\text{MN\_O\_5\_B4}}{\text{MN\_O\_5\_B1} \times \text{SM\_C\_7\_B3}}$	0.0109
14	HM_C_3_B1	0.0204	$\frac{\text{VR\_O\_7\_B2} - \text{MN\_O\_7\_B1}}{\text{VR\_O\_7\_B2} + \text{MN\_O\_7\_B1}}$	0.0156	$\frac{\text{MN\_O\_5\_B4}}{\text{SM\_C\_7\_B3} \times \text{MN\_O\_5\_B1}}$	0.0108
15	DR_O_3_B3	0.0202	MN_C_5_B4 − CR_C_7_B3	0.0148	$\frac{\text{MN\_O\_7\_B4}}{\text{HM\_C\_7\_B2} \times \text{MN\_CO\_3\_B3}}$	0.0101
16	SM_C_5_B3	0.0162	$\frac{\text{SM\_C\_7\_B3} - \text{VR\_O\_7\_B2}}{\text{SM\_C\_7\_B3} + \text{VR\_O\_7\_B2}}$	0.0142	$\frac{\text{MN\_O\_7\_B4}}{\text{MN\_C\_3\_B3} \times \text{HM\_C\_7\_B2}}$	0.0081
17	MN_C_5_B4	0.0152	$\frac{\text{VR\_O\_7\_B2} - \text{HM\_C\_7\_B4}}{\text{VR\_O\_7\_B2} + \text{HM\_C\_7\_B4}}$	0.0128	$\frac{\text{MN\_C\_7\_B4}}{\text{SM\_C\_7\_B3} \times \text{MN\_O\_5\_B1}}$	0.0053
18	DR_O_3_B1	0.0147	$\text{MN\_O\_7\_B4}/_{\text{CR\_C\_7\_B3}}$	0.0125	$\frac{\text{MN\_C\_3\_B4}}{\text{MN\_O\_5\_B1} \times \text{SM\_C\_7\_B3}}$	0.0049
19	MN_O_7_B3	0.0137	$\frac{\text{MN\_O\_7\_B1} - \text{VR\_O\_7\_B2}}{\text{MN\_O\_7\_B1} + \text{VR\_O\_7\_B2}}$	0.0118	$\frac{\text{MN\_C\_3\_B4}}{\text{SM\_C\_7\_B3} \times \text{MN\_O\_5\_B1}}$	0.0036
20	MN_O_3_B4	0.0128	$\text{MN\_O\_7\_B4}/_{\text{SM\_C\_7\_B3}}$	0.0117	$\frac{\text{MN\_C\_7\_B4}}{\text{MN\_O\_5\_B1} \times \text{SM\_C\_7\_B3}}$	0.0026

B1, B2, B3, B4: Band 1, Band 2, Band 3 and Band 4; HM: Homogeneity, EN: Entropy, SM, Second moment, MN: Mean, DR: Data range, CR: Correlation, VR: Variance; O: Occurrence, C: Co-occurrence; 3, 5, 7: 3x3, 5x5, 7x7.

After ranking the texture variables according to their importance, variable selection was conducted to identify the optimum number of variables for predicting forest AGB. Herein the RMSE of the calibration dataset (RMSEC) was used to select the optimum number of variables that yielded the lowest RMSE when predicting forest AGB. The variable selection results in Figure 2.5 depict that for raw band textures, eight variables were selected with the lowest RMSEC of 41.52 kg m^{-2} (15.45% of the mean), for two band texture combinations five variables were selected that produced a RMSEC of 30 kg m^{-2} (11.16% of the mean) and for three band texture combinations seven variables were selected that produced an RMSEC of 31.01 kg m^{-2} (11.53% of the mean). Generally the results indicate that the accuracy of all the texture models increased as the least important variables were progressively removed and finally the use of most important variables yielded the lowest RMSEC. The selected texture variables were used to fit the MLR and the RF algorithm in order to predict forest AGB.

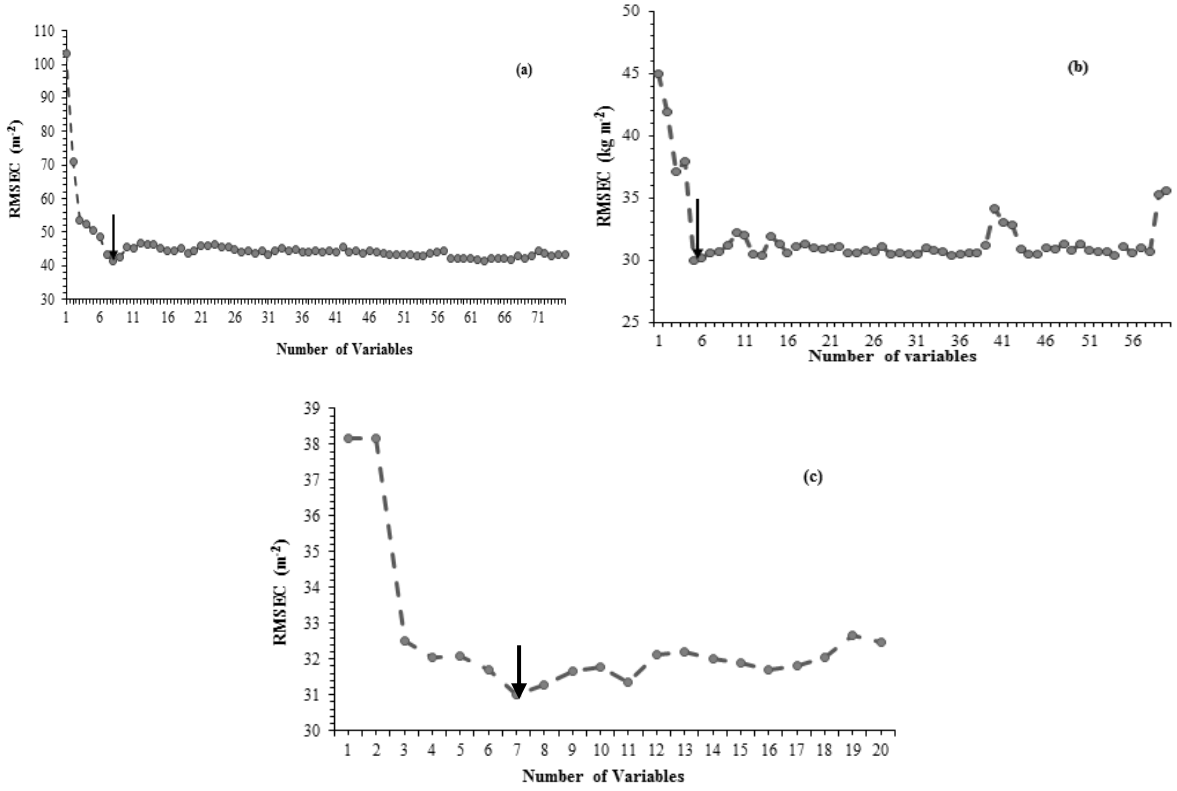

Figure 2.5: Selection optimum number of variables (texture) for predicting forest AGB using backward elimination search function, a, b and c, represent raw band texture, two band texture ratios and three band texture ratios, respectively.

2.3.6 Predictive performance of SPOT6 Texture Combination Models

The purpose of this analysis was to compare the accuracy of raw image texture bands against two-band texture combinations and three band texture combinations in predicting forest AGB. Table 2.11 shows the predictive accuracy results for the texture models. There were significantly high variations in accuracies obtained between the texture models with R^2 values ranging from 0.29 to 0.93. Generally, the RF algorithm outperformed the MLR as expected with an R^2 ranging from 0.53 to 0.93 for RF and 0.29 to 0.85 for MLR.

Table 2.11: Predictive Performance of the texture models

				Train dataset		**Test Dataset**		**10-Fold-CV**	
Texture variable	**Model**	**mtry**	**ntree**	R^2	**RMSE kg m^{-2} (RMSE %)**	R^2	**RMSE k m^{-2} (RMSE %)**	R^2	**RMSE kgm^{-2} (RMSE %)**
Raw band texture	MLR	-	-	0.51	91.40(34.00%)	0.67	89.92(33.45%)	0.29	114.722(42.68%)
	RF	1	893	0.79	42.70(15.89%)	0.64	94.13(35.02%)	0.53	92.82(34.53%)
two band texture ratio	MLR	-	-	0.67	75.24(27.99%)	0.82	67.09(24.96%)	0.63	82.89(30.84%)
	RF	1	860	0.90	40.20(14.96%)	0.85	60.65(22.56%)	0.67	77.56(28.86%)
three band texture ratio	MLR	-	-	0.76	63.69(23.70%)	0.85	59.77(22.24%)	0.75	68.48(25.48%)
	RF	1	939	0.93	32.59(12.12%)	0.88	54.54(20.29%)	0.77	65.66(24.43%)

Most interestingly the three textural processing methods used in this study produced significantly different results using both RF and MLR. The accuracy of texture variables increased from R^2 = 0.64, 0.85 to 0.88, and from 0.53, 0.67 to 0.77 using test dataset and 10-fold-crossvalidation method of single band texture, two band texture combinations and three band texture combinations, respectively. Figure 2.6 illustrates a linear relationship between measured and predicted biomass for all the texture variables. Since RF yielded the highest accuracy when compare to MLR, these scatterplots were developed using the RF algorithm. The graphs display the test dataset and the 10 fold cross-validation.

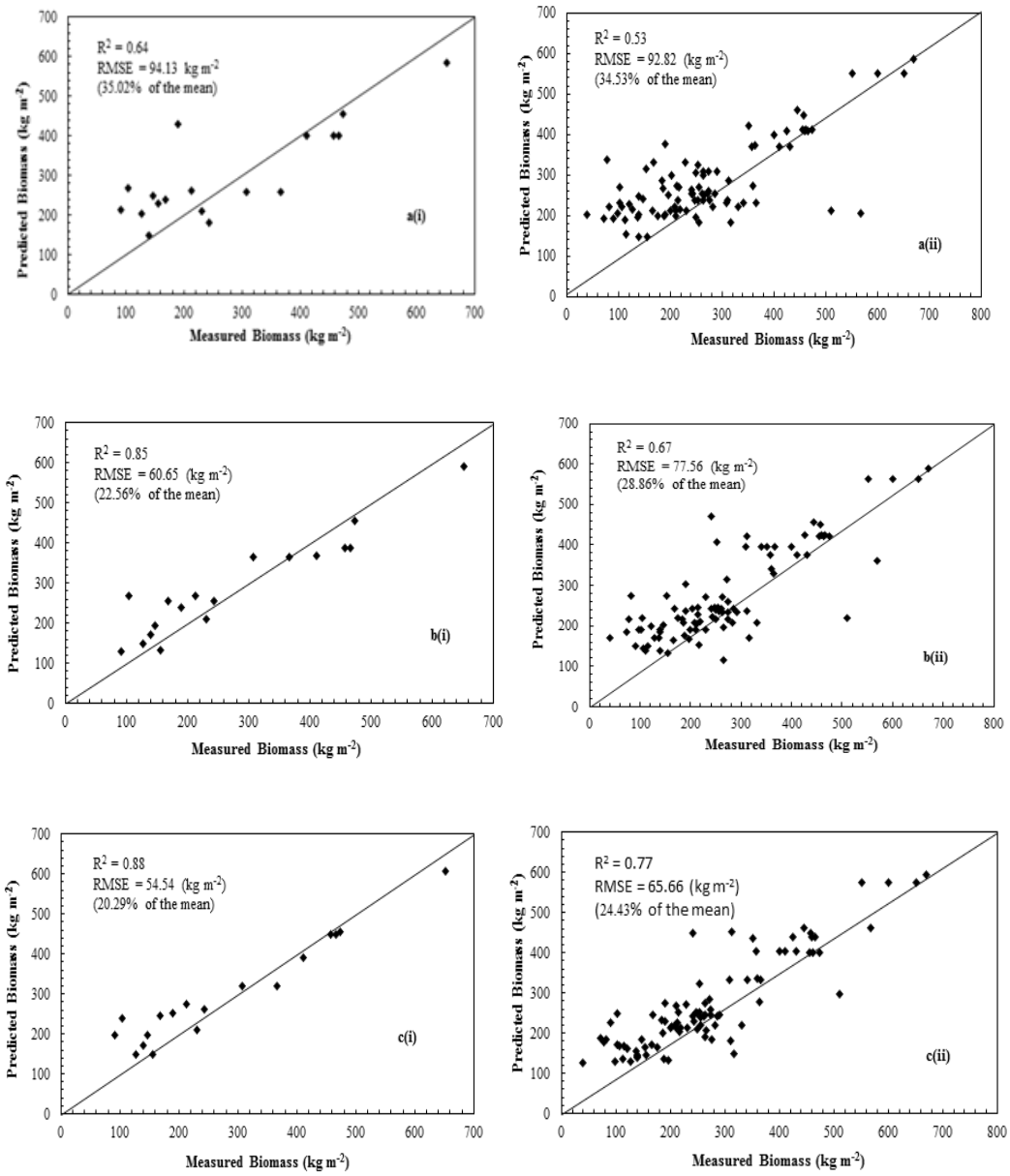

Figure 2.6: Shows that the three texture band models produced the overall highest predicted performance with a R^2 of 0.88 compared to both the two texture band ratios (R^2 = 0.85) and raw band texture (R^2 = 64) based on test dataset of the pan-sharpened image. Herein a, b, and c represent the raw texture bands, two band texture and three band texture combinations, respectively and i and ii represent test dataset and cross-validation dataset, respectively.

Figure 2.6 shows that the three band texture models produced the highest overall predicted performance with an R^2 of 0.88 and 0.77 compared to both the two band texture ratios (R^2 = 0.85 and 0.67) and raw texture bands (R^2 = 0.64 and 0.53) based on an independent test dataset and 10 fold cross-validation, respectively. The results showed an improved accuracy in the estimation of forest AGB using band texture ratios. The best selected three band texture ratios

were chosen for creating a predictive map showing forest AGB over the entire study area (see Table 2.12).

2.3.7 Frequency Analysis

Figure 2.7 shows a summary of the frequently occurring bands and texture measures using all the texture models (i.e. raw texture bands, two and three band texture combinations).

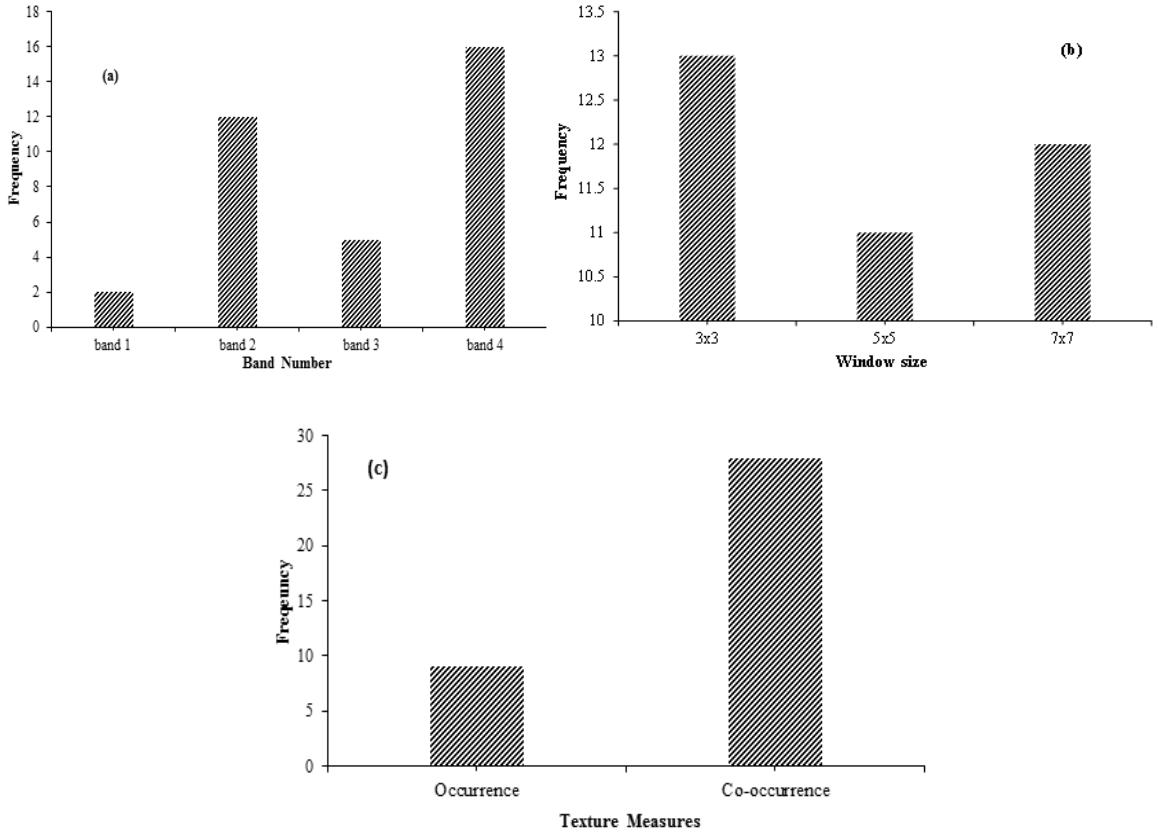

Figure 2.7: The frequencies of a: SPOT6 bands, b: window size, and c: texture measure in the selected models of single texture bands, 2 texture band ratios and 3 texture band ratios for the pan-sharpened image.

Results in Figure 2.7a shows that texture parameters computed from band 4 (NIR band) and band 2 (green band) contain the majority of forest AGB information. The window size that dominated the texture models was the 3x3 window size followed by the 7x7 window size (Figure 2.7b). In addition, Figure 2.7c shows that the co-occurrence of texture measures was

predominantly selected for model development compared to the occurrence texture measures. Table 2.12 shows the selected variables for all the pan-sharpened image texture models.

Table 2.12: Variables that were selected for constructing texture models using forward selection in RF regression.

Raw bands	Two band texture ratio	Three band texture ratio
SM_C_3_B2	$MN\_C\_3\_B4 - MN\_C\_5\_B2$	$\frac{MN\_C\_3\_B4}{CR\_C\_7\_B2 \times MN\_O\_5\_B2}$
CN_C_5_B4	$MN\_C\_3\_B4 - MN\_O\_5\_B2$	$\frac{MN\_C\_3\_B4}{MN\_O\_5\_B2 \times CR\_C\_7\_B2}$
MN_C_3_B4	$MN\_O\_5\_B4 / CR\_C\_7\_B3$	$\frac{MN\_C\_3\_B4}{HM\_C\_7\_B2 \times MN\_C\_3\_B3}$
EN_O_5_B1	$MN\_C\_5\_B2 - MN\_C\_3\_B4$	$\frac{MN\_C\_3\_B4}{MN\_C\_3\_B3 \times HM\_C\_7\_B2}$
MN_C_7_B4	$MN\_C\_5\_B2 - MN\_O\_3\_B4$	$\frac{MN\_O\_5\_B4}{MN\_C\_3\_B3 \times HM\_C\_7\_B2}$
EN_C_7_B4		$\frac{MN\_C\_3\_B4}{MN\_O\_5\_B1 \times CR\_C\_7\_B3}$
CN_C_7_B4		$\frac{EN\_O\_7\_B3}{CR\_C\_7\_B3 \times HM\_C\_7\_B4}$
DR_O_5_B4		

B1, B2, B3, B4: Band 1, Band 2, Band 3 and Band 4; HM: Homogeneity, EN: Entropy, SM, Second moment, MN: Mean, DR: Data range, CR: Correlation, VR: Variance; O: Occurrence, C: Co-occurrence; 3, 5, 7: 3x3, 5x5, 7x7.

The variables were selected using the RF selection method, with an optimum number. Through the RF selection process, eight variables were selected for raw texture bands, five variables were selected for two band texture combinations and seven variables were selected for the three band texture combination model.

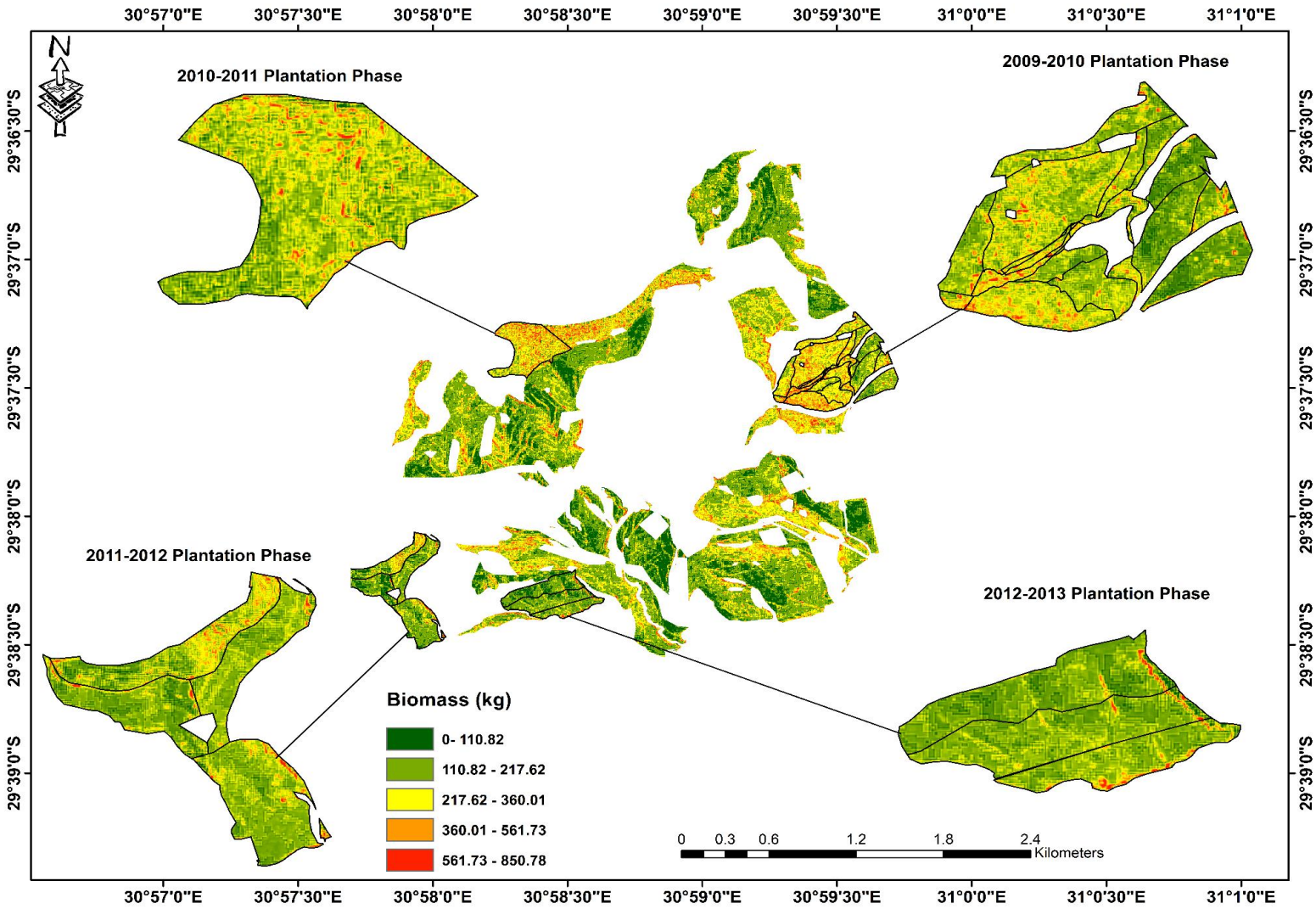

Figure 2.8: Above ground biomass map derived from the best performing three texture band combinations computed from the pan-sharpened image for the 2009-2013 plantation period.

Figure 2.8 shows the map of AGB produced in Python language using the three band texture combinations computed from the SPOT-6 pan-sharpened image that yielded the highest accuracy R^2 = 0.88, RMSE = 54.54 (20.29% of the mean) based on an independent test dataset. Furthermore, the map also shows that lower forest AGB values are located on younger tree plantations i.e. 2011-2012 and 2012-2013. Figure 2.9 shows the average forest AGB in kgm^{-2} for five years from 2009 to 2013 computed from the predicted map of biomass in Figure 2.8. The results in Figure 2.9 indicate that there is a significant successive variation in AGB with the highest mean AGB (325.83 kgm^{-2}) occurring in the 2009-2010 succession. Furthermore, the youngest succession 2012-2013 exhibited the lowest mean AGB (146.78 kgm^{-2}).

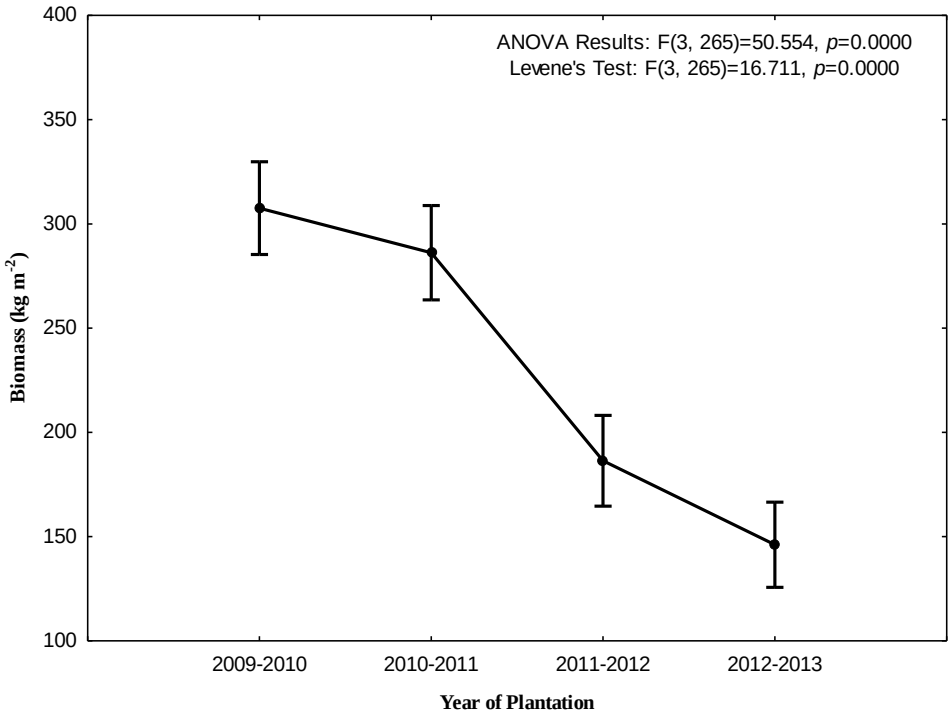

Figure 2.9: Mean AGB for the five year successional dates from 2009 to 2013 computed from the predicted above-ground biomass map shown in Figure 2.8.

An ANOVA was then conducted to assess the significance of the mean differences among the successional dates. The ANOVA results indicated that forest AGB varies significantly among the successional periods with F (3, 265) = 50.554, and $p < 0.001$ and the Levene's test of homogeneity of variance revealed that data did not meet the homogeneity of variance as a result Games Howell post hoc analysis was conducted. The Games Howell post hoc analysis revealed that the significant variation in mean forest AGB was observed in successional periods that are at least two years

apart. For example, trees planted in the 2009-2010 succession contained significantly higher AGB compared to trees planted in 2011-2012 and 2012-2013 with p-values of 0,001713 and 0,000010, respectively. Furthermore, trees planted in 2010-2011 contained significantly higher mean forest AGB compared to trees planted in 2011-2012 and 2012-2013 with p-values of 0,008628, and 0,000099, respectively.

2.4 Discussion

To establish a model for estimating forest AGB, we tested the performance of three image processing techniques. The use of texture combinations in this study proved that texture parameters are capable of estimating tree biomass more adequately when compared to raw texture bands. Multivariate analysis results for the three image processing techniques showed that single texture bands produced the lowest overall accuracy (R^2 = 0.64 and RMSE = 94.13 kg m^{-2}) followed by some improvements using the two band texture combination (R^2 = 0.85 and RMSE = 60.65 kg m^{-2}). However, the highest overall accuracy was obtained using three band texture combination (R^2 = 0.88 and RMSE = 54.54 kg m^{-2}).

The high performance of texture measures in predicting forest AGB was anticipated, because previous research has reported significantly higher correlations between forest AGB and texture parameters (Lu & Batistella, 2005; Eckert, 2012). Our results showed that co-occurrence texture parameters appeared frequently in the models, thus indicating that they contained the majority of forest AGB information. These results are similar to other studies that have shown that co-occurrence texture measures contain the highest vegetation information compared to occurrence texture measures (Yuan et al., 1991; Franklin et al., 2000; Lottering & Mutanga, 2012). Comparatively, our results yielded reasonable accuracies, considering the fact that we used freely available SPOT-6 imagery as opposed to using high-resolution satellite images that are expensive. This could be due to the fact that we optimised the window sizes, which improved the correspondence between percentage canopy cover and the remotely sensed data. In this study, higher accuracies were obtained using texture band combinations, computed from two and three texture bands.

These results coincide with the findings of Sarker and Nichol (2011), who achieved a 12% increase in model performance using simple ratio computed from 2 texture bands of AVNIR image (R^2 = 0.88), compared to single texture band models (R^2 = 0.76) for estimating forest AGB. Using Landsat-8, Dube and Mutanga (2015b), also found a high performance of texture band ratios (R^2 = 0.53) compared to raw texture bands (R^2 = 0.51) in predicting AGB of various forest plantation species. Of interest in this study, were significant improvements obtained from the three band texture combinations, which have not been previously reported in forest AGB estimations.

The high performance of the three band texture combinations can be attributed to 1) combining texture analysis with 2) band ratios and 3) the high spatial resolution of the pan-sharpened image. Texture parameters yielded good results due to their capability to simplify the tree canopy structure into homogenous pixels, by measuring the spatial distribution of image tone within a moving window using statistics (Wulder et al., 1998). We observed that our study area contained a high variation of canopy structure that resulted from a mixture of old trees with dense canopy structure and young trees with sparse canopy structure. According to Bastin et al. (2014), the canopy structure contributes to variation in grey-tone levels. Dense canopy structures produce coarse variation in grey-tone levels whereas, sparse canopy structures produce fine variations of grey-tones. This makes texture best suited for estimating forest AGB in this study as opposed to using band ratios computed from spectral reflectance, that lack the capability of simplifying complex canopy structures. Moreover, the use of three-band texture ratios improved the accuracy of forest AGB estimation by further enhancing the capability of band ratios, to minimise errors associated with sun illumination and topographic variations on upwelling radiance (Mather & Koch, 2011a; Nichol & Sarker, 2011; Dube & Mutanga, 2015b). Therefore, the combination of the three band ratio technique with texture analysis produced a model that is able to simplify complex canopy structures and thoroughly reduce topographic and sun illumination errors on upwelling radiance.

We also believe that the high performance of the three band texture combination is as a result of the integration of optimum window size selection and the high spatial resolution of the pan-sharpened image. This is due to the fact that image texture contains information about the ‘spatial’ distribution of tonal variations within a band (Haralick et al., 1973), and image tone variation is directly proportional to the resolution of the pixels (Lottering & Mutanga, 2012).

Therefore, selecting the optimum window sizes for forest AGB estimation increased the capability of texture models to detect and predict variation in forest AGB. These results are further affirmed by Lottering and Mutanga (2016) who found that selecting optimum spatial resolution for predicting *Gonipterus scutellatus* defoliation levels yielded higher accuracies compared to the normal spatial resolution of the Worldview-2 pan-sharpened image. Findings in this study emphasise the positive relationship between image texture and spatial resolution of an image.

Regardless of the high performance of the three texture band combinations in predicting forest AGB, we noted that accuracies achieved in this study were not higher than previously reported studies that used band texture combinations. For example, in this study, the highest accuracy obtained using three texture band combinations was $R^2 = 0.88$ and in Sarker and Nichol (2011) they obtained R^2 as high as 0.88 using a simple ratio (two band texture combinations) of texture parameters computed from ALOS AVNIR-2 image. We argue that results obtained in this study were reasonably high considering the complexity of the vegetation structure, composition and health conditions in our study area. Based on results obtained in this study, the presence of young trees planted in 2011-2013 affected the performance of the models. The results illustrated in Figure 5 showed that AGB of old trees planted in 2009-2011 yielded higher correlations with texture parameters $R^2 = 0.70$ compared to AGB prediction of young trees $R^2 = 0.25$. These results are explained better by Bingham and Sawyer (1992), who suggested that young trees have less dense canopy structure compared to old trees, therefore the discrete variation of grey-tone for individual young trees is not easily distinguishable, because more than one tree canopy occupies a single pixel. As a result, young trees contribute to wide variation in grey-tone that cannot be accurately measured using texture parameters, which explain why their forest AGB predictions were lower than old trees.

In addition, we observed that the trees were planted within senescing sugarcane and certain areas were characterised by high tree mortality rate. The presence of senescing vegetation could have affected the spatial distribution of grey tones, as a result inhibiting the formation of distinct forest textural patterns. This would mean that areas that contain high forest AGB with dead sugarcane would be poorly estimated by texture measures. Evidence of the presence of senescing vegetation in this study was verified by the results obtained in Figure 2.7, showing that the NIR band (band 4) followed by the green band (band 2) appeared most frequently on the overall texture

models investigated. The high sensitivity of the green waveband to forest AGB is indicative of the presence of senescing vegetation with high chlorophyll concentration variability. These results coincides with those of Carter (1993b), Gitelson et al. (1996) and Daughtry et al. (2000), who found the high sensitivity of the green waveband to chlorophyll content of senescing vegetation using spectral reflectance of hyperspectral sensors.

Furthermore, frequency analysis for optimal window size in our study indicated that the 3x3 window size was the most suitable for predicting AGB, followed by 7x7 window size. These results indicate that variability of AGB was best captured at higher resolution (smaller window size) where vegetation was less heterogeneous. We attribute these results to the high heterogeneity of the natural forest stand, consisting of shrubs such as *Chromolaena ordata* and tall grasses that could have affected the spatial distribution of grey tone of pixels in close proximity. These results are in concordance with those of Dye et al. (2012), who was able to map *Pinus patula* forest species better using variance texture measures computed from 3x3 window sizes. Findings from Dye et al. (2012), suggests that small window sizes contain detailed textural information of individual trees, whereas large window sizes contain texture information of the entire forest stands.

In summary, results of this study proved that the SPOT-6 pan-sharpened image is able to estimate forest AGB on the basis that effective image processing techniques are utilised. The utility of texture parameters in this study yielded promising results, however, outstanding results were achieved using three band texture combinations. Findings from this study show that the three band texture combination technique offers new opportunities for improving estimation of forest AGB, in areas with limited availability of very high spatial and spectral resolution imagery, such as Worldview-2 and hyperspectral sensors, and in areas with highly complex vegetation composition and stand structure. Based on our findings, other studies should use the 3x3, 5x5 and 7x7 window size for mapping forest AGB, however, they should test different texture combinations to suit their particular studies. Moreover, the results obtained in this study provided insight to the invaluable contribution of the Buffelsdraai reforestation program on above-ground biomass accumulation. The results indicated that forest AGB increases with the year of succession, thus illustrating that reforestation has a high potential to meet the objectives of maximising terrestrial carbon storage through forest AGB.

2.5 Conclusion

This study builds from previous research studies looking at estimation of forest AGB using image texture, however a new approach was introduced using three band texture combinations. This study has shown that the SPOT-6 pan-sharpened image is capable of predicting AGB in trees planted at various plantation phases using texture combinations. Furthermore, the study revealed that:

- The SPOT-6 pan-sharpened image was able to predict AGB of older trees planted in 2010-2011 better than younger trees planted in 2011-2013.
- The green band was highly sensitive to AGB variation, thus indicating the presence of senescing vegetation.
- The three band texture combination techniques yielded higher overall accuracy in predicting forest AGB, offering new opportunities for mapping forest AGB.

This new image processing technique has not been tested by researchers and therefore provides a new perspective and approach to mapping forest AGB in complex vegetation structure. In addition, we suggest that future studies should explore this invaluable technique of optimum window selection for texture computation, to enhance the detection and prediction of geographic entities. Overall, the study proved that there are significant benefits of reforestation, especially over a long term period as forest AGB was proven to increase over time.

CHAPTER 3

Evaluating the Effects of Forest Structural Diversity and Topography on Forest Above Ground Biomass using Three Band Texture Combinations and Advanced Machine Learning Algorithms

This chapter is based on: Hlatshwayo, S. T., Mutanga, O., Lottering, R.T, and Peerbhay, K., (In preparation). Evaluating the Effects of Forest Structural Diversity and Topography on Forest Above Ground Biomass using Three Band Texture Combinations Computed from SPOT-6 imagery and Advanced Machine Learning Algorithms: A case study of the Buffelsdraai Community Reforestation Project in KwaZulu-Natal, South Africa.

Abstract

Forest structural diversity and topographic variables play a significant role in determining forest above biomass (AGB) productivity and the flow of ecosystem goods and services in natural forest plantations. Poor selection of tree species in natural forest plantations can result in suppression of other species and these forests plantation can be less productive. The nature of the relationship between species interaction and forest AGB is further complicated by topographic variability. Topographic variables are pivotal in determining the distribution of tree species across the landscape and their growth rates. In this study we quantified the effects of forest structural diversity and topographic variables on forest AGB productivity using remote sensing and GIS models. Three band texture combinations derived from spatially optimised SPOT-6 image were used to derive models for predicting forest structural diversity attributes (such as species richness, species diversity, tree density, diameter diversity and the Gini coefficient). Furthermore, we tested the capability of two advanced machine learning algorithms, random forest and stochastic gradient boosting in predicting forest structural diversity attributes. The topographic variables were modelled using a digital elevation model derived from high resolution contour lines. The results revealed that the random forest algorithm was superior in predicting species diversity with R^2 of 0.88 and RMSE = 0.21 (15.22%), for species richness; R^2 = 0.86 and RMSE = 1.3 (21.35%), for diameter diversity; R^2 = 0.65 and RSME = 0.82 (32.54%) and for tree density R^2 = 0.85 and RMSE = 5.5 (16.52%). Whereas the stochastic gradient boosting algorithm yielded higher accuracies when predicting the Gini coefficient with R^2 of 0.64 and RMSE = 0.13 (28.26%). The results in this study further indicated that both forest structural diversity attributes and topographic variables have significant effects on forest AGB variability. Species diversity, species richness, tree density, slope and elevation yielded a negative relationship with forest AGB productivity. Conversely, diameter diversity, Gini coefficient, solar radiation and topographic wetness produced positive feedback with forest AGB productivity. Notably, diameter diversity, topographic wetness and solar radiation were principal in determining high forest AGB variability. However, species richness and diversity were principal in determining low forest AGB variability. The results in this study provide insight into the effects of spatial planning of randomly mixed natural forest plantations on forest AGB productivity. Furthermore, we advocate for the use of spatially optimised three texture band combinations for predicting and mapping forest structural attributes.

3.1 Introduction

The idea behind mixed forest plantations is arguably the best contemporary approach for increasing above ground biomass (AGB) and carbon sequestration. These forest plantations provide more ecological goods and services that include increasing biodiversity, restoration of ecosystem function and protecting the forest from pests and diseases (Lamb et al., 2005; Felton et al., 2010; Hulvey et al., 2013). Mixed plantations can be used for various objectives including commercial, arboriculture and sustainability needs (Forrester et al., 2005b). The success of mixed forest plantations is heavily dependent on the objectives of the reforestation project. The framework for evaluating the success of natural forest plantations suggest that, success is achieved when highly diverse tree plots have more AGB productivity than plots with very low diversity (Day et al., 2014). The spatial planning and design of mixed forest plantations intended for increasing AGB productivity is commonly a great challenge for ecologists (Erskine et al., 2006).

Optimising AGB productivity in mixed forests requires a clear understanding of tree species interaction and their response to variation in environmental gradients across the landscape (Erskine, 2002). The diversity of species in mixed forest plantations has been positively correlated with forest AGB (Forrester et al., 2005a; Kelty, 2006; Hulvey et al., 2013) and negatively correlated with forest AGB (Watt et al., 2003; Shirima et al., 2015). Tree species identity and size may act simultaneously to influence tree resource acquisition through dominance of the most productive species (selection hypothesis) and niche partitioning (complementarity hypothesis) in space and time (Cardinale et al., 2009b; Shirima et al., 2015). Therefore, diversity of forest species can either result in intense competition which forces niche restriction or reduced competition resulting from dominance of certain tree species (Huston, 1979). Overcoming species competition and dominance of certain species in mixed forests requires manipulation of both species diversity and functional diversity (Petchey & Gaston, 2002; Cardinale et al., 2009a).

Commonly mixtures consisting of a balance between nitrogen fixers and non-nitrogen fixers have been found to be more productive than monoculture forests (Debell et al., 1997; Khanna, 1997; Forrester et al., 2005b; Hulvey et al., 2013). Forrester et al. (2005a), found high forest AGB productivity in mixed plantations of *Eucaylptus globulus* and *Acacia mearnsii* in areas with low nitrogen (N) compared to monocultures of *Eucalyptus globulus*. Nitrogen fixing trees enhances

soil nutrient status by increasing N availability in the soil, thus reducing interspecies competition for soil nutrients where N is a limiting factor (Hoogmoed et al., 2014; Huang et al., 2014). Furthermore, studies have also indicated that selection of tree species with complementary height (Forrester, 2004; Forrester et al., 2005b) and diameter sizes can enhance forest AGB productivity (Mulder et al., 2004; Shirima et al., 2015). Selecting trees with compatible height growth dynamics can assist in avoiding the suppression of shade intolerant plants and to reduce competition for light, a suitable canopy stratification mixture consists of tall, shade intolerant trees and medium height shade tolerant trees. On the other hand, Shirima et al. (2015) found a decrease in species richness and tree density in mixtures with large and small diameter trees, which was attributed to the dominance of the most productive species. Elimination of large size trees and plantations of small size trees with compatible sizes has been found to increase both species diversity and tree density (Bengtsson et al., 1994; Loehle & Donald, 2000; Wright, 2002), therefore contributing to more stable forest AGB increases. However, according to research findings, species diversity and tree size are not the only factors that control forest AGB productivity. Environmental gradients which include; elevation, slope, light and water also come into play in determining forest AGB productivity (Wright, 2002).

Tree species assemblages and community structures of naturally grown forests are responsive to relative environmental gradient change. Tree species diversity and richness tend to decrease with elevation, due to local species adaptation to differences in edaphic and climatic conditions (Woollen et al., 2012). High elevations and steep slopes are characterised by limited soil nutrient availability, shallow soil depths and harsh climates. In response to these conditions, trees tend to be small and short with less diameter variability (Moser et al., 2011), which is due to high interspecies competition for resources. Elevation and slope also control the amount of incoming solar radiation and topographic wetness (Gracia et al., 2007; Saremi et al., 2014; Xu et al., 2015). These variables are responsible for slope drainage, soil water availability, slope temperature and light accessibility for plant growth. Studies have found a positive correlation between forest AGB with solar radiation and topographic wetness (Lin et al., 2012; Wang et al., 2014; Xu et al., 2015). The abundance of light and water in a forested ecosystem is crucial in reducing species competition for resources, therefore forest AGB tend to be high in areas where light and water are unlimited (Gracia et al., 2007). Quantifying the effects of forest structural diversity, species assemblages and

topography on forest AGB is therefore crucial in spatial planning of naturally grown forests with a high mixture of indigenous species.

The effects of forest structural diversity, species assemblages and topography on forest AGB are commonly quantified using forest inventory data measured at the field using conventional methods. These methods include, among others; species identification, height and diameter measurements (Forrester et al., 2005b; Erskine et al., 2006; Gracia et al., 2007). While these conventional methods for obtaining forest inventory data allow for accurate measurements of forest structure and diversity, obtaining this data is time consuming and is spatially restricted to plot level (Ozdemir & Karnieli, 2011; Lottering & Mutanga, 2012). Currently, conventional approaches to measuring forest inventory data are being side-lined in favor of remote sensing approaches (Dube et al., 2014). Remote sensing techniques are gaining popularity in forest inventory measurements, as they provide viable data collection techniques (Adjorlolo & Mutanga, 2013). The data collected using satellite imagery enables for a synoptic view of the earth's surface, at low cost and with minimum effort (Lottering & Mutanga, 2012). Therefore, remotely sensed data have become more favourable for collecting forest inventory data at local and regional levels.

Various remote sensing instruments have been utilised for forest inventory estimation, ranging from moderate resolution passive sensors (e.g. MODIS17, and Landsat) (Anderson et al., 1993; Dube & Mutanga, 2015a) to high resolution passive sensors (e.g. Spot-6, and Worldview-2) (Castillo-Santiago et al., 2010; Ozdemir & Karnieli, 2011; Eckert, 2012) and active sensors (e.g. light detection and ranging (LIDAR) and radar) (Lefsky et al., 2002; Hyde et al., 2006). Passive sensors measure forest structural attributes by recording irradiance reflected by vegetation (Mather & Koch, 2011a), whereas, active sensors such as LIDAR emit light like a pulse to measure ranges between the sensor and earth object. The pulse is responsive to variation in forest structural attributes (canopy architecture and total tree height) (Drake et al., 2002). As a result, LIDAR sensors provide more accurate information for measuring forest structural attributes. The use of LIDAR sensors, however, is limited by its availability and high cost in large areas that require high revisit rates (Ozdemir & Karnieli, 2011). In this regard, investigations tend to focus on the feasibility of using multispectral satellite sensor for forest structural attribute estimations.

Remote sensing studies commonly use vegetation indices derived from the red and near-infrared band (NIR) to estimate forest structural attributes at canopy level using multispectral images

(Anderson et al., 1993; Turner et al., 1999; Ingram et al., 2005; Shamsoddini et al., 2013). These indices are popular in remote sensing studies, due to their capability to eliminate errors associated with sun view angle, soil background and canopy architecture (Mather & Koch, 2011a). However, canopy shadows and high density vegetation results in poor performance of these vegetation indices, which limits their application in densely forested ecosystem (Lu et al., 2002; Mather & Koch, 2011a). As a result, more research is now geared towards the utility of image texture to characterise distinct variation in greyscales of satellite imagery (Dye et al., 2012; Eckert, 2012; Lottering & Mutanga, 2012; Bastin et al., 2014; Dube & Mutanga, 2015b). Image texture is able to detect distinct variation in canopy structure, tree density and tree height, making it suitable for estimating forest structural attributes and species variation (Franklin et al., 2001). Furthermore, texture measures are able to simplify complex canopy structures (Wulder et al., 1998), which are common in mixed forests. However, regardless of the success achieved using image texture, raw band textures are still subject to errors associated with irradiance from the sun, topographic effects on reflected radiance and sensor angle (Sarker & Nichol, 2011; Dube & Mutanga, 2015b).

Recent studies have taken advantage of the properties of vegetation indices and combined them with image texture properties to produce advanced texture band combinations that possess both qualities. These texture band combinations are immune to sun illumination effects, topographic effects and sensor angle similar to vegetation indices while possessing the strength of simplifying complex canopies provided by texture measures (Nichol & Sarker, 2011). This study builds up from Hlatshwayo et al. (Under revision), who used three band texture combinations to estimate forest AGB. Three band texture combinations have not been thoroughly tested in estimating forest structural attributes. Previous studies only used two texture band combinations. The advantage of using three band texture combinations is that they carry more information from texture bands that is otherwise impossible to obtain using restricted two texture bands (Hlatshwayo et al., under revision). In addition, establishing relationships between forest structural attributes can be complex due to existence of nonlinear correlations between texture measures and forest structural attributes (Dube et al., 2014). Fortunately, the development of advanced machine learning algorithms such random forest (RF) by Breiman (2001) and stochastic gradient boosting (SGB) by Friedman (2001), have made it possible to establish relationships between texture measures and forest structural attributes.

In this study, we evaluated the performance of RF and SGB in predicting forest structural attributes of mixed natural forest plantation in Durban KwaZulu-Natal. Both these algorithms have been favoured due to their capability to perform nonparametric statistics to establish correlations between datasets that are not linearly correlated (Dube et al., 2014). RF, in particular, performs well in datasets that have high collinearity, due to its variable importance technique (Mutanga et al., 2012). Predictor variables in RF are split during the tree growing process, such that each succeeding tree has its own important variables different from the previous tree (Freeman et al., 2015). This variable selection technique enables RF to overcome multicollinearity between predictor variables as their importance is split among the trees, making RF models more robust. Furthermore, RF is advantageous in its capability to provide internal error estimates, parameter tuning to provide more accurate models (Prasad et al., 2006; Elith et al., 2008). The SGB algorithm on the other hand has become more popular in regression analysis due to its capability to deal with inaccurate training data, outliers, unbalanced and missing data (Lawrence et al., 2004; De'Ath, 2007). In addition, the SGB is able to strengthen the predictive performance of weak learning algorithms using a stage-wise additive technique (Dube et al., 2014). There are currently contrasting views pertaining to the performance of SGB and RF in estimating forest structural attributes. Certain studies found a better performance of RF over SGB Freeman et al. (2015), while other studies found a better performance of the SGB model over the RF model (Dube et al., 2014) in modelling forests structural attributes using remotely sensed data. The objectives of this study were to assess the effects of forest structural diversity and topographic variables on forest AGB productivity. To achieve this objective we predicted forest structural diversity attributes using the spatially optimised three band texture combinations derived from SPOT-6 pan-sharpened image (1.5 m resolution). The relationship between the forest structural diversity attributes and three band texture combinations was modelled using the SGB and RF algorithms. We also used the digital elevation model to derive the topographic variables. Subsequently, using multivariate statistics we quantified the effects of forest structural diversity and topographic variables on forest AGB productivity.

3.2 Methods and materials

3.2.1 Study area

The Buffelsdraai landfill site community reforestation project is situated in Verulam, which is approximately 25 km north of Durban, KwaZulu-Natal South Africa (see Figure 3.1). The reforested area stretches across the buffer zone of the Landfill site owned by the eThekwini municipality. The study areas receive a mean annual precipitation of 766 mm/year, which primarily falls in summer. The mean annual temperatures of the study area for winter and summer are 27.5^{o}C and 22.5^{o}C, respectively. The topography of the area is generally undulating with elevation ranging from 200 m to 325 m above sea level (Mucina & Rutherford, 2006b). The area is host to approximately 50 indigenous tree species that grow on dry escarpment forests. Out of the 50 tree species 33 tree species were encountered during the field survey for this study.

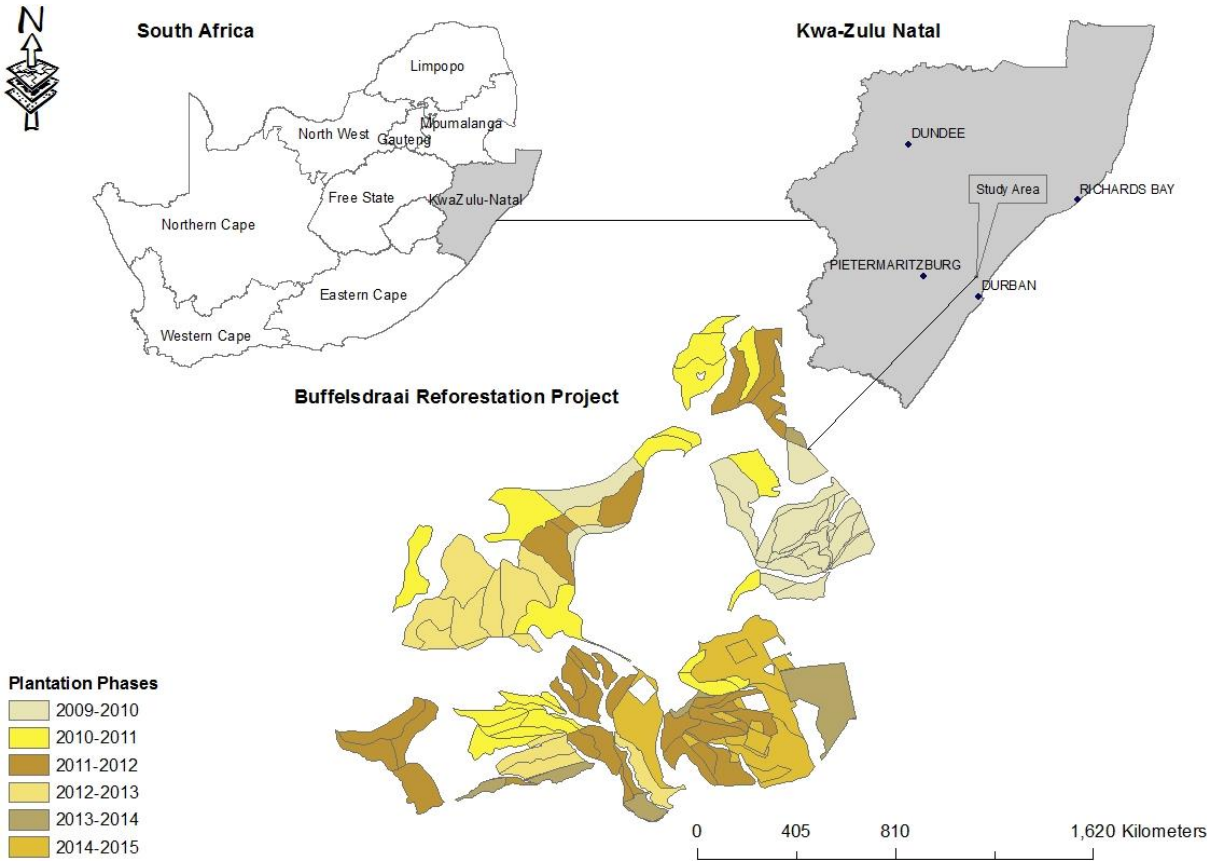

Figure 3.1: Location of study site and the plantation phases in the Buffelsdraai Reforestation site covering the study area.

3.2.2 Image acquisition

A high spatial resolution space-borne SPOT-6 satellite image covering the study area was acquired for 11th of April 2015 from South African Space Agency (SANSA). The SPOT-6 image consists of a panchromatic image with 1.5 m spatial resolution and four multispectral bands with 6 m spatial resolution. The panchromatic image is situated along the visible and near infrared region of the spectrum ranging from 0.450 µm to 0.745 µm and the four multispectral bands range from the blue (B1): 0.450-520 µm, green (B2): 0.530-0.590 µm, red (B3): 0.625-0.695 µm and near-infrared (B4): 0.760-0.890 µm. The images were ortho-projected by SANSA. Furthermore, the Fast Line-of-Sight Atmospheric Analysis of Spectral hypercubes (FLAASH) tool in ENVI 5.2 was used for radiometric corrections of the image, thus converting digital numbers (DN) of the image to absolute radiances using a constant factor determined during satellite launch.

3.2.3 Forest Inventory Data

Forest structure attributes (i.e. height, diameter and stem density) and tree species type were measured from 90 plots (35 m x 35 m) within the study area. The sampling plots were established randomly, using a tool called 'Create Random Samples' in ArcMap 10.3. The points were then transferred into a Trimble Geo 7x GPS with sub-meter accuracy and was used to identify the sampling plots in the field. Only sample plots with tall tree canopies were sampled to avoid spectral mixing from surrounding vegetation. In each plot, total tree height (H) was measured using a ranging rod, the diameter at ankle height (DAH) was measured using a graduated caliper, and the total number of trees and the name of all species occupying the plot were recorded.

3.2.4 Structural and Species Diversity Indices

Both the structural and species diversity indices were computed using data collected from the field. For the structural diversity indices, we assessed the horizontal structural diversity and the Gini coefficient (GC). The horizontal structural diversity was calculated using the standard deviation (σ) of DAH (cm), given by the equation:

$$\sigma_{DAH} = \sqrt{\frac{\sum_{i=1}^{n}(DAH_i - \overline{DAH})}{n-1}} \ldots (3.1)$$

where n is the total number of trees in the plot, i indexes individual tree DAH, and $\overline{DAH}$ represent the mean DAH of all trees in the plot (McRoberts et al., 2008).

In this study, we used the standard deviation of DAH (σ_{DAH}), because σ_{DAH} can ea

sily be calculated from common forest inventory data to produce tree density (TD) and allows for temporal changes to be easily detected (McRoberts et al., 2008; Mura et al., 2015). In addition, the framework of international research has chosen this method as the standard measure of structural diversity (Mura et al., 2015).

The GC is a measure of heterogeneity (Lexerød & Eid, 2006), Weiner and Solbrig (1984) adopted this method to measure plant population size hierarchies. To calculate the GC all trees measured within the plot need to be ranked according to size in an ascending order (Lexerød & Eid, 2006). The GC is given by the equation:

$$GC = \frac{\sum_{t=1}^{n}(2t - n - 1)ba_t}{\sum_{t=1}^{n} ba_t\,(n-1)} \ldots\ldots (3.2)$$

where n is the total number of trees, t indicates tree rank from 1…., n, (ranked according to DAH) and ba_t denotes the basal area of a tree in rank *t*.

The GC theoretically ranges from a minimum of 0 when all the tree sizes are equal, to a maximum of 1 when all trees are equal to 0 except for one individual tree (Ozdemir & Karnieli, 2011).

For forest diversity two widely used indices were adopted in this study; species richness (SR) and species diversity (SD) calculated using the Shannon Index (SI). Species richness was estimated as the total number of observed in the sampling plot (Gotelli & Colwell, 2001). Even though SR is relatively simple to calculate from sampled data, it does not however, take into account the relative abundance of species. For this reason, we also used the Shannon Index (H'), which is a commonly used index for species diversity (McRoberts et al., 2008). The Shannon Index is advantageous for SD measurements as it incorporates both the number of species observed and the relative abundance of each individual species observed in a sample plot. This index is given by the equation:

$$H' = -\sum\left[\left(\frac{n_i}{N}\right) \times \ln\left(\frac{n_i}{N}\right)\right]\ldots. (3.3)$$

Where n_i is the number of individuals per species (the i^{th} species), N denotes the total number of species in the plot and ln is the natural logarithm of a number.

Generally the Shannon Index ranges from 0 to 5.

3.2.5 Topographic Variables

The topographic variables utilised in this study were altitude, slope, topographic wetness index (TWI), and solar radiation. The digital elevation model (DEM) derived from a contour map (5 m resolution), was used to generate these topographic variables. Altitude is one of the important factors in determining forest AGB distribution, as it affects the temperature and atmospheric pressure (Gracia et al., 2007). In addition, altitude affects soil moisture depending on the height of the area above sea level and soil depth. Altitude was derived directly from the DEM in meters and reclassified using the reclassify tool in ArcMap 10.3. Figure 3.2 shows the altitude of the slopes and the reclassified altitude.

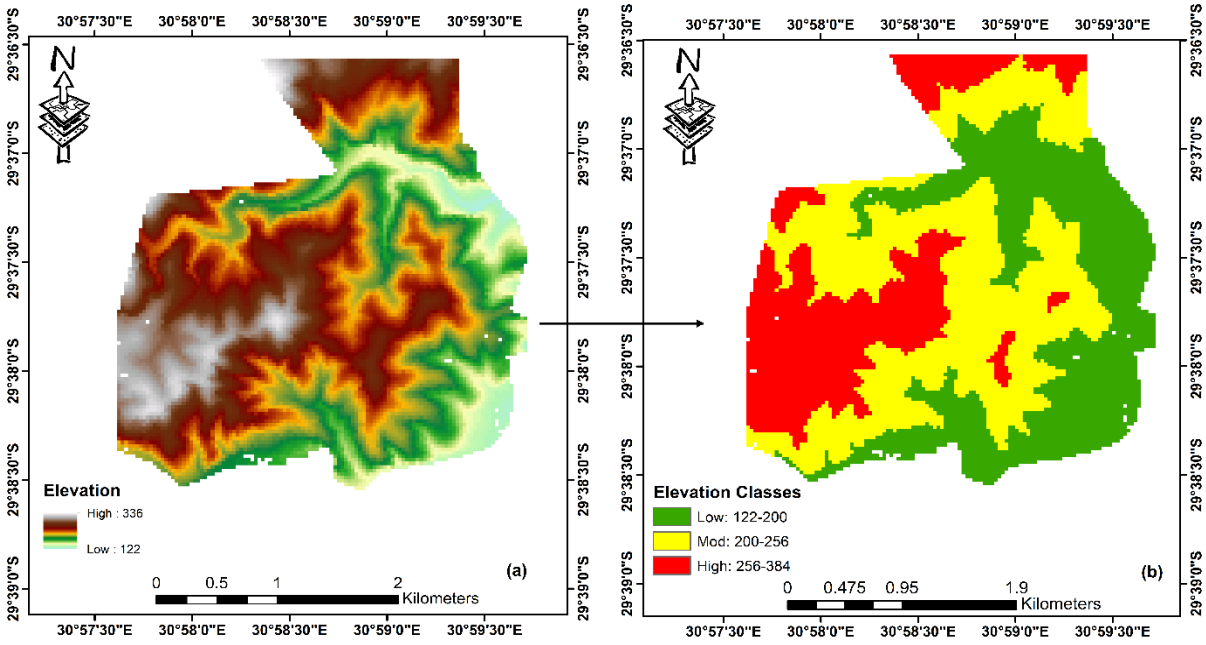

Figure 3.2: The digital elevation map of the study area computed from 5 m interval contour lines.

3.2.5.1 Slope

Slope steepness is one of the major determinants of forest AGB distribution, since slope affects soil drainage, nutrient availability and soil depth. Slope angle was derived from the DEM after Horn (1981), using the slope tool in ArcMap 10.3. Subsequently, the slope was smoothed using a

low 3x3 filter. The slope was then reclassified using the reclassify tool in ArcMap 10.3. Figure 3.3 shows the map of the derived slope angle in degrees and the reclassified slope.

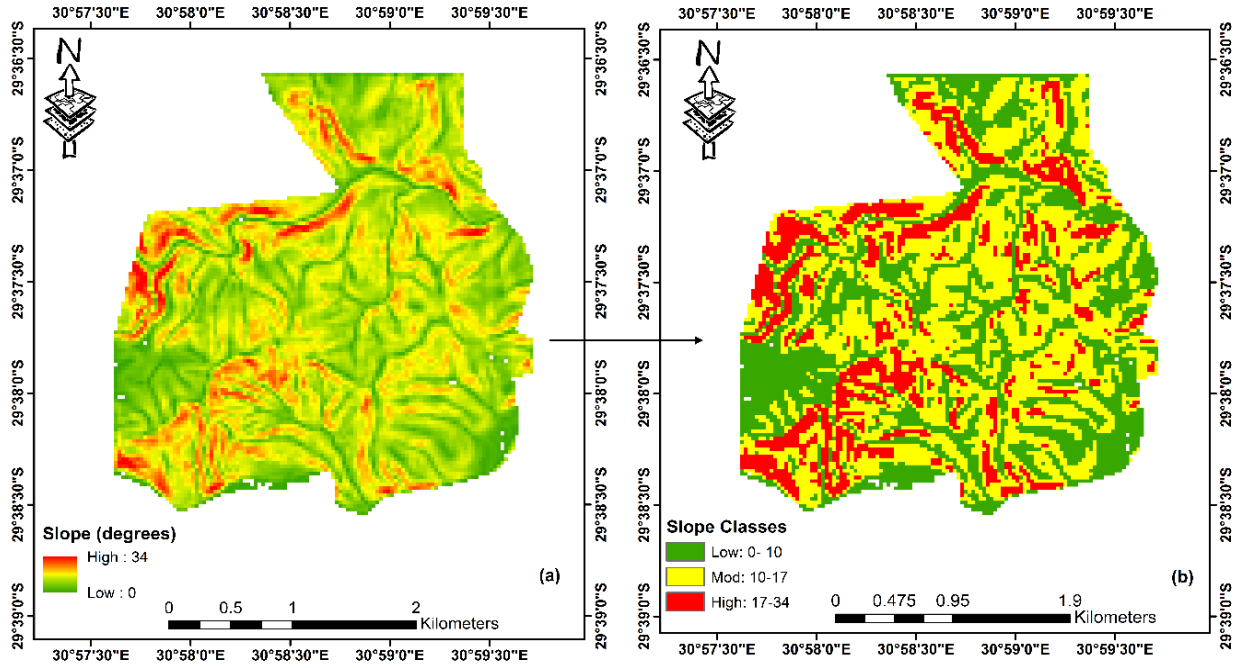

Figure 3.3: Slope steepness in angles derived from the DEM

3.2.5.2 Topographic Wetness Index

The wetness of the slope is vital in determining vegetation growth as it affects species distribution and the rate of forest AGB accumulation (Lin et al., 2012). Wet slopes are characterised by dense vegetation, while dry slopes are characterised by sparse vegetation. The topographic wetness index (TWI) was derived using Equation 3.4 below:

$$\text{TWI} = \ln\left(\frac{a}{tan\beta}\right) \ldots (3.4)$$

where β is the local slope and *a* is the local upslope area draining through a certain point per unit contour length (Quinn et al., 1995).

For the upslope area (the a variable), flow direction was initially derived from the sink filled DEM using the flow direction in ArcMap 10.3, 16 flow directions were used in this study. Subsequently, the flow direction was used to derive flow accumulation using the flow accumulation tool in

ArcMap 10.3, flow accumulation was used as the proxy for upslope area draining through a certain point per unit length. The local slope was derived from the DEM using the slope tool in ArcMap 10.3 as explained in section 3.2.5.1 above. Figure 3.4 shows the map for TWI as derived using the TWI equation and the reclassified image.

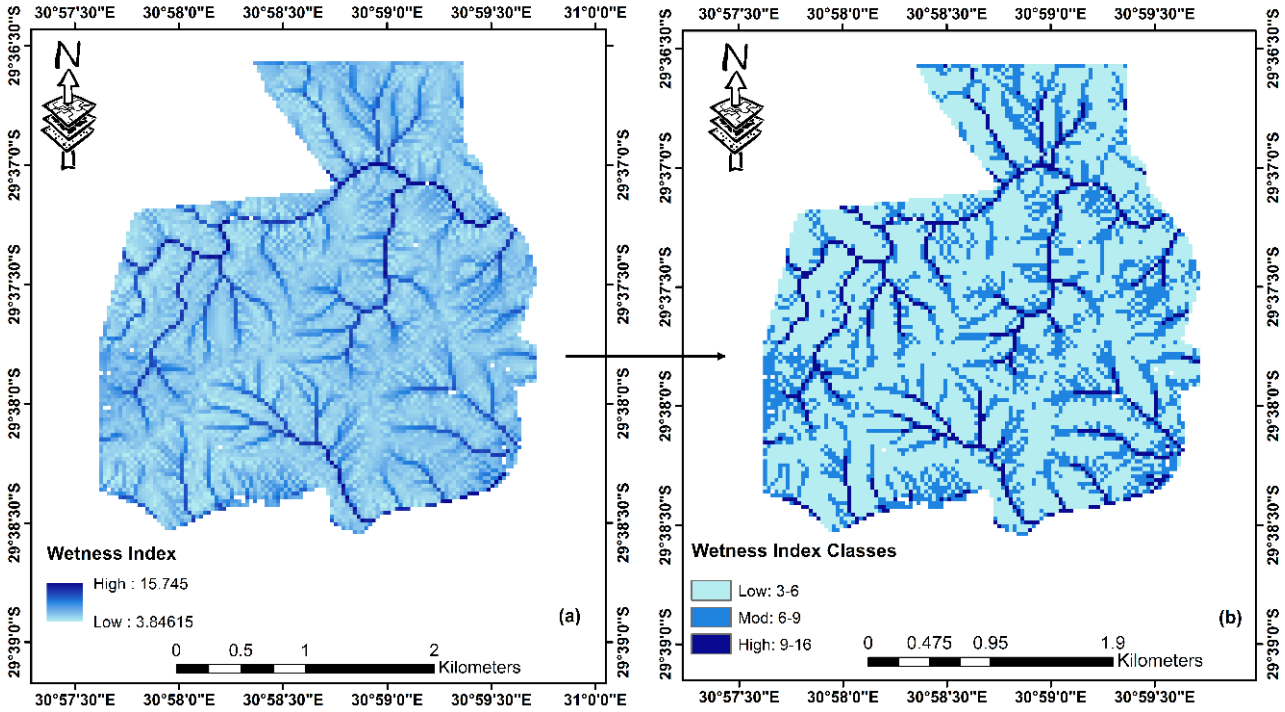

Figure 3.4: The topographic wetness of the study area based on the levitation and runoff accumulation areas.

3.2.5.3 Solar Radiation

The amount of incoming solar radiation in a specific area is critical in determining the rate of vegetation growth and the density of vegetation. This is due to light being the primary source for photosynthesis and thus is crucial in activating vegetation growth and development. In this study, solar radiation was estimated from the DEM using the area solar radiation tool in ArcMap 10.3 spatial analyst toolbox (www.esri.com). The solar radiation tool used the latitude of the location to derive solar declination and solar position, in this study one value of latitude was used as the study area is localised (Pachavo & Murwira, 2014). The area solar radiation calculation requires a sky size, which is based on the day intervals in this case (<14 days) and is localised, therefore a sky size of 512 was used (Kumar et al., 1997). The sky size is important in determining the

accuracy of the calculation of solar radiation, a large sky size has higher accuracies compared to a smaller sky size. To determine the view-shed of topography, slope direction is required and since the area consist of complex topography a 32 direction value was used. The cloud conditions of the sky are critical in determining solar radiation as they affect the amount of incoming and diffused radiation. In this study, sky conditions were generally clear, therefore a value of 0.3 was used (www.esri.com). The transmissivity of the sky is also an important factor in solar radiation calculation, and since general clear skies were used in this study a default value of 0.5 was used to account for transmissivity. Solar radiation was calculated for each month starting from January to December. Figure 3.5 shows the map of solar radiation (in MJ m^{-2} $year^{-1}$) and the reclassified map using the reclassify tool in ArcMap 10.3.

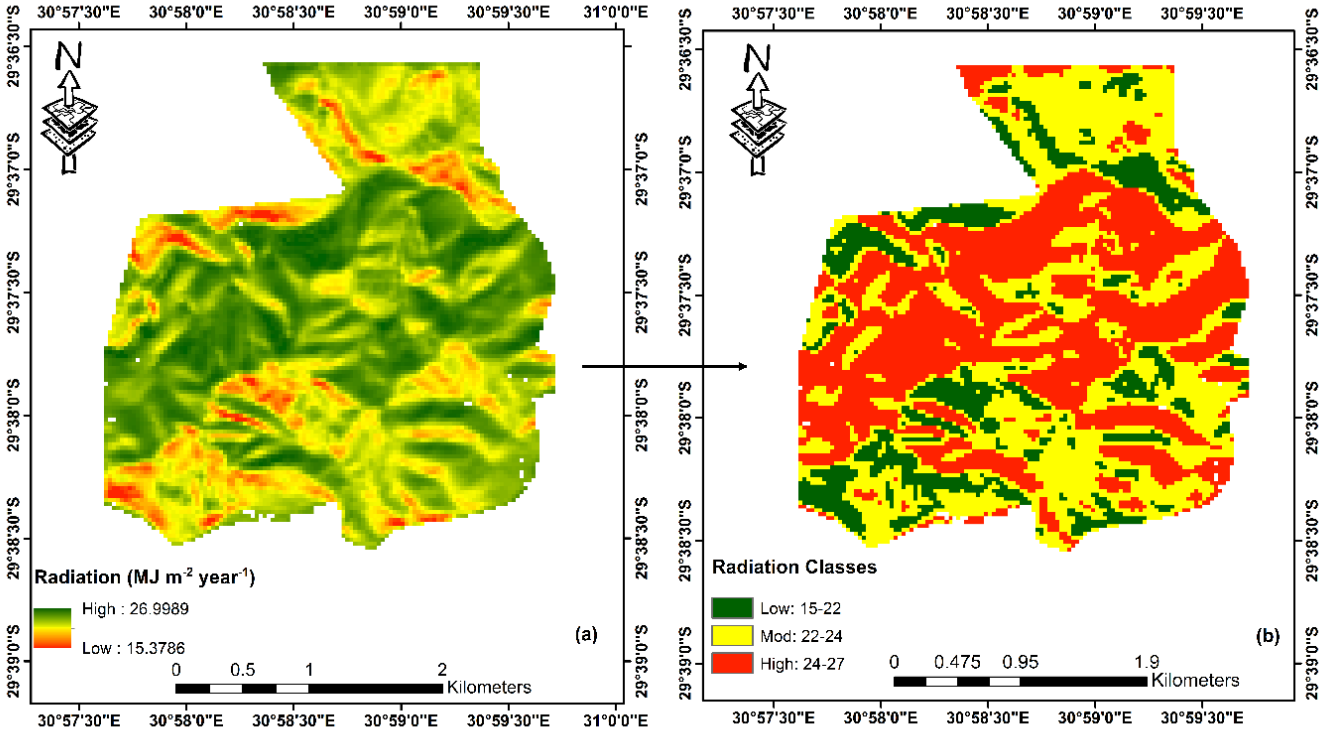

Figure 3.5: Annual shortwave radiation map (MJ m^{-2} $year^{-1}$) derived using Geographic Information System (GIS) and digital elevation model.

3.2.6 Remotely Sensed Biomass Model

In this study we used a forest biomass model developed by Hlatshwayo et al., (under revision), which predicted forest biomass using three texture band combinations. The model was created using random forest regression algorithm. Details of the model development are explained in Hlatshwayo et al., (under revision). The map of biomass is shown in Figure 3.6 below with accuracy of R^2 = 0.88 and RMSE = 54.54 kg m^{-2} (20.29% of the mean) (Hlatshwayo et al., under revision).

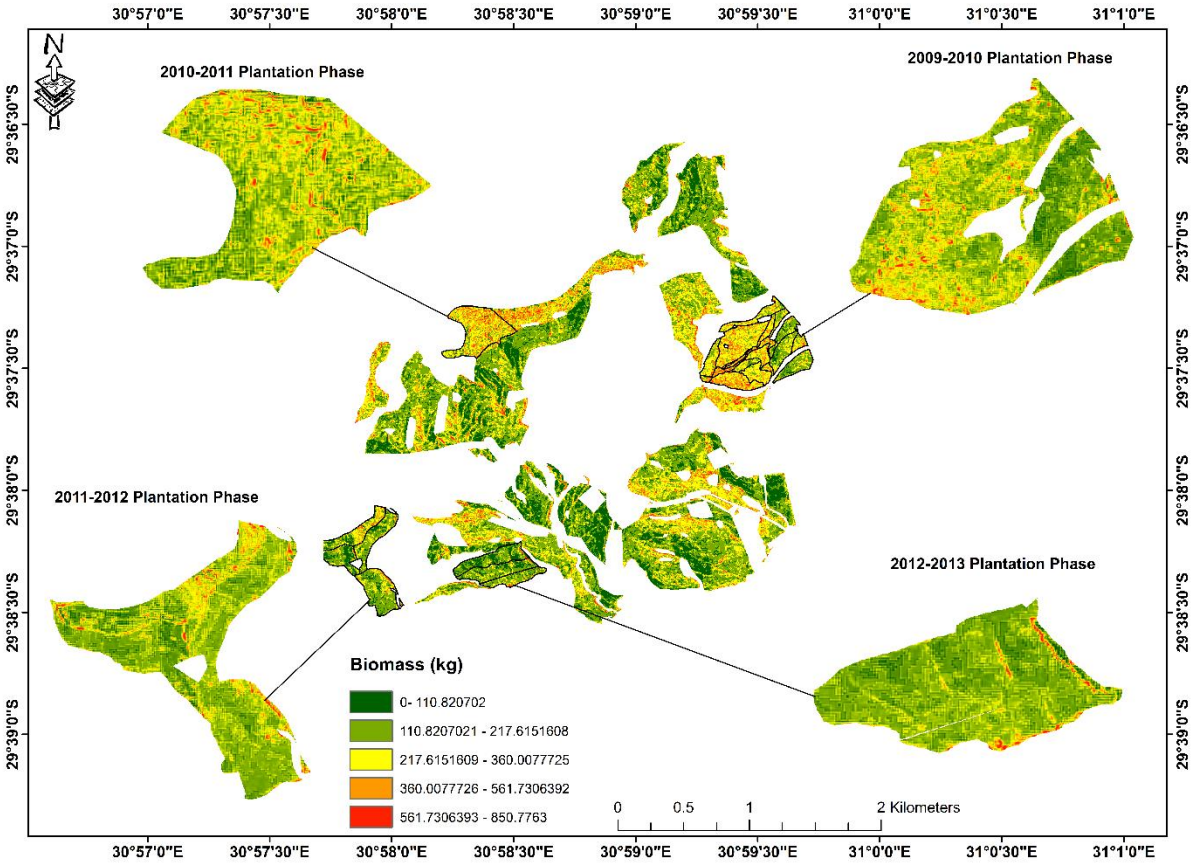

Figure 3.6: Above ground biomass map derived from the best performing three texture band combinations computed from the pan-sharpened image for the 2009-2013 plantation period.

3.2.7 Optimum Window Size Selection

Optimum window size selection was conducted in this study using the method proposed by Marceau et al. (1994), to identify the suitable windows capable of capturing forest structural and species diversity. This method requires segregation of the geographic entity under study into distinct classes. Followed by the resampling of the SPOT-6 pan-sharpened image using fixed neighbourhood window sizes (e.g. 3x3, 5x5, 7x7, 9x9, 11x11, 13x13, 15x15, 17x17, 19x19, 21x21). Finally, the minimum variance of each window size under each class was calculated to determine the minimum variance for that particular class.

3.2.8 Minimum variance

The minimum variance method after Marceau et al. (1994), has been applauded by researchers as an effective way of selecting optimum window size or spatial resolution for identifying geographic entities (Ismail et al., 2008; Lottering & Mutanga, 2016; Peerbhay et al., 2016). This method is used to indicate the window size that exhibited the lowest variance between pixels for each geographic entity (i.e. tree density, species diversity, species richness, diameter diversity and Gini coefficient). The spectral signals extracted from each window size were used to calculate variance of each window size. The window size that exhibited the lowest variance for each forest species and structural diversity class, was selected and used to extract image texture. Minimum variance was calculated using equations below:

$$Variance = \frac{\sum(x_{ij}-M)^2}{n-1} \quad \ldots (3.5)$$

$$Mean = \frac{\sum x_{ij}}{n} \ldots (3.6)$$

where M is the average of spectral signals, x_{ij} denotes the spectral signals and n represent the number of pixels in a moving window.

3.2.9 Image texture computation

Texture measures are used to characterise the spatial distribution of image grey-tone variation of individual bands (Moskal & Franklin, 2001). Variation in image tone is directly related to the structural assemblage of surfaces in relation to their neighbouring environments. These relationships are measured using texture variables (Franklin et al., 2001). In this paper, we used the grey level occurrence matrix (GLOM) and the grey level co-occurrences matrix (GLCM). The GLOM computes texture variables are based on the histogram of pixel intensities within a specified moving window and disregards the position of a pixel in relation to others (Lottering & Mutanga, 2012). However, The GLCM computes texture variables based on the spatial distribution of grey level pairs separated by a distant d in an angular direction θ (Haralick et al., 1973). We used five GLOM texture measure in this study namely data range, entropy, mean, skewness and variance, and eight GLCM texture measures namely contrast, correlation, dissimilarity, entropy, homogeneity, mean, second moment and variance. For a more detailed description of the GLOM texture measures refer to Lottering and Mutanga (2012), and Materka and Strzelecki (1998). A detailed description of the GLCM texture measures is provided by Materka and Strzelecki (1998) and (Haralick et al., 1973).

3.2.10 Extracting image texture

The GPS coordinates recorded in the field for all the sample plots (n = 90) were used to create point maps in ArcMap 10.3 software. The point maps were overlaid on the texture images to establish regions of interest (ROIs). These ROIs were subsequently used to extract image texture values. Zonal statistics tool in ArcGIS was used to extract the mean values of the image texture variables. The extracted image texture spectra were used to compute three texture band combinations using the Equations (3.7) and (3.8) below:

$$\frac{B1}{B2 \times B3} \ldots (3.7)$$

$$\frac{B1-B2}{B1+B3} \ldots (3.8)$$

where, B1, B2 and B3 represent the texture bands. The computed texture band combinations were then used to predict forest structure and diversity attributes.

3.2.11 Statistical Analysis

The relationship between forest structural diversity and the three band texture combinations was examined using stochastic gradient boosting and random forest regression, which are discussed below.

3.2.11.1 Random forest regression

The random forest machine learning algorithm is an advanced non-parametric ensemble extended from the bootstrap aggregation of regression and classification trees (Mutanga et al., 2012). Trees are grown in this algorithm by selecting a bootstrap of samples from the dataset to aggregate large numbers of trees (*ntree*) that are subsequently used to build the model (Freeman et al., 2015). The aggregated trees are grown maximally (not pruned), based on the bootstrap sample (Bassa et al., 2016). The samples that are not included in the bootstrapped sample are called the out-of-bag (OOB) sample. The advantage of RF is that it avoids overfitting by selecting a randomised sample of predictor variables (*mtry*) for growing trees, the *mtry* creates node splits for each tree in the ensemble (James et al., 2013). The OOB sample is used to test the predictive accuracy of each tree in the ensemble (Freeman et al., 2015). The RF algorithm requires optimisation of two parameters: (i) *ntree*, which is the number of regression trees aggregated using a bootstrapped sample of the observation (500 *ntree* is the default value), (ii) *mtry*, which is the number of predictors to incorporate at each split node.

3.2.11.2 Stochastic Gradient Boosting

Stochastic gradient boosting is a robust machine learning algorithm that uses tree ensemble to perform classifications and regressions (Friedman, 2001). The benefits of the SGB algorithm is that it can fit nonlinear relationships and it is insusceptible to statistical outliers, as a result it avoids overfitting, which improves the accuracy of the model (Dube et al., 2014). The intrinsic nature of the SGB is its capability to combine regression trees and boosting algorithms. The SGB model randomly selects a subset of the training dataset (50%) without replacement and uses a backwards stage-wise approach to fit regression trees to the model (Freeman et al., 2015). Contrary to the RF, the SGB uses total residual deviance to select the total number of trees (*ntree*) to incorporate in the model (i.e. trees are pruned). Therefore, the maximum number of trees are reached when the total residual deviance derived from data that is withheld remains constant (Freeman et al., 2015). The SGB algorithm requires optimisation of three parameters: (i) the maximum number of regression trees incorporated in the model (*ntree*); (ii) learning rate (*lr*) which determines the level of contribution of each tree to the model; (iii) tree complexity, which determines the split nodes of the trees using interaction of independent variables.

3.2.12 Model validation

The field and image texture dataset (n=90) were randomly split into 60% calibration dataset and 40% validation (Cho et al., 2009). The calibration data was used to train the predictive model. Subsequently, the model's reliability and quality were tested using the independent validation dataset.

3.3 Results

3.3.1 Descriptive statistics

The results in Table 3.1 illustrates the descriptive statistics of the forest structural diversity measures for the 90 sampling plots characterised. The Shapiro-Wilk test revealed that the forest structural diversity (FSD) dataset was normally distributed. The results indicate that generally SD and SR were very high in the study area with 1.38 and 6.09 averages for both SD and SR, respectively. In addition, DD was also very high in the study area with an average of 2.52 and a maximum of 9.10, indicating that there was differential tree growth in the study area. TD was varied across the study area with a standard deviation of 10.27 trees.m^{-2} and an average of 33 trees.m^{-2}, resulting from high mortality especially in poorly managed old trees (i.e. 2009-2011). The GC results indicated that there is a high tree size diversity in the study area with an average of 0.46 and maximum of 0.76, which indicated that there was possible seed germination among old trees.

Table 3.1: Descriptive statistics of the observed forest structural diversity measures.

Forest Structure Parameter	**Minimum**	**Maximum**	**Mean**	**Standard deviation**
Species Diversity (SD)	0	2.11	1.38	0.42
Diameter Diversity (DD)	0	9.10	2.52	1.51
Species Richness (SR)	1	12	6.09	2.26
Tree Density (TD) (m^{-2})	1	44	33.29	10.27
Gini Coefficient (GC)	0.25	0.76	0.46	0.14

3.3.2 Window Size Selection

Table 3.2 shows the optimum window sizes derived using the minimum variance technique. The chosen windows were based on the window size that exhibited the lowest variance. The results indicate that low species diversity levels were best captured by 13x13 window size, for moderate species diversity levels their window size was 15x15, and the high species diversity was best captured at 21x21 window size. The best window sizes for low, moderate and high diameter diversity were 3x3, 5x5, and 7x7, respectively. On the other hand, the optimum window sizes of 11x11, 15x15 and 17x17 were obtained for low, moderate, and high species richness. The results

also indicated that tree density and Gini coefficient variability were best captured at 3x3, 5x5 and 7x7 window sizes for low, moderate and high levels, respectively.

Table 3.2: Forest structural diversity attributes' classes used for optimum window size selection and their corresponding window sizes.

Forest Structure Parameter	Classes	Number of Plots	Window Size
Species Diversity	0-0.73	15	13x13
	0.74-146	33	15x15
	1.47-2.2	42	21x21
Diameter Diversity	0-1	31	3x3
	2-4	49	5x5
	5-9	7	7x7
Species Richness	1-4	19	11x11
	5-8	54	15x15
	9-13	17	17x17
Tree Density	1-15	9	3x3
	16-30	20	5x5
	31-45	61	7x7
Gini Coefficient	0.25-0.3.5	38	3x3
	0..36-0.55	41	5x5
	0.56-0.76	11	7x7

3.3.3 Correlation Analysis

The relationship between the three band texture ratios and the forest structural diversity measures were assessed using the Pearson's correlations test. The results for correlations test are illustrated in Table 3.3 for the three band texture combinations that yielded the highest correlation scores. The three band texture combinations that yielded the highest scores were subsequently used in the SGB and RF model to estimate the forest structural diversity measures. Generally, the Pearson's correlations test revealed that the three band texture combinations computed using the NDVI

equation yielded the highest correlations with the forest structural measures compared to the three band texture combinations computed using the simple ratio equation. Furthermore, the correlation tests showed that there was a stronger agreement between three band texture combinations with SD, SR and TD. Whereas, the Gini coefficient yielded the lowest correlation with the three band texture combination. In addition, the results also clearly demonstrate that the three texture band combinations that yielded higher agreements with the forest structural measures were derived from co-occurrence texture measures primarily from band 3 (red band) and band 2 (green band).

Table 3.3: Correlation between forest structural diversity attributes with three band texture combinations significant ($p < 0.05$).

Forest Structure Parameter	Three Band Texture Ratios	Correlation Coefficient (r)
Species Diversity	$\frac{CN\_C\_15\_B3 - VR\_C\_13\_B2}{CN\_C\_15\_B3 + SK\_O\_13\_B3}$	-0.84
	$\frac{VR\_C\_13\_B1 - CN\_C\_15\_B3}{VR\_C\_13\_B1 + SK\_O\_13\_B2}$	0.83
	$\frac{CR\_C\_13\_B2}{SK\_O\_13\_B1C \times R\_C\_13\_B3}$	0.82
Diameter Diversity	$\frac{HM\_C\_3\_B2}{CR\_C\_7\_B1 \times SM\_C\_3\_B3}$	0.70
	$\frac{HM\_C\_3\_B2 - VR\_O\_3\_B2}{HM\_C\_3\_B2 + DR\_O\_7\_B1}$	0.69
	$\frac{DR\_O\_5\_B1}{CR\_C\_7\_B1 \times SM\_C\_3\_B3}$	0.68
Species Richness	$\frac{DR\_O\_5\_B1 - HM\_C\_7\_B2}{DR\_O\_5\_B1 + MN\_O\_7\_B3}$	-0.85
	$\frac{VR\_O\_5\_B2 - HM\_C\_7\_B2}{VR\_O\_5\_B2 + MN\_O\_7\_B3}$	-0.84
	$\frac{HM\_C\_7\_B2 - DR\_O\_5\_B1}{HM\_C\_7\_B2 + MN\_O\_7\_B1}$	0.83
Tree Density	$\frac{HM\_C\_3\_B1}{SM\_C\_3\_B1 \times CR\_C\_7\_B3}$	-0.84
	$\frac{MN\_C\_7\_B4}{CR\_C\_7\_B3 \times CR\_C\_7\_B1}$	-0.83
	$\frac{DR\_O\_5\_B1 - HM\_C\_7\_B2}{DR\_O\_5\_B1 + HM\_C\_5\_B2}$	-0.82
Gini Coefficient	$\frac{MN\_C\_5\_B3}{HM\_CO\_5\_B1 \times DR\_O\_7\_B1}$	-0.50
	$\frac{DR\_O\_3\_B1}{VR\_O\_5\_B2 \times HM\_C\_5\_B1}$	-0.48
	$\frac{DR\_O\_7\_B1 - MN\_C\_5\_B1}{DR\_O\_7\_B1 + DR\_O\_5\_B4}$	0.47

B1, B2, B3, B4: Band 1, Band 2, Band 3 and Band 4; HM: Homogeneity, EN: Entropy, SM, Second moment, MN: Mean, DR: Data range, CR: Correlation, VR: Variance, CN: Contrast, SK: Skewness; O: Occurrence, C: Co-occurrence; 3, 5, 7, 13, 15: 3x3, 5x5, 7x7, 13x13, and 15x15.

3.3.4 Forest structural diversity predictions: RF and SGB Regression Model Performance

The RF and SGB regression were used to establish the best-fitting models for predicting forest parameters. Herein, the three band texture combinations that yielded the highest Pearson's

correlations score were used to compute models for predicting forest structure and diversity. The accuracy of the models was assessed using the RMSE and coefficient of determination (R^2), based on an independent test dataset. Table 3.4 summarises the major findings from the fitted models.

Table 3.4: Random forest and stochastic gradient boosting regression results illustrating the correlation between forest structure and band texture combinations with p-value<0.05.

Forest Structure Parameter +	Statistical Methods	tc	Ir	mtry	nt/ntree	R^2	RMSE (RMSE %)
Species Diversity	RF	-	-	9	500	0.88	0.21 (15.22%)
	SGB	10	0.001	-	500	0.85	0.26 (18.84%)
Diameter Diversity	RF	-	-	5	500	0.65	0.82 (32.54%)
	SGB	10	0.001	-	500	0.63	1.04 (41.27%)
Species Richness	RF	-	-	12	500	0.86	1.3 (21.35%)
	SGB	9	0.002	-	500	0.74	1.81 (29.72%)
Tree Density	RF	-	-	3	500	0.85	5.5 (16.52%)
	SGB	10	0.001	-	500	0.81	6.31 (18.95%)
Gini Coefficient	RF	-	-	17	500	0.54	0.21 (45.65%)
	SGB	10	0.001	-	500	0.64	0.13 (28.26%)

The results in Table 3.4 indicate that there was a strong agreement between the forest structure parameters and the three band texture combinations, with a high range of R^2 values (0.54-0.88) and RMSE mean% (15.22-45.65%) based on independent test datasets. Notably, the RF model incessantly yielded better results when compared to the SGB for the majority of the forest structural diversity parameters. More specifically, RF produced higher accuracies for SD, SR, TD and DD when fitted using the three band texture combinations. In contrast, the SGB model produced higher accuracies when modelling the GC with the R^2 value of 0.64 and RMSE value of 0.13 (28.26% of the mean measured GC). The scatterplots for all the best-fitted models is shown in Figure 3.7 indicating the distribution of the predicted and measured values.

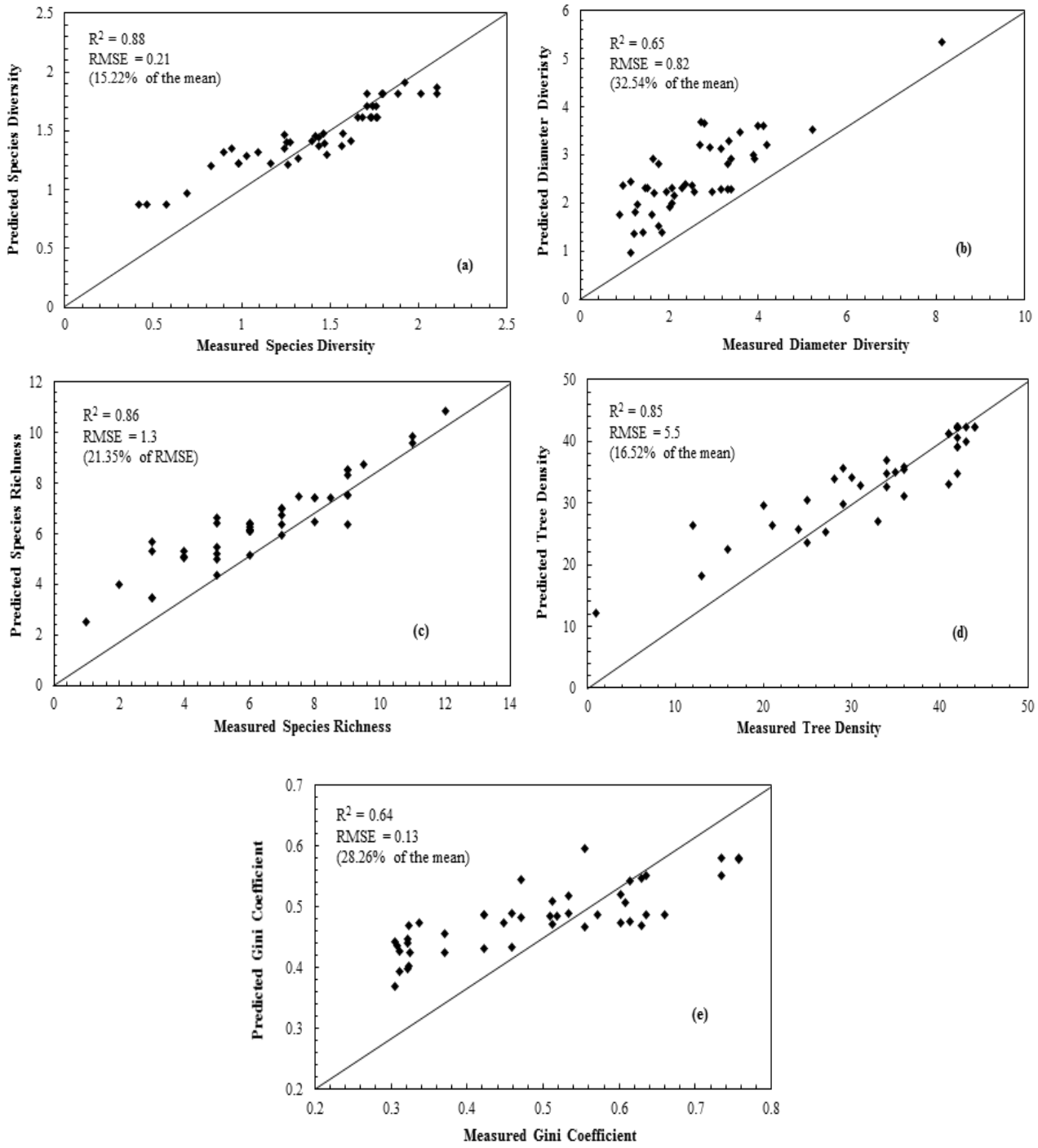

Figure 3.7: Shows one to one relationship between measured and predicted forest structural diversity attributes derived from three texture band models. Only best performing models are displayed here, a, b, c, d and e represent SD, DD, SR, TR and GC, respectively.

3.3.5 Best Performing Models' Variable Importance measurements

The relative contribution of the predictor variables (three band texture combinations) was measured using RF and SGB variable importance functions. These models use the out-of-bag (OOB) error rate to score the importance of the predictor variables. Figure 3.8 shows the important variables as selected by the RF and SGB models measured based on increasing OOB error rate. Herein, higher OOB error rate (%) indicates the higher importance of the predictor variable.

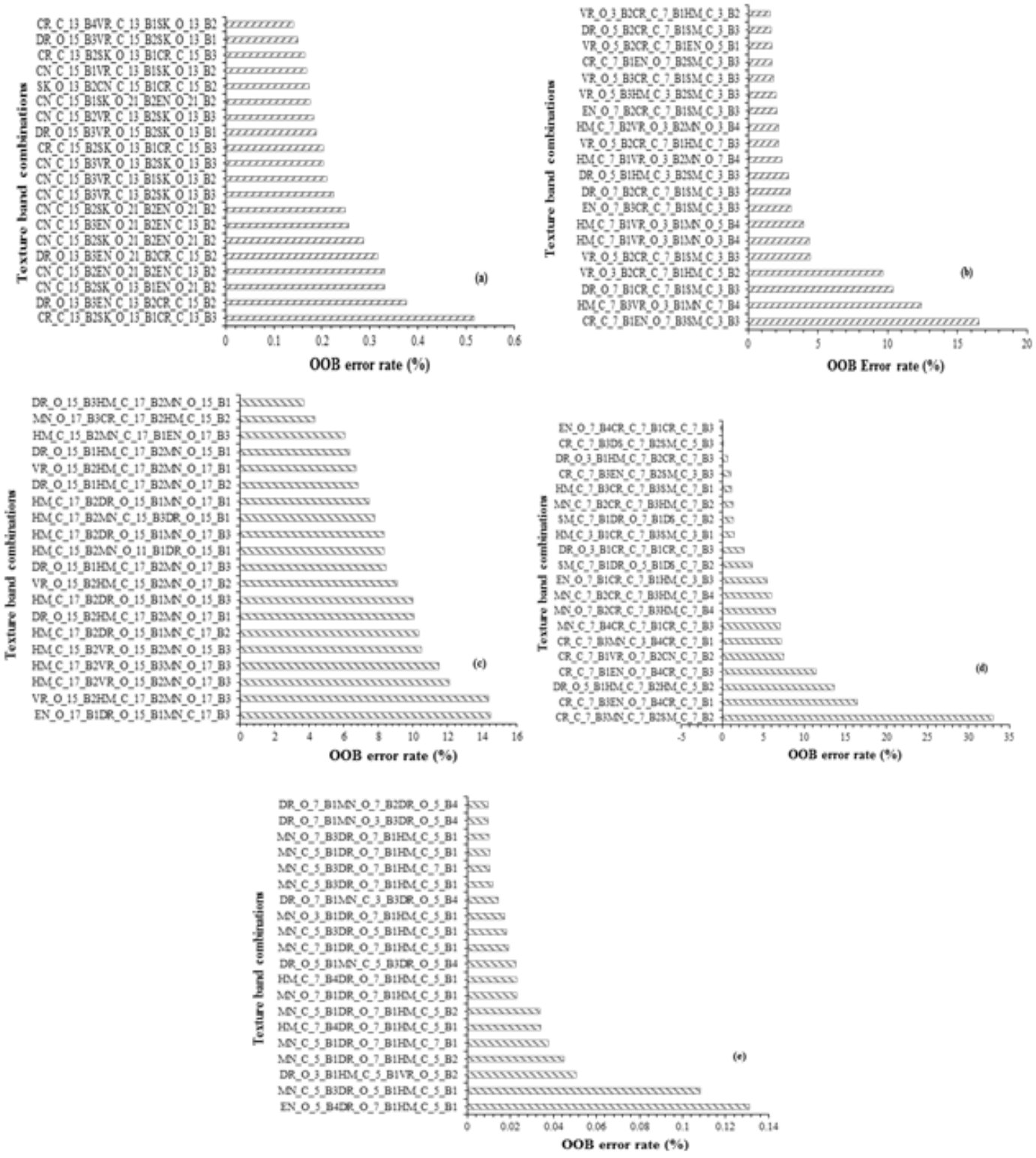

Figure 3.8: Variable importance measurements of texture models in predicting forest AGB using RF. Higher OOB error indicates high importance, a, b, c, d and e represent SD, DD, SR, TR and GC, respectively.

According to Figure 3.8, the co-occurrence variables appeared more frequently in three band texture combinations that contained greater importance compared to the occurrence texture measures. In addition, the results also indicate that Band 2 (green-band) and band 3 (red-band) appeared most frequently in the selected models for forest structure and diversity predictions. The models that yielded significant correlation results were subsequently used to generate maps of the forest structural diversity measures. Figure 3.9 shows maps of predicted forest structure variables using the best RF and SGB models developed using three-band texture combinations.

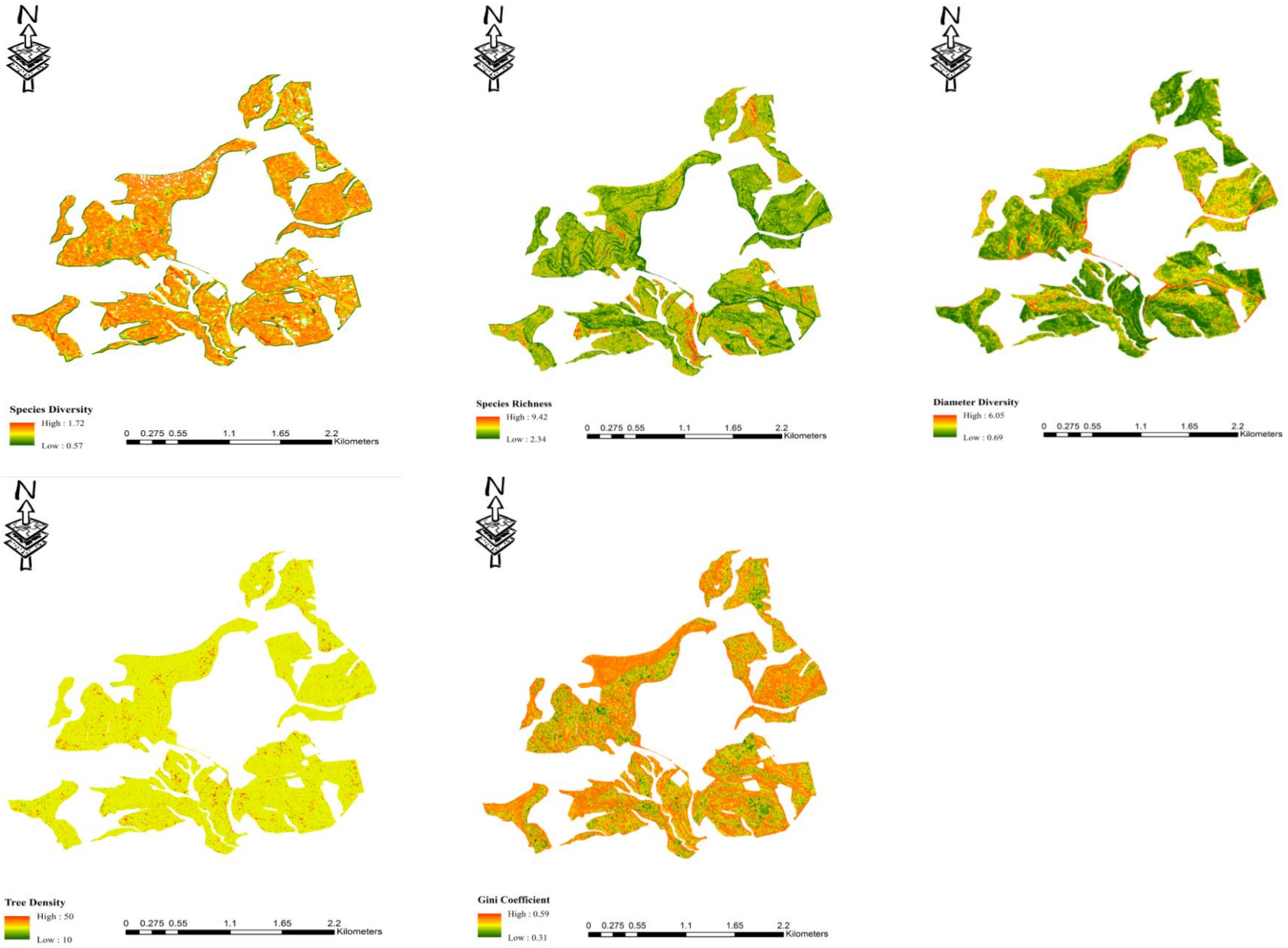

Figure 3.9: Map showing predicted forest structural diversity measures derived from the best performing three texture bands and regression algorithms.

The predicted maps of forest structure were subsequently reclassified using the tool ‘reclassify’ in ArcMap 10.3, to create three distinctive classes demonstrating the low, moderate and high levels of each predicted variable. The map showing the classes of the forest structural diversity is shown in Figure 3.10.

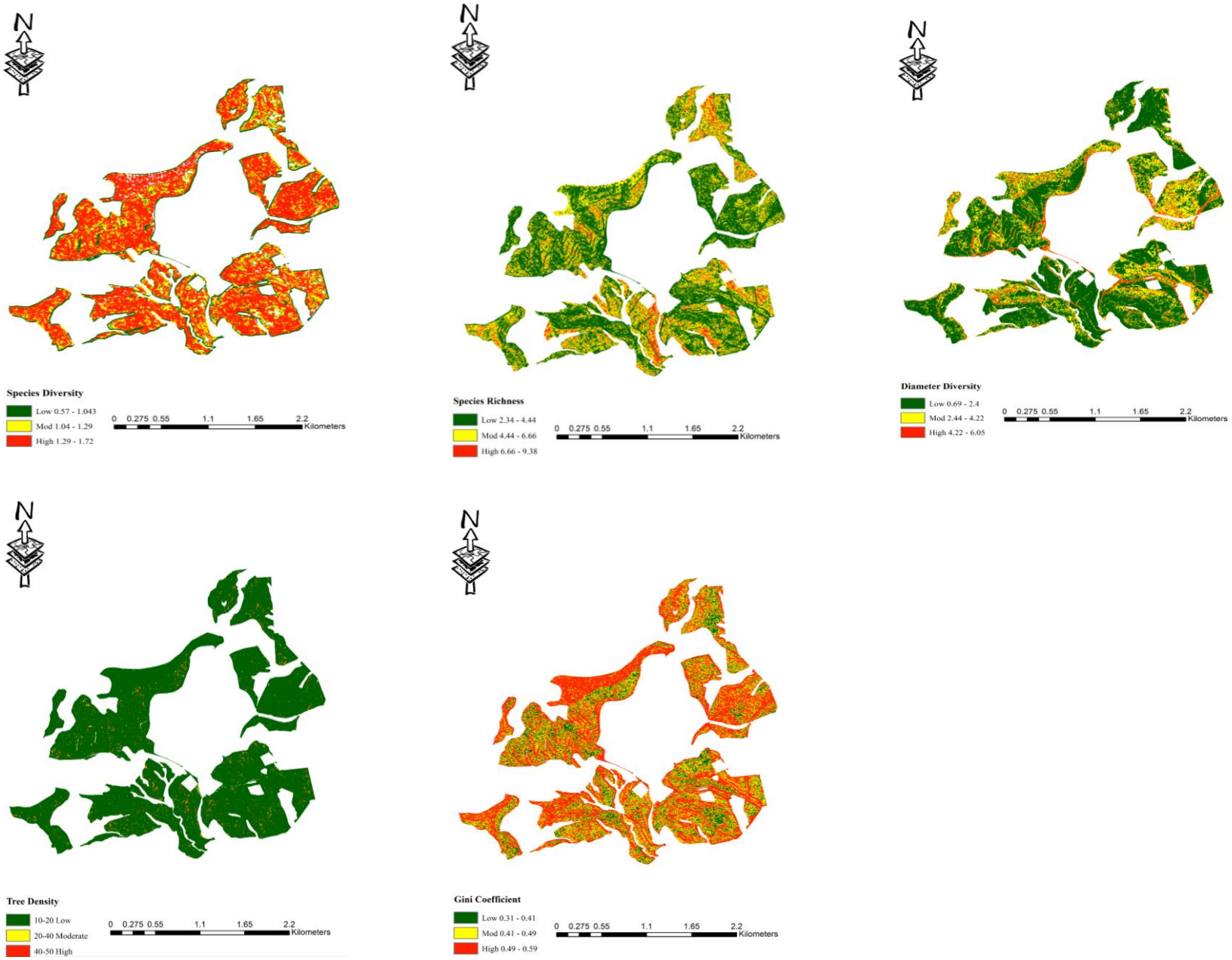

Figure 3.10: Reclassified map showing categorical groups of forest structural diversity measures derived from the predicted maps in figure 3.9.

3.3.6 Impact of forest structural diversity and topographic variables of forest AGB

The results in Figure 3.11 illustrate the effects of forest structural diversity on the spatial distribution of forest AGB. According to Figure 3.11(a), the mean of forest AGB in areas covered by low SD was higher than that of areas covered by moderate SD and high SD. Similarly, areas with low SR contained high average forest AGB as compared to areas with moderate to high SR. Conversely, areas with high DD contained higher mean forest AGB compared to areas with moderate DD and low DD. However, GC results revealed that areas with moderate GC contained greater mean forest AGB compared to areas with low and high GC. The TD results in Figure 3.11(d) yielded unexpected results in that, areas with low TD produced higher mean forest AGB compared to areas with moderate and low TD.

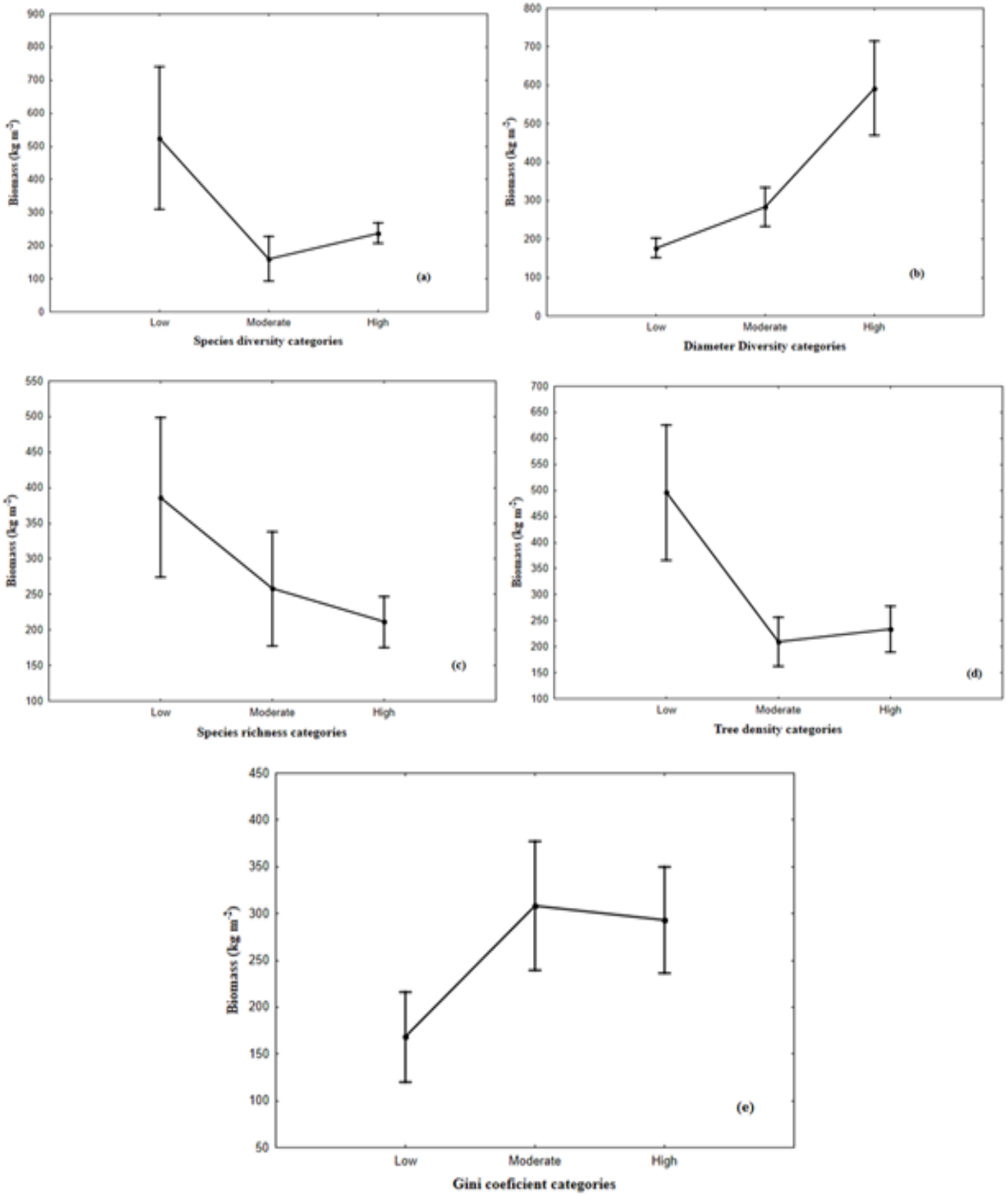

Figure 3.11: Mean forest AGB for the forest structural diversity categorical groups in figure, computed from the predicted above-ground biomass map.

The impacts of topographic variables on the spatial distribution of forest AGB are displayed in Figure 3.12. The results in Figure 3.12 revealed that altitude has an inversely proportional relationship with forest AGB, the mean forest AGB of low-lying altitude was the highest, followed by moderate and high-lying altitude. In contrast, forest AGB of the moderately steep slope was higher than that of gentle and very steep slopes. Furthermore, results for topographic

wetness index illustrated that areas with high moisture content contained higher forest AGB compared to areas with moderate and low moisture content, as contemplated. Similarly, there was a directly proportional relationship between solar radiation and forest AGB, areas exposed to high incoming solar radiation contained greater mean forest AGB compared to areas exposed to moderate and low solar radiation. One-way ANOVA statistical analyses were conducted to assess the significance of the differences in forest AGB between the categories (i.e. low, moderate and high) of each forest structural diversity measures and topographic variables.

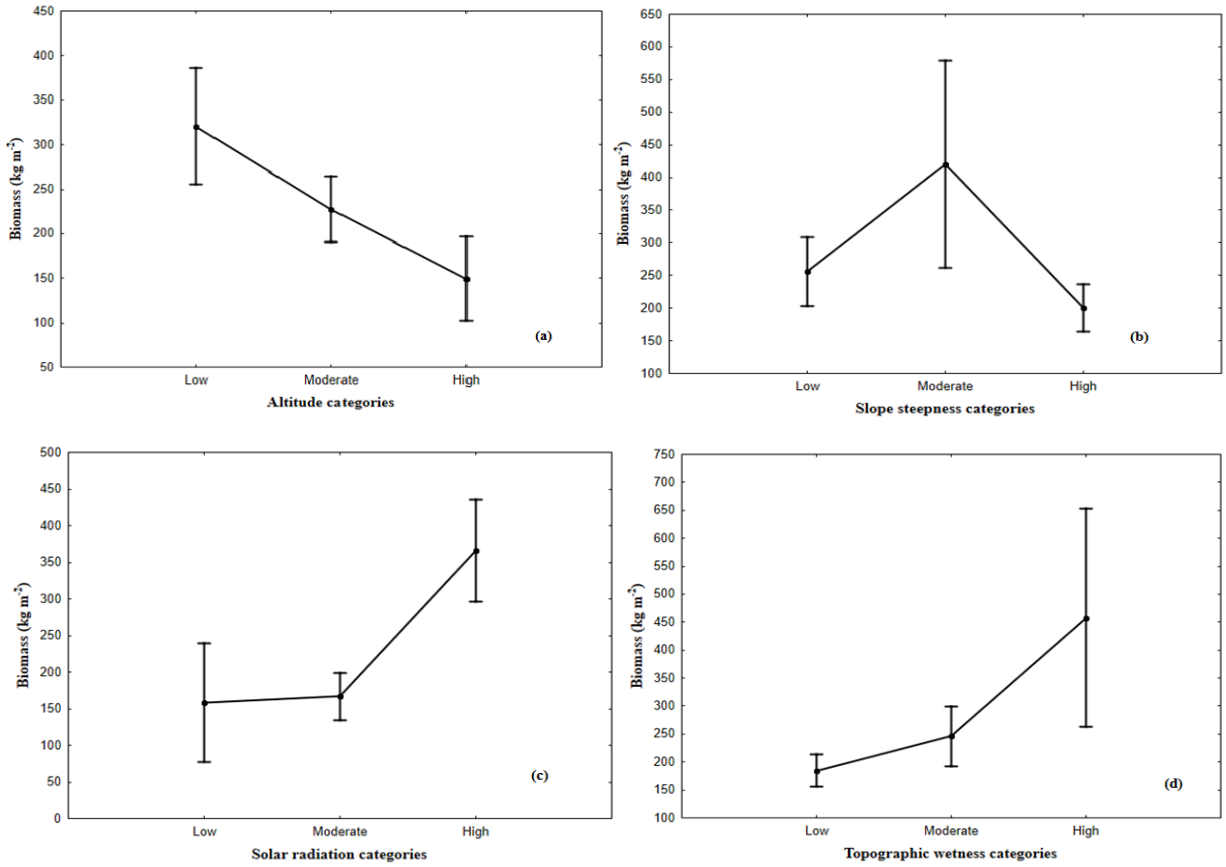

Figure 3.12: Mean forest AGB for the topographic variables' categorical groups extracted from the predicted above-ground biomass map shown in Figure 6, a, b, c and d represent altitude, slope, solar radiation and topographic wetness, respectively.

The one-way ANOVA results for forest structural diversity and the effects of topographic variables on forest AGB are displayed in Table 3.5. The results in Table 3.5 indicated that there were significant differences in mean forest AGB among the categories of all the forest structural diversity measures ($p < 0.05$). More specifically, the categorical groups (i.e. low, moderate and high) for species diversity, diameter diversity and tree density, yielded the highest statistically significant difference in mean forest AGB with ($p < 0.001$). In addition, the results further display significant differences in mean forest AGB among categorical groups of all the topographic variables ($p < 0.05$). Herein, slope steepness and topographic wetness

index produced the highest statistically significant difference in mean forest AGB among the categorical groups with ($p < 0.0001$). Further posthoc analyses were undertaken to identify the forest AGB means that were actually significantly different among the categorical groups.

Table 3.5: One-way Analysis of Variance results for the forest structural diversity attributes and the topographic variables

	Categorical Variables		
	Low	Moderate	High
Forest Structural Diversity			
Species Diversity	524.63±102.09	159.90±30.18	238.05±15.83
		F= 13.207, $p < 0.001$	
Diameter Diversity	177.01±12.94	283.83±23.58	592.04±55.02
		F = 65.543, $p < 0.001$	
Species Richness	386.44±54.72	257.88±38.06	211.16±17.59
		F = 5.687, $p < 0.01$	
Tree Density	495.80±63.16	209.05±23.04	233.92±21.20
		F = 15.084, $p < 0.001$	
Diameter Gini Coefficient	168.20±12.99	308.13±52.92	293.08±30.04
		F = 8.112, $p = .001$	
Topographic variable			
Altitude	320.81±32.55	227.64±16.87	149.75±22.44
		F = 6.798, $p < 0.01$	
Slope Steepness	255.78±25.63	419.95±74.39	200.20±17.77
		F = 9.396, $p < 0.001$	
Solar Radiation	158.36±36.31	166.54±15.39	365.96±34.71
		F = 8.041, $p < 0.01$	
Topographical Wetness Index	184.71±13.98	246.10±25.29	457.94±86.19
		F = 17.068, $p < 0.001$	

The Games-Howell posthoc analysis test was used to assess statistical differences of the forest AGB means among the categorical groups. Table 3.6 shows the results for the post-hoc analysis test. According to Table 3.6, it is clear that there is a significant difference in mean forest AGB among the categorical groups of SD ($p < 0.05$), with the highest difference occurring between the high and low SD levels t= 4.457368, $p < 0.001$. Similarly, the DD categorical groups had significantly different mean forest AGB, and the highest difference was between the high and low categorical groups, t = 11.08265, $p < 0.001$. However, means of forest AGB among the SR categorical groups were not all significantly different. In fact, only the high and low SR

categorical groups were significantly different, t = 3.136350, $p < 0.01$. Similarly, means of forest AGB among TD categorical groups were not all significantly different. The differences were among the high and low TD areas, t = 3.727416, $p < 0.001$ and the moderate and high TD areas, t = 2.695965, $p < 0.01$. The GC posthoc analyses revealed that there were significant differences in mean forest AGB between the low and high GC levels t = 4.215604, $p < 0.001$ and the low and moderate GC levels t = 3.411152, $p < 0.01$.

The posthoc tests for the altitude categorical groups revealed that mean forest AGB was significantly different between areas with low and high altitude, t = 3.339478, $p < 0.01$. However, mean forest AGB for areas with low and moderate altitude were not significantly different, t = 1.513042, $p = 0.14$. Slope steepness, on the other hand, yielded unexpected results, whereby areas with moderate and high slope steepness contained significantly different mean forest AGB, t = 3.936468, $p < 0.001$. In addition, areas with low and moderate slope steepness contained statistically different mean forest AGB, t = 2.347118, $p < 0.05$. However, mean forest AGB in areas with low and high slope steepness were marginally different, t = 1.83506, p = 0,07. Means of forest AGB among solar radiation categorical groups were not all significantly different. In fact, only low and high solar radiation and moderate and high solar radiation groups contained statistically significant mean forest AGB, t = 2.650024, $p < 0.01$ and t = 3.113561, $p < 0.01$, respectively. In contrast, the means of forest AGB among the topographic wetness index categorical groups were all significantly different. The results revealed that low and moderate topographic wetness and moderate and high topographic wetness contained significantly different means of forest AGB, t = 2.269731, $p < 0.05$, t = 5.537819, $p < 0.001$ and t = 2.990941, $p < 0.01$, respectively.

Table 3.6: Games Howell post-hoc analysis tests for the forest structural diversity and topographic variables' categorical groups

		Categorical Variables		
		Low	Moderate	High
Tree Diversity Measures				
Species Diversity	Low		t = 2.730387, $p < 0.05$	t= 4.457368, $p < 0.001$
	Moderate	t = 2.730387, $p < 0.05$		t = 2.074692, $p < 0.05$
	High	t= 4.457368, $p < 0.001$	t = 2.074692, $p < 0.05$	
Diameter Diversity				
	Low		t = 4.057479, $p < 0.001$	t = 11.08265, $p < 0.001$
	Moderate	t = 4.057479, $p < 0.001$		t = 5.836655, $p < 0.001$

	High	t = 11.08265 , $p < 0.001$	t = 5.836655, $p < 0.001$	
Species Richness				
	Low		t = 1.746706, $p = 0.087$	t = 3.136350, $p < 0.01$
	Moderate	t = 1.746706, $p = 0.087$		t = 1.24929, $p = 0.22$
	High	t = 3.136350, $p < 0.05$	t = 1.24929, $p = 0.22$	
Tree Density				
	Low		t = 1.615003, p = 0.11	t = 3.727416, $p < 0.01$
	Moderate	t = 1.615003, p = 0.11		t = 2.695965, $p < 0.01$
	High	t = 3.727416, p =<0.001	t = 2.695965, $p < 0.01$	
Diameter Gini Coefficient	Low		t = 3.411152, $p < 0.01$	t = 4.215604, $p < 0.001$
	Moderate	t = 3.411152, $p < 0.01$		t = 0.265128, $p = 0.79$
	High	t = 4.215604, $p < 0.001$	t = 0.265128, $p = 0.79$	
Topographic variables				
Altitude				
	Low		t = 1.513042, $p = 0.13$	t = 3.339478, $p < 0.01$
	Moderate	t = 1.513042, $p = 0.135893$		t = 2.507112, $p < 0.05$
	High	t = 3.339478, $p < 0.01$	t = 2.507112, $p < 0.05$	
Slope Steepness				
	Low		t = 2.347118, $p < 0.05$	t = 1.83506, p = 0,07
	Moderate	t = 2.347118, $p = 0.024538$		t = 3.936468, $p < 0.001$
	High	t = 1.83506, p = 0,071	t = 3.936468, $p < 0.001$	
Solar Radiation				
	Low		t = 0.232450, $p = 0.81$	t = 2.650024, $p < 0.01$
	Moderate	t = 0.232450, $p = 0.81$		t = 3.113561, $p < 0.01$
	High	t = 2.650024, $p < 0.01$	t = 3.113561, $p < 0.01$	
Topographical Wetness				
	Low		t = 2.269731, $p < 0.05$	t = 5.537819, $p < 0.001$
	Moderate	t = 2.269731, $p < 0.05$		t = 2.990941, $p < 0.01$
	High	t = 5.537819, $p < 0.001$	t = 2.990941, $p < 0.01$	

3.3.7 Principal Component Analysis

The principal component analyses (PCA) were conducted for both the topographic variables and forest structural diversity measures, to determine which of the factors strongly affected forest AGB. For these analyses forest AGB (extracted from the predicted forest biomass), was reclassified to low (100-300 kg m^{-2}), moderate (301-500 kg m^{-2}) and high (501-860 kg m^{-2}). The PCA results for the forest structural diversity measures are illustrated in Table 3.7. The

results show that the PCA analyses selected two principal components (axis 1 and 2) that explained 59.06% variation in forest AGB significant at (p-value < 0.05).

Table 3.7: Principal component analysis fitting forest structural diversity attributes as explanatory variables and forest AGB as the target categorical variable.

	Axis 1	**Axis 2**	**Axis 3**	**Axis 4**	**Axis 5**
Eigenvalues	2.538064	2,186814	1.131940	0.845334	0.740451
Explained variation (cumulative %)	31.72579	59.06097	73.21022	83.77690	93.03253
Forest structural diversity variables	correlation (r)				
Diameter diversity	0.663629	0.479295			
Species diversity	0.576132	0.689274			
Tree Density	0.517442	0.156099			
Species richness	0.434050	0.730300			
Gini coefficient	-0.177908	0.491162			

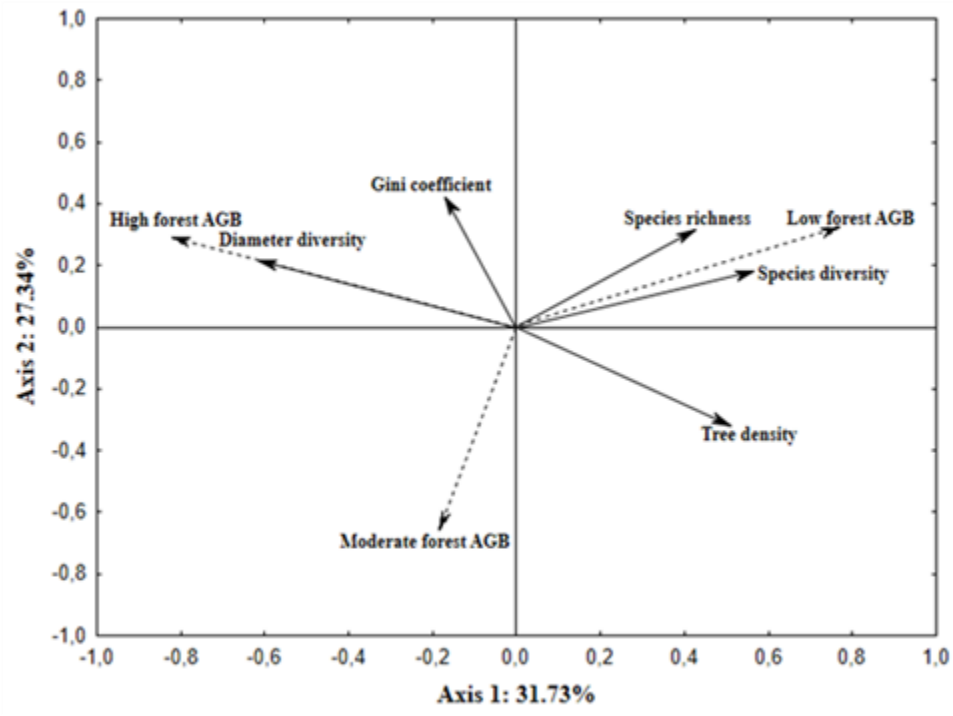

Figure 3.13: Principal component analysis plots for forest AGB categorical groups (dotted lines) and forest structural diversity attributes (solid lines) that had a significant influence (p < 0.01) on spatial distribution of forest AGB.

Notably, DD highly contributed to the spatial variation of high forest AGB levels r = 0.66 along axis 1 (see Fig 3.13). Whereas, SD and SR contributed to the spatial variation of low forest

AGB levels with r = 0.69 and 0.73, respectively. The PCA results for the topographic variables are illustrated in Table 3.8. The results in Table 3.8 showed that the topographic variables explained 52.99% spatial variation in forest AGB based on two principal components (axis 1 and 2) that were selected and were significant at (p-value < 0.05). The results in Figure 3.14 show the distribution of the topographic components along axis 1 and 2.

Table 3.8: Principal component analysis topographic variables as explanatory variables and forest AGB as the target categorical variable.

	Axis 1	**Axis 2**	**Axis 3**	**Axis 4**	**Axis 5**
Eigenvalues	2,249549	1,460170	1,214468	0,798371	0,898384
Explained variation (cumulative %)	32,1364	52,996	70.3455	81,7508	94,5849
Forest structural diversity variables	correlation (r)				
Solar radiation	0,656603	0,479295			
Topographic Wetness Index	0,473730	-0,365985			
Altitude	-0,452428	0,440493			
Slope	-0,433512	0,246748			

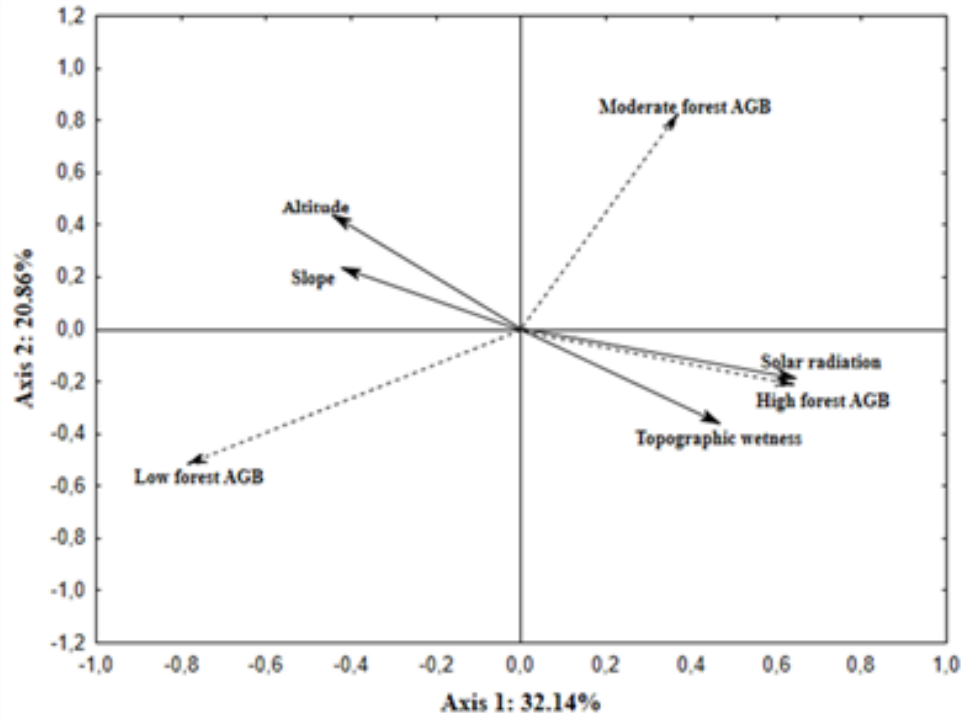

Figure 3.14: Principal component analysis plots for forest AGB categorical groups (dotted lines) and topographic variables (solid lines) that had a significant influence (P < 0.01) on spatial distribution of forest AGB.

Figure 3.14 illustrates that solar radiation and the topographic wetness index were responsible for explaining the spatial variation of high levels of forest AGB with r = 0.66 and 0.47, respectively along axis 1. However, none of the topographic variables were able to explain the spatial variation of either low or moderate forest AGB.

3.4 Discussion

The overall purpose of this study was to address issues pertaining the spatial planning of the Buffelsdraai reforested site (a dry escarpment forest) using advanced remote sensing 3 band image texture processing techniques and regression algorithms (SGB and RF). More specifically, the effects of topographic variables and forest structural diversity measures were studied using rigorous statistical test (PCAs and one-way ANOVAs). In summary, the results revealed that three band texture combinations were able to model the spatial distribution of forest structural diversity with plausible results using RF and SGB. In addition, the one-way ANOVA results revealed that there was significant differences among the forest structural diversity categorical groups and the topographic variables categorical groups as anticipated. More profoundly, were the PCA results that revealed that species richness and diversity explained a considerable amount of low forest AGB variation, whereas diameter diversity explained the variation of high forest AGB. The PCA results further revealed that solar radiation and topographic wetness were principal determinants of high forest AGB variation.

3.4.1 Predictive performance of RF and SGB Algorithm

The model performance evaluation results revealed that RF yielded higher accuracies for the majority of forest structural diversity measures (SD, SR, TD and DD) using three band texture ratios. However, the SGB algorithm produced higher accuracies for the GC variable. Notably, the differences in model performance of RF and SGB were almost negligible, the highest difference of ~10% was obtained when predicting the GC with R^2 = 0.54 and 0.64 for RF and SGB, respectively. The overall high performance of RF is attributed strongly to the manner in which the model utilises predictor variables that are highly correlated. The advantage of RF is

the fact that regression trees are grown independently (Freeman et al., 2015), therefore variable importance for highly correlated predictor variables is divided among trees (see Figure 3.8). Each tree is assigned one highly important variable, the remaining predictor variables are used in succeeding trees. This variable importance technique generates RF trees that are robust in their predictions (more predictor variables are used to generate the regression tree) and they are more resilient to overfitting (highly correlated predictors are divided among the trees to reduce multicollinearity within one tree).

The SGB algorithm, on the other hand, generates trees that are dependent on each other. In SGB the initial tree is built based on the original dataset, in successive trees, the predictors are used to predict the residuals of the previous tree (Friedman, 2001). This means that variable importance is highly dependent on the predictor variables that strongly contributed to the first tree and thus variable importance is concentrated on one variable (see Figure 3.8). Such a variable importance technique results in loss of information (only the first initial predictor variables contributes to the predictive power of the model) and this can weaken the predictive power of the model as illustrated in our study. These results are similar to those obtained by Freeman et al. (2015), who found that RF yielded overall higher accuracies when predicting forest structure compared to SGB based on the technique of variable importance ranking of the two models and the tree growing techniques of the models.

The variable importance measures for both the RF and SGB model ranked texture band combinations derived from co-occurrence texture measures as highly important (see Figure 3.8). These results confirm the findings by (Yuan et al., 1991; Franklin et al., 2000; Lottering & Mutanga, 2012), that co-occurrence texture measures contain the majority of vegetation information. Furthermore, the variable importance results in Figure 3.8 clearly demonstrated that band 2 (green band) and band 3 (red band) appeared more frequently on the high ranking predictor variables. The dominance of the green and red waveband indicates the high variability of chlorophyll and canopy structure in the study area similar to the findings of Carter (1993b), Gitelson et al. (1996) and Daughtry et al. (2000). This means that the vegetation structure is highly complex (with mixed species in close proximities) and the presence of senescing vegetation.

3.4.2 Ecological Interactions

The interaction between species yielded negative relationship with AGB in highly diverse areas and positive relationship with AGB in less diverse areas. Whereas the interaction between trees' sizes yielded positive relationship in areas with high tree size diversity and negative relationship with AGB in areas with low trees size diversity. Both these interactions were controlled by the availability of natural resources, such as light (solar radiation) and water (topographic wetness). These complex interactions and their relative contribution to forest AGB are discussed further in the sections below.

3.4.3 The effects of FSD measures on Forest AGB

The results in this study elucidated that species richness and diversity have a negative relationship with forest AGB productivity. The one-way ANOVA results revealed that areas that are highly mixed produced significantly lower forest AGB compared to areas with low levels of species mixture (see Table 3.5). Furthermore, the PCA results confirmed that species richness and diversity were principal in explaining low forest AGB spatial variability (see Figure 3.13). These results indicate that there was high competitive exclusion among species for natural resources such as light and water (Cardinale et al., 2009a; Shirima et al., 2015). This could be due to the mixture of species with high growth rates and productivity such as *Erythrina caffra* and *Dalbergia obovarta* with trees that have low growth rates (Bengtsson et al., 1994; Gracia et al., 2007). Furthermore, species such as *Acacia caffra*, *Dalbergia obovarta* and *Erythrina caffra*, have high competitive advantage, due to their adaptive capabilities to extreme dry conditions, cold winter temperatures and hot summer temperatures compared to *Millittia grandis*, *Trichilia dregeana* and *Syzigium cordatum* (Boon & Pooley, 2010). In addition, their capability to fix nitrogen into the soil gives them competitive advantage over non-nitrogen fixers for soil nutrients, especially in poorly nourished soils (Khanna, 1997; Forrester, 2004; Forrester et al., 2005a; Hoogmoed et al., 2014).

The poor relationship between SD and forest AGB productivity can also be explained by poor mixture of large tree sizes (such as *Acacia caffra*, *Dalbergia obovarta* and *Erythrina caffra)* with small tree sizes (such as *Brachyaleana discolour*, *Bredilia micrantha* and *Trema*

orientalis) (Forrester et al., 2005b; Shirima et al., 2015). In such instances, areas with large tree sizes tend to have low tree densities and species richness as a result of interspecies competition, therefore in these areas only trees with large sizes contributed to forest AGB productivity. Results in this study further indicated that high DD produced significantly higher forest AGB compared to areas with low DD (see Table 5). Furthermore, the PCA results revealed that DD was principal in explaining spatial variation of high forest AGB productivity. These results emphasised that large sized trees were dominant in areas with mixed variation of tree sizes. Another negative realtionship was identified between TD and forest AGB. The one-way ANOVA results illustrated that high TD areas produced lower forest AGB as compared to areas with low TD. This is a typical issue of interspecies competition which is intensified by close proximity of trees, thus increasing competition for resources that can inhibit species growth rates and survival (Shirima et al., 2015). Therefore, it can be concluded that spacing between planted trees plays a crucial role in forest AGB accrual rates.

3.4.4 The Effects of Topographic variables

The results in this study indicated that elevation and slope played a significant role in determining forest AGB spatial variability (see Table 3.5). More specifically, forest AGB was found to be higher in low altitude and moderately steep slopes compared to high altitude and very steep slopes. This negative relationship between forest AGB with altitude and slope is better explained by Woollen et al., (2012), who suggested that local tree species have different adaptation measures to varying climatic and edaphic conditions. High altitudes coupled with steep slopes tend to have insufficient soil nutrients, harsh climatic conditions and shallow soil depths (Moser et al., 2011), consequently trees tend to be smaller and short, with low diameter variability, in these areas. In addition, altitude and slope control the amount of soil water availability and solar radiation received by vegetation (Saremi et al., 2014). For example, north-facing slopes tend to receive higher solar radiation compared to south-facing slopes in the southern hemisphere as it is not shaded by the earth, and thus has more vegetation. Furthermore, low altitudes are characterised by high soil moisture and deep soils as they are zones of deposition and accumulation of soil debris, thus they support more vegetation compared to high altitudes (Gracia et al., 2007; Saremi et al., 2014; Xu et al., 2015).

The effects of solar radiation and topographic wetness on forest AGB were tested and results in this study revealed a positive feedback between these topographic variables and forest AGB. In fact, solar radiation and topographic wetness were found to be principal determinants of high forest AGB variability over altitude and slope. According to the one-way ANOVA analysis, high solar radiation and topographic wetness produced higher forest AGB compared to low solar radiation and topographic wetness. This is due to solar radiation and topographic wetness being responsible for slope drainage, soil water availability, slope temperature and light accessibility for plant growth. The higher the light and water available for plant growth, the lower is the interspecies competition for these primary resources (Lin et al., 2012; Wang et al., 2014; Xu et al., 2015). Interspecies competition can result in competitive exclusion of less productive species (Shirima et al., 2015), in this study such instances were encountered in mixtures of tree species with varying height and diameter size.

3.4.5 Implications for Spatial Planning of Reforestation sites

The results of this study showed a few ecological downfalls of the Buffelsdraai reforestation program. Natural forest restoration is arguably the best approach to combating climate change due to its multiple benefits that include high forest AGB accrual rates, biodiversity and ecological services. However, The spatial planning of mixed species reforestation requires rigorous experiments to assess whether the interaction between species and the ecological processes between the species will yield positive or negative feedback (Hulvey et al., 2013). The most common and successful approach that ecologists have identified for mixed forests, is the mixture of nitrogen fixing trees with non-nitrogen fixing trees (Forrester et al., 2005b). Such tree species mixture has been found to be successful especially in areas where nitrogen (N) is insufficient due to their symbiotic relationship. However, at sites with plenty N and a shortage of phosphorus (P) the interaction between nitrogen and non-nitrogen fixers may yield negative feedback with forest AGB, due to competition for P. In addition, other studies discovered that mixing tree species with varying tree sizes can reduce SD and TD, and can lead to dominance of the tree species with a larger size (Forrester et al., 2005b; Shirima et al., 2015).

Managing species interaction is vital for increasing forest AGB and ecosystem biodiversity. Species interaction can be managed using various spatial designs including; 1) thinning of dominant tree species, 2) planting species at different time scales depending on

their size and growth rates, and 3) by using varying TD or reducing trees at early growth stages, depending on availability of resources (Roberts & Gilliam, 1995; Getzin et al., 2006; Kelty, 2006). Canopy stratification has been found to be the best approach for mixed forests (Forrester et al., 2005b; Getzin et al., 2006). Successful canopy stratification requires mixture of short trees that are shade tolerant, with tall trees that are shade intolerant. This reduces interspecies competition for light (Getzin et al., 2006). Furthermore, the distance between tree species plays a significant role when studying their interactions. For example, tree species that are in close proximity can compete for light and water, however the litter from those tree species can be helpful to tree species slightly far from them by means of nutrient recycling of litter (Wright, 2002). Therefore, planting similar tree species in one row can be beneficial in terms of nutrient recycling for other species while reducing interspecies competition for light, as opposed to mixing tree species in one row. In addition, careful selection of tree sizes plays a significant role in overcoming interspecies competition for light, water and nutrients and for combating species dominance.

3.5 Conclusions

Forest structural diversity and topography, underpins forest AGB accrual rates in naturally grown forests. Species diversity and richness, is the foremost crucial step towards spatial planning for naturally grown forest. The success of mixed forest plantations is dependent on the ecological interaction of species that is determined by their functional groups and natural resource partitioning mechanisms. In this study we implemented multivariate statistical analysis to evaluate the effects of topography and forest structural diversity on forest AGB productivity. Results indicated that:

- There was a negative relationship between SD, SR and TD with forest AGB and positive relationship between tree size diversity and forest AGB.
- On the other hand, there was a negative relationship between altitude and slope steepness with forest AGB and a positive relationship between solar radiation and topographic wetness with forest AGB.

While more studies advocate for mixed forest plantations, we suggest that careful consideration should be given to the type of species planted within a plot. Most studies suggests the mixing

of nitrogen and non-nitrogen fixers. However the choice of these species should be based on the amount N they can fix and the rate of nutrient cycle they produce through leaf and fine-root litter. Furthermore, compatible tree size and height growth rates should be selected carefully to minimise interspecies competition for light and water. Overall, optimising AGB of mixed forest plantation is possible through careful spatial planning and designs that incorporate species interaction and the interaction of species with environmental gradients.

CHAPTER 4

Quantifying the Ecological Benefits of Above Ground Biomass using GIS and Remote Sensing Models: A synthesis

4.1 Introduction

Monitoring above ground biomass (AGB) in naturally growing forests is critical for climate change modelling. Natural forest plantations consist of complex ecological processes that are directly dependent on the interaction of tree species within the landscape and the interaction of environmental gradients with the species. Understanding the effects of these complex ecological processes on forest AGB productivity is crucial for optimising the overall benefits of reforestation on AGB accruals. Traditional methods of obtaining forest AGB and other forest inventory data have been side-lined in favour of remote sensing approaches that are viewed as viable and economical for obtaining forest inventory data.

While there are numerous attempts to estimate forest inventory data using satellite sensors with varying spatial and spectral resolution, there is no universal method for estimating forest structural attributes. In addition, models for forest inventory attributes developed using satellite imagery vary with study area, depending on the complexity of the vegetation and landscape structure, which affects the upwelling radiance. The challenge therefore in this study was to obtain a model that is capable of estimating forest structural attributes at various plantation phases and with complex canopy structure resulting from mixed species plantation. To develop these models in this study we assessed the potential of SPOT-6 imagery in predicting forest inventory attributes of trees planted at different time periods, using image texture combinations.

This studies objectives were set as follows:

1. Mapping the temporal and spatial distribution of forest AGB using three band texture combinations derived from the SPOT-6 image and random forest regression.
2. Predicting and mapping forest structural diversity using three band texture combinations derived from SPOT-6 pan-sharpened image and advanced machine learning algorithms (random forest and stochastic gradient boosting).
3. Assessing the effects of forest species and structural diversity on AGB productivity using GIS models and multivariate statistical analyses.
4. Determine the contribution of topographic variables on the spatial distribution of forest AGB using data derived from remote sensing and GIS models.

4.1.1 Estimating forest AGB of trees planted at different phases using texture band combinations and random forest regression.

To map the spatial and temporal distribution of forest AGB the spatial resolution of the SPOT-6 had to be optimised to detect suitable window sizes for detecting AGB variation. Forest AGB was divided into successional dates (i.e. 2009-2011 and 2011-2013), the classes were further divided into low, moderate and high for each successional date. The minimum variance technique was used to identify suitable window sizes for detecting forest AGB variation. The results in Table 4.1 show that older trees were best identified using larger window sizes (5x5 and 7x7), whereas the younger trees were best identified at using smaller window sizes (3x3 and 5x5).

Table 4.1: Canopy cover and wood density classes used for optimum window size selection for AGB estimation and their corresponding window sizes.

Succession Period	Biomass (kg m^{-2})	Number of Plots	Window Size
2009-2011	0-220	14	5x5
	221-440	12	7x7
	441-680	21	7x7
2011-2013	0-115	20	3x3
	116-230	18	5x5
	231-360	4	5x5

The selected optimum window sizes were used to derive texture measures that were used to model forest AGB. The texture bands were further combined using equations that enable two band and three combinations. The random forest (RF) and multiple linear regressions (MLR) were used and compared to model the relationship between forest AGB and texture parameters. Results in Table 4.2 show that the three band texture combination (R^2 = 0.88, and RMSE = 54.54 (20.29% of the mean)) performed better than the raw band textures (R^2 = 0.64 and RMSE = 94.13 (35.02% of the mean)) and the two band texture combinations (R^2 = 0.85 and RMSE = 60.65 (22.56% of the mean)), in predicting forest AGB using the random forest model.

Furthermore, the results also indicate that the RF model outperformed the MLR model in predicting forest model.

Table 4.2: Predictive Performance of the texture models

				Train dataset		Test Dataset		10-Fold-CV	
Texture variable	**Model**	**mtry**	**ntree**	R^2	**RMSE kg m^{-2} (RMSE %)**	R^2	**RMSE k m^{-2} (RMSE %)**	R^2	**RMSE kgm^{-2} (RMSE %)**
Raw band texture	MLR	-	-	0.51	91.40(34.00%)	0.67	89.92(33.45%)	0.29	114.722(42.68%)
	RF	1	893	0.79	42.70(15.89%)	0.64	94.13(35.02%)	0.53	92.82(34.53%)
two band texture ratio	MLR	-	-	0.67	75.24(27.99%)	0.82	67.09(24.96%)	0.63	82.89(30.84%)
	RF	1	860	0.90	40.20(14.96%)	0.85	60.65(22.56%)	0.67	77.56(28.86%)
three band texture ratio	MLR	-	-	0.76	63.69(23.70%)	0.85	59.77(22.24%)	0.75	68.48(25.48%)
	RF	1	939	0.93	32.59(12.12%)	0.88	54.54(20.29%)	0.77	65.66(24.43%)

The results from this study also indicated that forest AGB increased with age. According to Figure 4.1, concentration of high forest AGB were found in 2009-2010 and 2010-2011 plantation phases compared to younger trees. The one-way ANOVA results further proved that there was a significant ($p < 0.05$) variation in mean forest AGB across the successional dates. This chapter of the thesis proved that there are positive benefits of reforestation on ecosystem productivity over a period of time.

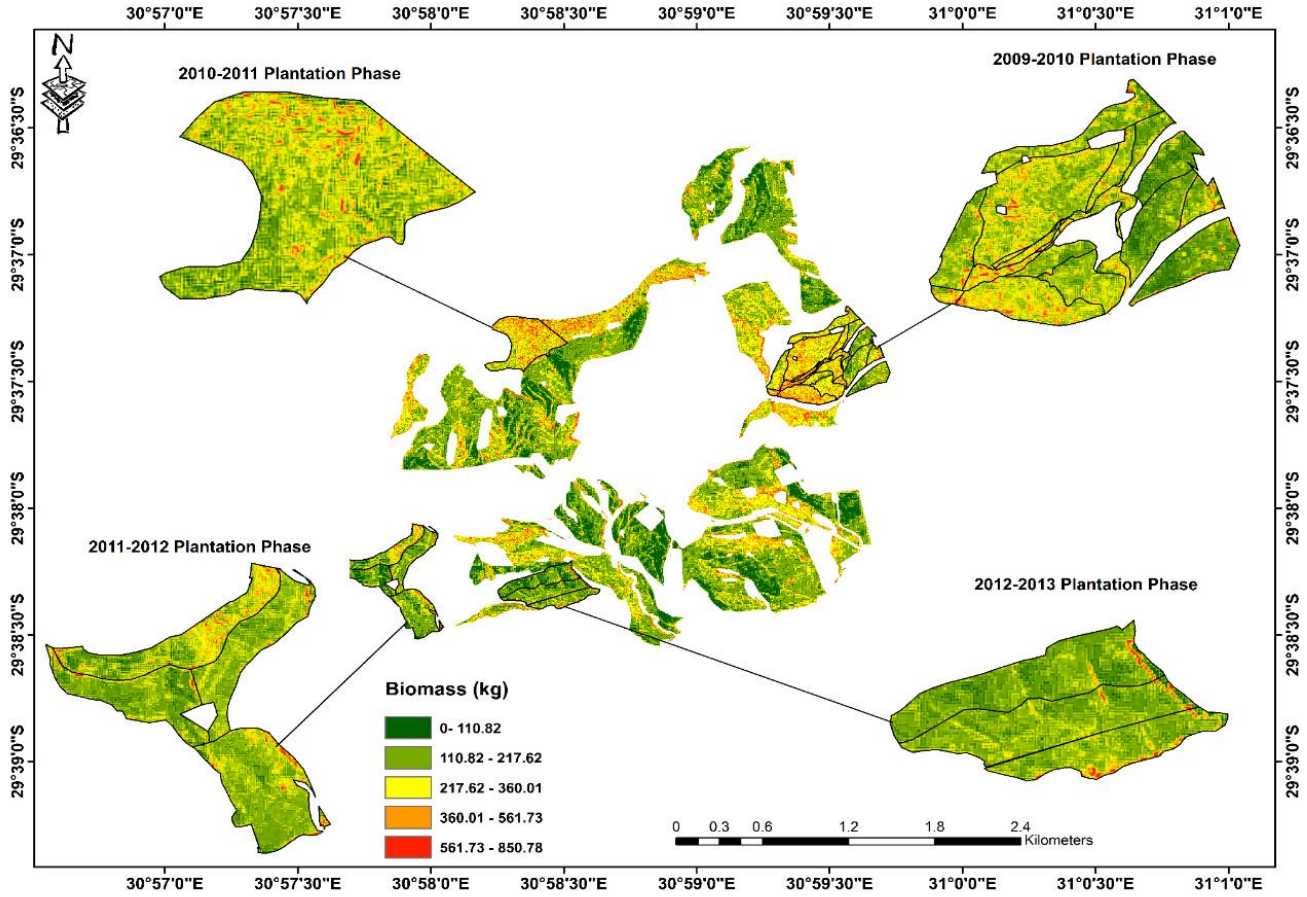

Figure 4.1: Above ground biomass map derived from the best performing three texture band combinations computed from the pan-sharpened image for the 2009-2013 plantation period.

4.1.2 Determining the effects of forest structural diversity and topographic variables on forest AGB productivity using remote sensing and GIS models.

Natural forest plantations are crucial in restoring various ecosystem goods and services, among them is stable forest AGB accrual. Research has proven that mixed forest plantations contribute more forest AGB compared to monocultures. However, the spatial planning of natural forest planning is currently a major challenge especially in cases where there are multiple objectives including increasing biodiversity and forest AGB productivity. The spatial design of forest with the above objectives requires thorough knowledge of species interaction according to their functional groups and the knowledge of their interaction with topographic variables. A challenge in this study was the fact that tree species were mixed randomly across the landscape and less attention was paid towards their interaction across the landscape. The purpose of this

chapter was to therefore quantify the effects of forest structural diversity and topographic variables on forest AGB accruals using remotely sensed and advanced machine learning algorithms.

Chapter 2 revealed that the three band texture combination provides valuable information in mapping forest AGB. Chapter 3 applied the same three band texture to model forest structural diversity attributes, using random forest and stochastic gradient boosting. Furthermore, this study also assessed the effects of topographic variables derived from a DEM on forest AGB distribution. The results in Table 4.3 show that random forest regression was able to predict species diversity (R^2 = 0.88, RMSE = 0.21 (15.22% of the mean)), species richness (R^2 = 0.86 and RMSE = 1.3 (21.35%)), diameter diversity (R^2 = 0.65 and RSME = 0.82 (32.54%)) and tree density (R^2 = 0.85 and RMSE = 5.5 (16.52%)) better than stochastic gradient boosting algorithm which only predicted the Gini coefficient (R^2 = 0.64 and RMSE = 0.13 (28.26%)) better than random forest.

Table 4.3: Random forest and stochastic gradient boosting regression results illustrating the correlation between forest structure and band texture combinations with *p*-value<0.05.

Forest Structure Parameter +	**Statistical Methods**	**tc**	**lr**	**mtry**	**nt/ntree**	**R^2**	**RMSE (RMSE %)**
Species Diversity	RF	-	-	9	500	0.88	0.21 (15.22%)
	SGB	10	0.001	-	500	0.85	0.26 (18.84%)
Diameter Diversity	RF	-	-	5	500	0.65	0.82 (32.54%)
	SGB	10	0.001	-	500	0.63	1.04 (41.27%)
Species Richness	RF	-	-	12	500	0.86	1.3 (21.35%)
	SGB	9	0.002	-	500	0.74	1.81 (29.72%)
Tree Density	RF	-	-	3	500	0.85	5.5 (16.52%)
	SGB	10	0.001	-	500	0.81	6.31 (18.95%)
Gini Coefficient	RF	-	-	17	500	0.54	0.21 (45.65%)
	SGB	10	0.001	-	500	0.64	0.13 (28.26%)

To determine the effects of predicted forest structural diversity and topographic variables on forest AGB productivity, multivariate statistical analysis were conducted. The one-way ANOVA results indicated that there were significant differences in mean forest AGB across the categorical groups of the forest structural diversity and topographic variables (p<0.05). The

PCA results in Figure 4.2 indicated that species diversity and species richness accounted for the variability of low forest AGB, whereas diameter diversity accounted for the variability of high forest AGB. On the other hand, results in Figure 4.2b revealed that solar radiation and topographic wetness index contributed to the spatial variation of high forest AGB.

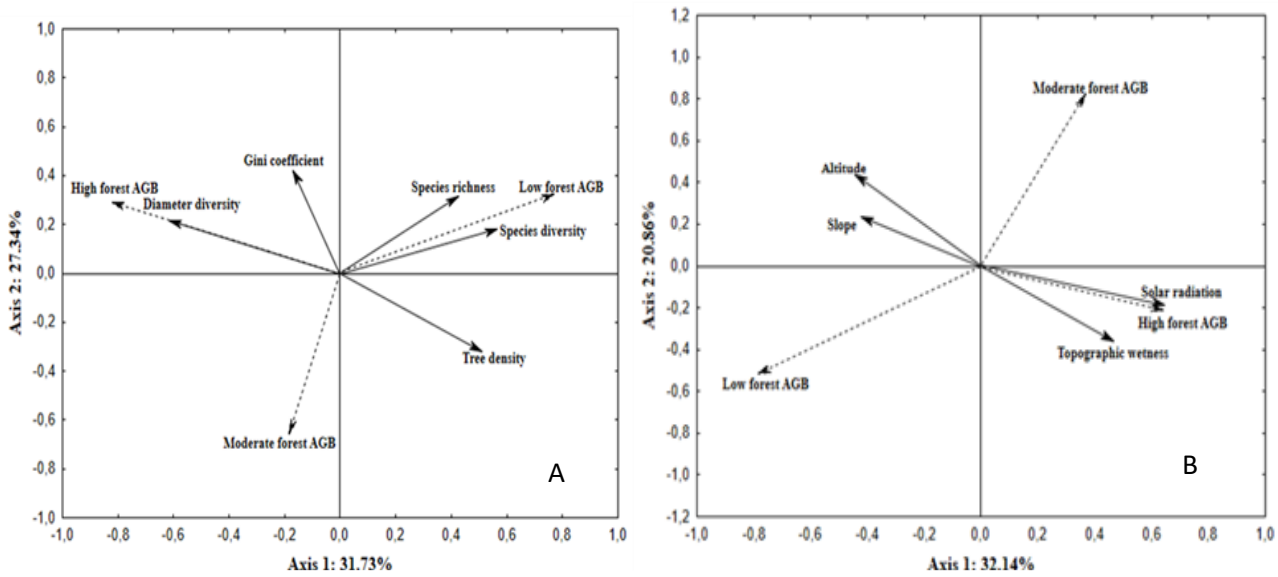

Figure 4.2: Principal component analysis plots for forest AGB categorical groups (dotted lines) with forest structural diversity (in A) and topographic variables (in B) that had a significant influence ($p < 0.01$) on spatial distribution of forest AGB.

Notably in this study, areas with high species diversity and richness contained lower forest AGB compared to areas with high species richness and diversity. Conversely, areas with high diameter diversity contained higher forest AGB compared to areas with low diameter diversity. These results indicate dominance of certain species that contributed heavily to forest AGB accruals hence the negative feedback in areas with high diversity, while large size trees appeared to dominate and outcompete tree species with low productivity, this is indicated by positive feedback between high diameter diversity and forest AGB. In addition, tree density also yielded negative feedback with forest AGB productivity, areas with high density contained relatively low mean AGB compared to areas with low tree density. This is a typical case of tree thinning as a result of high interspecies competition for resources in densely vegetated areas as opposed to areas with low tree density that have low interspecies competition. These results indicate the importance of spatial planning and design that manipulates the functional diversity and size compatibility to ensure that species are able to coexist and to increase both diversity and forest AGB accruals.

4.2 Conclusions

The intentions of this study were to utilise remotely sensed data to understand ecological processes that occur on a naturally grown forest. These ecological processes are controlled by tree species interaction and environmental variables across the landscape in space and time. To understand the effects of tree species interaction and environmental variables on ecological processes, we used forest AGB as an indicator of ecosystem productivity. Furthermore, forest AGB is part of the crucial ecological process that involves carbon sequestration and storage. In this study, we evaluated the accuracy of spatially optimised raw band textures, two and three band texture combinations derived from the SPOT-6 image, to predict forest AGB using random forest regression algorithm. The three band texture combinations were subsequently used to predict forest structural diversity attributes. The effects of forest diversity and environmental variables were assessed using GIS models and multivariate statistical analysis.

The conclusion drawn from this study were based on the research objectives stipulated in the introduction section and addressed in Chapter Two and Three:

1. The three texture band combinations computed from the SPOT-6 pan-sharpened image were able to predict forest AGB at various growth stages better than raw band textures and two band texture combinations. The results also proved that there is an increases in forest AGB over the years, high mean forest AGB was obtained in old trees and decreased with the age of the trees, thus indicating the potential
2. success of the reforestation programme over time.
3. The effects of forest structural diversity on forest AGB productivity was assessed using remotely sensed and GIS models. The three band texture models were able to predict forest structural diversity with promising results. Random forest regression algorithm was able to out-perform stochastic gradient boosting algorithm in predicting forest structural diversity.
4. The findings in this study revealed that there was a negative feedback between tree density, species diversity, and species richness with forest AGB. However, diameter diversity and the Gini coefficient were directly proportional to forest AGB. These results suggest that the spatial design of mixed forests need to manipulate both functional diversities and tree size to enable complementary tree species that are able to co-exist.

5. Using the DEM, the effects of topographic variables derived using GIS models were assessed. The results indicated that slope and elevation produced a negative feedback with forest AGB. Notably, the topographic wetness index and solar radiation produced a positive feedback with forest AGB. More importantly, PCA results revealed that solar radiation and topographic wetness were principal in determining spatial variation of high forest AGB.

The study generally advocates for the use of advanced image processing techniques, particularly the combination of minimum variance for window size optimisation with three band texture combinations. The study also demonstrated that reforestation produces positive feedback on AGB over a period of time. More importantly were the findings that suggested the importance of spatial planning and species selection in a mixed forest plantation. This research has the potential to contribute to a growing body of literature that addresses the benefits of mixed forest plantations using indigenous tree species on ecosystem productivity. More research addressing these complex ecological processes occurring in natural forests need to be undertaken, to further our understating of various forest spatial design and their effects on ecosystem productivity.

REFERENCES

Adjorlolo, C., & Mutanga, O. (2013). Integrating remote sensing and geostatistics to estimate woody vegetation in an African savanna. *Journal of Spatial Science, 58*(2), 305-322. doi: 10.1080/14498596.2013.815577

Anderson, G. L., Hanson, J. D., & Haas, R. H. (1993). Evaluating landsat thematic mapper derived vegetation indices for estimating above-ground biomass on semiarid rangelands. *Remote Sensing of Environment, 45*(2), 165-175. doi: http://dx.doi.org/10.1016/0034-4257(93)90040-5

Barbosa, J., Melendez-Pastor, I., Navarro-Pedreño, J., & Bitencourt, M., Dantas. (2014a). Remotely sensed biomass over steep slopes: An evaluation among successional stands of the Atlantic Forest, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing, 88*, 91-100. doi: http://dx.doi.org/10.1016/j.isprsjprs.2013.11.019

Barbosa, J. M., Broadbent, E., & Bitencourt, M. (2014b). Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. *International Journal of Forestry Research, 2014*, 14. doi: 10.1155/2014/715796

Bassa, Z., Bob, U., Szantoi, Z., & Ismail, R. (2016). Land cover and land use mapping of the iSimangaliso Wetland Park, South Africa: comparison of oblique and orthogonal random forest algorithms. *Journal of Applied Remote Sensing, 10*(1), 015017-015017.

Bastin, J.-F., Barbier, N., Couteron, P., Adams, B., Shapiro, A., Bogaert, J., & De Cannière, C. (2014). Aboveground biomass mapping of African forest mosaics using canopy texture analysis: toward a regional approach. *Ecological Applications, 24*(8), 1984-2001. doi: 10.1890/13-1574.1

Bengtsson, J., Fagerström, T., & Rydin, H. (1994). Competition and coexistence in plant communities. *Trends in Ecology & Evolution, 9*(7), 246-250. doi: http://dx.doi.org/10.1016/0169-5347(94)90289-5

Bingham, B. B., & Sawyer, J. O. (1992). Canopy structure and tree condition of young, mature, and old-growth Douglas-fir/hardwood forests. In R. R. Harris & D. E. Erman (Eds.), *Proceedings of the Symposium on Biodiversity of Northwestern California* (Vol. Report No. 29, pp. 141-149). Santa Rosa, CA. Berkeley: CA: University of California.

Boon, R., & Pooley, E. (2010). *Pooley's trees of eastern South Africa*: Flora and Fauna Publications Trust.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi: 10.1023/a:1010933404324

Brower, J. E., Zar, J. H., & von Ende, C. N. (1997). *Field and Laboratory Methods for General Ecology* (4th ed ed.). Dubuque, Iowa: Wm C. Brown Publishers.

Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental pollution, 116*(3), 363-372.

Cardinale, B. J., Hillebrand, H., Harpole, W. S., Gross, K., & Ptacnik, R. (2009a). Separating the influence of resource 'availability' from resource 'imbalance' on productivity–diversity relationships. *Ecology Letters, 12*(6), 475-487. doi: 10.1111/j.1461-0248.2009.01317.x

Cardinale, B. J., Hillebrand, H., Harpole, W. S., Gross, K., & Ptacnik, R. (2009b). Separating the influence of resource 'availability'from resource 'imbalance'on productivity–diversity relationships. *Ecology Letters, 12*(6), 475-487.

Carter, G. A. (1993a). Responses of Leaf Spectral Reflectance to Plant Stress. *American Journal of Botany, 80*(3), 239-243. doi: 10.2307/2445346

Carter, G. A. (1993b). Responses of leaf spectral reflectance to plant stress. *American Journal of Botany*, 239-243.

Castillo-Santiago, M. A., Ricker, M., & de Jong, B. H. J. (2010). Estimation of tropical forest structure from SPOT-5 satellite images. *International Journal of Remote Sensing, 31*(10), 2767-2782. doi: 10.1080/01431160903095460

Chave, Muller-Landau, H. C., Baker, T. R., Easdale, T. A., Steege, H. t., & Webb, C. O. (2006). REGIONAL AND PHYLOGENETIC VARIATION OF WOOD DENSITY ACROSS 2456 NEOTROPICAL TREE SPECIES. *Ecological Applications, 16*(6), 2356-2367. doi: 10.1890/1051-0761(2006)016[2356:RAPVOW]2.0.CO;2

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., . . . Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia, 145*(1), 87-99. doi: 10.1007/s00442-005-0100-x

Cho, M. A., Skidmore, A. K., & Sobhan, I. (2009). Mapping beech (Fagus sylvatica L.) forest structure with airborne hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation, 11*(3), 201-211. doi: http://dx.doi.org/10.1016/j.jag.2009.01.006

DAFF. (2012). *Integrated Growth and Development Plan*. South Africa: Department of Agriculture, Forestry and Fisheries.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey Iii, J. E. (2000). Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment, 74*(2), 229-239. doi: http://dx.doi.org/10.1016/S0034-4257(00)00113-9

Day, M., Baldauf, C., Rutishauser, E., & Sunderland, T. C. (2014). Relationships between tree species diversity and above-ground biomass in Central African rainforests: implications for REDD. *Environmental Conservation, 41*(01), 64-72.

De'Ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology, 88*(1), 243-251.

Debell, D. S., Cole, T. G., & Whitesell, C. D. (1997). Growth, development, and yield in pure and mixed stands of Eucalyptus and Albizia. *Forest Science, 43*(2), 286-298.

Díaz, S., Hector, A., & Wardle, D. A. (2009). Biodiversity in forest carbon sequestration initiatives: not just a side benefit. *Current Opinion in Environmental Sustainability, 1*(1), 55-60.

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., . . . Prince, S. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment, 79*(2), 305-319.

Dube, T., & Mutanga, O. (2015a). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing, 101*, 36-46.

Dube, T., & Mutanga, O. (2015b). Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS Journal of Photogrammetry and Remote Sensing, 108*, 12-32. doi: http://dx.doi.org/10.1016/j.isprsjprs.2015.06.002

Dube, T., Mutanga, O., Elhadi, A., & Ismail, R. (2014). Intra-and-inter species biomass prediction in a plantation forest: testing the utility of high spatial resolution spaceborne multispectral rapideye sensor and advanced machine learning algorithms. *Sensors, 14*(8), 15348-15370.

Dube, T., Mutanga, O., & Ismail, R. (2016). Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs. *Tropical Ecology, 57*(3), 393-405.

Dye, M., Mutanga, O., & Ismail, R. (2012). Combining spectral and textural remote sensing variables using random forests: predicting the age of Pinus patula forests in KwaZulu-

Natal, South Africa. *Journal of Spatial Science, 57*(2), 193-211. doi: 10.1080/14498596.2012.733620

Eckert, S. (2012). Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data. *Remote Sensing, 4*(4), 810-829.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813.

Erskine, P. D. (2002). Land clearing and forest rehabilitation in the Wet Tropics of north Queensland, Australia. *Ecological Management & Restoration, 3*(2), 135-137.

Erskine, P. D., Lamb, D., & Bristow, M. (2006). Tree species diversity and ecosystem function: can tropical multi-species plantations generate greater productivity? *Forest Ecology and Management, 233*(2), 205-210.

FAO. (2010). Global forest resources assessment 2010: Food and Agriculture Organization of the United Nations Roma.

Felton, A., Lindbladh, M., Brunet, J., & Fritz, Ö. (2010). Replacing coniferous monocultures with mixed-species production stands: An assessment of the potential benefits for forest biodiversity in northern Europe. *Forest Ecology and Management, 260*(6), 939-947. doi: http://dx.doi.org/10.1016/j.foreco.2010.06.011

Forrester, D. I. (2004). Mixed-species plantations of nitrogen-fixing and non-nitrogen-fixing trees.

Forrester, D. I., Bauhus, J., & Cowie, A. L. (2005a). Nutrient cycling in a mixed-species plantation of Eucalyptus globulus and Acacia mearnsii. *Canadian Journal of Forest Research, 35*(12), 2942-2950.

Forrester, D. I., Bauhus, J., & Cowie, A. L. (2005b). On the success and failure of mixed-species tree plantations: lessons learned from a model system of Eucalyptus globulus and Acacia mearnsii. *Forest Ecology and Management, 209*(1), 147-155.

Franklin, S. E., Hall, R. J., Moskal, L. M., Maudie, A. J., & Lavigne, M. B. (2000). Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing, 21*(1), 61-79. doi: 10.1080/014311600210993

Franklin, S. E., Waring, R. H., McCreight, R. W., Cohen, W. B., & Fiorella, M. (1995). Aerial and Satellite Sensor Detection and Classification of Western Spruce Budworm Defoliation in a Subalpine Forest. *Canadian Journal of Remote Sensing, 21*(3), 299-308. doi: 10.1080/07038992.1995.10874624

Franklin, S. E., Wulder, M. A., & Gerylo, G. R. (2001). Texture analysis of IKONOS panchromatic data for Douglas-fir forest age class separability in British Columbia. *International Journal of Remote Sensing, 22*(13), 2627-2632. doi: 10.1080/01431160120769

Freeman, E. A., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2015). Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance 1. *Canadian Journal of Forest Research, 46*(3), 323-339.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Getzin, S., Dean, C., He, F., A. Trofymow, J., Wiegand, K., & Wiegand, T. (2006). Spatial patterns and competition of tree species in a Douglas-fir chronosequence on Vancouver Island. *Ecography, 29*(5), 671-682. doi: 10.1111/j.2006.0906-7590.04675.x

Gitelson, A. A., Buschmann, C., & Lichtenthaler, H. K. (1999). The chlorophyll fluorescence ratio F 735/F 700 as an accurate measure of the chlorophyll content in plants. *Remote Sensing of Environment, 69*(3), 296-302.

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment, 58*(3), 289-298. doi: http://dx.doi.org/10.1016/S0034-4257(96)00072-7

Glenday, J. (2007). Carbon Storage and Sequestration Analysis for the eThekwini Environmental Services Management Plan Open Space System. *Report for the Environmental Planning and Climate Protection Department, eThekwini Municipality, Durban, South Africa*.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters, 4*(4), 379-391. doi: 10.1046/j.1461-0248.2001.00230.x

Gower, S. T., McMurtrie, R. E., & Murty, D. (1996). Aboveground net primary production decline with stand age: potential causes. *Trends in Ecology & Evolution, 11*(9), 378-382.

Gracia, M., Montané, F., Piqué, J., & Retana, J. (2007). Overstory structure and topographic gradients determining diversity and abundance of understory shrub species in temperate forests in central Pyrenees (NE Spain). *Forest Ecology and Management, 242*(2–3), 391-397. doi: http://dx.doi.org/10.1016/j.foreco.2007.01.056

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-3*(6), 610-621. doi: 10.1109/TSMC.1973.4309314

Hlatshwayo, S. T., Mutanga, O., Lottering, R., Kiala, Z., & Ismail, R. (Under revision). An Innovative Technique for Mapping Temporal and Spatial Variation of Forest Aboveground Biomass in the Reforested Buffelsdraai Landfill Site using Texture Combinations Computed from SPOT-6 Pan-sharpened Imagery. *Remote Sensing of Environment*.

Hoogmoed, M., Cunningham, S., Baker, P., Beringer, J., & Cavagnaro, T. (2014). N-fixing trees in restoration plantings: effects on nitrogen supply and soil microbial communities. *Soil Biology and Biochemistry, 77*, 203-212.

Horn, B. K. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE, 69*(1), 14-47.

Huang, X., Liu, S., Wang, H., Hu, Z., Li, Z., & You, Y. (2014). Changes of soil microbial biomass carbon and community composition through mixing nitrogen-fixing species with Eucalyptus urophylla in subtropical China. *Soil Biology and Biochemistry, 73*, 42-48.

Hulvey, K. B., Hobbs, R. J., Standish, R. J., Lindenmayer, D. B., Lach, L., & Perring, M. P. (2013). Benefits of tree mixes in carbon plantings. *Nature Clim. Change, 3*(10), 869-874. doi: 10.1038/nclimate1862

http://www.nature.com/nclimate/journal/v3/n10/abs/nclimate1862.html#supplementary-information

Huston, M. (1979). A general hypothesis of species diversity. *American naturalist*, 81-101.

Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment, 102*(1), 63-73.

Ingram, J. C., Dawson, T. P., & Whittaker, R. J. (2005). Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment, 94*(4), 491-507.

Ismail, R., Mutanga, O., Kumar, L., & Bob, U. (2008). DETERMINING THE OPTIMAL SPATIAL RESOLUTION OF REMOTELY SENSED DATA FOR THE DETECTION OF SIREX NOCTILIO INFESTATIONS IN PINE PLANTATIONS IN

KWAZULU-NATAL, SOUTH AFRICA. *South African Geographical Journal, 90*(1), 22-31. doi: 10.1080/03736245.2008.9725308

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 6): Springer.

Kayitakire, F., Hamel, C., & Defourny, P. (2006). Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sensing of Environment, 102*(3–4), 390-401. doi: http://dx.doi.org/10.1016/j.rse.2006.02.022

Kelsey, K. C., & Neff, J. C. (2014). Estimates of aboveground biomass from texture analysis of Landsat imagery. *Remote Sensing, 6*(7), 6407-6422.

Kelty, M. J. (2006). The role of species mixtures in plantation forestry. *Forest Ecology and Management, 233*(2–3), 195-204. doi: http://dx.doi.org/10.1016/j.foreco.2006.05.011

Khanna, P. K. (1997). Comparison of growth and nutrition of young monocultures and mixed stands of Eucalyptus globulus and Acacia mearnsii. *Forest Ecology and Management, 94*(1–3), 105-113. doi: http://dx.doi.org/10.1016/S0378-1127(96)03971-0

Kumar, L., Skidmore, A. K., & Knowles, E. (1997). Modelling topographic variation in solar radiation in a GIS environment. *International Journal of Geographical Information Science, 11*(5), 475-497.

Lamb, D., Erskine, P. D., & Parrotta, J. A. (2005). Restoration of degraded tropical forest landscapes. *Science, 310*(5754), 1628-1632.

Lawrence, R., Bunn, A., Powell, S., & Zambon, M. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment, 90*(3), 331-336.

Leblois, A., Damette, O., & Wolfersberger, J. (2017). What has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data. *World Development, 92*, 82-102. doi: http://dx.doi.org/10.1016/j.worlddev.2016.11.012

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience, 52*(1), 19-30.

Lexerød, N. L., & Eid, T. (2006). An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management, 222*(1–3), 17-28. doi: http://dx.doi.org/10.1016/j.foreco.2005.10.046

Lin, D., Lai, J., Muller-Landau, H. C., Mi, X., & Ma, K. (2012). Topographic Variation in Aboveground Biomass in a Subtropical Evergreen Broad-Leaved Forest in China. *PLoS One, 7*(10).

Loehle, C., & Donald, L. D. (2000). Strategy Space and the Disturbance Spectrum: A Life‐History Model for Tree Species Coexistence. *The American Naturalist, 156*(1), 14-33. doi: 10.1086/303369

Lottering, R., & Mutanga, O. (2012). Estimating the road edge effect on adjacent Eucalyptus grandis forests in KwaZulu-Natal, South Africa, using texture measures and an artificial neural network. *Journal of Spatial Science, 57*(2), 153-173. doi: 10.1080/14498596.2012.733617

Lottering, R., & Mutanga, O. (2016). Optimising the spatial resolution of WorldView-2 pan-sharpened imagery for predicting levels of Gonipterus scutellatus defoliation in KwaZulu-Natal, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing, 112*, 13-22. doi: http://dx.doi.org/10.1016/j.isprsjprs.2015.11.010

Lu, D. (2005). Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *International Journal of Remote Sensing, 26*(12), 2509-2525. doi: 10.1080/01431160500142145

Lu, D., & Batistella, M. (2005). Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon. *Acta Amazonica, 35*(2), 249-257.

Lu, D., Batistella, M., & Moran, E. (2005). Satellite estimation of aboveground biomass and impacts of forest stand structure. *Photogrammetric Engineering & Remote Sensing, 71*(8), 967-974.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2002). Above-Ground Biomass Estimation of Successional and Mature Forests Using TM Images in the Amazon Basin. In D. E. Richardson & P. van Oosterom (Eds.), *Advances in Spatial Data Handling: 10th International Symposium on Spatial Data Handling* (pp. 183-196). Berlin, Heidelberg: Springer Berlin Heidelberg.

Marceau, D. J., Gratton, D. J., Fournier, R. A., & Fortin, J.-P. (1994). Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment, 49*(2), 105-117. doi: http://dx.doi.org/10.1016/0034-4257(94)90047-7

Materka, A., & Strzelecki, M. (1998). Texture analysis methods–a review *Technical university of lodz, institute of electronics, COST B11 report, Brussels* (pp. 9-11).

Mather, P. M., & Koch, M. (2011a). *Computer processing of remotely-sensed images : an introduction*. Chichester, West Sussex, UK; Hoboken, NJ: Wiley-Blackwell.

Mather, P. M., & Koch, M. (2011b). *Computer processing of remotely-sensed images: an introduction*: John Wiley & Sons.

McRoberts, R. E., Winter, S., Chirici, G., Hauk, E., Pelz, D. R., Moser, W. K., & Hatfield, M. A. (2008). Large-scale spatial patterns of forest structural diversity. *Canadian Journal of Forest Research, 38*(3), 429-438.

Moser, G., Leuschner, C., Hertel, D., Graefe, S., Soethe, N., & Iost, S. (2011). Elevation effects on the carbon budget of tropical mountain forests (S Ecuador): the role of the belowground compartment. *Global Change Biology, 17*(6), 2211-2226. doi: 10.1111/j.1365-2486.2010.02367.x

Moskal, L. M., & Franklin, S. E. (2001). *Classifying multilayer forest structure and composition using high resolution, compact airborne spectrographic imager image texture*. Paper presented at the ASPRS Proceedings, St. Louis.

Mucina, L., & Rutherford, M. C. (2006a). *The vegetation of South Africa, Lesotho and Swaziland*. Strelitzia 19: South African National Biodiversity Institute: Pretoria.

Mucina, L., & Rutherford, M. C. (2006b). The vegetation of South Africa, Lesoto and Swaziland. Strelitzia 19. South African National Biodiversity Institute, Pretoria. *Memoirs of the Botanical Survey of South Africa*.

Mulder, C., Bazeley-White, E., Dimitrakopoulos, P., Hector, A., Scherer-Lorenzen, M., & Schmid, B. (2004). Species evenness and productivity in experimental plant communities. *Oikos, 107*(1), 50-63.

Mura, M., McRoberts, R. E., Chirici, G., & Marchetti, M. (2015). Estimating and mapping forest structural diversity using airborne laser scanning data. *Remote Sensing of Environment, 170*, 133-142.

Mutanga, O., Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation, 18*, 399-406.

Mutanga, O., & Skidmore, A. K. (2004a). Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment, 90*(1), 104-115. doi: http://dx.doi.org/10.1016/j.rse.2003.12.004

Mutanga, O., & Skidmore, A. K. (2004b). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing, 25*(19), 3999-4014. doi: 10.1080/01431160310001654923

Nasi, R., & Wunder, S. (2002). Forest ecosystem services: can they pay our way out of deforestation. Forestry Roundtable, UNFF II, March 11 2002, Costa Rica. *Center for International Forestry Research (CIFOR)*.

Nichol, J. E., & Sarker, L. R. (2011). Improved Biomass Estimation Using the Texture Parameters of Two High-Resolution Optical Sensors. *Geoscience and Remote Sensing, IEEE Transactions on, 49*(3), 930-948. doi: 10.1109/TGRS.2010.2068574

Noriko, H., Martin, H., Veronique De, S., Ruth, S. D. F., Maria, B., Louis, V., . . . Erika, R. (2012). An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters, 7*(4), 044009.

Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution, 116*(3), 381-389. doi: http://dx.doi.org/10.1016/S0269-7491(01)00214-7

Ozdemir, I., & Karnieli, A. (2011). Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *International Journal of Applied Earth Observation and Geoinformation, 13*(5), 701-710. doi: http://dx.doi.org/10.1016/j.jag.2011.05.006

Pachavo, G., & Murwira, A. (2014). Land-use and land tenure explain spatial and temporal patterns in terrestrial net primary productivity (NPP) in Southern Africa. *Geocarto International, 29*(6), 671-687. doi: 10.1080/10106049.2013.837101

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., . . . Canadell, J. G. (2011). A large and persistent carbon sink in the world's forests. *Science, 333*(6045), 988-993.

Peerbhay, K., Mutanga, O., Lottering, R., & Ismail, R. (2016). Mapping Solanum mauritianum plant invasions using WorldView-2 imagery and unsupervised random forests. *Remote Sensing of Environment, 182*, 39-48. doi: http://dx.doi.org/10.1016/j.rse.2016.04.025

Peichl, M., & Arain, M. A. (2006). Above- and belowground ecosystem biomass and carbon pools in an age-sequence of temperate pine plantation forests. *Agricultural and Forest Meteorology, 140*(1–4), 51-63. doi: http://dx.doi.org/10.1016/j.agrformet.2006.08.004

Petchey, O. L., & Gaston, K. J. (2002). Functional diversity (FD), species richness and community composition. *Ecology Letters, 5*(3), 402-411. doi: 10.1046/j.1461-0248.2002.00339.x

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems, 9*(2), 181-199.

Quinn, P., Beven, K., & Lamb, R. (1995). The in (a/tan/β) index: How to calculate it and how to use it within the topmodel framework. *Hydrological processes, 9*(2), 161-182.

Roberts, M. R., & Gilliam, F. S. (1995). Patterns and Mechanisms of Plant Diversity in Forested Ecosystems: Implications for Forest Management. *Ecological Applications, 5*(4), 969-977. doi: 10.2307/2269348

Rubner, Y., Puzicha, J., Tomasi, C., & Buhmann, J. M. (2001). Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Computer Vision and Image Understanding, 84*(1), 25-43. doi: http://dx.doi.org/10.1006/cviu.2001.0934

Saremi, H., Kumar, L., Turner, R., & Stone, C. (2014). Airborne LiDAR derived canopy height model reveals a significant difference in radiata pine (Pinus radiata D. Don) heights based on slope and aspect of sites. *Trees, 28*(3), 733-744.

Sarker, L., Nichol, J., & Mubin, A. (2013). Potential of Multiscale Texture Polarization Ratio of C-band SAR for Forest Biomass Estimation. In A. Abdul Rahman, P. Boguslawski, C. Gold & M. N. Said (Eds.), *Developments in Multidimensional Spatial Data Models* (pp. 69-83): Springer Berlin Heidelberg.

Sarker, L. R., & Nichol, J. E. (2011). Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sensing of Environment, 115*(4), 968-977. doi: http://dx.doi.org/10.1016/j.rse.2010.11.010

Shamsoddini, A., Trinder, J. C., & Turner, R. (2013). Pine plantation structure mapping using WorldView-2 multispectral image. *International Journal of Remote Sensing, 34*(11), 3986-4007. doi: 10.1080/01431161.2013.772308

Shirima, D. D., Totland, Ø., Munishi, P. K., & Moe, S. R. (2015). Relationships between tree species richness, evenness and aboveground carbon storage in montane forests and miombo woodlands of Tanzania. *Basic and Applied Ecology, 16*(3), 239-249.

Simpson, W. (1996). Method to estimate dry-kiln schedules and species grouping *Tropical and temperate hardwoods.* (Vol. Research Paper FPL-548). Madison, WI.

Sousa, A. M. O., Gonçalves, A. C., Mesquita, P., & Marques da Silva, J. R. (2015). Biomass estimation with high resolution satellite images: A case study of Quercus rotundifolia. *ISPRS Journal of Photogrammetry and Remote Sensing, 101*, 69-79. doi: http://dx.doi.org/10.1016/j.isprsjprs.2014.12.004

St-Louis, V., Pidgeon, A. M., Radeloff, V. C., Hawbaker, T. J., & Clayton, M. K. (2006). High-resolution image texture as a predictor of bird species richness. *Remote Sensing of Environment, 105*(4), 299-312. doi: http://dx.doi.org/10.1016/j.rse.2006.07.003

Strait, M., Rahmani, S., & Merkurev, D. (2008). Evaluation of pan-sharpening methods. UCLA Department of Mathematics: Los Angeles, USA.

Turner, D. P., Cohen, W. B., Kennedy, R. E., Fassnacht, K. S., & Briggs, J. M. (1999). Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sensing of Environment, 70*(1), 52-68.

Turner, W. R., Oppenheimer, M., & Wilcove, D. S. (2009). A force to fight global warming. *Nature, 462*(7271), 278-279.

Tuttle, E. M., Jensen, R. R., Formica, V. A., & Gonser, R. A. (2006). Using remote sensing image texture to study habitat use patterns: a case study using the polymorphic white-throated sparrow (Zonotrichia albicollis). *Global Ecology and Biogeography, 15*(4), 349-357. doi: 10.1111/j.1466-822X.2006.00232.x

UN. (2015). WMO confirms 2015 as hottest year on record. *Sustainable Development Knowledge Platform.* Retrieved 01/20/2016, 2016, from https://www.un.org/sustainabledevelopment/blog/2016/01/wmo-confirms-2015-as-hottest-year-on-record/

Van Vuuren, N., Banks, C., & Stöhr, H.-P. (1978). *Shrinkage and density of timbers used in the Republic of South Africa*. (Bulletin 57). Pretoria, South Africa.

Wang, Q., Xu, Y., Lu, Z., Bao, D., Guo, Y., Lu, J., . . . Jiang, M. (2014). Disentangling the effects of topography and space on the distributions of dominant species in a subtropical forest. *Chinese Science Bulletin, 59*(35), 5113-5122. doi: 10.1007/s11434-014-0453-9

Watt, M. S., Clinton, P. W., Whitehead, D., Richardson, B., Mason, E. G., & Leckie, A. C. (2003). Above-ground biomass accumulation and nitrogen fixation of broom (Cytisus scoparius L.) growing with juvenile Pinus radiata on a dryland site. *Forest Ecology and Management, 184*(1), 93-104.

Weiner, J., & Solbrig, O. T. (1984). The meaning and measurement of size hierarchies in plant populations. *Oecologia, 61*(3), 334-336. doi: 10.1007/bf00379630

Woollen, E., Ryan, C. M., & Williams, M. (2012). Carbon Stocks in an African Woodland Landscape: Spatial Distributions and Scales of Variation. *Ecosystems, 15*(5), 804-818. doi: 10.1007/s10021-012-9547-x

Wright, J. S. (2002). Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia, 130*(1), 1-14. doi: 10.1007/s004420100809

Wulder, M. A., LeDrew, E. F., Franklin, S. E., & Lavigne, M. B. (1998). Aerial Image Texture Information in the Estimation of Northern Deciduous and Mixed Wood Forest Leaf Area Index (LAI). *Remote Sensing of Environment, 64*(1), 64-76. doi: http://dx.doi.org/10.1016/S0034-4257(97)00169-7

Xu, Y., Franklin, S. B., Wang, Q., Shi, Z., Luo, Y., Lu, Z., . . . Jiang, M. (2015). Topographic and biotic factors determine forest biomass spatial distribution in a subtropical mountain moist forest. *Forest Ecology and Management, 357*, 95-103. doi: http://dx.doi.org/10.1016/j.foreco.2015.08.010

Yuan, X., King, D., & Vlcek, J. (1991). Sugar maple decline assessment based on spectral and textural analysis of multispectral aerial videography. *Remote Sensing of Environment, 37*(1), 47-54. doi: http://dx.doi.org/10.1016/0034-4257(91)90049-C