



UNIVERSITY OF TM
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

**MAPPING NATURAL FOREST COVER, TREE SPECIES
DIVERSITY AND CARBON STOCKS OF A SUBTROPICAL
AFROMONTANE FOREST USING REMOTE SENSING**

By

ENOCH GYAMFI-AMPADU

**Thesis submitted to the College of Agriculture, Engineering and
Sciences, at the University of KwaZulu-Natal, in fulfilment of the
academic requirement of the degree of Doctor of Philosophy
(PhD) in Environmental Science**

(Specialization: Remote Sensing)

Westville Campus

South Africa

May 2021

PREFACE

This research work was carried out at the Discipline of Environmental Sciences, School of Agriculture, Earth and Environmental Sciences (SEAS), University of KwaZulu-Natal, Westville Campus, Durban, South Africa from February 2018 to May 2021, under the supervision of Prof. Michael Gebreslasie of the School of Agriculture, Earth and Environmental Science of the University of KwaZulu-Natal, Westville Campus, Durban, South Africa.

I would like to declare that the contents of this research work have not been submitted in any form to any other university. It therefore represents my original work, except where due acknowledgements are made.

Enoch Gyamfi-Ampadu. Signed:



Date: 9 May 2021

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

Prof. Michael T. Gebreslasie. Signed:

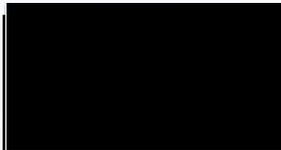


Date: 9 May 2021

DECLARATION 1: PLAGIARISM

I, Enoch Gyamfi-Ampadu, declare that:

- i. the research reported in this dissertation, except where otherwise indicated or acknowledged is my original work;
- ii. this dissertation has not been submitted in full or in part for any degree or examination to any other university;
- iii. this dissertation does not contain other persons' data, pictures, graphs or other information unless specifically acknowledged as being sourced from other persons;
- iv. this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - v. their words have been re-written but the general information attributed to them has been referenced;
 - vi. where their exact words have been used, their writing has been placed inside quotation marks, and referenced;
 - vii. where I have used material for which publications followed, I have indicated in detail my role in the work;
- viii. this dissertation is primarily a collection of material, published as journal articles. In some cases, additional material has been included;
- ix. this dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.



Signed: Enoch Gyamfi-Ampadu

Date: 9 May 2021

DECLARATION 2: PUBLICATIONS AND MANUSCRIPTS

1. **Gyamfi-Ampadu, E.;** Gebreslasie, M. (2021): Two Decades Progress on the Application of Remote Sensing for Monitoring Tropical and Sub-Tropical Natural Forests: A Review, *Forests*, (12), 739. <https://doi.org/10.3390/f12060739>
2. **Gyamfi-Ampadu, E.;** Gebreslasie, M.; Mendoza-Ponce, A. (2020): Mapping Natural Forest Cover Using Satellite Imagery of Nkandla Forest Reserve, Kwazulu-Natal, South Africa, *Remote Sensing Applications: Society and Environment*, (18). 100302. <https://doi.org/10.1016/j.rsase.2020.100302>
3. **Gyamfi-Ampadu, E.;** Gebreslasie, M.; Mendoza-Ponce, A. (2021): Multi-Decadal Spatial and Temporal Forest Cover Change Analysis of Nkandla Natural Reserve, South Africa, *Journal of Sustainable Forestry*. <https://doi.org/10.1080/10549811.2021.1891441>
4. **Gyamfi-Ampadu, E.;** Gebreslasie, M.; Mendoza-Ponce, A.: Mapping of Aboveground Carbon Stock in Sub-Tropical Natural Forest using Sentinel 2 Satellite Imagery and Random Forest Algorithm. (Under Review with Canadian Journal of Remote Sensing)
5. **Gyamfi-Ampadu, E.;** Gebreslasie, M.; Mendoza-Ponce, A. (2021): Evaluating Multi-Sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction. *Remote Sensing*, (13), 1033. <https://doi.org/10.3390/rs13051033>
6. **Gyamfi-Ampadu, E.;** Gebreslasie, M. Identifying the Best Season for Predicting Tree Species Diversity using Sentinel 2 Satellite Imagery and Random Forest Algorithm. (Under Review with Journal of Vegetation Science)



Signed: Enoch Gyamfi-Ampadu

Date: 9 May 2021

DEDICATION

This thesis is dedicated to my lovely wife, Wilhemina Gyamfi-Ampadu and beloved son, Jedidiah Gyamfi-Ampadu.

ACKNOWLEDGEMENTS

I am immensely thankful to the Most-High God for His of grace, wisdom, providence, and protection throughout my life and for leading me through my PhD studies. I am grateful to Him for taking care of my wife, son, parents and siblings all this while. Praise, honour and adoration be to His Holy Name for what He has done and yet to do in my life.

I am very grateful to my academic supervisor, Prof. Michael Gebreslasie, for his insights, constructive criticism and guidance during my research work. Your motivation, tolerance and guidance during my low moments all contributed to the success of my work. I cannot forget to mention your kindness towards me. I appreciate all that you have done for me.

I would also like to show my appreciation to Dr. Alma Mendoza-Ponce for her encouragement and insights on my research.

I extend my profound gratitude to my lovely wife, Wilhemina for holding the fort and taking care of our beloved son Jedidiah during my time of the study. You have shown me love and kindness since day one and I couldn't have made it without all your sacrifice and prayers.

I am thankful to my parents, Mr. Ampadu and Mad. Regina, and siblings Obed, Ellen and Flora for your love, prayers and care. I am grateful.

I would like to acknowledge all my friends and colleagues, Isaiah, Patrick, Jonathan, Eli, Amos, Gyasi, Frederick, Jimmy, Cosmos, Princess, Julius, Osa, Magugu, Jabulile, Gugu, Jemimah, Yemisi, Sinikiwe, Augustine, Armstrong, Charles, Phyllis, Vadhini, Charity, Sam, Perry, for your prayers and support.

I also thank Ms Sharon, Mr Zungu and all the field staff of Ezemvelo KZN Wildlife for their assistance during my data collection at the Nkandla Forest Reserve.

I acknowledge the support provided by Dr. Njoya for me during my study. I thank Mrs Susan, Dudu, Prievan, Simphiwe and other College and Departmental administrative staff for their assistance and support as well.

I thank the National Research Foundation (NRF) of South Africa for funding to my PhD studies.

ABSTRACT

Natural forests cover about a third of terrestrial landmass and provides benefits such as carbon sequestration, and regulation of biogeochemical cycles. It is essential that adequate information is available to support forest management. Remote Sensing imageries provide data for mapping natural forests. Hence, our study aimed at mapping the Nkandla Forest Reserve attributes with Remote Sensing imageries. Quantitative information on the forest attributes is non-existent for many of these forests, including the sub-tropical Afromontane Nkandla Forest Reserve. This does not support scientific and evidence based natural forest management. A review of literature revealed that progress has been made in Remote Sensing monitoring of natural forest attributes. The Random Forest (RF) and Support Vector Machine (SVM) were applied to Landsat 8 in classifying the land use land cover (LULC) classes of the forest. Each of the algorithms produced higher accuracy of above 95% with the SVM performing slightly better than the RF. The SVM, Markov Chain and Multi-Layer Perceptron Neural Network (MLPNN) were adopted for a spatiotemporal change detection over the last 30 years at decadal interval for the forest. There were consistent changes in each of the four LULC classes. The study further conducted a forecasting of LULC distribution for 2029. Aboveground carbon (AGC) estimation was carried out using Sentinel 2 imagery and RF modelling. Four models made up smade of Sentinel 2 products could successfully map the AGC with high accuracies. The last two studies focused on tree species diversity with the first evaluating the influence of spatial and spectral resolution on prediction accuracies by comparing the PlanetScope, RapidEye, Sentinel 2 and Landsat 8. Both the spatial and spectral resolution were found to influence accuracies with the Sentinel 2 emerging as the best imagery. The second aspect focused on identifying the best season for the prediction of tree species diversity. Summer imagery emerged as the best season and the winter being the least performer. Overall, our study indicates that Remote Sensing imageries could be used for successful mapping of natural forest attributes. The outputs of our studies could also be of interest to forest managers and Remote Sensing experts.

TABLE OF CONTENTS

PREFACE.....	i
DECLARATION 1: PLAGIARISM	ii
DECLARATION 2: PUBLICATIONS AND MANUSCRIPTS	iii
DEDICATION.....	iv
ACKNOWLEDGEMENTS.....	v
ABSTRACT.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	xii
LIST OF TABLES.....	xiv
CHAPTER 1. INTRODUCTION	1
1.1 Introduction.....	1
1.1 Specific objectives.....	2
1.2 Structure of the thesis.....	2
1.3 Brief description of chapters	3
1.3 Limitations of the study.....	4
References.....	4
CHAPTER 2: TWO DECADES PROGRESS ON THE APPLICATION OF REMOTE SENSING FOR MONITORING TROPICAL AND SUB-TROPICAL NATURAL FORESTS: A REVIEW	8
Abstract.....	9
2.1 Introduction	9
2.2 Methodology	10
2.3. Results	12
2.3.1 Aboveground biomass and carbon estimation using Remote Sensing	12
2.3.2 Tree species identification using Remote Sensing	15
2.3.3 Tree species diversity mapping using Remote Sensing.....	17
2.3.4 Forest cover mapping and change detection with Remote Sensing	18
2.4 Discussion	19
2.5 Conclusion	24
Acknowledgment	25
Funding	25
Conflict of Interest	25
References.....	25

CHAPTER 3. MAPPING NATURAL FOREST COVER USING SATELLITE IMAGERY OF NKANDLA FOREST RESERVE, KWAZULU-NATAL, SOUTH AFRICA.....	33
Abstract.....	34
3.1. Introduction.....	34
3.2 Materials and Methods.....	37
3.1 Study Area.....	37
3.2 Definition of forest cover classes.....	38
3.3 Data Used.....	39
3.3.1 Remote Sensing data acquisition and preprocessing.....	39
3.3.2 Reference data collection and processing.....	40
3.4 Image classification.....	41
3.4.1 Random Forest.....	41
3.4.2 Support Vector Machine.....	42
3.4.3 Accuracy Assessment.....	43
3.5 Results.....	44
3.5.1 Spatial extent of forest thematic cover.....	44
3.5.2 Comparison of RF and SVM mapping accuracy.....	47
3.5.3 General Variable Importance of Landsat 8 bands.....	48
3.5.4 Class by class variable importance.....	49
3.6 Discussion.....	51
3.7 Conclusions.....	55
Acknowledgment.....	55
Funding.....	55
Conflict of Interest.....	56
References.....	56
CHAPTER 4. MULTI-DECADAL SPATIAL AND TEMPORAL FOREST COVER CHANGE ANALYSIS OF NKANDLA NATURAL RESERVE, SOUTH AFRICA.....	65
Abstract.....	66
4.1 Introduction.....	66
4.2 Materials and methods.....	69
4.2.1 Study area.....	69
4.3 Data used.....	70
4.3.1 Field data.....	70
4.3.2 Remote Sensing imagery used and pre-processing.....	70
4.3.3 Environmental parameters data.....	71

4.4 Image classification and Accuracy Assessment.....	71
4.5 Historical Land Cover Change Detection	72
4.5.1 Transition Probability Matrices and Potential Maps	72
4.5.2 Land cover distribution forecasting.....	73
4.5.3 Validation and Projection	74
4.6 Results	74
4.6.1 Spatial distribution of cover types	74
4.7 Spatiotemporal Analysis	76
4.7.1 General forest cover changes.....	76
4.7.2 Decadal changes within forest cover types.....	78
4.7.3 Land cover predictions	79
4.8. Discussion.....	80
4.9. Conclusion	84
Acknowledgment	85
Funding	85
Conflict of Interest	85
References.....	85
Appendix 4.1.....	91
Appendix 4.2.....	92
Appendix 4.3.....	93
Appendix 4.4.....	94
CHAPTER 5: MAPPING THE ABOVEGROUND CARBON STOCK OF SUB-TROPICAL NATURAL FOREST USING SENTINEL 2 SATELLITE IMAGERY AND RANDOM FOREST ALGORITHM.....	95
Abstract.....	96
5.1 Introduction.....	96
5.2 Materials and Methods	98
5.2.1 Study Area.....	98
5.3 Field inventory and AGC estimation	100
5.3.1 Field data statistics.....	100
5.4 Remote Sensing imagery data	100
5.4.1 Spectral bands variables	101
5.4.2 Spectral Indices variables	101
5.4.3 Variable selection	103
5.5 Random Forest Regression Modelling.....	104

5.6 Models evaluation and variables importance	104
5.7 Results	105
5.7.1 Random Forest models performance	105
5.7.2 AGC map	106
5.7.3 AGC prediction variable importance	106
5.8 Discussion	107
5.9 Conclusion	111
Acknowledgment	111
Funding	112
Conflict of Interest	112
References	112
CHAPTER 6: EVALUATING MULTI-SENSORS SPECTRAL AND SPATIAL RESOLUTIONS FOR TREE SPECIES DIVERSITY PREDICTION	119
Abstract	120
6.1 Introduction	120
6.2. Materials and Methods	123
6.2.1. Study Area	123
6.2.2. Field Inventory and Diversity Indices Estimation	124
6.2.3. Field Inventory Data Analysis	125
6.2.4. Remote Sensing Data	126
6.2.5. Important Variables Selection	129
6.2.6. Random Forest Regression Modelling	129
6.2.7. Models Evaluation	130
6.2.8. Variable Importance	130
6.3. Results	130
6.3.2. Sensor Performance Evaluation	130
6.3.3. Predicting Important Variables	134
6.4. Discussion	136
6.5. Conclusions	140
Acknowledgment	140
Funding	140
Conflict of Interest	141
References	141
CHAPTER 7: IDENTIFYING THE BEST SEASON FOR PREDICTING TREE SPECIES DIVERSITY USING SENTINEL 2 SATELLITE IMAGERY AND RANDOM FOREST ALGORITHM	147

Abstract.....	148
7.1 Introduction.....	148
7.2 Materials and Methods.....	150
7.2.1 Study Area.....	150
7.3 Tree Data Collection and Diversity Estimation.....	151
7.4 Remote Sensing imagery data.....	152
7.5 Deriving of texture variables.....	153
7.5.1 Predicting Variables Selection.....	155
7.5.2 Random Forest regression models.....	155
7.5.3 Model evaluation and tree species diversity map production.....	156
7.6 Results.....	157
7.6.1 Field Measured Data Analysis.....	157
7.6.2 Image Model Performance.....	157
7.6.3 Important variables.....	159
7.7 Discussion.....	160
7.9 Conclusion.....	163
Acknowledgment.....	163
Funding.....	164
Conflict of Interest.....	164
References.....	164
CHAPTER 8: SYNTHESIS REPORT OF THE THESIS.....	168
8.1 Synthesis of the chapters.....	169
8.2 Two Decades Progress on The Application of Remote Sensing for Monitoring Tropical and Sub-Tropical Natural Forests: A Review.....	169
8.3 Mapping Natural Forest Cover Using Satellite Imagery of Nkandla Forest Reserve, Kwazulu-Natal, South Africa.....	170
8.4 Multi-Decadal Spatial and Temporal Forest Cover Change Analysis of Nkandla Natural Reserve, South Africa.....	170
8.5 Mapping of Aboveground Carbon Stock in Sub-Tropical Natural Forest using Sentinel 2 Satellite Imagery and Random Forest Algorithm.....	171
8.6 Evaluating Multi-Sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction.....	171
8.7 Identifying the Best Season for Predicting Tree Species Diversity using Sentinel 2 Satellite Imagery and Random Forest Algorithm.....	172

LIST OF FIGURES

Figure 2.1: Schematic diagram of the literature selection process	11
Figure 2.2: Number of studies carried out for each thematic area. Note: FCC and CD: Forest cover change and change detection, TSD: Tree species diversity, AGC and AGB: Aboveground carbon and Aboveground biomass, and TSI: Tree species identification.....	12
Figure 3.1: Map of the study area. Note: A is the Landsat 8 satellite image of the Nkandla Forest Reserve, and B is a map of South Africa indicating the location of the study area.	38
Figure 3.2: Thematic map of the Nkandla Forest Reserve. A is map produce by the SVM algorithm and B is a map produced by the RF algorithm.	46
Figure 3.3: General variable importance ranking of the Landsat 8 bands using the MDA. The bands B1(Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Short Wave infra-red 1), B7 (Short wave infra-red 2) and B9 (Cirrus) are normal numbering order.	49
Figure 3.4: Thematic class-specific variable importance ranking of the Landsat 8 spectral bands by the SVM algorithm. The bands B1(Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Short Wave infra-red 1), B7 (Short wave infra-red 2) and B9 (Cirrus) are ranked based on their importance value in mapping the classes.	50
Figure 3.5: Thematic class-specific variable importance ranking of the Landsat 8 bands by the RF algorithm. The bands B1(Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Short Wave infra-red 1), B7 (Short wave infra-red 2) and B9 (Cirrus) are ranked based on their importance value in mapping the classes.	51
Figure 4.1: Map of the study area. Note: Map A is a satellite image of the Nkandla Forest Reserve, and Map B is a map of South Africa indicating the location of the forest.	69
Figure 4.2: Land cover maps of the forest reserve produced from the classified images of 1989, 1999, 2009, and 2019.....	75
Figure 4.3: Predicted land cover distribution map of the forest for 2029.....	80
Figure 5.1: Map of the study area. Note: A is a map of the Nkandla Forest Reserve, B is a map of the KwaZulu-Natal (KZN) province and C is the map of South Africa indicating the location of KZN province and the Nkandla Forest Reserve.	99
Figure 5.2: Scatter plots indicating the relationship between the observed and predicted AGC. The line of best fit is represented by the blue line with the shaded region representing the 95% confidence interval. Note: A represents the Combined Variables model (A), B represents the Red-edge vegetation indices model, C represents the Spectral bands model and D represents the NIR vegetation indices.....	105
Figure 5.3: Final AGC distribution map produced for the Nkandla Forest Reserve.	106
Figure 5.4: Variable importance for the variables of the four RF model of models. A is for the Combined Variables model, B is for the Red-edge vegetation indices, C is for the Spectral bands model and D is for the NIR vegetation indices model. The variables have been ranked based on the least to the most important variable from left to right.	107

Figure 6.1: Map of the study area. Note: A is the Nkandla Forest Reserve, B is the map of South Africa indicating the location of KZN province and the forest and C is the map of Africa indicating the location of South Africa..... 124

Figure 6.2: Scatter plot for the Shannon Index prediction. (A) is for Sentinel 2, (B) is for RapidEye, (C) is for PlanetScope, and (D) is for Landsat 8. The blue line is the line of best fit and the dashed line is the 1:1 line as shown on the individual plots. 133

Figure 6.3: Scatter plot for the Simpson Index prediction. (A) is for Sentinel 2, (B) is for RapidEye, (C) is for PlanetScope, and (D) is for Landsat 8. The blue line is the line of best fit and the dashed line is the 1:1 line as shown on the individual plots. 133

Figure 6.4: Scatter plot for the Species richness predictions. (A) is for Sentinel 2, (B) is for RapidEye, (C) is for PlanetScope, and (D) is for Landsat 8. The blue line is the line of best fit and the dashed line is the 1:1 line as shown on the individual plots. 134

Figure 7.1: Map of the study area. Note: A represents the Nkandla Forest Reserve, B is a map of South Africa indicating the location of KZN province and the forest reserve and C is an Africa indicating the location of South Africa. 151

Figure 7.2: Scatter plots for the tree species diversity prediction. (A) is for the Summer imagery, (B) is for Spring imagery, (C) is for Autumn imagery, and (D) is for Winter imagery. The blue line is the line of best-fit while the dashed line is the 1:1 as illustrated in the individual plots..... 158

Figure 7.3: The diversity map produced for the Nkandla Forest Reserve. 158

LIST OF TABLES

Table 2.1: Aboveground biomass and carbon estimation using Remote Sensing	14
Table 2.2: Tree species identification using Remote Sensing	16
Table 2.3: Research carried out on tree species diversity	18
Table 2.4: Forest cover mapping and change detection.....	19
Table 3.1: Definition of thematic cover classes used in the study.....	39
Table 3.2: Training and validation data sets used in the study.	41
Table 3.3: Spatial coverage of each thematic cover classes produced by the SVM and RF Algorithms.	45
Table 3.4: Confusion Matrix for SVM and RF indicating the Overall Accuracy (OA), Producer’s Accuracy (PA), User’s Accuracy (UA) and kappa coefficient	47
Table 3.5: McNemar’s test results for comparison of Random Forest (RF) and Support Vector Machine (SVM) Algorithms.	48
Table 4.1: Characteristics of the Landsat images used.	71
Table 4.2: Accuracy assessment of forest cover classifications for 1989, 1999, 2009, and 2019 images.	76
Table 4.3: Decadal land cover changes that occurred for the Nkandla Forest Reserve from 1989 to 2019.	77
Table 4.4: Actual land cover of 2019 and predicted land cover distribution of the forest for 2029.....	80
Table 4.5: General land cover changes of the forest reserve between 1989 and 2019.	91
Table 4.6: Land Cover Transition Matrix of the Forest Reserve for 1989-1999.....	92
Table 4.7: Land Cover Transition Matrix of the Forest Reserve for 1999-2009.....	93
Table 4.8: Land Cover Transition Matrix of the Forest Reserve for 2009-2019.....	94
Table 5.1: Details of spectral bands.....	101
Table 5.2: Details of NIR and Red-edge region vegetation indices variables.....	102
Table 5.3: Descriptive statistics of field measured data.	100
Table 5.4: Model accuracies.	105
Table 6.1: Details of the spectral and spatial resolution of satellite imageries.....	128
Table 6.2: Descriptive statistics of Shannon Index (H'), Simpson Index (D_1) and Species richness (S) produced from the field inventory data.....	126
Table 6.3: RF sensor model accuracies for the Sentinel 2, RapidEye, PlanetScope and the Landsat 8 for Shannon Index, Simpson Index and the Species Richness.....	131
Table 6.4: The statistical analysis of the prediction made with the RF for each of the images under the Shannon Index, Simpson Index and the Species richness.	132

Table 6.5. Variable importance ranking for the Sentinel 2 image model under the Shannon Index, Simpson Index and the Species richness.	135
Table 6.6. Variable importance ranking for the RapidEye image model under the Shannon Index, Simpson Index and the Species richness.	135
Table 6.7: Variable importance ranking for the PlanetScope image model under the Shannon Index, Simpson Index and the Species richness.	136
Table 6.8. Variable importance ranking for the Landsat 8 image model under the Shannon Index, Simpson Index and the Species Richness.....	136
Table 7.1: Details of Sentinel 2 spectral bands used in the analysis.....	152
Table 7.2: Mathematical functions of the texture variables and their descriptions	154
Table 7.3: Descriptive statistics of the field inventory data.....	157
Table 7.4: RF model accuracies of the four seasonal imageries.....	157
Table 7.5: The variable importance ranking of the Summer imagery.	159
Table 7.6: The variable importance ranking of the Spring imagery.....	159
Table 7.7: Variable importance ranking for Autumn imagery.	160
Table 7.8: Variable importance ranking for Winter imagery.....	160
Table 6.3: RF sensor model accuracies for the Sentinel 2, RapidEye, PlanetScope and the Landsat 8 for Shannon Index, Simpson Index and the Species Richness.....	131
Table 6.4: The statistical analysis of the prediction made with the RF for each of the images under the Shannon Index, Simpson Index and the Species richness.	132
Table 6.5. Variable importance ranking for the Sentinel 2 image model under the Shannon Index, Simpson Index and the Species richness.	135
Table 6.6. Variable importance ranking for the RapidEye image model under the Shannon Index, Simpson Index and the Species richness.	135
Table 6.7: Variable importance ranking for the PlanetScope image model under the Shannon Index, Simpson Index and the Species richness.	136
Table 6.8. Variable importance ranking for the Landsat 8 image model under the Shannon Index, Simpson Index and the Species Richness.....	136
Table 7.1: Details of Sentinel 2 spectral bands used in the analysis.....	152
Table 7.2: Mathematical functions of the texture variables and their descriptions	154
Table 7.3: Descriptive statistics of the field inventory data.....	157
Table 7.4: RF model accuracies of the four seasonal imageries.....	157
Table 7.5: The variable importance ranking of the Summer imagery.	159
Table 7.6: The variable importance ranking of the Spring imagery.....	159
Table 7.7: Variable importance ranking for Autumn imagery.	160
Table 7.8: Variable importance ranking for Winter imagery.	160

CHAPTER 1. INTRODUCTION

1.1 Introduction

Natural forest ecosystems are home to about 80% of terrestrial biodiversity (Aerts and Honnay, 2011) and provide both tangible and non-tangible benefits locally and nationally. The tangible benefits include timber, plant medicine and non-timber forest products (NTFPs) whereas non-tangible benefits include carbon sequestration, amelioration of local climate, regulation of water and biogeochemical cycles and protection of habitats. Forests are also vital for the socio-economic and the socio-cultural wellbeing of society that either depends on them directly or indirectly (Agrawal et al., 2013). Many natural forests are declining due to natural phenomena such as climate change and human disturbances such as deforestation (Nkonya et al., 2016). Hence, the development of robust and cost-effective mapping research methods that provide adequate information on status and attributes of natural forest is essential for current sustainable forest management demands.

Remote Sensing mapping of natural forests is a means of obtaining the prerequisite information that deepens understanding of forest dynamics. Furthermore, the information provided on the natural forest through Remote Sensing mapping facilitates decision making, conservation initiatives, resources restoration and mitigation measures by local managers, government agencies, civil society organizations, advocacy groups and scientists (Keenan et al., 2015). The specific forest attributes for which Remote Sensing data and methods are applied for mapping includes forest structure (Ahmed et al., 2015, St-Onge and Grandin, 2019), forest cover (Kanniah et al., 2016, Schlund et al., 2017, Suzuki et al., 2018, Xie et al., 2019, Yin et al., 2017), the chemical component of canopy and leaves [for instance foliar Nitrogen] (Mutowo et al., 2018a), deforestation and change detection (Nandy et al., 2011, Münch et al., 2017), biomass and carbon estimation (Asner and Mascaro, 2014, Chenge and Osho, 2018, Gizachew et al., 2016, Shen et al., 2020), invasive plant monitoring (Müllerová et al., 2013, Rajah et al., 2019), tree species diversity (Fundisi et al., 2020, Grabska et al., 2019a, Mutowo and Murwira, 2012a), and individual tree species identification (Cho et al., 2015, Cross et al., 2019, Fang et al., 2020, Ferreira et al., 2019) and urbanization (Rimal et al., 2018, Addae and Oppelt, 2019).

Many advances have been observed in recent years in new generation sensors [optical and active] (Fassnacht et al., 2016); for instance, improvement in the spectral and spatial resolution (Omer et al., 2015), the sensitivity of spectral band to vegetation (Xu et al., 2018, Roy et al., 2014, Mutanga et al., 2016), temporal and radiometric resolution that enables them to capture forest growth stages [phenological monitoring] (Phiri and Morgenroth, 2017, Mutanga et al., 2016). As a result of these advances, there has been an increase in Remote Sensing applications to the mapping of natural forests across many climatic zones. It likely that more advances will be made in sensors based on specific demands and more research may be expected moving forward in a bid to further improve Remote Sensing studies on natural forests.

Machine learning algorithms and their application to forest monitoring and mapping have increased in recent years (Adam et al., 2014a). They are normally grouped into parametric (assumes a normal distribution of data) and non-parametric (does not assume a normal distribution of data) algorithms (Fassnacht et al., 2016). Examples of the parametric ones include Maximum Likelihood (ML), various types of Discriminant Analysis and Generalized Linear Model whereas examples of non-parametric algorithms include RF, SVM and Neural

Networks. The amalgamation of these algorithms and Remote Sensing data and products helps to improve the capabilities and accuracies of classification and predicting models in mapping operations for natural forests.

Although the number of studies in Remote Sensing mapping of natural forests has increased in many regions and countries with enhanced management and conservation measures, the same cannot be said for South Africa. Quantitative and qualitative information of natural forest attributes and characteristics like forest cover extent, carbon stocks estimation, tree species diversity data and spatial maps, are non-existent for many natural forests, which is a major hindrance to sustainable forest management planning and decision making. As such, management systems that adopt a scientific based approach based on Remote Sensing mapping information are lacking. The mapping of many units of natural forests across the country could culminate into databases on carbon levels, forest cover extent, tree species diversity scales and other important forest attributes which would enhance management and conservation measures. Hence, our study aimed to apply Remote Sensing imagery and technology to map the forest cover, estimate carbon stocks and predict tree species diversity of the Nkandla Forest Reserve, which is an Afromontane sub-tropical natural forest in the KwaZulu-Natal province of South Africa. The output of our study is expected to provide the requisite information that will contribute immensely to scientifically based forest resource management and conservation. It is also expected to contribute knowledge to future Remote Sensing applications research that would be conducted for similar sub-tropical and other forest types locally, nationally and globally. Modelling and classification methods as well as spectral products that contributed to satisfactory accuracies in the various aspects of our studies could also be adopted in other studies.

To achieve the aim of the study, the specific objectives below are adopted.

1.1 Specific objectives

1. To review existing literature on Remote Sensing application to natural forest monitoring.
2. To map the land use and land cover (LULC) of the Nkandla Forest Reserve.
3. To determine the LULC changes of the Nkandla Forest Reserve (1989-2019) and predict the land cover distribution for 2029.
4. To predict the aboveground carbon (AGC) stocks of the Nkandla Forest Reserve.
5. To predict the tree species diversity of the Nkandla Forest Reserve.
6. To identify the best season for tree species diversity prediction.

1.2 Structure of the thesis

This thesis is composed of eight chapters presented in a manuscript format. It has a general introduction of the study (Chapter one), followed by six manuscripts that is a systematic review (Chapter Two), mapping of forest cover (Chapter Three), change detection and forecasting of

land cover distribution for 2029 (Chapter Four), aboveground carbon estimation (Chapter Five), tree species diversity prediction (Chapter Six), identifying of the best season for tree species diversity predictions (Chapter Seven) and the last chapter (Chapter Eight) is a synthesis of the thesis. Four of the chapters (chapters two, three, four and six) are published in international peer reviewed journals (see declaration 2). While chapters five and seven are currently being revised and subsequently submitted to journals. Each of the manuscripts addresses one of the specific objectives of the study. The content and format of the peer reviewed manuscripts are maintained in the thesis compilation. So, each of the manuscripts has an introduction, materials and methods (methodology), results, discussion, conclusion and reference. It is worth mentioning that there are some sections of the manuscripts that have overlaps and repetitions, as they were unavoidable, for instance under the materials and methods (description of area). However, since some of the manuscripts have already been published and others are under peer review, the repetitions are deemed trivial as they did not compromise on the overall content and context of the thesis.

1.3 Brief description of chapters

Chapter 1

The chapter provides an overview of the research, the gap identified and justification.

Chapter 2

The chapter is a reviews literature on progress made in the application of Remote Sensing to topical and sub-tropical natural forests monitoring. It was based on the exploration of predefined thematic areas of research and the specific assessment for the reviewed studies includes the country of research, the Remote Sensing imageries used, the machine learning algorithms utilised and the accuracies obtained. Recommendations were made for the gaps and limitations identified.

Chapter 3

The chapter is a research on the application of Landsat 8 and Random Forest (RF) and Support Vector Machine (SVM) to the mapping of land cover classes of the Nkandla Forest Reserve. Four land cover classes of the forest, which are closed canopy forest, open canopy forest, grassland and bare sites were successfully mapped for the forest. As part of the analysis, the performances of the RF and SVM in the classification (delineation) of four land cover classes of the forest were evaluated.

Chapter 4

The chapter presents the findings of decadal change detection of the Nkandla Forest Reserve from 1989 to 2019 and forecasting the spatial distribution of the land cover classes for 2029. Landsat satellite imageries, SVM, Markov Chain Modelling and Multi-layer Perceptron Neural Network are used in the land cover classification, post-classification change detection and forecasting for the land cover distribution.

Chapter 5

The chapter is a research on the prediction of aboveground carbon (AGC) of the Nkandla Forest Reserve using Sentinel 2 and Random Forest machine learning algorithm. Four separate RF models were developed for the prediction. The predicting variables for the models were composed of spectral bands only, near infrared vegetation indices only, red edge vegetation indices only and the combined variables model. The performance RF models of each set of variables were evaluated and results presented.

Chapter 6

The chapter presents findings on a multi-sensor evaluation on the prediction of tree species diversity of the Nkandla Forest Reserve. The effects of spatial and spectral resolution on the performance and accuracies of satellite imageries in the prediction were evaluated. The satellite imageries used were the Sentinel 2, RapidEye, PlanetScope and Landsat 8. The modelling was done with the RF algorithm.

Chapter 7

The chapter is research on identifying the best season for the prediction of tree species diversity using Sentinel 2 imagery and RF algorithm. An imagery was obtained for each of the years and used for the analysis. The best imagery was identified, and recommendations was made for other future research.

Chapter 8

The chapter is a synthesis of the overall thesis.

1.3 Limitations of the study

The limitations of the study have been listed below;

1. There was a plan to engage the managers of the Nkandla Forest Reserve, the fringe communities and other stakeholders in a workshop and present the outcomes of the study to them. It was envisaged to help them have credible information on the forest and how each can contribute to protecting and conserving it. However, it could not be help due to lack of funds and time.
2. A baseline LULC map was not available for the forest. As such we had to determine LULC classes before the data collection and mapping. As such the maps produced through our study could not be compared with any other map. We had to rely on our accuracy assessment to authenticate our maps.
3. Tree species diversity and carbon maps are also not available for the Nkandla Forest Reserve. Similarly, we had to determine the accuracy of our maps based of the Random Forest models.

References

ADAM, E., MUTANGA, O., ABDEL-RAHMAN, E. M. & ISMAIL, R. 2014. Estimating standing biomass in papyrus (*Cyperus papyrus* L.) swamp: exploratory of in situ

- hyperspectral indices and random forest regression. *International Journal of Remote Sensing*, 35, 693-714.
- ADDAE, B. & OPPELT, N. 2019. Land-Use/Land-Cover Change Analysis and Urban Growth Modelling in the Greater Accra Metropolitan Area (GAMA), Ghana. *Urban Science*, 3.
- AERTS, R. & HONNAY, O. 2011. Forest restoration, biodiversity and ecosystem functioning. *J BMC ecology*, 11, 29.
- AGRAWAL, A., CASHORE, B., HARDIN, R., SHEPHERD, G., BENSON, C. & MILLER, D. 2013. Economic contributions of forests. *Background Paper*, 1.
- AHMED, O. S., FRANKLIN, S. E., WULDER, M. A. & WHITE, J. C. 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 89-101.
- ASNER, G. P. & MASCARO, J. 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, 140, 614-624.
- CHENGE, I. B. & OSHO, J. S. A. 2018. Mapping tree aboveground biomass and carbon in Omo Forest Reserve Nigeria using Landsat 8 OLI data. *Southern Forests: a Journal of Forest Science*, 80, 341-350.
- CHO, M. A., MALAHLELA, O. & RAMOELO, A. 2015. Assessing the utility WorldView-2 imagery for tree species mapping in South African subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *International Journal of Applied Earth Observation and Geoinformation*, 38, 349-357.
- CROSS, M., SCAMBOS, T., PACIFICI, F., VARGAS-RAMIREZ, O., MORENO-SANCHEZ, R. & MARSHALL, W. 2019. Classification of Tropical Forest Tree Species Using Meter-Scale Image Data. *Remote Sensing*, 11.
- FANG, F., MCNEIL, B. E., WARNER, T. A., MAXWELL, A. E., DAHLE, G. A., EUTSLER, E. & LI, J. 2020. Discriminating tree species at different taxonomic levels using multi-temporal WorldView-3 imagery in Washington D.C., USA. *Remote Sensing of Environment*, 246.
- FASSNACHT, F. E., LATIFI, H., STEREŃCZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- FERREIRA, M. P., WAGNER, F. H., ARAGÃO, L. E. O. C., SHIMABUKURO, Y. E. & DE SOUZA FILHO, C. R. 2019. Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 119-131.
- FUNDISI, E., MUSAKWA, W., AHMED, F. B. & TESFAMICHAEL, S. G. 2020. Estimation of woody plant species diversity during a dry season in a savanna environment using the spectral and textural information derived from WorldView-2 imagery. *PLoS One*, 15, e0234158.
- GIZACHEW, B., SOLBERG, S., NAESSET, E., GOBAKKEN, T., BOLLANDSAS, O. M., BREIDENBACH, J., ZAHABU, E. & MAUYA, E. W. 2016. Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. *Carbon Balance Manag*, 11, 13.
- GRABSKA, E., HOSTERT, P., PFLUGMACHER, D. & OSTAPOWICZ, K. 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, 11.
- KANNIAH, K. D., MOHD NAJIB, N. E. & VU, T. T. 2016. Forest Cover Mapping in Iskandar Malaysia Using Satellite Data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W1, 71-75.

- KEENAN, R. J., REAMS, G. A., ACHARD, F., DE FREITAS, J. V., GRAINGER, A. & LINDQUIST, E. 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, 352, 9-20.
- MÜLLEROVÁ, J., PERGL, J. & PYŠEK, P. 2013. Remote Sensing as a tool for monitoring plant invasions: Testing the effects of data resolution and image classification approach on the detection of a model plant species *Heracleum mantegazzianum* (giant hogweed). *International Journal of Applied Earth Observation and Geoinformation*, 25, 55-65.
- MÜNCH, Z., OKOYE, P., GIBSON, L., MANTEL, S. & PALMER, A. 2017. Characterizing Degradation Gradients through Land Cover Change Analysis in Rural Eastern Cape, South Africa. *Geosciences*, 7.
- MUTANGA, O., DUBE, T. & AHMED, F. 2016. Progress in Remote Sensing: vegetation monitoring in South Africa. *South African Geographical Journal*, 98, 461-471.
- MUTOWO, G. & MURWIRA, A. 2012. Relationship between remotely sensed variables and tree species diversity in savanna woodlands of Southern Africa. *International Journal of Remote Sensing*, 33, 6378-6402.
- MUTOWO, G., MUTANGA, O. & MASOCHA, M. 2018. Evaluating the Applications of the Near-Infrared Region in Mapping Foliar N in the Miombo Woodlands. *Remote Sensing*, 10.
- NANDY, S., KUSHWAHA, S. P. S. & DADHWAL, V. K. 2011. Forest degradation assessment in the upper catchment of the river Tons using Remote Sensing and GIS. *Ecological Indicators*, 11, 509-513.
- NKONYA, E., MIRZABAEV, A. & VON BRAUN, J. 2016. *Economics of land degradation and improvement: a global assessment for sustainable development*, Springer.
- OMER, G., MUTANGA, O., ABDEL-RAHMAN, E. M. & ADAM, E. 2015. Performance of Support Vector Machines and Artificial Neural Network for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 4825-4840.
- PHIRI, D. & MORGENROTH, J. 2017. Developments in Landsat land cover classification methods: A review. *Remote Sensing*, 9, 967.
- RAJAH, P., ODINDI, J., MUTANGA, O. & KIALA, Z. 2019. The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. *Nature Conservation*, 35, 41-61.
- RIMAL, B., ZHANG, L., KESHTKAR, H., HAACK, B. N., RIJAL, S. & ZHANG, P. 2018. Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information*, 7, 154.
- ROY, D. P., WULDER, M. A., LOVELAND, T. R., C.E, W., ALLEN, R. G., ANDERSON, M. C., HELDER, D., IRONS, J. R., JOHNSON, D. M., KENNEDY, R., SCAMBOS, T. A., SCHAAF, C. B., SCHOTT, J. R., SHENG, Y., VERMOTE, E. F., BELWARD, A. S., BINDSCHADLER, R., COHEN, W. B., GAO, F., HIPPLE, J. D., HOSTERT, P., HUNTINGTON, J., JUSTICE, C. O., KILIC, A., KOVALSKYY, V., LEE, Z. P., LYMBURNER, L., MASEK, J. G., MCCORKEL, J., SHUAI, Y., TREZZA, R., VOGELMANN, J., WYNNE, R. H. & ZHU, Z. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172.
- SCHLUND, M., SCIPAL, K. & DAVIDSON, M. W. J. 2017. Forest classification and impact of BIOMASS resolution on forest area and aboveground biomass estimation. *International Journal of Applied Earth Observation and Geoinformation*, 56, 65-76.

- SHEN, G., WANG, Z., LIU, C. & HAN, Y. 2020. Mapping aboveground biomass and carbon in Shanghai's urban forest using Landsat ETM+ and inventory data. *Urban Forestry & Urban Greening*, 51.
- ST-ONGE, B. & GRANDIN, S. 2019. Estimating the Height and Basal Area at Individual Tree and Plot Levels in Canadian Subarctic Lichen Woodlands Using Stereo WorldView-3 Images. *Remote Sensing*, 11.
- SUZUKI, K., RIN, U., MAEDA, Y. & TAKEDA, H. 2018. Forest Cover Classification Using Geospatial Multimodal Data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 1091-1096.
- XIE, Z., CHEN, Y., LU, D., LI, G. & CHEN, E. 2019. Classification of Land Cover, Forest, and Tree Species Classes with ZiYuan-3 Multispectral and Stereo Data. *Remote Sensing*, 11.
- XU, Z., CHEN, J., XIA, J., DU, P., ZHENG, H. & GAN, L. 2018. Multisource Earth Observation Data for Land-Cover Classification Using Random Forest. *IEEE Geoscience and Remote Sensing Letters*, 15, 789-793.
- YIN, H., KHAMZINA, A., PFLUGMACHER, D. & MARTIUS, C. 2017. Forest cover mapping in post-Soviet Central Asia using multi-resolution Remote Sensing imagery. *Sci Rep*, 7, 1375.

**CHAPTER 2: TWO DECADES PROGRESS ON THE APPLICATION OF
REMOTE SENSING FOR MONITORING TROPICAL AND SUB-TROPICAL
NATURAL FORESTS: A REVIEW**



Review

**Two Decades Progress on The Application of Remote Sensing
for Monitoring Tropical and Sub-Tropical Natural Forests:
A Review**

Enoch Gyamfi-Ampadu *  and Michael Gebreslasie 

School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus,
Private Bag X54001, Durban 4000, South Africa; gebreslasie@ukzn.ac.za

* Correspondence: egampadu@gmail.com

Abstract

Forest covers about a third of terrestrial land surface, and tropical and subtropical zones form a major part. These natural forests provide multiple benefits to society, the economy, and the environment of the countries. Hence, to maintain these benefits, an effective and efficient forest monitoring, and management system is vital. Remote Sensing applications constitute a significant approach to monitoring forests, and several studies have demonstrated the potential. Thus, this paper reviews the progress made by the Remote Sensing imagery on the applications to the tropical and sub-tropical natural forests monitoring over the last two decades (2000-2020). The review focused on thematic areas such as aboveground biomass and carbon estimations, tree species identification, tree species diversity, and forest cover and changes mapping. A systematic search of relevant articles was made on Web of Science, Science Direct, and Google Scholar by applying a Boolean operator and using keywords related to the thematic areas. We identified fifty peer-reviewed articles that studied the tropical and subtropical natural forests using Remote Sensing data. Asian and Southern American natural forests are the most researched natural forests, while African natural forests are not. Medium spatial resolution imagery (Landsat, ASTER, and Sentinel 2 imageries) was extensively utilized for forest cover and change mapping and aboveground biomass and carbon estimation. High spatial resolution imagery (Quickbird, WorldView-2 & -3, and Active and passive sensor Aerial imageries) were used for tree species identification. Linear regression (LR) was the most used parametric algorithm whereas the Random Forest (RF) and Support Vector Machine (SVM) were among the most non-parametric algorithms. There is potential in utilising emerging satellite information, as they hold promise to further enhance monitoring and management of natural forests. We recommend more research to identify approaches that can help to overcome challenges of Remote Sensing applications to thematic areas so that further and sustainable progress can be made to effectively monitor and manage sustainable forest benefits.

Keyword: Forests, Remote Sensing, Satellite, Monitoring, Application, Algorithm,

2.1 Introduction

Forests cover about one-third of the earth's land surface area (Khan et al., 2020), and tropical and sub-tropical forests form a major component of the total area. Natural forests host diverse plant and animal species (Ruiz-Benito et al., 2014), cater for forage resources to insect pollinators (Goulson, 1999), and help alleviate the effects of climate change by atmospheric carbon sequestration (Saatchi et al., 2011a, Pan et al., 2011). As many authors have pointed out, forests provide various functions that include ecological, economic, social, and recreational functions at local, regional, and global scales (Miura et al., 2015). Moreover, in developing countries, forests support millions of rural people's direct livelihood by providing food, medicine, fuel, fibre, non-timber forest products, and social and cultural functions (Mayaux et al., 2013, Agrawal et al., 2013). The world's largest tropical and sub-tropical forests are located in the Amazon region, followed by the tropical forests of Central and West African, termed as the Guineo-Congolian region, while the third-largest tropical forest region is located in Southeast Asia (Malhi et al., 2013).

In recent decades, natural forests have been declining at alarming rates in most parts of the world (Keenan et al., 2015). As documented in many studies, humans are continuously changing the land use to access the planet's resources through the clearance of forests for

agricultural activities and urban expansion (Ranagalage et al., 2019). Furthermore, in developing countries, drivers of deforestation, including timber and fuel extraction, have been used along with evidence concerning underlying causes such as economy, political instability, and governance (Maynard and Royer, 2004). These threats could modify the forest ecosystems (Brose and Hillebrand, 2016), in so doing reduce their functions (Hautier et al., 2014) and ultimately result in the extinction of tree species and species homogenization (Solar et al., 2015, Wang and Loreau, 2016). Thus, timely and consistent monitoring of the natural forest is critical because their location and condition affect local, regional and global climate, and have significant consequences for biodiversity and well-being of millions of rural and urban people. Regular and accurate monitoring of natural forests spatial extent, species composition, physiological characteristics, forest cover change and drivers of change, and carbon content would be essential for providing information for policy formulation, implementation of climate change-related agreements, monitoring sustainable schemes for timber extraction, reporting duties and conservation management measures of natural forests (MacDicken, 2015, Laurin et al., 2016a).

Remote Sensing has been utilized in a wide variety of applications confronting the forest management and conservation sectors. The relevance and application of Remote Sensing for natural forest studies have been widely demonstrated by various authors. These applications include questions linked to forest biophysical parameters inventory, forest biochemical mapping (Vasudeva et al., 2020), tree species discrimination and mapping (Wagner et al., 2018, Fang et al., 2020), carbon stocks (Asner et al., 2011, Asner and Mascaro, 2014, De Moraes et al., 2010), forest land cover change mapping (Fokeng et al., 2020, Margono et al., 2014, Martinez del Castillo et al., 2015, Mayaux et al., 2005), biodiversity assessment and monitoring (Ferreira et al., 2016, Fundisi et al., 2020, Grabska et al., 2019a), assessment of forest extent (Qin et al., 2015), and tree crown delineation (Dalponte et al., 2008, Dalponte et al., 2012, Holmgren and Persson, 2004, Immitzer et al., 2012).

An example of a review that has been of importance to Remote Sensing practitioners is Mutanga et al. (2016), who showed the progress made in the application of Remote Sensing for the monitoring of vegetation in South Africa. It considered studies done from 1996 to 2015 on sensors and the type of vegetation with a focus on biomass, species discrimination, and land cover, and vegetation quality. Although certain important aspects were covered in the review, it failed to assess studies for tropical and sub-tropical areas. Since many advancements have been seen in sensors, machine learning algorithms, and many studies for the natural forest in tropical and sub-tropical natural forests, it will be worth conducting such a review that captures these advancements and areas of limitation. Therefore, our review covers a two-decade (2000-2020) progress made on the application of Remote Sensing in monitoring the tropical and subtropical natural forests. The output of this review will provide information on advances made and the shortfall observed on the use of Remote Sensing imageries for monitoring natural forests in the tropical and subtropical zones for forest managers, ecologists, and Remote Sensing experts and how monitoring approaches can be improved.

2.2 Methodology

A systematic search was conducted on the World Catalog, ISI Web of Science, Google Scholar databases to retrieve relevant articles. The search was conducted with a Boolean operator, “AND” and a combination of keywords which was “Remote Sensing” AND “Forest cover”

AND “classification” AND "mapping" AND "natural" AND "forest tree species identification" AND “biomass” AND “carbon” AND "diversity." This search returned 6820 articles that generally relate to keywords used for the search. The search was further conducted using similar keywords but restricted to studies published between 2000 and 2020. This resulted in 1060 published articles. Thereafter, the titles and abstracts of the articles were assessed to determine their relevance to the study before downloading. Furthermore, non-natural forest such as plantations and duplicated studies were removed. Also, studies of global forests, urban forests, mangroves, savanna, and dry forests were excluded, which led to the selection of 156 potential articles. The full text of the 156 articles was downloaded for further screening through abstract and full-text reading and subjection to the objectives of the review. The final screening resulted in 50 articles that fully met the criteria of article selection. The search strategy, screening, and selection processes of the relevant articles are provided in a schematic diagram in Figure 2.1.

The 50 articles were further grouped into four thematic areas of Remote Sensing monitoring, which are 1) Biomass and carbon stock estimation, 2) Individual tree species identification, 3) Tree species diversity prediction, and 4) Forest cover mapping and change detection. In each of the articles, the focus was placed on the country of research, the type of Remote Sensing data employed, the algorithm used for the mapping and modelling and the accuracy produced.

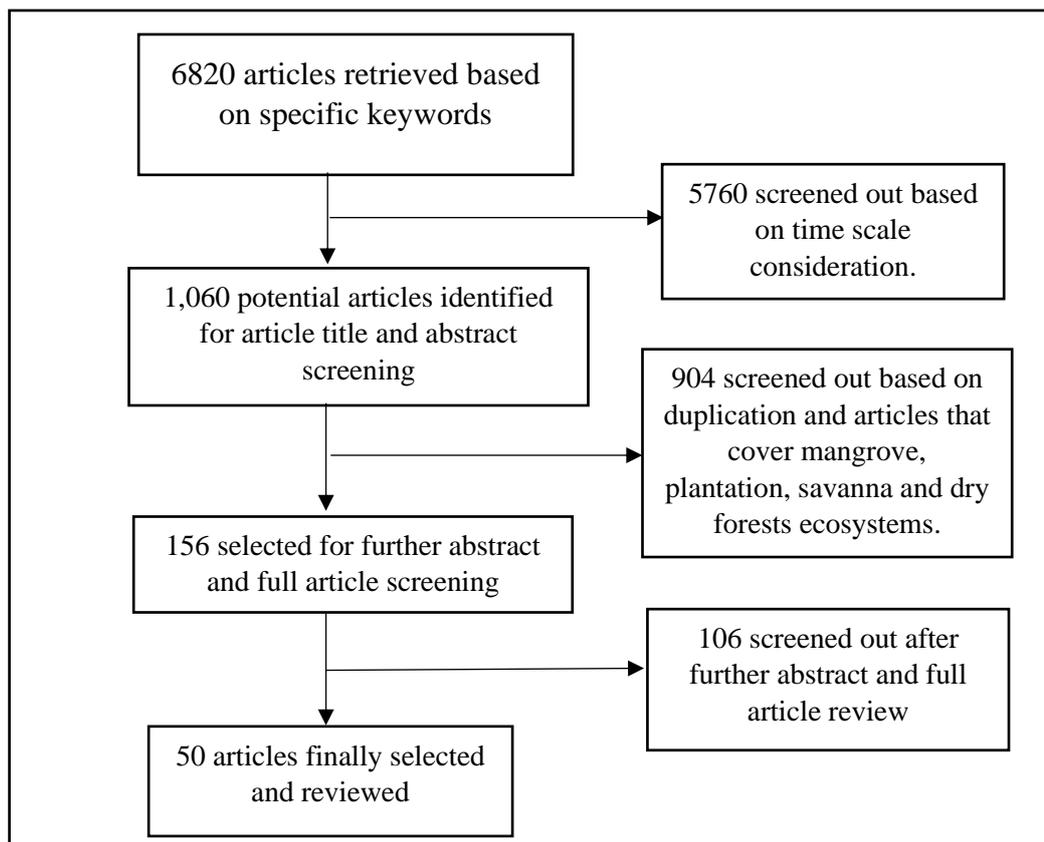


Figure 2.1: Schematic diagram of the literature selection process

2.3. Results

Figure 2.2 illustrate the number of studies carried out over the last 20 years under the four thematic areas considered under our review.

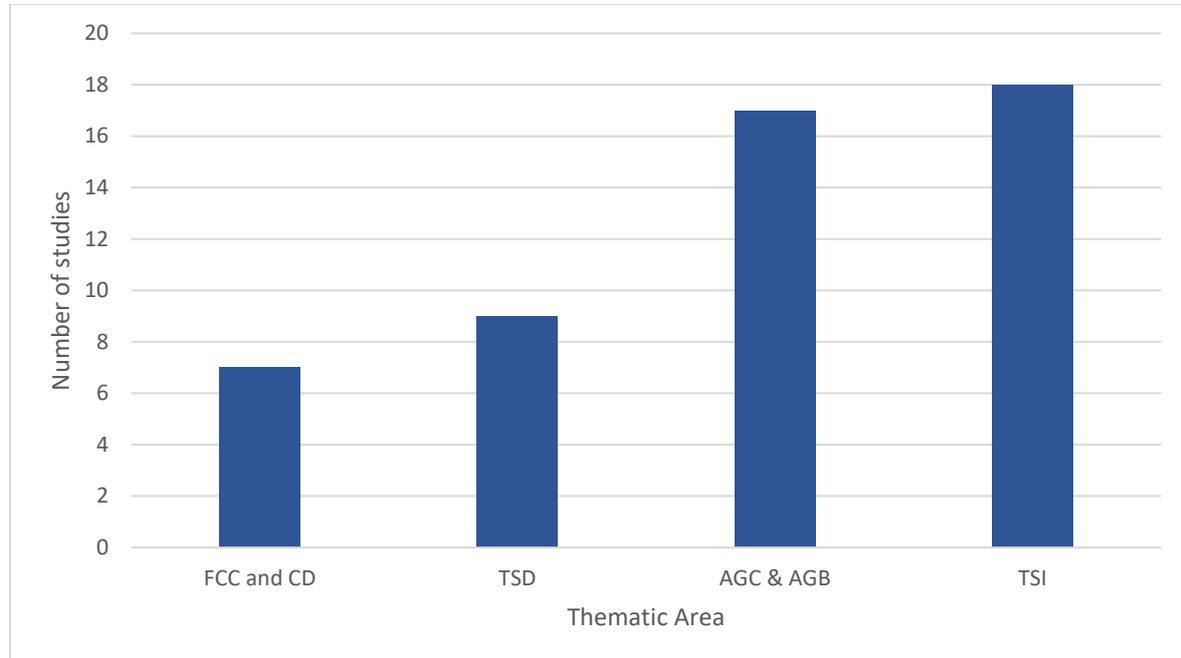


Figure 2.2: Number of studies carried out for each thematic area. Note: FCC and CD: Forest cover change and change detection, TSD: Tree species diversity, AGC and AGB: Aboveground carbon and Aboveground biomass, and TSI: Tree species identification.

2.3.1 Aboveground biomass and carbon estimation using Remote Sensing

Several studies utilized different sensors to predict aboveground biomass (Broadbent et al., 2008, Hernández-Stefanoni et al., 2014, Pandit et al., 2018a) and aboveground carbon (Lin et al., 2016, Mbaabu et al., 2014, Wangda et al., 2019), [Table 2.1]. The majority of these studies have been conducted in the tropical and sub-tropical forests region of Asian countries such as China (Chen et al., 2018, Chen et al., 2019, Lin et al., 2016) and Nepal (Mbaabu et al., 2014, Wangda et al., 2019). Southern America placed second in the number of studies conducted (Clark et al., 2011, González-Jaramillo et al., 2018, Hernández-Stefanoni et al., 2014) with African tropical and subtropical forest region lagging behind. We noted that the area and extent of the study sites range between 4 ha (Broadbent et al., 2008) and 6292.68 ha (Pandey et al., 2019).

The majority of the studies identified by the selection criteria utilized optical sensor imagery to predict AGB and AGC. For example, Landsat 8 was utilized by many studies (Li et al., 2019, Pandit et al., 2018a, Wallis et al., 2019, Zhang et al., 2019). Other studies reviewed have utilized high spatial resolution optical sensor imagery such as Quickbird (Broadbent et al., 2008); and Sentinel 2 (Vasudeva et al., 2020). O

other studies reviewed in this study have explored the capabilities of active sensor imagery such as LiDAR (González-Jaramillo et al., 2018, Hernández-Stefanoni et al., 2014, Lin et al., 2016). We also identified studies that fused optical and active sensor imagery; for example Wangda et al. (2019) and Mbaabu et al. (2014), fused Lidar data and GeoEye imagery to predict AGB. Sentinel-1 SAR and Sentinel 2 fusion is also utilized by a number of studies (Chen et al., 2018, Chen et al., 2019, Ghosh and Behera, 2018).

Non-parametric machine learning algorithms such as ANN (Chen et al., 2018), RF (Zhang et al., 2019); SVR (Chen et al., 2018); XGBoost (Li et al., 2019) were most commonly utilized. In contrast, others used parametric algorithms such as LR (Mbaabu et al., 2014); MLR (Lin et al., 2016); and OLS (Hernández-Stefanoni et al., 2014). Overall, both types of algorithms produced satisfactory accuracy, but Non-parametric approaches produced high accuracies for studies that adopted them.

Accuracies reported in the reviewed articles range from a coefficient of determination as weak as 0.31 (Wallis et al., 2019) to a coefficient of determination as good as 0.97 (Chen et al., 2019). Studies that utilized a fusion of two sensors reported better results (Chen et al., 2019, Clark et al., 2011).

Table 2.1: Aboveground biomass and carbon estimation using Remote Sensing

Reference	Country	Sensor names	Algorithms /Methods	Area	Accuracy
(Broadbent et al., 2008)	Bolivia	Quickbird	LR	4 ha	$R^2 = 0.70$
(Sarker and Nichol, 2011)	Hong Kong	AVNIR-2	SWR, LR	100 km ²	$R^2 = 0.88$ RMSE=32 t/ha
(Clark et al., 2011)	Costa Rica	LiDAR, HYDICE	OLS, GLS,	Not specified	$R^2 = 0.90$ RMSE =38.3 Mg/ha
(Hernández-Stefanoni et al., 2014)	Yucatan Peninsula	LiDAR	OLS	9 ha	$R^2 = 0.89$
(Mbaabu et al., 2014)	Nepal	LiDAR, GeoEye-1	LR	5,821 ha	$R^2 = 0.81$
(Lin et al., 2016)	Taiwan	LiDAR	MLR	Not specified	$R^2 = 0.91$ RMSE = 15 ton/ha - 210 ton/ha
(Pandit et al., 2018b)	Nepal	Landsat 8	RF, MLR	Not specified	$R^2 = 0.95$ RMSE = 13.3 t/ha
(González-Jaramillo et al., 2018)	Ecuador	LiDAR	LR	~85 km ²	$R^2 = 0.91$
(Ghosh and Behera, 2018)	India	Sentinel 1 SAR, Sentinel 2, Sentinel 1	RF, SGB	400 km ²	$R^2 = 0.71$ RMSE = 105.027 t/ha
(Chen et al., 2018)	China	SAR, Sentinel 2, Landsat 8,	RF, ANN, GWR, SVR	Not specified	$r = 1$ RMSE = 0.08 Mg/ha
(Zhang et al., 2019)	China	Landsat TM	RF	6.06 million ha	$R^2 = 0.73$ RMSE = 6.66 Mg/ha
(Chen et al., 2019)	China	Sentinel 1 SAR, Sentinel 2	SWR, GWR, ANN, SVR, and RF.	17,481 ha	$R^2 = 0.97$ RMSE = 61.11 Mg/ha
(Li et al., 2019)	China	Landsat 8	LR, RF, XGBoost	13.00×10 ⁴ km ²	$R^2 = 0.37$
(Pandey et al., 2019)	India	MODIS	LR	6292.68 km ²	$R^2 = 0.94$
(Wallis et al., 2019)	Ecuador	Landsat 8	PLSR		$R^2 = 0.31$
(Wangda et al., 2019)	Nepal	GeoEye-1, RapidEye, LiDAR	MLR	1888 ha	$R^2 = 0.88$ RMSE = 44 kg/tree

(Vasudeva et al., 2020)	India	Sentinel 2	RF, ANN, SVM	84.46 km ²	R ² = 0.86 RMSE = 0.26%
-------------------------	-------	------------	--------------	-----------------------	---------------------------------------

Note: HYDICE: Hyperspectral Digital Imagery Collection Experiment, SAR: Synthetic Aperture Radar, LiDAR: Light Detection and Ranging, AVNIR: Advanced Visible and Near Infrared Radiometer type 2. SWR: Stepwise Regression, GWR: Geographically Weighted Regression, LR: Linear Regression, MLR: Multiple Linear Regression, PLSR: Partial Least Squares Regressions; OLS: Ordinary Least Squares Regression, GLS: Generalized Least Squares Regression, ANN: Artificial Neural Network, SVR: Support Vector Machine for Regression, RF: Random Forest (RF), XGBoost: Extreme Gradient Boosting, SGB: Stochastic Gradient Boosting

2.3.2 Tree species identification using Remote Sensing

The majority of the tree species detection studies were carried out in Southern American countries, including Brazil (Ferreira et al., 2013, Ferreira et al., 2016, Ferreira et al., 2019), Costa Rica (Clark et al., 2005, Zhang et al., 2006, Clark and Roberts, 2012) and Panama (Garzon-Lopez et al., 2013). Most of the studies found in African tropical and subtropical forests were conducted in South Africa (Cho et al., 2015, Omer et al., 2015, van Deventer et al., 2017) and Ghana (Laurin et al., 2016a). We found fewer studies carried out in Asia (Lin et al., 2015, Cao et al., 2016). In most cases, the studied forest cover area ranges from 70 ha (Féret and Asner, 2012, Feret and Asner, 2013) to 6000 ha (Cho et al., 2015).

Several studies used very high-resolution optical sensors, including WorldView-2 (Cho et al., 2015, Omer et al., 2015, Wagner et al., 2018), and WorldView-3 (Ferreira et al., 2019). Other studies used hyperspectral data such as the HYDICE (Clark et al., 2005, Zhang et al., 2006, Clark and Roberts, 2012), AISA EAGLE (Ferreira et al., 2016, Laurin et al., 2016a), AISA HAWK ((Ferreira et al., 2016), and imaging spectrometer (Féret and Asner, 2012, Feret and Asner, 2013, Ferreira et al., 2013, Somers and Asner, 2014). Fewer studies have used the LiDAR (Cao et al., 2016) and digital aerial photography (Garzon-Lopez et al., 2013) for tree species identification.

Non-parametric and parametric statistical approaches were also utilized for tree species identification using Remote Sensing imagery. Machine learning algorithms such as the RF (Clark and Roberts, 2012, Cao et al., 2016), SVM (Féret and Asner, 2012, Laurin et al., 2016a, Wagner et al., 2018) while others also used parametric types such the ML (Clark et al., 2005) and the LDA, QDA, and RDA (Feret and Asner, 2013, Ferreira et al., 2013).

The studies recorded high accuracy that ranged from 70% (Ferreira et al., 2019) and 96% (Ferreira et al., 2013, Wagner et al., 2018). Generally, the very high-resolution images had the high accuracies than most of the other image types.

Table 2.2: Tree species identification using Remote Sensing

Reference	Country	Sensor/ Data Set	Algorithm	Area	Average Accuracy
(Clark et al., 2005)	Costa Rica	HYDICE	LDA, ML, and SAM	Not specified	$\geq 92\%$
(Zhang et al., 2006)	Costa Rica	HYDICE	SMA	Not specified	Not specified
(Clark and Roberts, 2012)	Costa Rica	HYDICE	RF	Not specified	≥ 85
(Féret and Asner, 2012)	USA	CAO-Alpha System, LiDAR	SVM	70 ha	$\geq 90\%$
(Feret and Asner, 2013)	USA	CAO-Alpha System	LDA, RDA, QDA, Linear-SVM, Radial-SVM, ANN, KNN	70 ha	$\geq 73\%$
(Ferreira et al., 2013)	Brazil	ASD Spectroradiometer	LDA	Not specified	96%
(Garzon-Lopez et al., 2013)	Panama	DAP	LR and Visual Analysis	150 ha	76%
(Somers and Asner, 2014)	Hawaii, USA	EO-1 Hyperion	MESMA, WASMA	1500	$R^2 = 0.74$ $KC = 0.65$
(Cho et al., 2015)	South Africa	WorldView-2	SVM	6000 ha	≥ 89
(Omer et al., 2015)	South Africa	WorldView-2	SVM, ANN	Not specified	$\geq 77\%$
(Lin et al., 2015)	Taiwan	QuickBird	ML	Not specified	Not Specified
(Ferreira et al., 2016)	Brazil	AISA EAGLE, AISA HAWK, WorldView-3	LDA, Radial-SVM, L-SVM, and RF	Not specified	$\geq 84\%$
(Cao et al., 2016)	China	LiDAR	RF	Not specified	86.2%
(Laurin et al., 2016a)	Ghana	AISA EAGLE, Sentinel 2	SVM, ML	815 km ²	92.34%
(van Deventer et al., 2017)	South Africa	ASD Spectroradiometer, WorldView-2, RapidEye,	PLS-RF	Not specified	$> 92\%$
(Wagner et al., 2018)	Brazil	WorldView-2	SVM	237.6 ha	96%

Note: CAO: Carnegie Airborne Observatory, SMA: Spectral Mixture Analysis, ML: Maximum Likelihood, SAM: Spectral Angle Mapper, LDA: Linear Discriminant Analysis, SVM: Support Vector Machine, RDA: Radial Discriminant Analysis, QDA: Quadratic Discriminant Analysis, AISA: Airborne Imaging Spectrometer for Application, DAP: Digital Aerial Photography, KC: Kappa Coefficient, SCKC: Species Conditional Kappa Coefficients, J-M: Jeffrei–Matusita, ASD: Analytical Spectral Device, PLS: Partial Least Square, MESMA: Multiple Endmember Spectral Mixture Analysis, WESMA: Wavelength Endmember Spectral Mixture Analysis.

2.3.3 Tree species diversity mapping using Remote Sensing

Africa had the highest number of tree species studies, with studies conducted in Sierra Leone (Laurin et al., 2014) and Kenya (Maeda et al., 2014, Schäfer et al., 2016). Studies in the Asian region had the second-highest number of studies which were conducted in India (Nagendra et al., 2010), Kyrgyzstan (Feilhauer and Schmidtlein, 2009b), and China (Zhao et al., 2018). Southern America had the least number of studies carried out in Panama (Gillespie et al., 2009). Area of coverage was reported by three studies ranging between 90 km² (Feilhauer and Schmidtlein, 2009b) and 850 km² (Maeda et al., 2014).

Medium and high spatial resolution optical multispectral imageries were used for the tree species diversity mapping, for example, Landsat (Gillespie et al., 2009, Nagendra et al., 2010, Maeda et al., 2014, Foody and Cutler, 2006), ASTER (Feilhauer and Schmidtlein, 2009b), and IKONOS (Nagendra et al., 2010). Some other studies also explored Hyperspectral data for tree species mapping, for example, AISA EAGLE (Laurin et al., 2014, Schäfer et al., 2016), and AVIRIS (Carlson et al., 2007). Gillespie et al. (2009) fused Landsat ETM+ and AIRSAR to map tree species diversity.

Almost all the studies reviewed for this thematic area used the LR, with the exceptions being GRNN, MLPNN (Foody and Cutler, 2006), and RF (Laurin et al., 2014, Zhao et al., 2018). The accuracies reported in these reviewed articles for the tree species diversity mapping ranged from a coefficient of correlation (r) of 0.36 (Nagendra et al., 2010) to R^2 of 0.85 (Carlson et al., 2007).

Table 2.3: Research carried out on tree species diversity

Reference	Country	Sensor names	Algorithms	Area	Accuracy
(Foody and Cutler, 2006)	Malaysia	Landsat TM	GRNN, MLPNN	300km ²	r =0.69
(Carlson et al., 2007)	Hawaii, USA	AVIRIS	LR, MCS	Not specified	R ² = 0.85
(Feilhauer and Schmidlein, 2009b)	Kyrgyzstan	ASTER	DCA	90 km ²	R ² = 0.61
(Gillespie et al., 2009)	Panama	Landsat ETM+, AIRSAR	LR	Not specified	R ² = 0.51
(Nagendra et al., 2010)	India	IKONOS and Landsat ETM+	Not specified	Not specified	r = 0.33
(Laurin et al., 2014)	Sierra Leone	AISA EAGLE	RF	Not specified	R ² = 0.84.9 RMSE = 0.30
(Maeda et al., 2014)	Kenya	Landsat-5 TM, and Landsat-7 ETM+	LR	850 km ²	R ² = 0.36
(Schäfer et al., 2016)	Kenya	AISA EAGLE	K Means clustering, LR	Not specified	R ² = 0.50 RMSE = 3
(Zhao et al., 2018)	China	PHI-3, LiDAR	RF	Not specified	R ² = 0.83, RMSE=0.25

Note: AVIRIS: Airborne Visible and Infrared Imaging Spectrometer, AIRSAR: Airborne Synthetic Aperture Radar, MCS: Monte-Carlo Simulation, DCA: Detrended Correspondence Analysis (DCA), PHI: Pushbroom Hyperspectral Imager, TM: Thematic Mapper, ETM: Enhanced Thematic Mapper Plus, GRNN: Generalised Regression Neural Networks, MLPNN: Multi-layer Perceptron Neural Network, TM: Thematic Mapper.

2.3.4 Forest cover mapping and change detection with Remote Sensing

As presented in Table 2.4, two studies were identified in Africa, which was carried out specifically in Nigeria (Ochege and Okpala-Okaka, 2017) and South Africa (Gyamfi-Ampadu et al., 2020). The other studies were carried out one article in Southern America, Belize (Voight et al., 2019), and three articles in Asia, Bhutan (Bruggeman et al., 2016), and Bangladesh (Rahman and Sumantyo, 2010), and India (Joshi et al., 2006).

It was observed that the Landsat ETM+ and Landsat 8 was the most used multispectral imageries (Bruggeman et al., 2016, Ochege and Okpala-Okaka, 2017, Voight et al., 2019, Gyamfi-Ampadu et al., 2020) whereas Rahman and Sumantyo (2010) applied an active sensor. Most of the studies under this thematic area employed the ML (Rahman and Sumantyo, 2010, Bruggeman et al., 2016, Ochege and Okpala-Okaka, 2017). The non-parametric statistical approach was adopted by Gyamfi-Ampadu et al. (2020) for forest cover mapping and change detection. High accuracies were reported by these studies, with the least being 83% (Rahman

and Sumantyo, 2010) and the highest being 97% (Ochege and Okpala-Okaka, 2017, Voight et al., 2019).

Table 2.4: Forest cover mapping and change detection

Reference	Country	Sensor names	Algorithms	Area	Accuracy
(Joshi et al., 2006)	India	IRS-1 C WiFS	<i>K</i> -means clustering	Not Specified	85%
(Rahman and Sumantyo, 2010)	Bangladesh	SIR-C, ALOS PALSAR	ML	Not specified	83%
(Bruggeman et al., 2016)	Bhutan	Landsat ETM+	ML	Not specified	87.5%
(Ochege and Okpala-Okaka, 2017)	Nigeria	Landsat 7 ETM+	ML	Not specified	97%
(Voight et al., 2019)	Belize	Landsat 8	CART	Not specified	97%
(Gyamfi-Ampadu et al., 2020)	South Africa	Landsat 8	RF, SVM	2218	95%

Note: SIR-C: Shuttle Imaging Radar-C, ALOS: Advanced Land Observation Satellite PALSAR, WiFS: Wide Field Scanner, ML: Maximum Likelihood, HyMap: Hyperspectral MAPper, CART: Classification and Regression Trees.

2.4 Discussion

Information-driven and evidence-based forest management and conservation are required to deal with the complex and dynamic nature of forests. Remote Sensing approaches help in bridging the gap between science and practice for monitoring and managing natural forests by providing resource information (Bustamante et al., 2016). As a means of science approaches meeting practical needs, monitoring outcomes must be able to inform researchers, policymakers, funding agencies to develop pragmatic and well-adapted conservation and governance initiatives and contribute to strengthening management actions and policy (Gaston et al., 2006, Bustamante et al., 2016).

Our paper generally reviews studies on progress made over the last two decades on Remote Sensing applications to monitor tropical and subtropical natural forests. Based on the applied search criteria, the majority of the study was conducted in Southern America, followed by Asia, while the least number of the study was found in Africa. Specifically, this paper reviews Remote Sensing studies for aboveground biomass and carbon estimation, tree species identification, tree species diversity, and forest cover mapping and change detection. Questions such as what Remote Sensing sensor types have been used, the methodology developed, and the accuracy achieved is investigated. The accurate and timely estimation of AGB and AGC are vital for carbon accounting and climate change policy direction, support CO₂ emission monitoring, and forest management (Asner and Mascaro, 2014).

The outcomes of the AGB and AGC have a significant impact on local, regional and global climate change policies (Chave et al., 2014). Hence, natural forest AGB and AGC research

have increased markedly over the last few years at various levels. Asia and Latin America produced more research, with Africa lagging behind, which is detrimental to the execution of projects such as the Reduce Emissions from Deforestation and Forest Degradation (REDD+). Africa is one of the areas to be affected by climate changes, thus it makes it vital for accurate information on the AGB and AGC to support such decision making.

Many studies accessed and used optical Remote Sensing data for AGB and AGC predictions. The data type ranges from low, medium to very high resolution, such as Landsat (Li et al., 2019, Wallis et al., 2019), Sentinel 2 (Vasudeva et al., 2020), and GeoEye-1 (Wangda et al., 2019), respectively. Recent advancements in some sensors such as the Landsat 8 have increased its sensitivity to vegetation (Dube and Mutanga, 2015, Roy et al., 2014), making it a suitable sensor for AGB and AGC estimation. The highly informative three red-edge bands included in the spectral bands of the Sentinel 2 increases its capabilities in carbon and biomass predictions. It is worth noting that the number of studies that used Landsat 8 and Sentinel 2 increased over time, and their wide-scale of coverage and free accessibility and affordability could be reasons for the increase (Dube et al., 2016). Regarding hyperspectral data, the numerous spectral bands provide it with many capabilities for predictions across different regions (Koch, 2010, Clark et al., 2011, Sarker and Nichol, 2011). Although hyperspectral data has many capabilities, it has a problem of saturation in dense natural forests and band redundancy that can potentially affect its predictive ability negatively (Anu et al., 2011, Fassnacht et al., 2014b).

Active sensors have been extensively used for AGB and AGC estimations with the LiDAR being the most utilized (Sarker et al., 2013, Hernández-Stefanoni et al., 2014, Lin et al., 2016). This is because LiDAR data can provide tree height data for a reasonable estimation of tree volume and delineation of tree crowns. Hence, and its combination with regression analysis and machine learning techniques produces better prediction outcomes (Lin et al., 2016). The height and structural attributes data have the potential of supporting other data types for accurate estimates. In light of that, recent studies have employed the fusion of other earth observation data and LiDAR with the anticipation of improving results (Clark et al., 2011). A limitation of AGC and AGB estimation observed is the covering of between 1-15 tree by emergent tree species in forest ecosystems (Broadbent et al., 2008). Such trees are covered from the nadir of sensors and it could likely lead to incorrect and underestimation of AGC and AGB. Fusion methods involving LiDAR and other earth observation data could likely help to deal with such problems. However, the commercial nature of LiDAR and the hyperspectral data make them largely inaccessible and hinder their extensive use, especially in Africa (Lu et al., 2014a).

The algorithms and statistical analysis employed are some of the factors that influence the accuracies of AGB and AGC predictions. The LR and RF algorithm remains the most commonly used in AGC and AGB studies (Sarker and Nichol, 2011, Mbaabu et al., 2014, González-Jaramillo et al., 2018). The parametric nature of LR makes it assume a normal distribution of data, whereas the non-parametric RF does not assume a normal distribution of data (Fassnacht et al., 2016). High accuracy has been seen in all the studies, and as more advancements are being realized in methodological approaches, accuracy will improve and continue to be high. More multi-temporal AGC and AGB research are recommended since they help to monitor changes in local carbon stock and biomass.

Tree species identification has become necessary due to factors such as species extinction and invasiveness. The accurate identification of tree species is vital for forest ecosystem management and conservation, especially for tropical and subtropical forests that are highly complex and diverse (Cao et al., 2016). Moreover, it is of critical importance to the modelling

of tree growth (Falkowski et al., 2010), and correct estimation of biomass and tree species diversity mapping (Jones et al., 2010). Interestingly, Africa made much progress in identifying individual tree species research, which is vital for natural forest management and conservation.

Individual tree species identification is made possible with the advancement of very high-spatial resolution imageries and the availability of LiDAR. The spatial resolution of Remote Sensing imageries is key in discriminating and identifying individual tree species. In selecting imagery for the identification process, the optimal resolution is likely to depend on the forest type and the methods applied (Fassnacht et al., 2016). There is a maximization of the tree species discrimination when the pixel size of the utilized data gives room to depict the intrinsic spatial characteristics of the trees being examined (Marceau et al., 1994). Furthermore, the spectral properties based on which tree species identification is carried out requires adequate spectral dissimilarity among species (Schäfer et al., 2016). Unique spectral signatures are usually exhibited by tree species and are often linked to their biochemical and structural properties (Asner and Martin, 2009, Asner et al., 2009). Hence, there is an increase in the species identification accuracy when spatial and spectral information is combined (Feret and Asner, 2013).

The recent advanced development in the spatial resolution of multispectral imageries has helped researchers to move beyond the community-level mapping of tree species to individual species level mapping (Cho et al., 2015). Progress was made in the number of researches that employed VHR multispectral satellite imageries for individual tree species identification over the last decade. However, VHR satellite imageries such as the WorldView -2 and Worldview-3 are highly used multispectral data for this kind of research and other types such as the SkySat and Pleiades 1 are not making many breakthroughs in their application to tree species identification research despite their high spatial resolution. It is likely that they are not known, or little information is available on them. It would be recommended that the SkySat and Pleiades 1 be explored for individual tree species identification, to ascertain their capabilities to discriminate between tree species and produce high accuracies.

It was observed that hyperspectral data was employed in a majority of the studies and produced high accuracies. It is likely to be a result of the ability of the hyperspectral data to discriminate among various tree species because of its numerous narrow range bands that makes them sensitive to trees and vegetation in general. Multispectral or hyperspectral data fusion with LiDAR is also one of the approaches that produced higher accuracies in tree species identification. This method is used due to the combined ability of the LiDAR and the hyperspectral data. The LiDAR can provide structural information and as identified; the hyperspectral data have high tree species separability abilities.

Different kinds of machine learning were applied, and they have become the main means of developing models for individual tree species identification. The SVM was the most used and could be due to its robustness to noise, ability to deal with high dimensional data, less training sample requirements and fast prediction (Fassnacht et al., 2016). The non-parametric algorithms such as the RF and the SVM are found to perform better than parametric ones such as the ML (Castro-Esau et al., 2004). As such, some studies that use mixed input data variables, including spectral bands, vegetation indices, and texture variables, usually prefer to utilize non-parametric algorithms (Immitzer et al., 2012). It has led to an increase in the use of non-parametric machine learning algorithms over time. Improved computational competencies

enhance this trend in freely available new software like the R and Python statistical packages (Fassnacht et al., 2016).

Tree species diversity mapping is also of high priority for natural forest management and conservation research as well as policy development (Carlson et al., 2007). Biodiversity is of broader sense than a count of the species present, as the species composition and their relative abundance are of equal importance (Purvis and Hector, 2000). These components of biodiversity are encompassed in the concept of tree species diversity. The monitoring and measuring of tree species diversity is a requirement for mitigating loss of biodiversity and sustainable forest management (Chrysafis et al., 2020). Liang et al. (2007) found a strong relationship between tree species diversity and basal area growth. It was observed to be related to the recruitment in stands of higher tree species diversity. Remote Sensing has become a good source of information on tree species diversity at the landscape level over the years. The spatial distribution of the community of tree species is captured through modelling and prediction with majority of the reviewed studies over the last two decades were conducted in Asian and African countries with a few in South America.

The medium spatial resolution of the Landsat imageries can map the community of tree species although it cannot identify individual tree species just like other multispectral imageries. The Sentinel 2 was not used in any study but most recent studies including Mallinis et al. (2020) have shown its robustness for modelling tree species diversity. The authors of that study found the Sentinel 2 performing better than the RapidEye which has a higher spatial resolution. Thus, it is recommended that the Sentinel 2 could be utilized for tree species diversity prediction due to its capabilities. The utilization of hyperspectral data and data fusion methods was an improvement in tree species diversity prediction, which could be due to their high sensitivity to tree species. New generation imageries such as the PlanetScope, WorldView 2, WorldView 3, TripleSat could also be adopted for tree species diversity for various natural forests as they may have the capability to produce improved prediction accuracies.

Concerning the prediction algorithms, there was not much difference in those used in other thematic areas mapping as the LR was much preferred. They were able to produce satisfactory accuracies and the preference for it could be related to their performance in many studies over the years. Despite the successes achieved over the years in this thematic area, most studies have not considered phenological stages that manifest in different seasons and how they could enhance the prediction of tree species diversity. Accuracies are likely to be improved through this means and hence it is worth considering such an approach. A limitation observed is the lack of information on the best season for tree species diversity in tropical or sub-tropical natural forests. Emerging studies may explore the identification of the best season when the condition of tree species captured in imageries could improve the prediction accuracies.

Natural forest cover mapping and change detection monitoring are other thematic areas of research that are vital to forest management due to climate change, declining forest cover and increasing human population that puts pressure on forests. Mapping the extent of forests provides information on their status that could support decision-making and initiatives meant to protect and conserve the forest for the continual provision of ecosystem goods and services (Gyamfi-Ampadu et al., 2020). Remote Sensing mapping of forest cover is restricted to spatially explicit broad classes of vegetation cover but not necessarily the individual tree species (Gudex-Cross et al., 2017). The forest cover mapping enhances the understanding of

carbon sequestration and stocks, level of biodiversity, sustainability in natural resource utilization and global change (Reddy et al., 2015). Information from forest cover mapping serves as the baseline for spatio-temporal change detection analysis of forest ecosystems. Similarly, increasing trends in deforestation, forest degradation, and fragmentation increase atmospheric carbon dioxide (CO₂) emissions, which contribute to climate change and global warming. Forest cover changes also have implications on the level of carbon stocks, biodiversity, and habitats (Asner, 2009). Therefore, it is important for change detection analysis that could help to determine the extent of change over time and mitigation initiatives that could be adopted.

Researchers preferred the use of Landsat satellite imageries for both forest cover mapping and change detection analysis. The Landsat system has inherent uniqueness in the application to land cover mapping due to its longest uninterrupted Earth Observation programme and first to offer free global images (Woodcock et al., 2008, Wulder et al., 2016). The long history offers researchers the opportunity to gain vital insights on current and past change trends in land cover (Wulder et al., 2016). Its medium spatial and spectral resolution facilitates the detection of natural and anthropogenic changes at both local and regional scales (Prince et al., 2009). Furthermore, it has a wide swath width of 185 km which makes it a good image for landscape-level applications (Mutanga et al., 2016). Hence, the preference to use it may be related to these reasons as well as the free availability which enables financial resource-constrained researchers who cannot afford commercial imageries to have access to data that could enhance their research on forest cover mapping and change detection (Phiri and Morgenroth, 2017). Forest cover mapping or change detection analysis that is intended to be done within the last decade could explore the use of imageries like the Sentinel 2 and RapidEye as they are also beginning to have long and good archival imageries that can support such studies.

Concerning the classification algorithms, the ML was applied in most of the studies and high accuracies were produced (Rahman and Sumantyo, 2010, Bruggeman et al., 2016, Ochege and Okpala-Okaka, 2017). The RF and the SVM which have proving to be robust for vegetation studies also produced high accuracies in other studies (Gyamfi-Ampadu et al., 2020).

Most of the studies conducted aboveground biomass (AGB) and carbon (AGC) estimation with few of them considering belowground biomass and carbon. Although AGB and AGC forms the higher percentage of both biomass and carbon stocks, it may be worth estimating the below ground as well to be able to ascertain the total biomass and carbon especially for tropical and sub-tropical natural forest. The detrimental effects of deforestation and degradation on forest ecosystems to AGB and AGC as was rarely determine by researchers. Forest cover loss is on the rise in most tropical and sub-tropical areas and relating it to the biomass and carbon stocks could inform forest management and conservation measures. Spatial and temporal analysis of biomass and carbon studies are limited or not emphasized in studies. Such analysis is also key to knowing the spatial distribution of these forest attributes and at what point they change.

Many researchers continued to use common sensors such as Worldview2 and 3 for individual tree species identification. It will be important for emerging studies to use sensors such SkySat and Plaeides 1 to be used to ascertain their performance. The effect of phenological stages on individual tree species identification was not made clear by researchers. Also, the maximum number of tree species that can be separated and identified by sensors and models is vital information, but it is only one study that mentions and discusses its effects. There could be data

saturation of variables in highly diverse forest such as tropical and sub-tropical forests which affect accuracies. Hence, researchers could assess this and provided information for the scientific communities for initiatives to be put in place in modelling.

New generation and advanced imageries such as WorldView 2, WorldView 3, TripleSat, PlanetScope are yet to be fully adopted for tree species diversity prediction for tropical and sub-tropical natural forest ecosystems. These have high spatial resolution and could likely lead to higher accuracies as observed in some studies. Furthermore, researchers failed to determine which diversity indices would be appropriate and contribute to higher accuracies. Information on this may help other researchers know which ones to apply in prediction. In order for science to meet the current forest management demands, forest cover mapping and change detection should be related to factors and drivers of forest cover losses and changes by researchers. There could be communities-forest proximity and population density analysis and the influences they have on forest cover changes. It would be a more practical way of assessing forest changes. Forest cover mapping could be used to improve biomass and carbon estimation. Such information was not made available by most researchers. Relating forest cover mapping to biomass and carbon stocks can help to effectively assess any stock levels and how they are improving or declining.

2.5 Conclusion

Our review of progress made in Remote Sensing application to forest monitoring over the past two decades presented interesting observations based on the thematic areas. The natural forest carbon and biomass, tree species identification, tree species diversity prediction and forest cover mapping and change detection were observed to be key areas of Remote Sensing monitoring. The country of research, Remote Sensing data utilized, machine learning algorithm applied for the modelling, prediction and classification and the accuracy produced were all assessed. More research would be needed in Africa on carbon and biomass as these are directly related to climate change. This is because Africa has been identified as one of the zones to be affected most by climate change.

Advancement was observed in the types of Remote Sensing data applied to the monitoring of the various thematic areas. More freely available data such as the Landsat and Sentinel 2 were used much in African countries where there is less research funding that is hindering the utilization of commercial very high resolution, hyperspectral and active data for natural forest monitoring research. The machine learning algorithms that were used for the classification, modelling and predictions contributed much to the high accuracies observed for most of the studies.

The outcome of this review is of importance to Remote Sensing researchers who are researching tropical and sub-tropical natural forests. The research outputs can guide the selection of Remote Sensing data and machine learning algorithms that can enhance research outputs. More research is recommended in these thematic areas and other relevant ones to provide adequate and credible information to forest managers and ecologists towards efficient conservation and protection initiatives.

Acknowledgment

The authors thank the funders of this research.

Funding

This work is based on the research supported wholly by the National Research Foundation (NRF) of South Africa (Grant Number: 132924). The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest

References

- AGRAWAL, A., CASHORE, B., HARDIN, R., SHEPHERD, G., BENSON, C. & MILLER, D. 2013. Economic contributions of forests. *Background Paper*, 1.
- ANU, S., RALPH, D., DAR, R., MICHELLE, H. & J. BRYAN, B. 2011. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sensing of Environment* 115, 2917–2930.
- ASNER, G. P. 2009. Automated mapping of tropical deforestation and forest degradation: CLASlite. *Journal of Applied Remote Sensing*, 3.
- ASNER, G. P., HUGHES, R. F., MASCARO, J., UOWOLO, A. L., KNAPP, D. E., JACOBSON, J., KENNEDY-BOWDOIN, T. & CLARK, J. K. 2011. High-resolution carbon mapping on the million-hectare Island of Hawaii. *Frontiers in Ecology and the Environment*, 9, 434-439.
- ASNER, G. P. & MARTIN, R. E. 2009. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, 7, 269-276.
- ASNER, G. P., MARTIN, R. E., FORD, A. J., METCALFE, D. J. & LIDDELL, M. J. 2009. Leaf chemical and spectral diversity in Australian tropical forests. *Ecological applications*, 19, 236-253.
- ASNER, G. P. & MASCARO, J. 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, 140, 614-624.
- BROADBENT, E. N., ASNER, G. P., PEÑA-CLAROS, M., PALACE, M. & SORIANO, M. 2008. Spatial partitioning of biomass and diversity in a lowland Bolivian forest: Linking field and Remote Sensing measurements. *Forest Ecology and Management*, 255, 2602-2616.
- BROSE, U. & HILLEBRAND, H. 2016. Biodiversity and ecosystem functioning in dynamic landscapes. *Philos Trans R Soc Lond B Biol Sci*, 371.
- BRUGGEMAN, D., MEYFROIDT, P. & LAMBIN, E. F. 2016. Forest cover changes in Bhutan: Revisiting the forest transition. *Applied Geography*, 67, 49-66.
- BUSTAMANTE, M., ROITMAN, I., AIDE, T. M., ALENCAR, A., ANDERSON, L. O., ARAGÃO, L., ASNER, G. P., BARLOW, J., BERENGUER, E. & CHAMBERS, J. 2016. Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity. *Global change biology*, 22, 92-109.

- CAO, L., COOPS, N. C., INNES, J. L., DAI, J., RUAN, H. & SHE, G. 2016. Tree species classification in subtropical forests using small-footprintfull-waveform LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 49, 39–51.
- CARLSON, K. M., ASNER, G. P., HUGHES, R. F., OSTERTAG, R. & MARTIN, R. E. 2007. Hyperspectral Remote Sensing of Canopy Biodiversity in Hawaiian Lowland Rainforests. *Ecosystems*, 10, 536-549.
- CASTRO-ESAU, K. L., SÁNCHEZ-AZOFEIFA, G. & CAELLI, T. J. R. S. O. E. 2004. Discrimination of lianas and trees with leaf-level hyperspectral data. 90, 353-372.
- CHAVE, J., REJOU-MECHAIN, M., BURQUEZ, A., CHIDUMAYO, E., COLGAN, M. S., DELITTI, W. B., DUQUE, A., EID, T., FEARNESIDE, P. M., GOODMAN, R. C., HENRY, M., MARTINEZ-YRIZAR, A., MUGASHA, W. A., MULLER-LANDAU, H. C., MENCUCCINI, M., NELSON, B. W., NGOMANDA, A., NOGUEIRA, E. M., ORTIZ-MALAVASSI, E., PELISSIER, R., PLOTON, P., RYAN, C. M., SALDARRIAGA, J. G. & VIEILLEDENT, G. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob Chang Biol*, 20, 3177-90.
- CHEN, L., REN, C., ZHANG, B., WANG, Z. & XI, Y. 2018. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. *Forests*, 9.
- CHEN, L., WANG, Y., REN, C., ZHANG, B. & WANG, Z. 2019. Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data. *Remote Sensing*, 11.
- CHO, M. A., MALAHLELA, O. & RAMOELO, A. 2015. Assessing the utility WorldView-2 imagery for tree species mapping in South African subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *International Journal of Applied Earth Observation and Geoinformation*, 38, 349-357.
- CHRYSAFIS, I., KORAKIS, G., KYRIAZOPOULOS, A. P. & MALLINIS, G. 2020. Predicting Tree Species Diversity Using Geodiversity and Sentinel-2 Multi-Seasonal Spectral Information. *Sustainability*, 12.
- CLARK, M., ROBERTS, D. & CLARK, D. 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96, 375-398.
- CLARK, M. L. & ROBERTS, D. A. 2012. Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sensing*, 4, 1820-1855.
- CLARK, M. L., ROBERTS, D. A., EWEL, J. J. & CLARK, D. B. 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*, 115, 2931-2942.
- DALPONTE, M., BRUZZONE, L. & GIANELLE, D. 2008. Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 1416-1427.
- DALPONTE, M., BRUZZONE, L. & GIANELLE, D. 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment*, 123, 258-270.
- DE MORAES, J. F. L., SEYLER, F., CERRI, C. C. & VOLKOFF, B. 2010. Land cover mapping and carbon pools estimates in Rondonia, Brazil. *International Journal of Remote Sensing*, 19, 921-934.
- DUBE, T. & MUTANGA, O. 2015. 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., SHOKO, C., SAMUEL, A. & BANGIRA, T. 2016. Remote Sensing of aboveground forest biomass: A review. *Tropical Ecology*, 57, 125-132.
- FALKOWSKI, M. J., HUDAK, A. T., CROOKSTON, N. L., GESSLER, P. E., UEHLER, E. H. & SMITH, A. M. J. C. J. O. F. R. 2010. Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating LiDAR data. 40, 184-199.
- FANG, F., MCNEIL, B. E., WARNER, T. A., MAXWELL, A. E., DAHLE, G. A., EUTSLER, E. & LI, J. 2020. Discriminating tree species at different taxonomic levels using multi-temporal WorldView-3 imagery in Washington D.C., USA. *Remote Sensing of Environment*, 246.
- FASSNACHT, F. E., LATIFI, H., STERENĆZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- FASSNACHT, F. E., NEUMANN, C., FORSTER, M., BUDDENBAUM, H., GHOSH, A., CLASEN, A., JOSHI, P. K. & KOCH, B. 2014. Comparison of Feature Reduction Algorithms for Classifying Tree Species With Hyperspectral Data on Three Central European Test Sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7, 2547-2561.
- FEILHAUER, H. & SCHMIDTLEIN, S. 2009. Mapping continuous fields of forest alpha and beta diversity. *Applied Vegetation Science*, 12, 429-439.
- FERET, J.-B. & ASNER, G. P. 2013. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 73-84.
- FÉRET, J.-B. & ASNER, G. P. 2012. Semi-Supervised Methods to Identify Individual Crowns of Lowland Tropical Canopy Species Using Imaging Spectroscopy and LiDAR. *Remote Sensing*, 4, 2457-2476.
- FERREIRA, M. P., GRONDONA, A. E. B., ROLIM, S. B. A. & SHIMABUKURO, Y. E. 2013. Analyzing the spectral variability of tropical tree species using hyperspectral feature selection and leaf optical modeling. *Journal of Applied Remote Sensing*, 7.
- FERREIRA, M. P., WAGNER, F. H., ARAGÃO, L. E. O. C., SHIMABUKURO, Y. E. & DE SOUZA FILHO, C. R. 2019. Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 119-131.
- FERREIRA, M. P., ZORTEA, M., ZANOTTA, D. C., SHIMABUKURO, Y. E. & DE SOUZA FILHO, C. R. 2016. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sensing of Environment*, 179, 66-78.
- FOKENG, R. M., FORJE, W. G., MELI, V. M. & BODZEMO, B. N. 2020. Multi-Temporal Forest Cover Change Detection in the Metchie-Ngoum Protection Forest Reserve, West Region of Cameroon. *The Egyptian Journal of Remote Sensing Space Science*, 23, 113-124.
- FOODY, G. M. & CUTLER, M. E. J. 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological Modelling*, 195, 37-42.
- FUNDISI, E., MUSAKWA, W., AHMED, F. B. & TESHAMICHAEL, S. G. 2020. Estimation of woody plant species diversity during a dry season in a savanna environment using the spectral and textural information derived from WorldView-2 imagery. *PLoS One*, 15, e0234158.
- GARZON-LOPEZ, C. X., BOHLMAN, S. A., OLFF, H. & JANSEN, P. A. 2013. Mapping tropical forest trees using high-resolution aerial digital photographs. *Biotropica*, 45, 308-316.

- GASTON, K. J., CHARMAN, K., JACKSON, S. F., ARMSWORTH, P. R., BONN, A., BRIERS, R. A., CALLAGHAN, C. S., CATCHPOLE, R., HOPKINS, J. & KUNIN, W. E. 2006. The ecological effectiveness of protected areas: the United Kingdom. *Biological Conservation*, 132, 76-87.
- GHOSH, S. M. & BEHERA, M. D. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*, 96, 29-40.
- GILLESPIE, T. W., SAATCHI, S., PAU, S., BOHLMAN, S., GIORGI, A. P. & LEWIS, S. 2009. Towards quantifying tropical tree species richness in tropical forests. *International Journal of Remote Sensing*, 30, 1629-1634.
- GONZÁLEZ-JARAMILLO, V., FRIES, A., ZEILINGER, J., HOMEIER, J., PALADINES-BENITEZ, J. & BENDIX, J. 2018. Estimation of Above Ground Biomass in a Tropical Mountain Forest in Southern Ecuador Using Airborne LiDAR Data. *Remote Sensing*, 10.
- GOULSON, D. 1999. Foraging strategies of insects for gathering nectar and pollen, and implications for plant ecology and evolution. *Perspectives in Plant Ecology, Evolution Systematics*, 2, 185-209.
- GRABSKA, E., HOSTERT, P., PFLUGMACHER, D. & OSTAPOWICZ, K. 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, 11.
- GUDEX-CROSS, D., PONTIUS, J. & ADAMS, A. 2017. Enhanced forest cover mapping using spectral unmixing and object-based classification of multi-temporal Landsat imagery. *Remote Sensing of Environment*, 196, 193-204.
- GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2020. Mapping natural forest cover using satellite imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, 18, 1-14.
- HAUTIER, Y., SEABLOOM, E. W., BORER, E. T., ADLER, P. B., HARPOLE, W. S., HILLEBRAND, H., LIND, E. M., MACDOUGALL, A. S., STEVENS, C. J., BAKKER, J. D., BUCKLEY, Y. M., CHU, C., COLLINS, S. L., DALEO, P., DAMSCHEN, E. I., DAVIES, K. F., FAY, P. A., FIRN, J., GRUNER, D. S., JIN, V. L., KLEIN, J. A., KNOPS, J. M., LA PIERRE, K. J., LI, W., MCCULLEY, R. L., MELBOURNE, B. A., MOORE, J. L., O'HALLORAN, L. R., PROBER, S. M., RISCH, A. C., SANKARAN, M., SCHUETZ, M. & HECTOR, A. 2014. Eutrophication weakens stabilizing effects of diversity in natural grasslands. *Nature*, 508, 521-5.
- HERNÁNDEZ-STEFANONI, J., DUPUY, J., JOHNSON, K., BIRDSEY, R., TUN-DZUL, F., PEDUZZI, A., CAAMAL-SOSA, J., SÁNCHEZ-SANTOS, G. & LÓPEZ-MERLÍN, D. 2014. Improving Species Diversity and Biomass Estimates of Tropical Dry Forests Using Airborne LiDAR. *Remote Sensing*, 6, 4741-4763.
- HOLMGREN, J. & PERSSON, Å. 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, 90, 415-423.
- IMMITZER, M., ATZBERGER, C. & KOUKAL, T. 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, 4, 2661-2693.
- JONES, T. G., COOPS, N. C. & SHARMA, T. J. R. S. O. E. 2010. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. 114, 2841-2852.
- JOSHI, P. K. K., ROY, P. S., SINGH, S., AGRAWAL, S. & YADAV, D. 2006. Vegetation cover mapping in India using multi-temporal IRS Wide Field Sensor (WiFS) data. *Remote Sensing of Environment*, 103, 190-202.

- KEENAN, R. J., REAMS, G. A., ACHARD, F., DE FREITAS, J. V., GRAINGER, A. & LINDQUIST, E. 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, 352, 9-20.
- KHAN, I. A., KHAN, M. R., BAIG, M. H. A., HUSSAIN, Z., HAMEED, N. & KHAN, J. A. 2020. Assessment of forest cover and carbon stock changes in sub-tropical pine forest of Azad Jammu & Kashmir (AJK), Pakistan using multi-temporal Landsat satellite data and field inventory. *PLoS One*, 15, e0226341.
- KOCH, B. 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral Remote Sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65, 581-590.
- LAURIN, G. V., PULETTI, N., HAWTHORNE, W., LIESENBERG, V., CORONA, P., DARIO PAPAIE, D., CHEN, Q. & VALENTINI, R. 2016. Discrimination of tropical forest types, dominant species and mapping function guilds by hyperspectral and simulated Sentinel 2. *Remote Sensing of the Environment*, 164-176.
- LAURIN, V. G., CHEUNG-WAI CHAN, J., CHEN, Q., LINDSELL, J. A., COOMES, D. A., GUERRIERO, L., DEL FRATE, F., MIGLIETTA, F. & VALENTINI, R. 2014. Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. *PLoS One*, 9, e97910.
- LI, LI, LI & LIU 2019. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. *Forests*, 10.
- LIANG, J., BUONGIORNO, J., MONSERUD, R. A., KRUGER, E. L. & ZHOU, M. 2007. Effects of diversity of tree species and size on forest basal area growth, recruitment, and mortality. *Forest Ecology and Management*, 243, 116-127.
- LIN, C., POPESCU, S. C., THOMSON, G., TSOGT, K. & CHANG, C. I. 2015. Classification of Tree Species in Overstorey Canopy of Subtropical Forest Using QuickBird Images. *PLoS One*, 10, e0125554.
- LIN, C., THOMSON, G. & POPESCU, S. 2016. An IPCC-Compliant Technique for Forest Carbon Stock Assessment Using Airborne LiDAR-Derived Tree Metrics and Competition Index. *Remote Sensing*, 8.
- LU, D., CHEN, Q., WANG, G., LIU, L., LI, G. & MORAN, E. 2014. A survey of Remote Sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, 9, 63-105.
- MACDICKEN, K. G. 2015. Global Forest Resources Assessment 2015: What, why and how? *Forest Ecology and Management*, 352, 3-8.
- MAEDA, E. E., HEISKANEN, J., THIJS, K. W. & PELLIKKA, P. K. E. 2014. Season-dependence of Remote Sensing indicators of tree species diversity. *Remote Sensing Letters*, 5, 404-412.
- MALHI, Y., ADU-BREDU, S., ASARE, R. A., LEWIS, S. L. & MAYAUX, P. 2013. African rainforests: past, present and future. *Philosophical Transactions of the Royal Society B*, 368, 20120312.
- MALLINIS, G., CHRYSAFIS, I., KORAKIS, G., PANA, E. & KYRIAZOPOULOS, A. P. 2020. A Random Forest Modelling Procedure for a Multi-Sensor Assessment of Tree Species Diversity. *Remote Sensing*, 12.
- 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, 105-117.
- MARGONO, B. A., POTAPOV, P. V., TURUBANOVA, S., STOLLE, F. & HANSEN, M. C. 2014. Primary forest cover loss in Indonesia over 2000–2012. *Nature Climate Change*, 4, 730-735.

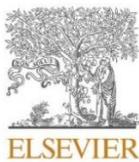
- MARTINEZ DEL CASTILLO, E., GARCÍA-MARTIN, A., LONGARES ALADRÉN, L. A. & DE LUIS, M. 2015. Evaluation of forest cover change using Remote Sensing techniques and landscape metrics in Moncayo Natural Park (Spain). *Applied Geography*, 62, 247-255.
- MAYAUX, P., HOLMGREN, P., ACHARD, F., EVA, H., STIBIG, H. J. & BRANTHOMME, A. 2005. Tropical forest cover change in the 1990s and options for future monitoring. *Philos Trans R Soc Lond B Biol Sci*, 360, 373-84.
- MAYAUX, P., PEKEL, J.-F., B. D. E., DONNAY, F., LUPI, A., ACHARD, F., CLERICI, M., BODART, C., BRINK, A., NASI, R. & BELWARD, A. 2013. State and Evolution of the African rainforests between 1990 and 2010. *Phil Trans R Soc B*.
- MAYNARD, K. & ROYER, J.-F. 2004. Sensitivity of a general circulation model to land surface parameters in African tropical deforestation experiments. *Climate Dynamics*, 22, 555-572.
- MBAABU, P. R., HUSSIN, Y. A., WEIR, M. & GILANI, H. 2014. Quantification of carbon stock to understand two different forest management regimens in Kayar Khola Watershed Nepal. *Indian Society of Remote Sensing* 42, 745–754.
- MIURA, S., AMACHER, M., HOFER, T., SAN-MIGUEL-AYANZ, J., ERNAWATI & THACKWAY, R. 2015. Protective functions and ecosystem services of global forests in the past quarter-century. *Forest Ecology and Management*, 352, 35-46.
- MUTANGA, O., DUBE, T. & AHMED, F. 2016. Progress in Remote Sensing: vegetation monitoring in South Africa. *South African Geographical Journal*, 98, 461-471.
- NAGENDRA, H., ROCCHINI, D., GHATE, R., SHARMA, B. & PAREETH, S. 2010. Assessing Plant Diversity in a Dry Tropical Forest: Comparing the Utility of Landsat and Ikonos Satellite Images. *Remote Sensing*, 2, 478-496.
- OCHEGE, U. F. & OKPALA-OKAKA, C. 2017. RS of vegetation cover changes in the humid tropical rainforests of Southeastern Nigeria 1984 2014. *Cogent Geoscience*, 3.
- OMER, G., MUTANGA, O., ABDEL-RAHMAN, E. M. & ADAM, E. 2015. Performance of Support Vector Machines and Artificial Neural Network for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 4825-4840.
- PAN, Y., BIRDSEY, R. A., FANG, J., HOUGHTON, R., KAUPPI, P. E., KURZ, W. A., PHILLIPS, O. L., SHVIDENKO, A., LEWIS, S. L. & CANADELL, J. G. 2011. A large and persistent carbon sink in the world's forests. *Science*, 1201609.
- PANDEY, P. C., SRIVASTAVA, P. K., CHETRI, T., CHOUDHARY, B. K. & KUMAR, P. 2019. Forest biomass estimation using Remote Sensing and field inventory: a case study of Tripura, India. *Environ Monit Assess*, 191, 593.
- PANDIT, S., TSUYUKI, S. & DUBE, T. 2018a. Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sensing*, 10.
- PANDIT, S., TSUYUKI, S. & DUBE, T. 2018b. Landscape-Scale Aboveground Biomass Estimation in Buffer Zone Community Forests of Central Nepal: Coupling In Situ Measurements with Landsat 8 Satellite Data. *Remote Sensing*, 10.
- PHIRI, D. & MORGENROTH, J. 2017. Developments in Landsat land cover classification methods: A review. *Remote Sensing*, 9, 967.
- PRINCE, S., BECKER-RESHEF, I. & RISHMAWI, K. 2009. Detection and mapping of long-term land degradation using local net production scaling: Application to Zimbabwe. *Remote Sensing of Environment*, 113, 1046-1057.
- PURVIS, A. & HECTOR, A. 2000. Getting the measure of biodiversity. *Nature*, 405, 212-219.
- QIN, Y., XIAO, X., DONG, J., ZHANG, G., SHIMADA, M., LIU, J., LI, C., KOU, W. & MOORE, B. 2015. Forest cover maps of China in 2010 from multiple approaches and

- data sources: PALSAR, Landsat, MODIS, FRA, and NFI. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 1-16.
- RAHMAN, M. M. & SUMANTYO, J. T. S. 2010. Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. *Applied Geomatics*, 2, 113-121.
- RANAGALAGE, M., WANG, R., GUNARATHNA, M. H. J. P., DISSANAYAKE, D. M. S. L. B., MURAYAMA, Y. & SIMWANDA, M. 2019. Spatial Forecasting of the Landscape in Rapidly Urbanizing Hill Stations of South Asia: A Case Study of Nuwara Eliya, Sri Lanka (1996–2037). *Remote Sensing*, 11.
- REDDY, C. S., JHA, C. S., DIWAKAR, P. G. & DADHWAL, V. K. 2015. Nationwide classification of forest types of India using Remote Sensing and GIS. *Environ Monit Assess*, 187, 777.
- ROY, D. P., WULDER, M. A., LOVELAND, T. R., C.E, W., ALLEN, R. G., ANDERSON, M. C., HELDER, D., IRONS, J. R., JOHNSON, D. M., KENNEDY, R., SCAMBOS, T. A., SCHAAF, C. B., SCHOTT, J. R., SHENG, Y., VERMOTE, E. F., BELWARD, A. S., BINDSCHADLER, R., COHEN, W. B., GAO, F., HIPPLE, J. D., HOSTERT, P., HUNTINGTON, J., JUSTICE, C. O., KILIC, A., KOVALSKYY, V., LEE, Z. P., LYMBURNER, L., MASEK, J. G., MCCORKEL, J., SHUAI, Y., TREZZA, R., VOGELMANN, J., WYNNE, R. H. & ZHU, Z. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172.
- RUIZ-BENITO, P., GÓMEZ-APARICIO, L., PAQUETTE, A., MESSIER, C., KATTGE, J. & ZAVALA, M. A. 2014. Diversity increases carbon storage and tree productivity in Spanish forests. *Global Ecology and Biogeography*, 23, 311-322.
- SAATCHI, S. S., HARRIS, N. L., BROWN, S., LEFSKY, M., MITCHARD, E. T., SALAS, W., ZUTTA, B. R., BUERMANN, W., LEWIS, S. L. & HAGEN, S. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108, 9899-9904.
- SARKER, L. R. & NICHOL, J. E. 2011. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sensing of Environment*, 115, 968-977.
- SARKER, M. L. R., NICHOL, J., IZ, H. B., AHMAD, B. B. & RAHMAN, A. A. 2013. Forest Biomass Estimation Using Texture Measurements of High-Resolution Dual-Polarization C-Band SAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 3371-3384.
- SCHÄFER, E., HEISKANEN, J., HEIKINHEIMO, V. & PELLIKKA, P. 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. *Ecological Indicators*, 64, 49-58.
- SOLAR, R. R. D. C., BARLOW, J., FERREIRA, J., BERENGUER, E., LEES, A. C., THOMSON, J. R., LOUZADA, J., MAUÉS, M., MOURA, N. G. & OLIVEIRA, V. H. 2015. How pervasive is biotic homogenization in human-modified tropical forest landscapes? *Ecology Letters*, 18, 1108-1118.
- SOMERS, B. & ASNER, G. P. 2014. Tree species mapping in tropical forests using multi-temporal imaging spectroscopy: Wavelength adaptive spectral mixture analysis. *International Journal of Applied Earth Observation and Geoinformation*, 31, 57-66.
- VAN DEVENTER, H., CHO, M. A. & MUTANGA, O. 2017. Improving the classification of six evergreen subtropical tree species with multi-season data from leaf spectra simulated to WorldView-2 and RapidEye. *International Journal of Remote Sensing*, 38, 4804-4830.
- VASUDEVA, V., NANDY, S., PADALIA, H., SRINET, R. & CHAUHAN, P. 2020. Mapping spatial variability of foliar nitrogen and carbon in Indian tropical moist deciduous forest

- using sentinel data. *INTERNATIONAL JOURNAL OF REMOTE SENSING*, 42 1139–1159.
- VOIGHT, C., HERNANDEZ-AGUILAR, K., GARCIA, C. & GUTIERREZ, S. 2019. Predictive Modeling of Future Forest Cover Change Patterns in Southern Belize. *Remote Sensing*, 11.
- WAGNER, F. H., FERREIRA, M. P., SANCHEZ, A., HIRYE, M. C. M., ZORTEA, M., GLOOR, E., PHILLIPS, O. L., DE SOUZA FILHO, C. R., SHIMABUKURO, Y. E. & ARAGÃO, L. E. O. C. 2018. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 362-377.
- WALLIS, C. I. B., HOMEIER, J., PEÑAC, J., BRANDL, R., FARWIG, N. & BENDIX, J. 2019. Modeling tropical montane forest biomass, productivity and canopy traits with multispectral Remote Sensing data. *Remote Sensing of Environment*, 77–92.
- WANG, S. & LOREAU, M. 2016. Biodiversity and ecosystem stability across scales in metacommunities. *Ecol Lett*, 19, 510-8.
- WANGDA, P., HUSSIN, Y. A., BRONSVELD, M. C. & KARNA, Y. K. 2019. Species stratification and upscaling of forest Carbon estimates to landscape scale using GeoEye Nepal. *International Journal of Remote Sensing*, 40, 7941-7965.
- WOODCOCK, C. E., ALLEN, R., ANDERSON, M., BELWARD, A., BINDSCHADLER, R., COHEN, W., GAO, F., GOWARD, S. N., HELDER, D. & HELMER, E. 2008. Free access to Landsat imagery. *Science*, 320.
- WULDER, M. A., WHITE, J. C., LOVELAND, T. R., WOODCOCK, C. E., BELWARD, A. S., COHEN, W. B., FOSNIGHT, E. A., SHAW, J., MASEK, J. G. & ROY, D. P. 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, 185, 271-283.
- ZHANG, J., RIVARD, B., SÁNCHEZ-AZOFEIFA, A. & CASTRO-ESAU, K. 2006. Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. *Remote Sensing of Environment*, 105, 129-141.
- ZHANG, M., DU, H., ZHOU, G., LI, X., MAO, F., DONG, L., ZHENG, J., LIU, H., HUANG, Z. & HE, S. 2019. Estimating Forest Aboveground Carbon Storage in Hang-Jia-Hu Using Landsat TM/OLI Data and Random Forest Model. *Forests*, 10.
- ZHAO, Y., ZENG, Y., ZHENG, Z., DONG, W., ZHAO, D., WU, B. & ZHAO, Q. 2018. Forest species diversity mapping using airborne LiDAR and hyperspectral data in a subtropical forest in China. *Remote Sensing of Environment*, 213, 104-114.

CHAPTER 3. MAPPING NATURAL FOREST COVER USING SATELLITE IMAGERY OF NKANDLA FOREST RESERVE, KWAZULU-NATAL, SOUTH AFRICA

Remote Sensing Applications: Society and Environment 18 (2020) 100302



Contents lists available at [ScienceDirect](#)

Remote Sensing Applications: Society and Environment

journal homepage: <http://www.elsevier.com/locate/rsase>



Mapping natural forest cover using satellite imagery of Nkandla forest reserve, KwaZulu-Natal, South Africa



Enoch Gyamfi-Ampadu^{a,*}, Michael Gebreslasie^a, Alma Mendoza-Ponce^{b,c}

^a School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban, 4000, South Africa

^b International Institute of Applied System Analysis, Schloßpl. 1, 2361, Laxenburg, Austria

^c Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, Mexico City, Mexico

Abstract

Natural forest ecosystems are vital environmental resources that provide multiple benefits to society, making it imperative to be monitored and mapped for practical management purposes. Satellite Remote Sensing technology is a new source of data and information for forest management and conservation. This study, therefore, applied Support Vector Machine (SVM) and Random Forest (RF) algorithms to Landsat 8 image for mapping a natural forest in South Africa. The objectives were to classify the forest into specific thematic cover classes that indicate its condition, compare the classification performance of the two algorithms based on their default parameters, and determine the most important variables that contributed to the mapping accuracy. The closed canopy forest was determined as the dominant thematic class, followed in descending order by the open canopy forest, grassland, and bare sites. Both algorithms obtained high classification accuracies of above 95%, although the SVM was slightly superior to the RF. The McNemer test indicated that the difference in performance between the two algorithms was statistically insignificant. The most important variables that contributed to the accuracy were the red, blue, green, Near Infrared and Short-Wave Infrared bands, which is attributed to their sensitivity to vegetation. The information provided through the study can be utilized for the planning, management and prioritization initiatives aimed at the protection and conservation of the forest reserve and similar forest ecosystems. The mapping approach could be used for other natural forest ecosystems to ascertain the spatial coverage of the specific thematic cover for conservation purposes. The SVM is recommended for forest ecosystem mapping as it optimally utilized the capabilities of the spectral bands that reflect their actual importance in the mapping of each cover class. The bands identified as important variables can be incorporated as part of input variables when using Landsat 8 satellite imagery for natural forest mapping.

Keywords: Forest ecosystem, Remote Sensing, Support Vector Machine, Random Forest, Mapping, Conservation.

3.1. Introduction

Natural forest ecosystems are a key resource serving a multitude of functions in the provision of goods and services (Cardinale et al., 2012, Gilroy et al., 2014, Lohbeck et al., 2016, Pedro et al., 2015) which contributes immensely to the wellbeing of society. These include the provision of habitat for wildlife species and maintenance of biodiversity, amelioration of local climate, carbon sequestration, provision of aesthetic values, protection of watersheds as well serving as a source of food, medicine, timber, and non-timber forest products [NTFPs] (Miura et al., 2015, Jucker et al., 2016, Mori et al., 2017, Jactel et al., 2017). These multiple benefits can be sustained when forests are intact and undisturbed. However, forest ecosystems are threatened by deforestation and degradation over time mainly due to human activities (Nkonya et al., 2016).

Forest cover mapping and monitoring are therefore essential to provide adequate data and information on the status and condition of the forest to assist in planning and initiatives geared

towards sustainable forest management. Some of this information might include the spatial coverage and scale (Chen and Bradshaw, 1999), functional composition (Laurin et al., 2016b), afforestation and deforestation rates (Hirose et al., 2016, Omruuzun et al., 2015), forest types and successional stages (Laurin et al., 2013), and tree species information (Immitzer et al., 2012). The generation of such data and information on natural forest ecosystems is vital as they could be proxy or input data for carbon stock estimation (Nogueira et al., 2015), forest species distribution modelling (Foody et al., 2003), forest cover change detection and ecosystem services assessment (Balthazar et al., 2015).

Requisite forest cover information can be obtained by either traditional inventory or Remote Sensing approaches. However, traditional methods are time-consuming and may cover only a small extent of the area. Remote Sensing methods, on the other hand, cover large areas (Turner et al., 2003, Turner, 2014, Skidmore and Pettorelli, 2015, Jetz et al., 2016), provides cost-effective temporal data (Wulder et al., 2004, dBozkaya et al., 2015) and able to measure and estimate the richness and distribution for forests (Rocchini et al., 2005, He et al., 2011, Pettorelli et al., 2014). Furthermore, the data and information generated could easily be updated, transferred and shared. These make Remote Sensing approaches have an advantage over the traditional methods and hence its adoption for obtaining and providing data and information for sustainable forest management.

Remote Sensing hyperspectral or multispectral data are adopted for the classification and prediction of forest conditions and attributes. Hyperspectral sensors produce a high amount of spectral information from land surface objects due to their numerous narrow spectral bands (Dalponte et al., 2008). This enables them to produce high accuracies as demonstrated by some studies. For instance, Petropoulos et al. (2012) employed the Hyperion hyperspectral imagery to optimally map land cover using the support vector machine (SVM) and object-oriented method. In other studies, Luo et al. (2015) fused Light Detection and Ranging (LiDAR) and Compact Airborne Stereographic Imager (CASI) to map land cover classification. It proved useful as the height and ground-based information contributed to obtaining high accuracy. The SVM outperformed the maximum likelihood algorithm (ML) when they were applied to the fused data. Ghamisi et al. (2015) similarly fused LiDAR and AISA Eagle data and LiDAR and Airborne Laser Mapper (ALS) data for mapping land cover in two different regions. This also yielded high accuracy while the RF outperformed the SVM when applied to the data.

Multispectral satellite data are alternative to hyperspectral data as they are also capable of mapping land cover classes accurately. For instance, in a comparative study, Jia et al. (2014) employed Landsat 8 and Landsat 7 data sets and SVM and ML algorithms for land cover mapping using a texture-based approach. Landsat 8 was more capable in the mapping process due to its high spectral and spatial resolution while the SVM also proved superior to the ML. Another comprehensive comparison was carried out by applying object and pixel-based classification approach and ML, SVM, k-Nearest Neighbour (KNN), Feature Analyst (FA) and Spectral Mixture Analysis (SMA) algorithms to Landsat 8 and Landsat 5 (Poursanidis et al., 2015). The study found the pixel based SVM and the Landsat 8 to have produced a more accurate land cover map due to the ability of the SVM to utilize the pixels and the resolution of the Landsat 8, respectively.

In other land cover mapping, the multitemporal and angular MODIS data was found to be more efficient in estimating tree cover extent as compared to nadir-view multispectral data (Heiskanen and Kivinen, 2008). The potential of Multispectral Airborne Laser Scanning (MALS) which provides 3D point clouds have also been assessed for land cover mapping (Matikainen et al., 2017). It turned out that the MALS was optimal in mapping both ground level and elevated classes and was indicated to have the potential to undertake automated land-use change detection. Over the years, other multispectral data such as IKONOS (Kim et al., 2011), SPOT (Lobo et al., 2010), RapidEye (Adam et al., 2014b) have equally demonstrated to be good data types that could be utilized for both local and regional land cover mapping.

Prediction of forest attributes has also been undertaken using both hyperspectral and multispectral data. Hyperspectral data have been used for predicting tree species (Clark and Roberts, 2012, Dalponte et al., 2013, Feret and Asner, 2013, Ghosh et al., 2014) and they were more efficient in identifying individual species as well as discriminating variations among species of similar spectral characteristics. This was attributed to its numerous spectral bands allowing for identification of minimal variation among spectral attributes of species. Advanced multispectral satellite data with improved spectral resolution have also been used in predicting species diversity and richness (Gillespie et al., 2009, Nagendra et al., 2010, Omer et al., 2015, Sheeren et al., 2016). All these data types have contributed to enhancing the understanding of forest and environmental systems for sustainable management.

It is acknowledged that in land cover and landscape mapping, the accuracy is not based on the use of the right data or image alone, but is equally dependent on the classification algorithm (Lu and Weng, 2007). In view of that, machine learning algorithms have hence been employed by Remote Sensing researchers and scientists for land cover mapping in recent years (Thanh Noi and Kappas, 2017). These algorithms are non-parametric, which do not assume normality in data distribution and are effective and efficient in terms of process time and have the ability to produce high accuracies than parametric algorithms (Fassnacht et al., 2016). The algorithm learns the characteristics of the object through the training samples and uses it to identify the characteristics of the unclassified data (Belgiu and Drăguț, 2016).

The RF (Breiman, 2001) and the SVM (Vapnik, 1995) have been two top machine learning algorithms applied to the hyperspectral and multispectral data for spatial and spectral analysis. The use and popularity of RF and SVM have increased significantly among the other non-parametric algorithms (Adam et al., 2014b, Thanh Noi and Kappas, 2017) mainly due to their superiority in handling complex Remote Sensing data. Apart from already mentioned research, other studies involving forest and other landscapes classification that have employed the RF (Rodriguez-Galiano et al., 2012, Stefanski et al., 2013, Nitze et al., 2015, Tsutsumida and Comber, 2015, Ahmed et al., 2015) and the SVM (Petropoulos et al., 2012, Paneque-Gálvez et al., 2013, Singh et al., 2013) affirms their capabilities. They are mostly used either together for comparative studies or on a single basis for mapping purposes.

Our study set out to use Landsat 8 satellite image for mapping the natural forest cover of the Nkandla Forest Reserve in KwaZulu-Natal, South Africa with the RF and SVM algorithms. The objectives of the study are hence to; i) map the thematic cover classes of Nkandla Forest Reserve with the SVM and RF algorithms, ii) assess the performances of the RF and SVM

algorithms, and iii) determine the important Landsat 8 bands valuable for accurate forest cover mapping. The study was necessitated because there is no quantitative information that provide details on the land use land cover classes of the Nkandla Forest Reserve. It important to indicate that the performance of Landsat 8 imagery and the RF and SVM algorithm has been provided in many studies. That notwithstanding, their performances could be influenced by the type of forest, the modelling approach and the expertise of the researcher in their application. Therefore, it is vital that further studies are carried out in different forest types and good modelling approach utilised to ascertain their performances. Hence, the need for our study to be conducted to assess the performance of the Landsat 8 as well as the RF and SVM algorithms for the sub-tropical Afromontane forest. With regards to the choice of study area, the Nkandla Forest Reserve was used for the study due to its unique attribute as an Afromontane natural forest. Furthermore, it is among the limited natural forests coverage in South Africa and information on its land use land cover (LULC) is non-existent. Thus, the outcome of our study will provide information to forest managers and policymakers for the management and prioritization of initiatives aimed at conserving the forest in the wake of increasing climate change and habitat degradation. It is also expected that findings could contribute to the knowledge base on Landsat 8 spectral bands and the two machine learning algorithms to enhance their use in other studies.

3.2 Materials and Methods

3.1 Study Area

The Nkandla Forest Reserve is a natural forest of unique attributes and among the limited natural forest ecosystems in South Africa (Figure 3.1). It is an Afromontane Sub-tropical forest in the KwaZulu-Natal Province and located on latitude 28° 43' 50.88" S and longitude 30° 7' 9.84" E. It has a temperate, subtropical climate with the highest average temperature of 27°C in December and January and the lowest average of 2°C in the winter months of June and July (Ezemvelo KZN Wildlife, 2015a). The topography is generally steep and undulating, with an altitude extending up to about 1300 m. The South African Government notice of 1st March 1918 established it as a forest reserve with an estimated area of about 2218 ha. It was re-proclaimed as a forest reserve and gazetted under the KwaZulu Government Conservation Act 1992. The forest is of socio-cultural importance to the fringing communities and it is formally managed under a semi-protected management system. Surrounding communities are allowed to obtain non-timber forest products (NTFPs) and domestic animals including cattle are allowed into the reserve to graze. The reserve has recorded some invasive plants and trees such as bugweed (*Solanum mauritianum*), American bramble (*Rubus cuneifolius*), Mauritius thorn (*Caesalpinia decapatala*), Australian blackwood (*Acacia melanoxylon*) and Gum tree (*Eucalyptus spp*). The patches of grassland are found mostly on and along hills with some patches found between areas of forest tree cover. Also, certain boundaries are surrounded by grassland which experiences frequent fires. The fringing and adjacent communities are fast expanding and transforming due to developmental activities, plantation development, and crop and animal farming. There is a single asphalted road that runs through the forest reserve linking

the communities. There is also office space and staff residences that cover some portions of the reserve area.

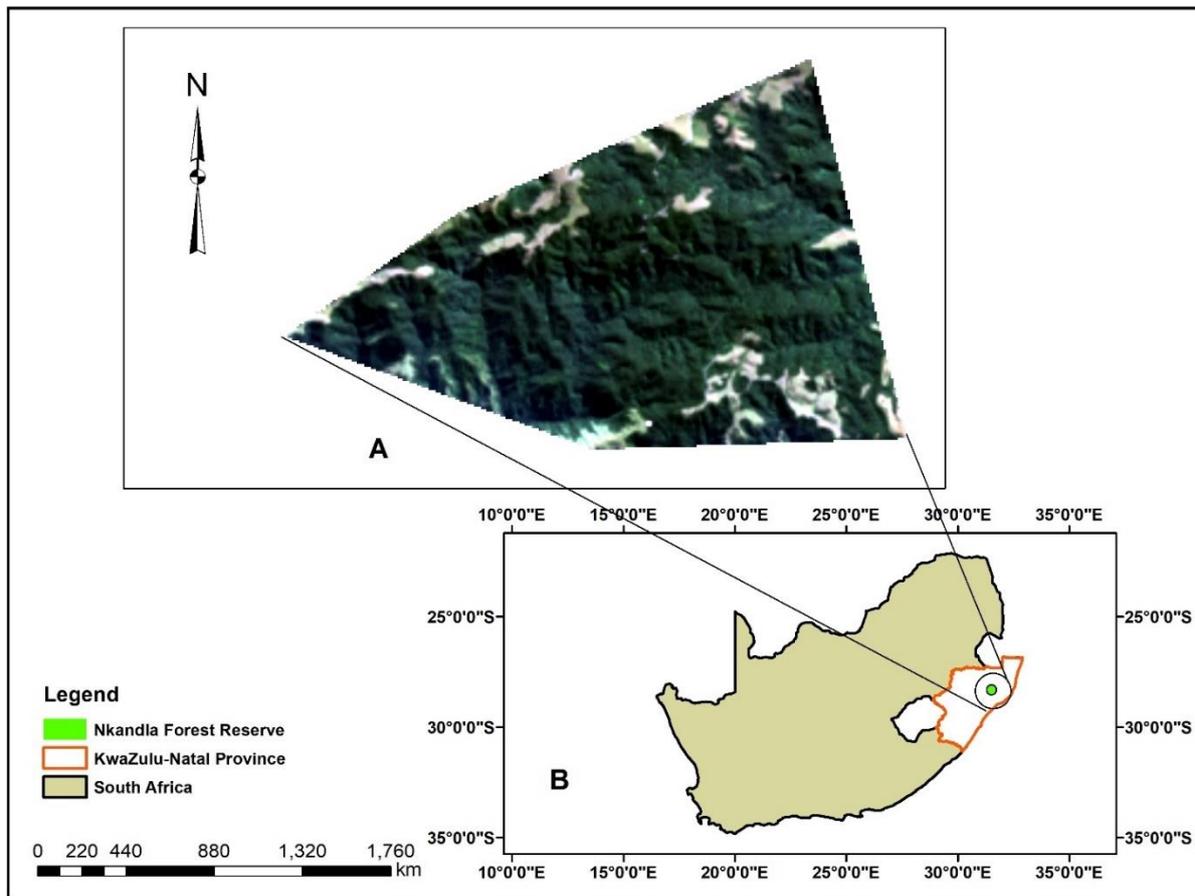


Figure 3.1: Map of the study area. Note: A is the Landsat 8 satellite image of the Nkandla Forest Reserve, and B is a map of South Africa indicating the location of the study area.

3.2 Definition of forest cover classes

Four thematic cover classes were defined as deemed relevant to study the condition of the Nkandla Forest Reserve. The classes are closed canopy forest, open canopy forest, grassland, and bare site areas (Table 3.1). With regards to the determining the four thematic classes, a reconnaissance survey was carried out in the forest to have an overview of the thematic cover classes before the actual data collection was undertaken. This served as basis for the land use land cover classification.

Table 3.1: Definition of thematic cover classes used in the study

Forest class	Code	Definition
Closed canopy forest	CCF	This forest class has a canopy of trees touching each other and forming a continuous canopy layer (70% to 100%) that does not allow much sunlight onto the forest floor. It has low vegetation on the forest floor, and visibility under the canopy could be up to about 20 m and beyond.
Open canopy forest	OCF	This forest class does not have a continuous formation of the canopy (30% to 70%). It has gaps in the crown of trees and allows much sunlight onto the forest floor. This area has much growth of seedlings, saplings, herbaceous plants, shrubs, and weeds. Visibility under the canopy could be less than 20 m.
Grassland	GL	This class is characterized by the continuous growth of grass and herbaceous layer.
Bare sites	BS	This class comprises sites within the forest that do not have vegetation cover at all.

3.3 Data Used

The Remote Sensing and the field data used for the study have been described below.

3.3.1 Remote Sensing data acquisition and preprocessing

A Landsat 8 satellite imagery was used for the study. The recently launched Landsat 8 sensor is among satellites providing multispectral images that are used for different forest and land cover classification (Poursanidis et al., 2015, Fassnacht et al., 2015, Sothe et al., 2017, Chen et al., 2017, Pastor-Guzman et al., 2015). The Landsat 8 multispectral images have improved radiometric sensitivity (overall noise reduction) which enhances the characterization of forest conditions as well as the refined spectral range of bands including the near-infrared (NIR) that are vital for improved spectral response to vegetation (Pahlevan et al., 2014, El-Askary et al., 2014). Furthermore, it has improved signal to noise ratio (SNR) and increased spectral record precision resulting from the increased sensor-dwell time at each ground pixel (Jia et al., 2014, Irons et al., 2012).

A cloud-free image captured on 8th May 2019 was downloaded from the United States Geological Survey (USGS) website. The date of the image was chosen to closely follow the time of field data collection as discussed in the section below. The image was atmospherically and geometrically corrected. Eight spectral bands of the image were then selected based on their spatial resolution and used as input variables for the classification algorithms. The selected bands were Coastal aerosol (Band 1), Blue (Band 2), Green (Band 3), Red (Band 5),

Near infra-red (Band 5), Short Wave infra-red 1 (Band 6), Short wave infra-red 2 (Band 7) and Cirrus (Band 9).

3.3.2 Reference data collection and processing

Ground-based field techniques were used in collecting reference data sets (Lowry et al., 2007) for developing and validating the mapping model. It was based on predefined thematic cover classes and done in a manner that is representative of each cover class of the forest reserve (Foody, 2004, Foody and Mathur, 2004). The fieldwork was conducted from 22nd April 2019 to 7th May 2019 where reference points were collected through a random data collection approach. Collecting randomly placed reference data across a study area for each thematic cover helps to avoid opportunistic biased classification (Hammond and Verbyla, 1996, Friedl et al., 2000, Zhen et al., 2013). Furthermore, this approach was used because apart from the grassland areas and the bare sites which could be clearly identified from afar, that of the closed and open canopy forest covers cannot be determined as such. A minimum area of 25m² and an ocular estimation of site composition and structural parameters were hence used to determine which area falls under each class especially for the closed and open canopy forest covers when they were opportunistically encountered. This was done through traversing along transects across the forest into the different areas in the middle, north, south east and western parts.

Global Position System (GPS) points were recorded for a class when it met the minimum area (25m²), site composition and structural parameters. The number of points collected at a site was dependent on the extent of a cover class. More than one point was recorded for a cover class within a site that had wide coverage to ensure that it is well delineated in the mapping process. The number of points to be collected for each cover class was however not predetermined which gives reasons for the differences in the number of points for each of them. The points were superimposed on the satellite image and random-sized polygons were digitized around them on a class-by-class basis in the ArcMap 10.6.1 mapping environment. The satellite image and the cover class reference polygons were then imported into the R statistical package (Team, 2017). The pixel values of the cover class reference polygons were extracted and subsequently divided into a training data set (70%) and independent validation data set (30%) [Table 3.2] for each cover class using a randomized approach. The SVM and RF algorithms were trained using the training data set, while the accuracy assessment was done using the independent validation data set.

Table 3.2: Training and validation data sets used in the study.

Thematic cover category	Reference points	No. of Polygons	Training set pixels	Validation set pixels	No. of pixels
Closed canopy forest	93	63	160	69	229
Open canopy forest	81	76	120	45	165
Grassland	78	69	108	43	151
Bare sites	49	20	20	11	31
Total	301	228	408	168	576

3.4 Image classification

The pixel-based approach was used for the classification of the thematic cover classes. These processes were carried out in the R statistical package environment (Team, 2017) with the RF and SVM algorithms. The design mechanism of the RF and SVM algorithms, as well as the packages (libraries) and functions used for each in the classification, have been detailed below.

3.4.1 Random Forest

The RF is an ensemble classifier that makes provision for growing a large number of trees after which the class is determined through a vote for the most popular class (Breiman, 2001). The RF uses the bagging approach where trees are created from a subset of the training dataset through a replacement process. The bagging ensemble approach has been demonstrated to achieve higher accuracy than single tree classifiers such as the Decision Tree (DT) and is stable and not sensitive to noise in the training data or overtraining (Briem et al., 2002, DeFries and Chan, 2000, Pal and Mather, 2003, Chan and Paelinckx, 2008).

In the process, about two-thirds of the sample data are used to train the classifier as a training or calibration set. The other one third, which is called the out-of-bag (OOB) set that was not part of the training set are reserved and used for internal cross-validation that provides an estimate on the performance of the RF model (Breiman, 2001). It is considered to be an unbiased estimation of the generalization error, which is contributed by the proportion between the misclassification and the total number of OOB elements. The RF does not overfit the data (Rodríguez-Galiano et al., 2011) because as the number of trees increases, the generalization error of a forest converges (Breiman, 2001). Furthermore, the generalization error generated depends on the strength of every single tree and the correlation between them.

The RF classifier uses two development parameters known as the *Ntree* and the *Mtry*, which are used when building the model for the classification. The *Ntree* is the number of decision trees while the *Mtry* is the number of variables selected and tested for the best split when growing the trees. The *Ntree* value could be set as large as possible, and this will not result in

overfitting of the data (Kulkarni and Sinha, 2012, Ghosh and Joshi, 2014). Several used 500 *Ntree* value since it is the default number (Belgiu and Drăguț, 2016) and also probably because there is a stabilization of the errors even before this number of classification trees is achieved (Lawrence et al., 2006). Several studies have indicated that the use of the default value produced satisfactory results (Duro et al., 2012, Immitzer et al., 2012, Zhang and Roy, 2017). Breiman (2001), mentions that increasing the number of trees than the required number might not be essential, although it does not affect the model. Feng et al. (2015) also stated that with an *Ntree* value of 200, the model could achieve accurate results. Concerning the *Mtry*, the default value, which is the square root of the number of input variables is used by many studies (Gislason et al., 2006, Duro et al., 2012). The total number of variables could be used (Ghosh and Joshi, 2014), but it may increase the computational time (Belgiu and Drăguț, 2016) when building the model.

Furthermore, the RF provides a function to determine the important variable that contributed most to the classification accuracy. The important variable can be determined either by the Mean Decrease Accuracy (MDA) or the Mean Decreased Gini [(MDG] (Breiman, 2001, Liaw and Wiener, 2002). The MDA presents an estimate of accuracy by quantifying the degree to which an input variable in the model provides a decreased mean squared error. The MDG, on the other hand, is an impurity metric (Belgiu and Drăguț, 2016), which measures the node impurity or the degree to which an input variable produces a terminal node in the forest from the classification trees (Ahmed et al., 2015). The MDG could be indicated to be the variable that provides the best split at the node of trees in the forest created for the classification.

In the R package, the “*caret*” package (library) was used alongside the “*rf*” function in the model building and prediction (classification). The 70% training data set of each of the thematic cover classes (Table 3.1) was used to train and build the model. The input variables for the model were the eight selected spectral bands. The default *Ntree* and *Mtry* values were used in the modeling. No model tuning was therefore done during modeling as means of a possible means of comparing the performance RF algorithm with SVM using the default parameters. The model was then applied to the Landsat 8 image of the forest to complete the classification. The variable importance function of the RF was employed to determine and rank the general most important variables that contributed to the classification accuracy. Furthermore, each variable was analyzed on a class-by-class basis to evaluate their contribution to the classification of the specific thematic cover classes.

3.4.2 Support Vector Machine

The SVM algorithm undertakes classification using a statistical learning theory (Vapnik, 1995, Fauvel et al., 2006, Licciardi et al., 2009). The classifier is trained to find an optimal hyperplane through the minimization of the upper bound of the classification error (Cortes and Vapnik, 1995). The SVM was initially developed as a binary classifier that was meant to separate only two classes (Huang et al., 2002). In that process, the classifier finds a separating hyperplane that will best separate two classes in a multidimensional feature space, hence the hyperplane is the decision-making surface upon which the class separation takes place (Petropoulos et al., 2012). The decision boundary selected will be the one that leaves the greatest distance between

the hyperplane and the closest vector of the two classes (Vapnik, 1995). The closest data points to the hyperplane are termed as ‘support vectors’ as they are used in measuring the margin. The number of support vectors is consequently small (Vapnik, 1995). This applies generally to linearly separable data and so the linear function is used for the classification.

However, there are multiclass data such as remotely sensed data (e.g. multispectral satellite images), which may be difficult to separate using normal linear separable SVMs. This is because such data points or types overlap with each other and hence linear separable decision boundaries may not provide high accuracy (Mountrakis et al., 2011). As a means to classify such non-linear data types, a set of new SVM kernels has been developed including the radial basis function (RBF), polynomial and sigmoid have been developed (Fauvel et al., 2006). Among these, the RBF and the polynomial kernels are mostly used for remotely sensed data (Huang et al., 2002, Oommen et al., 2008). In such multiclass data, the SVM classifiers use a one-against-one (pairwise) or one-against-all classification approach, where the correct class is determined following a voting mechanism (Karatzoglou et al., 2006, Mazzoni et al., 2007, Ghosh and Joshi, 2014). Two parameters normally tuned during model building are the cost of constraints violation (C) and sigma (σ). The C parameter accounts for the overfitting of the model while the σ parameter controls the shape of the hyperplane (Ghosh et al., 2014).

In this study, the radial based function SVM (RBF-SVM) kernel is subsequently used for the classification of the forest cover classes. At the modeling stage, the “*caret*” package provides an estimate that is appropriate for the σ parameter using the “*sigest*” function based on the training data (Ghosh et al., 2014). The “*caret*” package in the R statistical package. The “*svmRadial*” function was used for the RBF-SVM (subsequently called SVM). The model training and building approach employed for the RF was used for SVM. The modeling and classifications for SVM algorithms were done using the default parameters for the purpose of comparison. No model tuning was carried out in the model building. In a similar fashion to the RF, each variable was analyzed on how they contributed to the mapping of the specific thematic cover classes.

3.4.3 Accuracy Assessment

Generally, the accuracy of land cover mapping is assessed by the degree to which the classification agrees with the validation reference data (Zhen et al., 2013). The training set used for the classification must be statistically independent of the validation set (Hammond and Verbyla, 1996) for credibility purposes. The 30% independent validation data that was set aside from the total reference data (Table 3.2) was used for the accuracy assessment for the thematic cover mapping by the SVM and RF algorithms. The accuracy assessment was conducted through the computation of accuracy estimates of the confusion matrix for each classifier, including the overall accuracy (OA), producer’s accuracy (PA) and user’s accuracy (UA) and kappa coefficient. The overall accuracy is a division of the total correctly classified pixels (the diagonal number in the confusion matrix) by the total pixels in the confusion matrix (Congalton, 1991, Congalton, 2001). The producer’s accuracy is obtained by dividing the total correctly classified pixels in a category by the total number of pixels in that category (column numbers in the confusion matrix) and is known as the omission error. It indicates how an area

is well classified. On the other hand, the user's accuracy is a division of the total correctly classified pixels in a category by the total number of classified pixels (row numbers in the confusion matrix) and known as commission error (Congalton, 1991, Congalton, 2001). It is indicative of the probability that a classified pixel in an image represents the actual category on the ground (Story and Congalton, 1986). The kappa coefficient measures the difference of the agreement between reference data and the algorithm employed for the classification against the likelihood of agreement between the reference data and a random algorithm (Adam et al., 2014b).

As a means of testing the difference in the performance of the RF and SVM classifiers, McNemar's test which judges the significant difference between two proportions (McNemar, 1947, de Leeuw et al., 2006, Petropoulos et al., 2012) was performed. It is a non-parametric test that is based on the confusion matrices of different classifiers using 2 by 2 dimensions (Foody, 2004). The 2 by 2 dimension used eliminates the constraints of large size matrices as they converted to that dimension with a focus on the binary distinction between correctly and misclassified classes (Foody, 2004). As indicated by Foody (2004) and de Leeuw et al. (2006), it is based on a standardized normal test static which is expressed as;

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}, \quad (1)$$

where f_{12} is the total number of pixels that were misclassified by the first algorithm but correctly classified by the second algorithm, while f_{21} is the total number of pixels that were correctly classified by the first algorithm but misclassified by the second algorithm (Manandhar et al., 2009, Adam et al., 2014b). Using the 5% level of significance ($p \leq 0.05$), the difference in accuracies of the confusion matrices of the two classifiers will be statistically significant if the Z value is greater than 1.96 (Congalton and Mead, 1983).

3.5 Results

3.5.1 Spatial extent of forest thematic cover

The SVM and the RF produced different spatial extents for each of the thematic classes with some differences being wide and others minimal (Table 3.3 and Figure 3.2). The SVM classified the closed canopy forest as the most dominant cover class with a spatial extent of 1059.23 ha while the bare site which the least dominant with a coverage of 20.97 ha. The open canopy forest and the grassland cover types were the second and third dominant cover with 910.60 ha and 226.55ha respectively. Contrary to the outcome of the SVM algorithm, the RF determined the open canopy forest as the most dominant class with an estimated coverage of

1195.45 ha while the closed canopy forest was determined as the second dominant with total coverage of 741.17 ha. The spatial extent of the grassland and the bare site was estimated to be 264.32 ha and 16.42 ha.

Both algorithms determined a similar geographic location for each of the thematic cover classes with the only differences being in the extent of each specific thematic cover. Most of the closed canopy forest is distributed in the middle of the forest and towards the north, south, southwestern and northeastern sections of the forest reserve. The open canopy forest lies in the northeastern, southeastern and southwestern parts of the reserve while small patches of open canopy forest are found in the middle site of the reserve. The grassland is mostly distributed in patches at the northeastern, northwestern and southeastern boundaries of the reserve. The bare sites are also clustered around the southeastern, northeastern boundaries and some dotted around the southwestern boundary. Most of the bare sites are found within the grassland areas.

Table 3.3: Spatial coverage of each thematic cover classes produced by the SVM and RF Algorithms.

Thematic cover/classes	Code	SVM		RF	
		Area (Ha)	Percentage cover (%)	Area (Ha)	Percentage cover (%)
Open canopy forest	OCF	910.60	41.07	1195.45	53.91
Closed canopy forest	CCF	1059.23	47.77	741.17	33.43
Grassland	GL	226.55	10.22	264.32	11.92
Bare sites	BS	20.97	0.94	16.42	0.74
Total Area		2217.36	100	2217.36	100

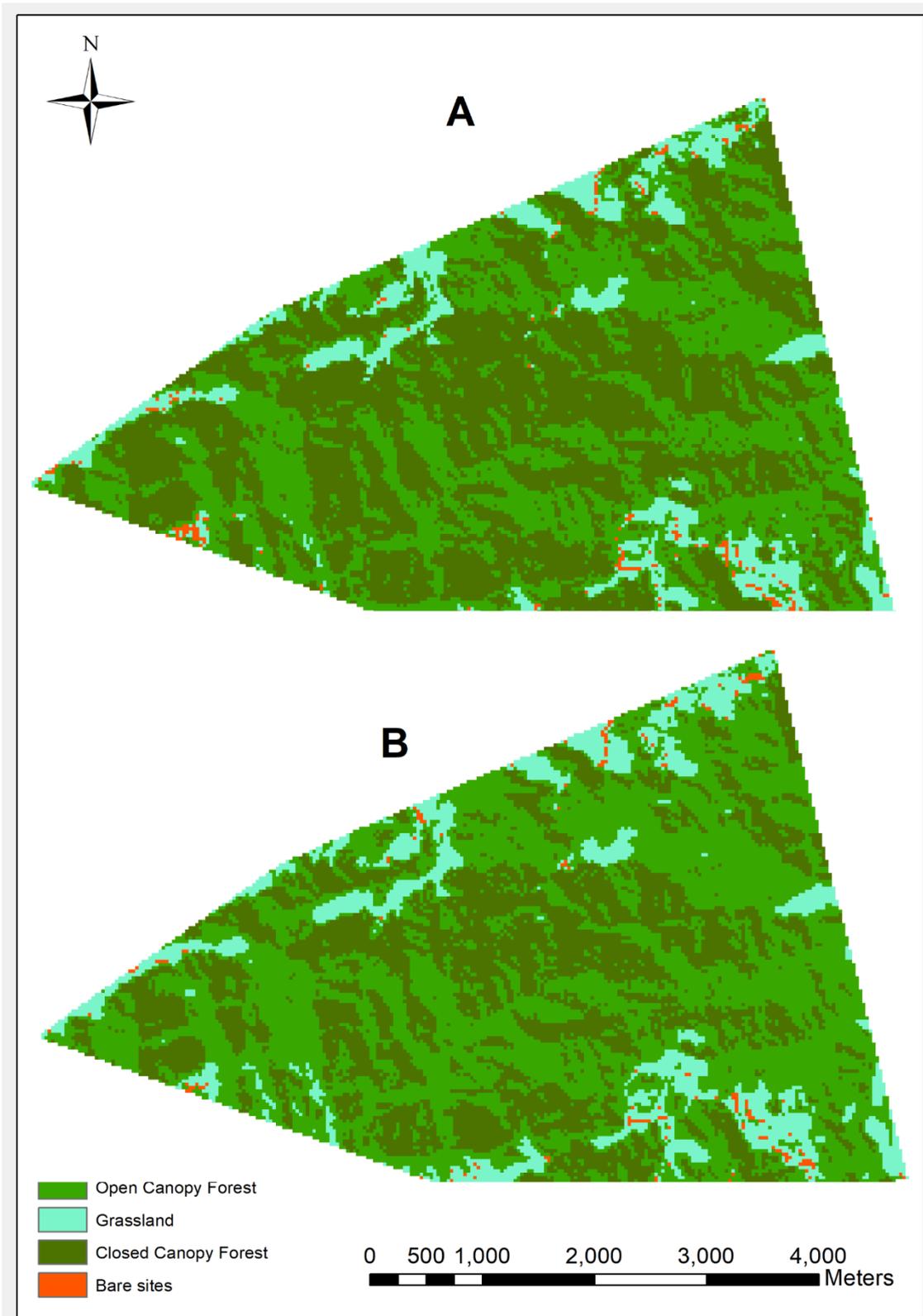


Figure 3.2: Thematic map of the Nkandla Forest Reserve. A is map produce by the SVM algorithm and B is a map produced by the RF algorithm.

3.5.2 Comparison of RF and SVM mapping accuracy

The classification performance of the SVM and RF algorithms was evaluated using the independent validation data set. The confusion matrices of the SVM and the RF (Table 3.4) provide details on the estimates of the user's accuracy and producer's accuracy for each thematic cover class as well as the overall accuracy and kappa coefficient. The SVM recorded a slightly higher overall accuracy of 95.83% and a kappa coefficient of 0.94 than the RF which had an overall accuracy of 95.24% with a kappa coefficient of 0.93.

The SVM produced higher estimates for almost all the producer's accuracies than RF of the thematic cover classes. It was only in the producer's accuracy of the open canopy forest where the RF had a higher estimate than the SVM. Both algorithms recorded the highest producer's accuracy for the closed canopy forest with estimates of 100% for the SVM and 98.55% for the RF. Similarly, they also recorded the lowest for the bare site with SVM obtaining 81.81% and the RF obtaining 72.72%. The SVM again had higher estimates in the user's accuracy than the RF except for that of the closed canopy forest where the RF outperformed the SVM. Both recorded the highest values for the open canopy forest and the lowest values for the bare sites.

Table 3.4: Confusion Matrix for SVM and RF indicating the Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA) and kappa coefficient

		SVM						RF						
		Reference data						Reference data						
		BS	CCF	GL	OCF	Row total	UA (%)	BS	CCF	GL	OCF	Row total	UA (%)	
Classified data	BS	9	0	2	0	11	81.81	BS	8	0	2	0	10	80.00
	CCF	0	69	0	3	72	95.83	CCF	0	68	0	2	70	97.14
	GL	2	0	41	0	43	95.35	GL	3	0	41	0	44	93.18
	OCF	0	0	0	42	42	100	OCF	0	1	0	43	44	97.73
Column total	11	69	43	45	168		Column total	11	69	43	45	168		
	PA (%)	81.81	100	95.35	93.33		PA (%)	72.72	98.55	93.35	95.56			
	OA	95.83					OA	95.24						
	Kappa	0.94					Kappa	0.93						

Further in this study, the McNemar test was conducted to compare the difference between the RF and SVM algorithm in relation to the accuracy parameters in the confusion matrices. As illustrated in Table 3.5, the two algorithms agreed on 161 pixels out the 168 pixels, with 157 correctly classified (bottom right) and 4 pixels misclassified (top left) by both algorithms. The

SVM and RF, however, disagreed on 7 pixels out of the 168 pixels. SVM correctly classified 4 pixels that were misclassified by RF (bottom left), while RF correctly classified 3 pixels, which were misclassified by SVM (top right). The McNemar test produced a Z-value of 0.37 corresponding to a P-value of 1.0 at a 5% significance level. The values obtained indicate that there is no significant difference in accuracy between the SVM and the RF algorithms.

Table 3.5: McNemar’s test results for comparison of Random Forest (RF) and Support Vector Machine (SVM) Algorithms.

		RF		
		Misclassified	Correctly classified	Total
SVM	Misclassified	4	3	7
	Correctly classified	4	157	161
	Total	8	160	168
Z – value		0.377		

3.5.3 General Variable Importance of Landsat 8 bands

The important variables that contributed most to the overall accuracy of the classification were determined and evaluated based on the Mean Decrease Accuracy (MDA) of the RF. On the overall (general) scale, the red band (B4) was determined as the most important variable while Cirrus (B9) was the least important variable (Figure 3.3). All the visible ranges of bands, that are red (B4), green (B3) and blue (B2), together with the near-infrared (NIR) and shortwave infrared 1 (SWIR 1) were ranked among the first five that contributed much to the accurate classification. Although the B9 was determined to be the least important variable, dropping it in the subsequent iteration in the modelling process did not change the initial accuracy obtained or the order of importance for the rest of the seven bands.

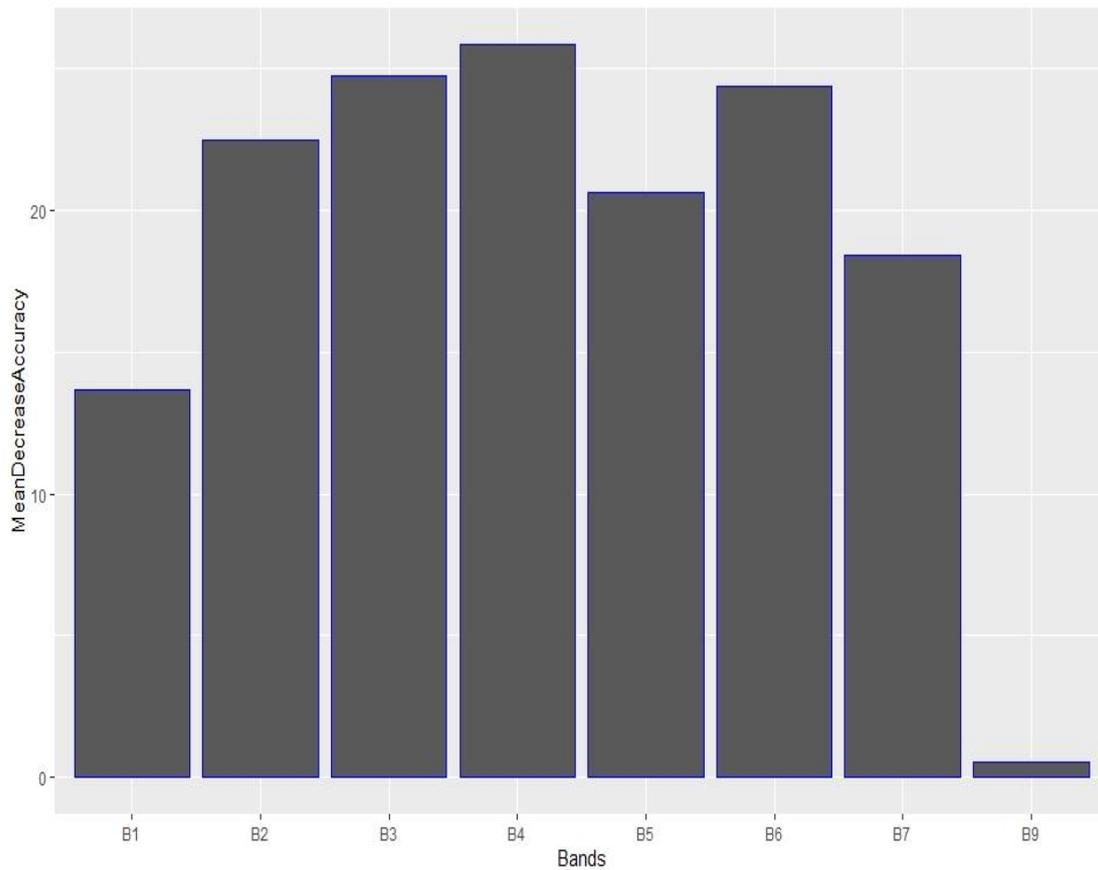


Figure 3.3: General variable importance ranking of the Landsat 8 bands using the MDA. The bands B1(Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Short Wave infra-red 1), B7 (Short wave infra-red 2) and B9 (Cirrus) are normal numbering order.

3.5.4 Class by class variable importance

The contribution of each spectral band to the mapping of each thematic cover class was determined for both the SVM and the RF algorithms. Under the SVM, all the spectral bands recorded 100% importance in the mapping of the bare sites (X1), closed canopy forest (X2) and the grassland (X3) except for the NIR and Cirrus (Figure 3.4). In the case of the open canopy (X4), it was the coastal aerosol, red, green and blue bands that had 100% importance value. Relatively lower importance values were however obtained by the NIR (61.60%), SWIR 1 (70.27%) and SWIR 2 (99.78%) for the open canopy. The Cirrus did not have any contribution to the mapping of the open canopy forest as it had a 0% importance value. The specific thematic cover class variable importance measure for the SVM algorithm was much uniform with fewer variations in values.

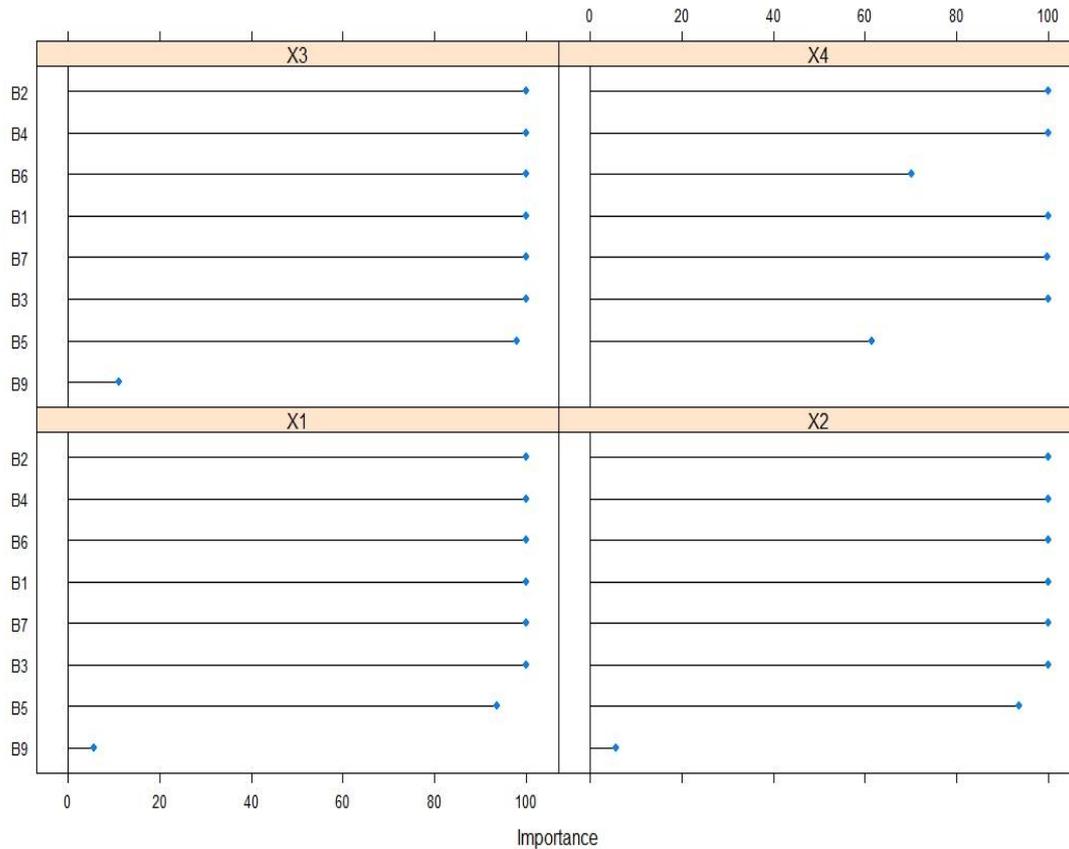


Figure 3.4: Thematic class-specific variable importance ranking of the Landsat 8 spectral bands by the SVM algorithm. The bands B1(Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Short Wave infra-red 1), B7 (Short wave infra-red 2) and B9 (Cirrus) are ranked based on their importance value in mapping the classes.

There were however many variations in the importance values among the spectral bands for each of the thematic cover classes by the RF (Figure 3.5). For the bare site (1), the coastal aerosol was the most important band with an importance value of 83.19% while the blue and red bands were the second and third with values of 68.21% and 49.64% respectively. The SWIR 1, however, had a 0% contribution, while the green, SWIR 2, cirrus and NIR bands contributed different degrees although equal to or below 40%. The three most important bands to the closed canopy forest (2) were the visible range bands, with the green (62.45%), blue (62.02%) and red (60.59%) appearing in descending order. The other bands had an importance value of below 50% with the least important being the cirrus. All the bands displayed improved importance in the mapping of the grassland cover (3). The most important was the SWIR 1 with 100% value followed by the red band with 92% value. The rest of the bands had importance values between 50% and 80% except for the coastal aerosol and 9 which had values below 40%. It is only in this class that band 9 could achieve a value slightly above 20%. Similarly, to the closed canopy forest, the bands in the visible range were important with the red, green and blue appearing in descending order obtaining values of 81.74%, 56.51%, and 56.20% respectively. The other bands obtained an importance of below 40% in mapping the thematic cover class.

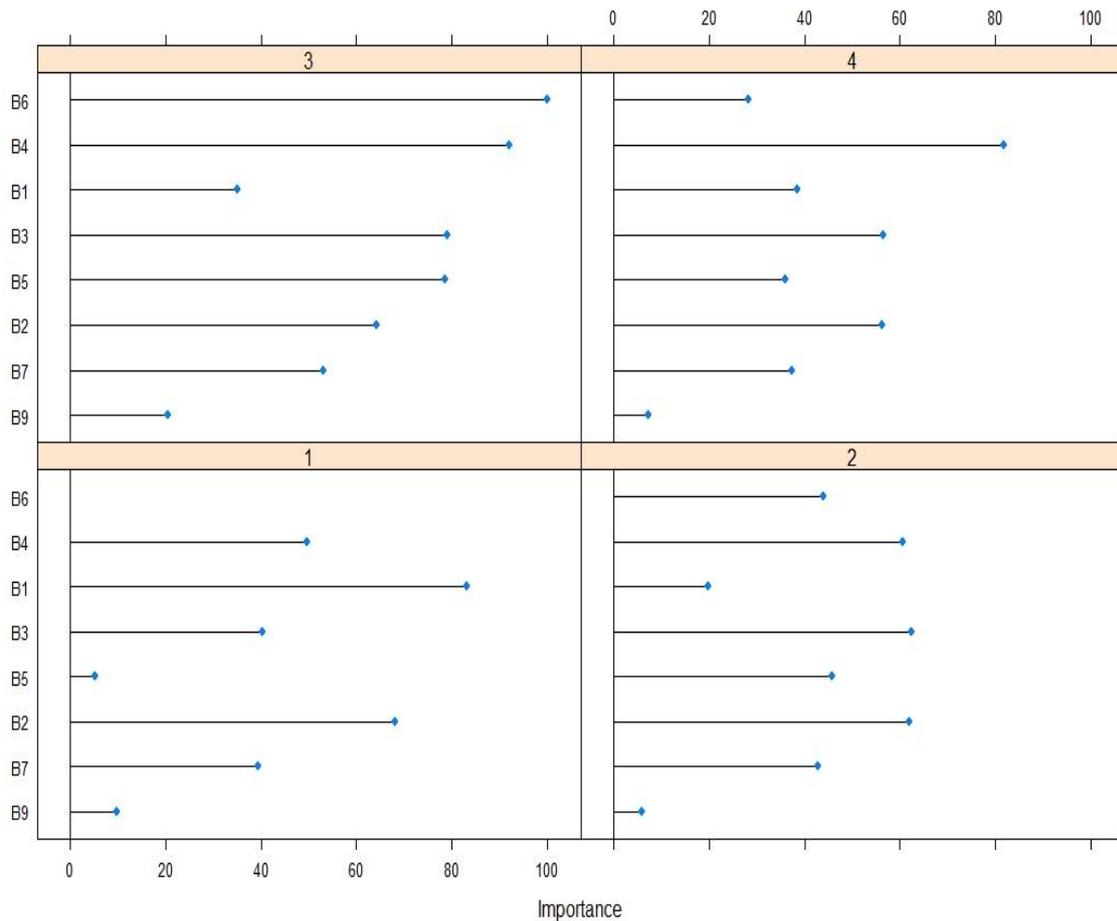


Figure 3.5: Thematic class-specific variable importance ranking of the Landsat 8 bands by the RF algorithm. The bands B1(Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Short Wave infra-red 1), B7 (Short wave infra-red 2) and B9 (Cirrus) are ranked based on their importance value in mapping the classes.

3.6 Discussion

The mapping of the Nkandla Forest Reserve through the application of SVM and RF algorithms to Landsat 8 satellite imagery facilitated a provision of an understanding of the forest cover status, the extent of each thematic class for management and conservation purposes. A process that also contributes to deepening the knowledge in the application use of the algorithms and their inherent capabilities as well as an overview of how each band contributed to the mapping specific thematic forest cover class.

The SVM and RF algorithms utilized the pixel values of the training sites from the Landsat 8 data to define and delineate the four important natural forest cover classes. The closed canopy forest was identified to be the dominant forest cover. Spatially, it covers about 47.76% of the forest reserve area with its areas of distribution and geographic locations clearly identified. The closed canopy forest is characterized by an unbroken or continuous layer of the tree canopy and have less sunlight reaching the forest floor (Arihafa and Mack, 2013) as compared to the

open canopy forests. As a result, regeneration may be less induced in such portions of forest ecosystems (Yamamoto, 2000). However, such areas are still able to sequester carbon and serve as carbon sinks (Kline et al., 2016) and thus play a major role in climate change mitigation (Köhl et al., 2015). For a continual realization of such benefits, forest managers will have to ensure that there are no disturbances in locations where they are distributed, especially negative human interventions.

The open canopy forest was the second dominant forest class covering about 41.07% of the forest and distributed around and within the closed canopy forest. This area is characterized by broken canopy or gaps which might have resulted from past forest disturbance caused by human interventions such as the felling of trees or natural occurrences such as tree falls. The gaps allow sunlight to get to the forest floor (Marthews et al., 2008) that initiates a number of ecological and biological activities that are not common in areas without gaps (Sanford et al., 1986). These include variations in carbon and nutrient cycles (Feeley et al., 2007), alteration in competition for sun among species (Rüger et al., 2009) and variations in microclimate (Marthews et al., 2008). These gaps, however, may be critical for enhanced regeneration through the advanced growth of tree seeds from the soil seed banks (Sanford et al., 1986, Yamamoto, 2000). These ecological processes could account for the observed regeneration of tree seedlings within the gaps in these portions of the forest. The regeneration process could enhance the productive and protective functions of the forests because of the continuous biological and ecological process. For instance, these areas of active tree regrowth and regeneration may result in increased carbon sequestration through the accumulation of biomass (Feeley et al., 2007), which are mainly stored in above-ground and below-ground biomass (Vicharnakorn et al., 2014). This process enhances the contribution of the forest in climate change mitigation and amelioration of the local climate.

It must, however, be indicated that as a result of the gaps, the species composition of the forest could be altered due to variation in individual species shade tolerance capacity (de Römer et al., 2007, Tanaka et al., 2008). Moreover, there could be increased growth in lianas and climbing trees, which can increase competition for space and nutrients with regenerating trees (Toledo-Aceves and Swaine, 2008). The lianas and climbers could cause over shading and high tree mortality, and eventually lead to reduced forest biomass and carbon sequestration (Körner, 2004, Granados and Körner, 2002). Furthermore, the canopy gaps could lead to the growth of ticks and other invasion species (Baret et al., 2008) that can colonize the gaps. This is ascertained in the observed presence of such climbers, lianas, thorns, and thickets found in the open canopy area of the forest. Therefore, it will be worth it for forest management to pay close attention to the open canopy areas and have strategies to mitigate the negative effects while facilitating forest regeneration to restore the area into a closed canopy forest.

The grassland, which is the third dominant thematic cover in the forest reserve area has no tree cover and a total spatial extent of about 10.22%. This might have been caused by frequent fires especially around and along the forest boundaries. The presence of no trees might have been caused by the presence of rocks or impervious layers underneath the soil that cannot support tree growth. In addition, the thick growth of grass may suppress the growth of any tree seed dispersed by wind to that location. Another reason could be a result of browsing by grazing

cattle in the forest reserve area. The grasses could be fuel for wildfire (Cheney and Gould, 1995, Snyder et al., 2006, Davies et al., 2015), and hence effective fire management initiatives are required to prevent such occurrences. Furthermore, the grasslands could be a source of weeds and alien herbaceous species (Chambers et al., 2007, Davies et al., 2011a), which could possibly colonize or spread into the open and closed canopy forest areas. This is evident in the presence of bugweed (*Solanum mauritianum*) commonly spotted in certain portions of the forest especially around the western part of the forest reserve. There is also the presence of Australian blackwood (*Acacia melanoxylon*) which is capable of taking over grassland areas. The invasive weeds and plants may hence alter the natural ecosystem and ecological process of the forest and possibly limits its functional ability in the course of time. Efforts must hence be in place by forest managers to monitor and control the spread of grassland, weed, or alien species in these areas clearly mapped and delineated. Notwithstanding these possible threats, the covering of the soil by grassland may be vital in preventing soil erosion and soil loss in those portions of the forest reserve.

The bare sites constitute the least dominant area, which covers 0.7% of the total area and has no vegetation cover. It includes areas developed into offices, accommodation, road network and sites with exposed soil. Areas with no vegetation that have bare soils are prone to soil erosion (Keesstra et al., 2016) and soil loss (Labrière et al., 2015). Although it has the smallest spatial cover as compared to the other cover types, conservation managers could still put in efforts to promote tree growth if only they are not rocky sites or covered with impervious layers. Soil erosion and soil loss must be monitored in such areas to prevent possible detrimental effects to adjacent cover types. It will be important for management to reduce the development of permanent infrastructure in the forest as it also contributes to the loss of tree cover and interferences with natural systems.

The mapping process which has clearly delineated the various thematic forest cover classes, their distribution and spatial extent will serve as important information to forest managers to where to implement conservation management strategies.

On evaluating the performance of the algorithms, the SVM was slightly superior in the mapping of the natural forest cover by producing high accuracies than the RF. This was evident in the estimates obtained for the overall, producer's and user's accuracies as well as the kappa coefficient where some differences were observed. Furtherance to evaluating their performance, the McNemar test (McNemar, 1947, Foody, 2004, de Leeuw et al., 2006) revealed that the difference between mapping accuracies of the SVM and the RF was statistically insignificant. This is in line with Adam et al. (2014b) who also had the SVM producing a slightly higher accuracy than the RF but the difference was statistically insignificant. It is important to note that the two algorithms were applied using their default design parameters with no tuning undertaken. Their ability to produce high accuracies above 95% could, therefore, be attributed to their inherent design capabilities (Fassnacht et al., 2016, Thanh Noi and Kappas, 2017). This is evident in studies that found the SVM (Petropoulos et al., 2012, Paneque-Gálvez et al., 2013) the RF (Yin et al., 2017, Matasci et al., 2018, Pelletier et al., 2016) to have demonstrated much robustness in forest cover and other land cover classification and mapping.

The general variable importance variable evaluated using the MDA of RF algorithm, enabled the study to provide an overview of the importance of each of the Landsat 8 bands (input variables) used in the mapping of the forest cover. The red band was determined to be the most important band in the mapping and contributed much to the accuracy with the second most important being the green band. The three other bands which add up to form the top five were the SWIR 1, the blue and the NIR. Their determination as the top five bands, especially the visible bands and the NIR could be attributed to their sensitivity to vegetation and making them useful for vegetation mapping (Dube and Mutanga, 2015, Li et al., 2013, Roy et al., 2014, Xu et al., 2018). This is evidently and practically affirmed as the forest reserve has a substantial vegetation cover above 99% (open canopy, closed canopy, and grassland area) as determined by the algorithms. El-Askary et al. (2014) and Pahlevan et al. (2014) further indicated that the Landsat 8 has more refined spectral bands that makes them important in characterizing forest conditions. This information is vital since it will provide insight into the bands that could be incorporated into input variables of the Landsat 8 satellite for mapping natural forest ecosystems. Knowing the important variable will enhance feature selection, which has been indicated as a valuable technique that contributes to reducing redundancy, enhancing computation and improving classification accuracy (Pal and Foody, 2010, Onojeghuo and Blackburn, 2011, Millard and Richardson, 2013, Nitze et al., 2015).

The evaluation of the spectral bands on how each contributed to the mapping of the specific thematic cover class provided an in-depth knowledge of their importance to the classes. It demonstrates their sensitivity to each cover class which may have a direct influence on the accuracy obtained as shown in the producer's and user's accuracies. It must, however, be acknowledged that the performance or importance of the spectral bands may not be mutually exclusive of the algorithm that is used based on the findings of this study. The algorithm could influence how the spectral bands perform and the level of importance it might have in the classification or mapping of cover classes. For instance, the spectral bands had importance values of above 90% to 100% in almost all the classes under the SVM algorithm mapping except for cirrus which had very low value. The outcome of the class-specific importance for the SVM algorithm, therefore, may be a true reflection of the inherent design capabilities of the individual Landsat 8 bands in detecting land surface cover objects such as vegetation. This is because the NIR band is designed to be sensitive to growing vegetation like forest while the SWIR bands are sensitive to land surfaces with exposed soils (Li et al., 2013). These bands were among those that had very high importance values in the mapping of the closed and open canopy forest as well as the bare sites. This is contrary to that of the RF algorithm, where the level of importance obtained for the NIR and SWIR bands was very low, with many differences and somehow not reflecting their true design capabilities. This was the same for all the other bands except for the grassland cover where the SWIR1 had 100% and the red band had 92%. The SVM may hence be recommended for use in similar studies involving natural forest ecosystem mapping based on the outcomes of this study. This is because it optimally utilized the spectral bands in a way that reflects their actual importance and design capabilities in specific thematic forest cover mapping. This outcome is vital as it could inform the selection of features, algorithms and the application of machine learning methods in forest cover mapping.

3.7 Conclusions

Natural forest ecosystems are an important resource that produces ecosystem goods and services beneficial to the wellbeing of society. The Nkandla Forest Reserve is among the natural forests in South Africa that contribute to the provision of such services. Remote Sensing forest cover mapping information outputs are necessary to inform and assist in forest management planning. The use of Landsat 8 satellite imagery to map the thematic cover classes proved useful. The thematic cover mapping was made possible by the SVM and the RF algorithms, however, the SVM was slightly superior to the RF considering all the accuracy parameters. The McNemer test further revealed that the difference between the accuracies of the two algorithms is statistically insignificant. The obtaining of high accuracies by the two algorithms even under default conditions may indicate that their internal design parameter makes them much more robust for forest cover mapping based on the outcomes of this study. However, the SVM may be recommended for forest ecosystem mapping involving the use of the Landsat 8 satellite image as it optimally utilized the spectral bands in a manner that demonstrates their design capabilities in mapping specific cover classes.

Generally, the three visible range bands, the SWIR 1 and the NIR were the top five important variables of Landsat 8 that contributed to accurate classifications. This is attributed to their sensitivity to vegetation and hence these bands could be incorporated into bands selected for forest cover classification. The detailed spatial information provided through the forest cover mapping could be vital to forest managers in prioritization and prescribing interventions such as enhanced natural regeneration reforestation, fire management, carbon stock assessments and other conservation initiatives for the forest. Moreover, the data and information could be assessed and integrated into the national database of forest ecosystems for effective planning on issues such as climate change mitigation by policymakers and other relevant stakeholders.

Acknowledgment

The authors appreciate the support provided by Ms. Sharon Louw (Ecologist) of Ezemvelo KwaZulu-Natal Wildlife (EKZNW) in securing a permit to conduct the study in the Nkandla Forest Reserve. We thank Mr. Elliackim Zungu (Conservation Manager), Mr. Simon Makhaye, Mr. Sandebulawa Biyela and other supporting staff of EKZNW for assisting in field data collection. The authors also thank Mr. Charles Zungunde (University of KwaZulu-Natal) for driving the field team across the forest as well as assisting in data collection. We further thank the journal editors and reviewers for their efforts, time and insights.

Funding

This research and study received a block grant from the South African System Analysis Centre (SASAC) through the National Research Foundation (NRF) of South Africa, with grant number 118770. The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest

References

- ADAM, E., MUTANGA, O., ODINDI, J. & ABDEL-RAHMAN, E. M. 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35, 3440-3458.
- AHMED, O. S., FRANKLIN, S. E., WULDER, M. A. & WHITE, J. C. 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 89-101.
- ARIHAFI, A. & MACK, A. L. 2013. Treefall Gap Dynamics in a Tropical Rain Forest in Papua New Guinea. *Pacific Science*, 67, 47-58.
- BALTHAZAR, V., VANACKER, V., MOLINA, A. & LAMBIN, E. F. 2015. Impacts of forest cover change on ecosystem services in high Andean mountains. *Ecological Indicators*, 48, 63-75.
- BARET, S., COURNAC, L., THÉBAUD, C., EDWARDS, P. & STRASBERG, D. 2008. Effects of canopy gap size on recruitment and invasion of the non-indigenous *Rubus alceifolius* in lowland tropical rain forest on Réunion. *Journal of Tropical Ecology*, 24, 337-345.
- BELGIU, M. & DRĂGUȚ, L. 2016. Random forest in Remote Sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- BREIMAN, L. 2001. Random forests. *Machine learning*, 45, 5-32.
- BRIEM, G. J., BENEDIKTSSON, J. A. & SVEINSSON, J. R. 2002. Multiple classifiers applied to multisource Remote Sensing data. *IEEE transactions on geoscience and Remote Sensing*, 40, 2291-2299.
- CARDINALE, B. J., DUFFY, J. E., GONZALEZ, A., HOOPER, D. U., PERRINGS, C., VENAIL, P., NARWANI, A., MACE, G. M., TILMAN, D., WARDLE, D. A., KINZIG, A. P., DAILY, G. C., LOREAU, M., GRACE, J. B., LARIGAUDERIE, A., SRIVASTAVA, D. S. & NAEEM, S. 2012. Biodiversity loss and its impact on humanity. *Nature*, 486, 59-67.
- CHAMBERS, J. C., ROUNDY, B. A., BLANK, R. R., MEYER, S. E. & WHITTAKER, A. 2007. What makes Great Basin sagebrush ecosystems invulnerable to *Bromus tectorum*? *Ecological Monographs*, 77, 117-145.
- CHAN, J. C.-W. & PAELINCKX, D. 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, 2999-3011.
- CHEN, B., XIAO, X., LI, X., PAN, L., DOUGHTY, R., MA, J., DONG, J., QIN, Y., ZHAO, B., WU, Z., SUN, R., LAN, G., XIE, G., CLINTON, N. & GIRI, C. 2017. A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131, 104-120.
- CHEN, J. & BRADSHAW, G. A. 1999. Forest structure in space: a case study of an old growth spruce-fir forest in Changbaishan Natural Reserve, PR China. *Forest Ecology and Management*, 120, 219-233.

- CHENEY, N. & GOULD, J. 1995. Fire growth in grassland fuels. *International Journal of Wildland Fire*, 5, 237-247.
- CLARK, M. L. & ROBERTS, D. A. 2012. Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sensing*, 4, 1820-1855.
- CONGALTON, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of environment*, 37, 35-46.
- CONGALTON, R. G. 2001. Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire*, 10, 321-328.
- CONGALTON, R. G. & MEAD, R. A. 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, 49, 69-74.
- CORTES, C. & VAPNIK, V. 1995. Support-vector networks. *Machine learning*, 20, 273-297.
- DALPONTE, M., BRUZZONE, L. & GIANELLE, D. 2008. Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 1416-1427.
- DALPONTE, M., ORKA, H. O., GOBAKKEN, T., GIANELLE, D. & NAESSET, E. 2013. Tree Species Classification in Boreal Forests With Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 2632-2645.
- DAVIES, K. W., BOYD, C. S., BATES, J. D. & HULET, A. 2015. Dormant season grazing may decrease wildfire probability by increasing fuel moisture and reducing fuel amount and continuity. *International Journal of Wildland Fire*, 24, 849-856.
- DAVIES, K. W., BOYD, C. S., BECK, J. L., BATES, J. D., SVEJCAR, T. J. & GREGG, M. A. 2011. Saving the sagebrush sea: an ecosystem conservation plan for big sagebrush plant communities. *Biological Conservation*, 144, 2573-2584.
- DBOZKAYA, A. G., BALCIK, F. B., GOKSEL, C. & ESBAH, H. 2015. Forecasting land-cover growth using remotely sensed data: a case study of the Igneada protection area in Turkey. *Environmental monitoring and assessment*, 187, 59.
- DE LEEUW, J., JIA, H., YANG, L., LIU, X., SCHMIDT, K. & SKIDMORE, A. 2006. Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing*, 27, 223-232.
- DE RÖMER, A. H., KNEESHAW, D. D. & BERGERON, Y. 2007. Small gap dynamics in the southern boreal forest of eastern Canada: Do canopy gaps influence stand development? *Journal of Vegetation Science*, 18, 815-826.
- DEFRIES, R. & CHAN, J. C.-W. 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74, 503-515.
- DUBE, T. & MUTANGA, O. 2015. 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.
- DURO, D. C., FRANKLIN, S. E. & DUBÉ, M. G. 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of environment*, 118, 259-272.
- EL-ASKARY, H., ABD EL-MAWLA, S., LI, J., EL-HATTAB, M. & EL-RAEY, M. 2014. Change detection of coral reef habitat using Landsat-5 TM, Landsat 7 ETM+ and Landsat 8 OLI data in the Red Sea (Hurghada, Egypt). *International journal of Remote Sensing*, 35, 2327-2346.

- EZEMVELO KZN WILDLIFE 2015. Nkandla Forest Complex: Protected Area Management Plan. *Unpublished*, Version 1.0, 1-173.
- FASSNACHT, F. E., LATIFI, H., STERENĆZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- FASSNACHT, F. E., LI, L. & FRITZ, A. 2015. Mapping degraded grassland on the Eastern Tibetan Plateau with multi-temporal Landsat 8 data — where do the severely degraded areas occur? *International Journal of Applied Earth Observation and Geoinformation*, 42, 115-127.
- FAUVEL, M., CHANUSSOT, J. & BENEDIKTSSON, J. A. Evaluation of kernels for multiclass classification of hyperspectral Remote Sensing data. 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 2006. IEEE, II-II.
- FEELEY, K. J., DAVIES, S. J., ASHTON, P. S., BUNYAVEJCHEWIN, S., NUR SUPARDI, M., KASSIM, A. R., TAN, S. & CHAVE, J. M. 2007. The role of gap phase processes in the biomass dynamics of tropical forests. *Proceedings of the Royal Society B: Biological Sciences*, 274, 2857-2864.
- FENG, Q., LIU, J. & GONG, J. 2015. UAV Remote Sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing*, 7, 1074-1094.
- FERET, J.-B. & ASNER, G. P. 2013. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 73-84.
- FOODY, G. M. 2004. Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, 70, 627-633.
- FOODY, G. M., BOYD, D. S. & CUTLER, M. E. 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of environment*, 85, 463-474.
- FOODY, G. M. & MATHUR, A. 2004. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93, 107-117.
- FRIEDL, M., WOODCOCK, C., GOPAL, S., MUCHONEY, D., STRAHLER, A. & BARKER-SCHAAF, C. 2000. A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data.
- GHAMISI, P., BENEDIKTSSON, J. A. & PHINN, S. 2015. Land-cover classification using both hyperspectral and LiDAR data. *International Journal of Image and Data Fusion*, 6, 189-215.
- GHOSH, A., FASSNACHT, F. E., JOSHI, P. K. & KOCH, B. 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, 26, 49-63.
- GHOSH, A. & JOSHI, P. K. 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26, 298-311.
- GILLESPIE, T. W., SAATCHI, S., PAU, S., BOHLMAN, S., GIORGI, A. P. & LEWIS, S. 2009. Towards quantifying tropical tree species richness in tropical forests. *International Journal of Remote Sensing*, 30, 1629-1634.
- GILROY, J. J., WOODCOCK, P., EDWARDS, F. A., WHEELER, C., BAPTISTE, B. L., URIBE, C. A. M., HAUGAASEN, T. & EDWARDS, D. P. 2014. Cheap carbon and

- biodiversity co-benefits from forest regeneration in a hotspot of endemism. *Nature Climate Change*, 4, 503.
- GISLASON, P. O., BENEDIKTSSON, J. A. & SVEINSSON, J. R. 2006. Random forests for land cover classification. *Pattern Recognition Letters*, 27, 294-300.
- GRANADOS, J. & KÖRNER, C. 2002. In deep shade, elevated CO₂ increases the vigor of tropical climbing plants. *Global Change Biology*, 8, 1109-1117.
- HAMMOND, T. & VERBYLA, D. 1996. Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, 17, 1261-1266.
- HE, K. S., ROCCHINI, D., NETELER, M. & NAGENDRA, H. 2011. Benefits of hyperspectral Remote Sensing for tracking plant invasions. *Diversity and Distributions*, 17, 381-392.
- HEISKANEN, J. & KIVINEN, S. 2008. Assessment of multispectral, -temporal and -angular MODIS data for tree cover mapping in the tundra-taiga transition zone. *Remote Sensing of Environment*, 112, 2367-2380.
- HIROSE, K., OSAKI, M., TAKEDA, T., KASHIMURA, O., OHKI, T., SEGAH, H., GAO, Y. & EVRI, M. 2016. Contribution of hyperspectral applications to tropical peatland ecosystem monitoring. *Tropical Peatland Ecosystems*. Springer.
- HUANG, C., DAVIS, L. S. & TOWNSHEND, J. R. G. 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23, 725-749.
- IMMITZER, M., ATZBERGER, C. & KOUKAL, T. 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing*, 4, 2661-2693.
- IRONS, J. R., DWYER, J. L. & BARSİ, J. A. 2012. The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sensing of Environment*, 122, 11-21.
- JACTEL, H., BAUHUS, J., BOBERG, J., BONAL, D., CASTAGNEYROL, B., GARDINER, B., GONZALEZ-OLABARRIA, J. R., KORICHEVA, J., MEURISSE, N. & BROCKERHOFF, E. G. 2017. Tree Diversity Drives Forest Stand Resistance to Natural Disturbances. *Current Forestry Reports*, 3, 223-243.
- JETZ, W., CAVENDER-BARES, J., PAVLICK, R., SCHIMEL, D., DAVIS, F. W., ASNER, G. P., GURALNICK, R., KATTGE, J., LATIMER, A. M., MOORCROFT, P., SCHAEPMAN, M. E., SCHILDHAUER, M. P., SCHNEIDER, F. D., SCHRODT, F., STAHL, U. & USTIN, S. L. 2016. Monitoring plant functional diversity from space. *Nat Plants*, 2, 16024.
- JIA, K., WEI, X., GU, X., YAO, Y., XIE, X. & LI, B. 2014. Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto International*, 29, 941-951.
- JUCKER, T., AVĂCĂRIȚEI, D., BĂRNOAIEA, I., DUDUMAN, G., BOURIAUD, O., COOMES, D. A. & GILLIAM, F. 2016. Climate modulates the effects of tree diversity on forest productivity. *Journal of Ecology*, 104, 388-398.
- KARATZOGLOU, A., MEYER, D. & HORNIK, K. 2006. Support vector machines in R. *Journal of Statistical Software*, 15, 1-28.
- KESSTRA, S., PEREIRA, P., NOVARA, A., BREVIK, E. C., AZORIN-MOLINA, C., PARRAS-ALCÁNTARA, L., JORDÁN, A. & CERDÀ, A. 2016. Effects of soil management techniques on soil water erosion in apricot orchards. *Science of the Total Environment*, 551, 357-366.
- KIM, S. R., LEE, W. K., KWAK, D. A., BIGING, G. S., GONG, P., LEE, J. H. & CHO, H. K. 2011. Forest cover classification by optimal segmentation of high resolution satellite imagery. *Sensors (Basel)*, 11, 1943-58.
- KLINE, J. D., HARMON, M. E., SPIES, T. A., MORZILLO, A. T., PABST, R. J., MCCOMB, B. C., SCHNEKENBURGER, F., OLSEN, K. A., CSUTI, B. & VOGELER, J. C. 2016.

- Evaluating carbon storage, timber harvest, and habitat possibilities for a Western Cascades (USA) forest landscape. *Ecological applications*, 26, 2044-2059.
- KÖHL, M., LASCO, R., CIFUENTES, M., JONSSON, Ö., KORHONEN, K. T., MUNDHENK, P., DE JESUS NAVAR, J. & STINSON, G. 2015. Changes in forest production, biomass and carbon: Results from the 2015 UN FAO Global Forest Resource Assessment. *Forest Ecology and Management*, 352, 21-34.
- KÖRNER, C. 2004. Through enhanced tree dynamics carbon dioxide enrichment may cause tropical forests to lose carbon. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359, 493-498.
- KULKARNI, V. Y. & SINHA, P. K. Pruning of random forest classifiers: A survey and future directions. 2012 International Conference on Data Science & Engineering (ICDSE), 2012. IEEE, 64-68.
- LABRIÈRE, N., LOCATELLI, B., LAUMONIER, Y., FREYCON, V. & BERNOUX, M. 2015. Soil erosion in the humid tropics: A systematic quantitative review. *Agriculture, Ecosystems & Environment*, 203, 127-139.
- LAURIN, G. V., LIESENBERG, V., CHEN, Q., GUERRIERO, L., DEL FRATE, F., BARTOLINI, A., COOMES, D., WILEBORE, B., LINDSELL, J. & VALENTINI, R. 2013. Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. *International Journal of Applied Earth Observation and Geoinformation*, 21, 7-16.
- LAURIN, G. V., PULETTI, N., HAWTHORNE, W., LIESENBERG, V., CORONA, P., PAPALE, D., CHEN, Q. & VALENTINI, R. 2016. Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment*, 176, 163-176.
- LAWRENCE, R. L., WOOD, S. D. & SHELEY, R. L. 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, 100, 356-362.
- LI, P., JIANG, L. & FENG, Z. 2013. Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors. *Remote Sensing*, 6, 310-329.
- LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. *R news*, 2, 18-22.
- LICCIARDI, G., PACIFICI, F., TUIA, D., PRASAD, S., WEST, T., GIACCO, F., THIEL, C., INGLADA, J., CHRISTOPHE, E. & CHANUSSOT, J. 2009. Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3857-3865.
- LOBO, A., LEGENDRE, P., REBOLLAR, J. L. G., CARRERAS, J. & NINOT, J. M. 2010. Land cover classification at a regional scale in Iberia: separability in a multi-temporal and multi-spectral data set of satellite images. *International Journal of Remote Sensing*, 25, 205-213.
- LOHBECK, M., BONGERS, F., MARTINEZ-RAMOS, M. & POORTER, L. 2016. The importance of biodiversity and dominance for multiple ecosystem functions in a human-modified tropical landscape. *Ecology*, 97, 2772-2779.
- LOWRY, J., RAMSEY, R., THOMAS, K., SCHRUPP, D., SAJWAJ, T., KIRBY, J., WALLER, E., SCHRADER, S., FALZARANO, S. & LANGS, L. 2007. Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: a case study of the Southwest Regional Gap Analysis Project (SWReGAP). *Remote Sensing of Environment*, 108, 59-73.

- LU, D. & WENG, Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote Sensing*, 28, 823-870.
- LUO, S., WANG, C., XI, X., ZENG, H., LI, D., XIA, S. & WANG, P. 2015. Fusion of Airborne Discrete-Return LiDAR and Hyperspectral Data for Land Cover Classification. *Remote Sensing*, 8.
- MANANDHAR, R., ODEH, I. & ANCEV, T. 2009. Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sensing*, 1, 330-344.
- MARTHEWS, T. R., BURSLEM, D. F., PHILLIPS, R. T. & MULLINS, C. E. 2008. Modelling direct radiation and canopy gap regimes in tropical forests. *Biotropica*, 40, 676-685.
- MATASCI, G., HERMOSILLA, T., WULDER, M. A., WHITE, J. C., COOPS, N. C., HOBART, G. W. & ZALD, H. S. J. 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sensing of Environment*, 209, 90-106.
- MATIKAINEN, L., KARILA, K., HYYPPÄ, J., LITKEY, P., PUTTONEN, E. & AHOKAS, E. 2017. Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128, 298-313.
- MAZZONI, D., GARAY, M. J., DAVIES, R. & NELSON, D. 2007. An operational MISR pixel classifier using support vector machines. *Remote Sensing of Environment*, 107, 149-158.
- MCNEMAR, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 153-157.
- MILLARD, K. & RICHARDSON, M. 2013. Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. *Canadian Journal of Remote Sensing*, 39, 290-307.
- MIURA, S., AMACHER, M., HOFER, T., SAN-MIGUEL-AYANZ, J., ERNAWATI & THACKWAY, R. 2015. Protective functions and ecosystem services of global forests in the past quarter-century. *Forest Ecology and Management*, 352, 35-46.
- MORI, A. S., LERTZMAN, K. P., GUSTAFSSON, L. & CADOTTE, M. 2017. Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *Journal of Applied Ecology*, 54, 12-27.
- MOUNTRAKIS, G., IM, J. & OGOLE, C. 2011. Support vector machines in Remote Sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 247-259.
- NAGENDRA, H., ROCCHINI, D., GHATE, R., SHARMA, B. & PAREETH, S. 2010. Assessing Plant Diversity in a Dry Tropical Forest: Comparing the Utility of Landsat and Ikonos Satellite Images. *Remote Sensing*, 2, 478-496.
- NITZE, I., BARRETT, B. & CAWKWELL, F. 2015. Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation*, 34, 136-146.
- NKONYA, E., MIRZABAEV, A. & VON BRAUN, J. 2016. *Economics of land degradation and improvement: a global assessment for sustainable development*, Springer.
- NOGUEIRA, E. M., YANAI, A. M., FONSECA, F. O. & FEARNESIDE, P. M. 2015. Carbon stock loss from deforestation through 2013 in Brazilian Amazonia. *Global change biology*, 21, 1271-1292.
- OMER, G., MUTANGA, O., ABDEL-RAHMAN, E. M. & ADAM, E. 2015. Performance of Support Vector Machines and Artificial Neural Network for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 4825-4840.

- OMRUUZUN, F., BASKURT, D. O., DAGLAYAN, H. & CETIN, Y. Y. Utilizing hyperspectral Remote Sensing imagery for afforestation planning of partially covered areas. *Image and Signal Processing for Remote Sensing XXI*, 2015. International Society for Optics and Photonics, 96432N.
- ONOJEGHUO, A. O. & BLACKBURN, G. A. 2011. Optimising the use of hyperspectral and LiDAR data for mapping reedbed habitats. *Remote Sensing of Environment*, 115, 2025-2034.
- OOMMEN, T., MISRA, D., TWARAKAVI, N. K., PRAKASH, A., SAHOO, B. & BANDOPADHYAY, S. 2008. An objective analysis of support vector machine based classification for Remote Sensing. *Mathematical geosciences*, 40, 409-424.
- PAHLEVAN, N., LEE, Z., WEI, J., SCHAAF, C. B., SCHOTT, J. R. & BERK, A. 2014. On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic Remote Sensing. *Remote Sensing of Environment*, 154, 272-284.
- PAL, M. & FOODY, G. M. 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 2297-2307.
- PAL, M. & MATHER, P. M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of environment*, 86, 554-565.
- PANEQUE-GÁLVEZ, J., MAS, J.-F., MORÉ, G., CRISTÓBAL, J., ORTA-MARTÍNEZ, M., LUZ, A. C., GUÈZE, M., MACÍA, M. J. & REYES-GARCÍA, V. 2013. Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *International Journal of Applied Earth Observation and Geoinformation*, 23, 372-383.
- PASTOR-GUZMAN, J., ATKINSON, P., DASH, J. & RIOJA-NIETO, R. 2015. Spatiotemporal Variation in Mangrove Chlorophyll Concentration Using Landsat 8. *Remote Sensing*, 7, 14530-14558.
- PEDRO, M. S., RAMMER, W. & SEIDL, R. 2015. Tree species diversity mitigates disturbance impacts on the forest carbon cycle. *Oecologia*, 177, 619-630.
- PELLETIER, C., VALERO, S., INGLADA, J., CHAMPION, N. & DEDIEU, G. 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156-168.
- PETROPOULOS, G. P., KALAITZIDIS, C. & PRASAD VADREVU, K. 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences*, 41, 99-107.
- PETTORELLI, N., SAFI, K. & TURNER, W. 2014. Satellite Remote Sensing, biodiversity research and conservation of the future. The Royal Society.
- POURSANIDIS, D., CHRYSOULAKIS, N. & MITRAKA, Z. 2015. Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping. *International Journal of Applied Earth Observation and Geoinformation*, 35, 259-269.
- ROCCHINI, D., BUTINI, S. A. & CHIARUCCI, A. 2005. Maximizing plant species inventory efficiency by means of remotely sensed spectral distances. *Global Ecology and Biogeography*, 14, 431-437.
- RODRÍGUEZ-GALIANO, V. F., ABARCA-HERNÁNDEZ, F., GHIMIRE, B., CHICA-OLMO, M., ATKINSON, P. M. & JEGANATHAN, C. 2011. Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environmental Sciences*, 3, 44-49.
- RODRIGUEZ-GALIANO, V. F., GHIMIRE, B., ROGAN, J., CHICA-OLMO, M. & RIGOL-SANCHEZ, J. P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.

- ROY, D. P., WULDER, M. A., LOVELAND, T. R., C.E. W., ALLEN, R. G., ANDERSON, M. C., HELDER, D., IRONS, J. R., JOHNSON, D. M., KENNEDY, R., SCAMBOS, T. A., SCHAAF, C. B., SCHOTT, J. R., SHENG, Y., VERMOTE, E. F., BELWARD, A. S., BINDSCHADLER, R., COHEN, W. B., GAO, F., HIPPLE, J. D., HOSTERT, P., HUNTINGTON, J., JUSTICE, C. O., KILIC, A., KOVALSKYY, V., LEE, Z. P., LYMBURNER, L., MASEK, J. G., MCCORKEL, J., SHUAI, Y., TREZZA, R., VOGELMANN, J., WYNNE, R. H. & ZHU, Z. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172.
- RÜGER, N., HUTH, A., HUBBELL, S. P. & CONDIT, R. 2009. Response of recruitment to light availability across a tropical lowland rain forest community. *Journal of Ecology*, 97, 1360-1368.
- SANFORD, R. L., BRAKER, H. E. & HARTSHORN, G. S. 1986. Canopy openings in a primary neotropical lowland forest. *Journal of Tropical Ecology*, 2, 277-282.
- SHEEREN, D., FAUVEL, M., JOSIPOVIĆ, V., LOPES, M., PLANQUE, C., WILLM, J. & DEJOUX, J.-F. 2016. Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series. *Remote Sensing*, 8.
- SINGH, S. K., SRIVASTAVA, P. K., GUPTA, M., THAKUR, J. K. & MUKHERJEE, S. 2013. Appraisal of land use/land cover of mangrove forest ecosystem using support vector machine. *Environmental Earth Sciences*, 71, 2245-2255.
- SKIDMORE, A. K. & PETTORELLI, N. 2015. Agree on biodiversity metrics to track from space: ecologists and space agencies must forge a global monitoring strategy. *Nature*, 523, 403-406.
- SNYDER, R. L., SPANO, D., DUCE, P., BALDOCCHI, D. & XU, L. 2006. A fuel dryness index for grassland fire-danger assessment. *Agricultural and Forest Meteorology*, 139, 1-11.
- SOTHE, C., ALMEIDA, C., LIESENBERG, V. & SCHIMALSKI, M. 2017. Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. *Remote Sensing*, 9.
- STEFANSKI, J., MACK, B. & WASKE, B. 2013. Optimization of object-based image analysis with random forests for land cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6, 2492-2504.
- STORY, M. & CONGALTON, R. G. 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52, 397-399.
- TANAKA, H., SHIBATA, M., MASAKI, T., IIDA, S., NIIYAMA, K., ABE, S., KOMINAMI, Y. & NAKASHIZUKA, T. 2008. Comparative demography of three coexisting *Acer* species in gaps and under closed canopy. *Journal of Vegetation Science*, 19, 127-138.
- TEAM, R. D. C. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>.
- THANH NOI, P. & KAPPAS, M. 2017. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel)*, 18.
- TOLEDO-ACEVES, T. & SWAINE, M. D. 2008. Effect of lianas on tree regeneration in gaps and forest understorey in a tropical forest in Ghana. *Journal of Vegetation Science*, 19, 717-728.
- TSUTSUMIDA, N. & COMBER, A. J. 2015. Measures of spatio-temporal accuracy for time series land cover data. *International Journal of Applied Earth Observation and Geoinformation*, 41, 46-55.
- TURNER, W. 2014. Conservation. Sensing biodiversity. *Science*, 346, 301-2.

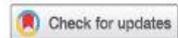
- TURNER, W., SPECTOR, S., GARDINER, N., FLADELAND, M., STERLING, E. & STEININGER, M. 2003. Remote Sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18, 306-314.
- VAPNIK, V. 1995. The nature of statistical learning theory. . Springer-Verlag.
- VICHARNAKORN, P., SHRESTHA, R., NAGAI, M., SALAM, A. & KIRATIPRAYOON, S. 2014. Carbon Stock Assessment Using Remote Sensing and Forest Inventory Data in Savannakhet, Lao PDR. *Remote Sensing*, 6, 5452-5479.
- WULDER, M. A., HALL, R. J., COOPS, N. C. & FRANKLIN, S. E. 2004. High spatial resolution remotely sensed data for ecosystem characterization. *AIBS Bulletin*, 54, 511-521.
- XU, Z., CHEN, J., XIA, J., DU, P., ZHENG, H. & GAN, L. 2018. Multisource Earth Observation Data for Land-Cover Classification Using Random Forest. *IEEE Geoscience and Remote Sensing Letters*, 15, 789-793.
- YAMAMOTO, S.-I. 2000. Forest gap dynamics and tree regeneration. *Journal of forest research*, 5, 223-229.
- YIN, H., KHAMZINA, A., PFLUGMACHER, D. & MARTIUS, C. 2017. Forest cover mapping in post-Soviet Central Asia using multi-resolution Remote Sensing imagery. *Sci Rep*, 7, 1375.
- ZHANG, H. K. & ROY, D. P. 2017. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sensing of environment*, 197, 15-34.
- ZHEN, Z., QUACKENBUSH, L. J., STEHMAN, S. V. & ZHANG, L. 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *International Journal of Remote Sensing*, 34, 6914-6930.

**CHAPTER 4. MULTI-DECADAL SPATIAL AND TEMPORAL FOREST
COVER CHANGE ANALYSIS OF NKANDLA NATURAL RESERVE, SOUTH
AFRICA**

JOURNAL OF SUSTAINABLE FORESTRY
<https://doi.org/10.1080/10549811.2021.1891441>



ORIGINAL ARTICLE



**Multi-Decadal Spatial and Temporal Forest Cover Change
Analysis of Nkandla Natural Reserve, South Africa**

Enoch Gyamfi-Ampadu^a, Michael Gebreslasie ^a, and Alma Mendoza-Ponce^{b,c}

^aSchool of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Durban, South Africa; ^bEcosystem Services and Management, International Institute of Applied System Analysis, Laxenburg, Austria; ^cCentro De Ciencias De La Atmósfera, Universidad Nacional Autónoma De México, Mexico City, Mexico

Abstract

Forest cover change analyses have an essential role in forest management. Thus, this study adopted Landsat satellite imageries to assess the decadal spatiotemporal forest cover changes that occurred between 1989 and 2019 and predicted the 2029 land cover distribution of the Nkandla Forest Reserve, facing encroachment threats. The support vector machine algorithm and Land Change Modelling were utilized to classify and detect changes that occurred between 1989-1999, 1999-2009, 2009-2019. The Markov Chain Model and Multi-Layer Perceptron were adopted for future land cover prediction. Consistent changes through inter-transitioning between the land cover types (closed canopy forest, open canopy forest, grassland, and bare sites) were detected. The closed canopy forest increased from 883.46 ha to 1059.23 ha, whereas the open canopy forest declined from 1091.89 ha to 910.60 ha between 1989 and 2019. Generally, the observed changes were caused by ecological processes and human disturbances. The future cover prediction indicated that the closed canopy forest will decline between 2019 and 2029, whereas the open canopy forest, grassland, and bare sites will increase. The information provided through this study will support the management of the Nkandla forest to ensure its continual supply of ecosystem services of national and global importance.

Keywords: Natural Forest, Remote Sensing, Change Detection, Sustainable, Support Vector, Modelling.

4.1 Introduction

Natural forest ecosystems contribute significantly to the maintenance of biodiversity (Arroyo-Rodriguez et al., 2020, Boedhihartono, 2017, Kumar et al., 2019) and the provision of other ecosystem services such as mitigation of climate change (Gilroy et al., 2014, Lohbeck et al., 2015). Natural forests also have social and cultural benefits (Agrawal et al., 2013, Ahammad et al., 2019), such as traditional medicines (Tugume et al., 2016) which are important for societal wellbeing. The forest species, density, and spatial extent are vital in ensuring the provision of these services. These characteristics are determined by the ecological and biological process and interdependence that takes place within the natural forest ecosystem.

In many places, natural forests are changing and dwindling due to deforestation, degradation, and fragmentation. Anthropogenic factors such as deforestation and natural elements such as regeneration, insect attack, and tree mortality are among the major causes of forest cover changes (Gómez et al., 2016, Ngwira and Watanabe, 2019, Barlow et al., 2016). Natural forest cover changes have severe implications on biodiversity richness, habitat conservation, carbon sequestration and storage, climate change regulation, soil conservation, water supplies, and other ecosystem services (Reddy et al., 2018, Zhu et al., 2018, Chaudhary et al., 2016, Sharma et al., 2019). Thus, forest cover change detection and future forest cover prediction studies using Remote Sensing technologies are essential for providing information that supports effective forest management strategies to ensure these benefits are sustained.

Many studies have utilized Remote Sensing imagery for change detection and forecasting of future forest cover distribution. This is because Remote Sensing imagery covers a large area, it is cost-effective (Khatami et al., 2016), and serves as a source of spatial data for undertaking forest cover change detection (Hansen and Loveland, 2012, Vittek et al., 2014, Rawat and

Kumar, 2015). Land cover change detection methods are a highly variable and evolving area of application (Tewkesbury et al., 2015, Wessels et al., 2016). User discretion is usually employed in selecting the best change detection approach (Tewkesbury et al., 2015), as there are no universal methods for such assessments. However, the methods adopted could be dependent on the Remote Sensing data available, the topography of the study area, and the scope of the study (Lu et al., 2014b). Hence, most studies develop a hybrid methodological approach for the change detection operation. Supervised and unsupervised image classification, Markov Chain and Cellular Automata models, Multi-Layer Perceptron Neural Networks, and image differencing are among the approaches mostly used in forest cover change detection and forecasting.

Da Ponte et al. (2017) conducted a forest cover loss analysis for the Paraguay Atlantic forest from 2003 to 2013 using 15 image scenes of Landsat Enhanced Thematic Mapper (ETM) and the Landsat 8. The landcover classification based on which the change detection was conducted had overall accuracies ranging from 81% to 95% and kappa coefficient ranging from 0.62 to 0.90. The study found that a total of 6000 km² forest was lost between 2003 and 2013 from a total forest area of 33, 000 km². This was found to be caused by the clearing of land for mechanized agriculture land use systems and has left most areas with fragmented landscapes. The situation has compromised the integrity of the forest and affected its functionality. Mihai et al. (2017) also carried out a study in the Iezer Mountains of Romania to detect total forest cover change that occurred between 2002 and 2015. The Remote Sensing data used were the Landsat ETM, Landsat 8 and Sentinel 2. The Maximum Likelihood algorithm was used for the image classification of seven land cover classes in the forest and the overall accuracies ranged higher than 91% while the kappa coefficient was above 0.90. The image differencing technique was used for the change detection and it was discovered that a total of 6100 ha was lost between 2002 and 2015. This area of forest was converted to pasture and barren land. In another study, a multi-temporal change detection analysis was undertaken to identify forest cover losses that took place between 1984 and 2015. Through a post image classification analysis, it was observed that the year 2000 had the most change with 1431.6 ha of forest land converted to oil palm plantations, built-up area and human settlements. However, there was a slight increase of about 2.59% for the primary forest and 5.06 for the secondary forest between 2000 and 2015. This increase was a positive trend and maintenance of such phenomenon, could improve forest cover.

In Mexico, Vázquez-Quintero et al. (2016) developed a hybrid methodological approach that incorporated a Maximum likelihood supervised classification with Markov Chain and Cellular Automation model approach in classifying two main forest types (Pine and Oak forests) for change detection and future cover prediction for 2028. This was carried out using archival Landsat images which were Multispectral Scanner (MSS) of 1973, Thematic Mapper (TM) of 1990, and Landsat 8 of 2014. The land cover classification had overall accuracies of 93%, 94%, and 92% for 1973, 1990, and 2014 images respectively. Their findings indicated that the pine forest was the most affected over the period while the oak forest will further decline in its extent by 2028. Similarly, Voight et al. (2019) in Southern Belize employed Landsat 8 imagery of 2014, 2016, and 2017 in modelling forest cover change and predict future distribution. The Google Earth Engine was used for the land cover classification and obtained an overall accuracy of 88% for 2014, 94% for 2016, and 95% for 2017. A post-classification future forest cover prediction for 2026 was carried out using the Markov Chain Model. The prediction

indicated that the extent of forest cover in 2016 will decrease by 3.1% in 2026 mainly due to expansion in agricultural lands. These dynamics of changes are indicated to have negative implications that will require effective forest management to possibly initiate mitigation measure and reduce the vulnerability of the forest landscape.

Apart from change detection and future projection of forest landscapes, other studies have been carried out in urban landscapes with similar modelling approaches. These studies analyzed the rapid expansion in urban lands at the expense of vegetative areas and the associated negative implications. A change detection to identify how urban areas in Greater Accra, Ghana revealed that there has been a 277% increase in built-up areas, while forest areas have declined from 34% to 6.5% between 1991 and 2015 (Addae and Oppelt, 2019). The future projections indicated that the built-up area would increase from 44% of the total landmass in 2015 to 70% in 2025. This study too was carried out with Landsat 4 TM, Landsat 7 ETM, Landsat 7 ETM+, and Landsat 8 for the years 1991, 2000, 2009, and 2015, respectively. Similarly, a Maximum Likelihood classification and Multi-Layer Perceptron Markov Chain Model were used for the change detection and future land distribution prediction. In another study, the effect of rapid urbanization was assessed and it revealed that an increased rate of urbanization happened between 1996 and 2017 (Ranagalage et al., 2019). The support vector machine was used to classify the various land use land cover (LULC) with the use of Landsat 5TM image of 1996 and 2006, and Landsat 8 image of 2017. The overall accuracies were 85% for 1996, 93% for 2006, and 92% for 2017 images. A post-classification change detection revealed that there had been a 1791 ha increase in built-up areas and a 1919 ha decline in agricultural lands between 1996 and 2017. The study further carried out two-time step future predictions for 2027 and 2037 which indicated the built-up will increase by 13.4% and 17.7%, respectively. On the other hand, the agricultural land will have a decline of 3.2% in 2027 and 2.5 in 2037.

Most of these studies conducted for forest cover or urban landscapes are observed to have provided the information needed to assist in the forest and other land use management and conservation. However, such studies are lacking in the South African scenario. To date, there has not been a study that assesses the past and present status of natural forests, as well as its possible future implications. Therefore, a forest cover change modelling and forecasting of future cover distribution are important for natural forests in a country that has only 0.4% of the total land area as natural forests (DAFF, 2015). As a result of the increasing threat to natural forests cover, comprehensive and credible information on coverage and likely future coverage is critical for effective natural forest management strategy. Thus, this study aimed to undertake spatial and temporal analysis of the Nkandla Natural Forest Reserve between 1989 and 2019 at a decadal interval and predict the likely forest cover distribution for 2029. The Nkandla Forest Reserve which is the study area in context faces possible threats due to expanding fringe and adjacent communities as well as increasing human activities in the forest (Ezemvelo KZN Wildlife, 2015b). Issues such as wildfires, agricultural activities, exotic species plantations, and uncontrolled grazing are among the immediate threats. The quantitative and qualitative outcomes of this study will provide an understanding of the change dynamics, trends, and distribution of land use cover of the forest, which will be beneficial for forest management and conservation. Furthermore, the outcomes have the potential to contribute to national policy directions on climate change mitigation and forest protection, which can also feed into international initiatives.

4.2 Materials and methods

4.2.1 Study area

The Nkandla Forest Reserve was established in 1918. It is an Afromontane sub-tropical forest in the KwaZulu-Natal (KZN) province of South Africa and has a total area of 2,217 ha. It is located on 28° 43' 50.88" S and 30° 7' 9.84" E (Figure 4.1). The forest experiences a peak average temperature of 27°C between December and January, and the lowest average temperature of 2°C in the winter months of June and July (Ezemvelo KZN Wildlife, 2015b). It has a generally steep and undulating topography and an altitude of a minimum level of 500 m and exceeding 1300 m. It has four main land cover types made up of closed canopy forest (1,059.23 ha), open canopy forest (910.60 ha), grassland (226.55 ha), and bare sites [20.97 ha] (Gyamfi-Ampadu et al., 2020). The grasslands are found on hilltops and downhills with some patches interspersing the areas of the closed and open canopy forests. Some areas around the forest boundaries experience frequent fires due to the activities of the forest fringe communities. The fringe communities graze their domestic animals such as cattle in the grasslands of the forest and obtain some non-timber forest products (NTFPs). An increase in commercial plantations and expanding communities possess some level of threat to the forest.

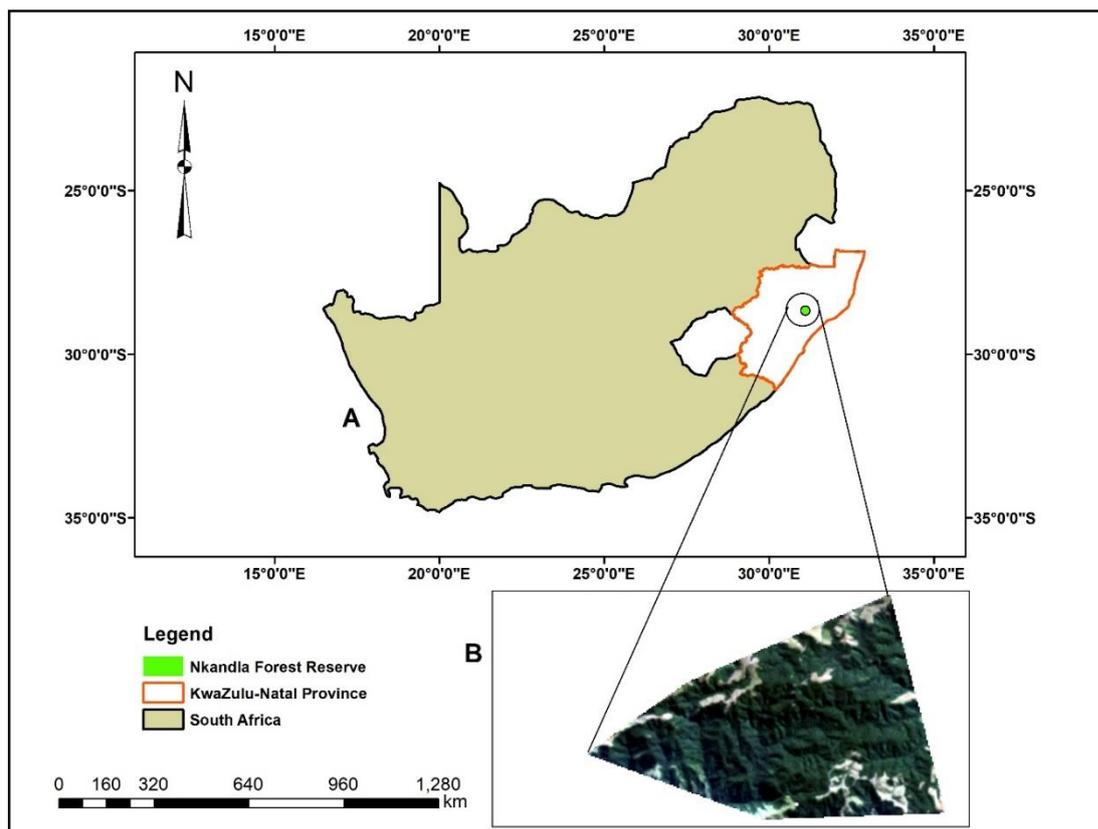


Figure 4.1: Map of the study area. Note: Map A is a satellite image of the Nkandla Forest Reserve, and Map B is a map of South Africa indicating the location of the forest.

4.3 Data used

4.3.1 Field data

Ground-based reference data were collected for the forest cover classification. Predefined thematic land cover types were determined to enhance the reference data collection. These land cover types are closed canopy forest, open canopy forest, grassland, and bare site. The closed canopy forest has the tree canopy touching each other to form a continuous layer (70% to 100%) with less sunlight reaching the forest floor. It has low vegetation on the forest floor, and visibility may be up to about 20 m under the canopy. The open canopy forest is characterized by a broken canopy, a high level of gaps, and no continuous canopy layer (30% to 70%). It has an undergrowth of seedlings and other herbaceous plants with visibility of up to less than 20 m. The grassland has a continuous growth of the grass and herbaceous layer. The bare sites are made up of sites that do not have any vegetation cover.

A random data collection approach was employed to avoid opportunistic biased classification (Zhen et al., 2013, Millard and Richardson, 2015, McRoberts and Westfall, 2016). Moreover, it also served best especially for the closed and open canopy forest types which cannot be easily identified from afar. A 25 m² minimum area criteria and ocular estimation of site composition and structural parameters were used in determining each land cover type when they were opportunistically encountered through traversing along transects. When a site met the defined parameters, Global Position System (GPS) coordinates were recorded and placed under the appropriate cover type. The GPS coordinates of each land cover type were superimposed on the 2019 image and random-sized polygons were digitized around them in the ArcMap 10.6.1 mapping environment. A total of 63, 76, 69, and 20 random sized polygons were digitized for the closed canopy forest, open canopy forest, grassland, and bare sites cover types respectively.

4.3.2 Remote Sensing imagery used and pre-processing

The Remote Sensing data used for the study was Landsat satellite imagery as it has proven useful in many land use land cover (LULC) analyses (Voight et al., 2019, Da Ponte et al., 2017, Fokeng et al., 2020). Cloud free Landsat 4 TM, Landsat 5 TM, and Landsat 8 images were obtained from the United States Geological Survey (USGS) online data portal (www.earthexplorer.usgs.gov). The images were for 1989, 1999, 2009, and 2019 and the details have been illustrated in Table 4.1 The images were preprocessed to optimal parameters to facilitate their use for the required analysis. The apparent reflectance function in ArcGIS 10.6.1 was used to atmospherically correct the image and transform the image radiance to spectral reflectance. All the images had the same extent as well as columns and rows to ensure their effective use for the analysis.

Table 4.1: Characteristics of the Landsat images used.

Acquisition date	Landsat type	Sensor	Spatial resolution (m)	Number of bands used
19/04/1989	Landsat 4	TM	30	6
09/01/1999	Landsat 4	TM	30	6
12/05/2009	Landsat 5	ETM	30	6
08/05/2019	Landsat 8	OLI	30	7

Note: TM is Thematic Mapper, ETM is Enhanced Thematic Mapper and OLI is Operational Land Imager.

4.3.3 Environmental parameters data

The Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model (DEM) data were downloaded from the United States Geological Survey (USGS) online data portal (www.earthexplorer.usgs.gov). The DEM of the reserve was delineated and used to generate the slope gradient and the aspect using ArcGIS 10.6.1. They were processed into raster files and projected to the coordinate systems of the images and as well the same dimensions. They were subsequently used as environmental data in the change modelling for the production of transitional potential maps.

4.4 Image classification and Accuracy Assessment

The satellite images and the digitized reference polygons were imported into the R statistical package environment (Team, 2017). The pixel values representing each of the land cover types were extracted from the digitized reference polygons. The pixel values were then partitioned into a training set (70%) and a validation set (30%) for each land cover type. Subsequently, the radial function support vector machine (SVM) algorithm was used for classifying the images. The SVM has proven to be a robust image classification algorithm in many studies (Cervantes et al., 2020, Wang et al., 2017, Thanh Noi and Kappas, 2018). The “caret” package which contains several classification functions was employed for the process (Ghosh et al., 2014). The “svmRadial” function was applied and used for the training of the classification model. The model was trained with the training data set to classify the 1989, 1999, 2009, and 2019 images. The default parameters of the algorithm were used with no tuning or optimization as it is possible to obtain optimal accuracies with default parameters (Gyamfi-Ampadu et al., 2020). The common spectral bands that were selected for each of the images were the red, green, blue, near infra-red (NIR), and the shortwave infrared bands (SWIR 1& 2). These bands are sensitive to vegetation (Roy et al., 2014, Dube and Mutanga, 2015) and the application of the SVM algorithm reflected the importance of each in classifying and mapping the land cover types. The validation set was subsequently used for the classification accuracy assessment and confusion matrices were derived for them. The matrix depicts the estimates of the parameters used for the accuracy assessment which are the overall accuracy, producer’s accuracy (omission error) and user’s accuracy (commission error) and the kappa coefficient (Pal and Ziaul, 2017, Rousta et al., 2018).

4.5 Historical Land Cover Change Detection

The Land Change Modeler [LCM] (Eastman, 2015) was implemented in the TerrSet Geospatial Modelling system to model the land cover changes that occurred at decadal intervals from 1989-1999, 1999-2009, and 2009-2019 as well as the overall change that occurred between 1989 and 2019. The changes were assessed based on the coverage of each land cover type at each decade. The areas of gain and losses, areas of no change, spatial trends, and extent for each of the land cover types were mapped for each time interval. The gains and losses analysis estimated the changes while the areas of no change were identified through the persistence analysis.

4.5.1 Transition Probability Matrices and Potential Maps

The transition potentials provide an estimation of the likely future changes a land cover type may experience between two historical periods. Transition probability matrices were created with Markov Chain Model (MCM) using the 1989, 1999, 2009, and 2019 forest cover maps. The MCM works based on the Markovian stochastic approach that predicts the possibility of change from one state to the other by using transition probability metrics of land cover changes experienced over a certain period (Rimal et al., 2018, Wu et al., 2019). It relies on the transition probabilities to control the temporal dynamics among the land cover types (Kamusoko et al., 2009). It then facilitates the estimation of states of conversion and transfer rates among the types (Sang et al., 2011). The MCM is not capable of simulating spatially distributed cover change (Sang et al., 2011, Mishra and Rai, 2016), but it can determine and predict land cover change quantities effectively (Yang et al., 2012). The integration of the MCM with algorithms such as the Multi-Layer Perceptron Neural Networks (MLPNN) provides a robust method to quantify and model the temporal and spatial changes in land cover (Mishra and Rai, 2016).

The transition potential maps were hence created with the MLPNN algorithm. The MLPNN is made up of interconnected nodes that respond to weighted inputs it receives from other nodes in the network (Addae and Oppelt, 2019). It is trained to reduce errors in a network by weight adjustment between the nodes. It has the capability of making the best generalization for each cover transition and simulation (Maithani, 2015). In conducting land cover modelling, the MLPNN algorithm creates the maps by using transition sub-models made up of transition between cover types over a time interval. The potential maps are then used for the simulation and prediction of future land cover distribution.

In the actual production of the potential maps, the MLPNN trains the model with the sample pixels that experienced a transition from one land cover type to another and a different pixel set that did not experience any change in the two land cover maps. With the use of default parameters, the sample size is set equal to the smallest number of the set of pixels that transitioned from one land cover type to another. The selected set of pixels are assigned to two classes by the MLPNN consisting of a set for the training of the model and another set for validating the accuracy of the MLPNN model. The MLPNN uses the training set of pixels as an example in developing a multivariate function that predicts the transition potential based on the values at any location for the assigned explanatory variables (Eastman, 2015). When the process is completed, the MLPNN produces several statistics that provide information regarding the power of the explanatory variables as well as the accuracy of predicting the land cover type transitions and persistence.

There are three layers which are inputs, outputs, and hidden layers in the MLPNN algorithm structure. Concerning this study, the 1999 and 2009 land cover maps were used as base maps for the algorithm. The other inputs consisted of 10 transitional sub-models of the cover types and three environmental variables. The environmental variables were the DEM, slope gradient, and aspect generated for the study area. The transition potential maps were the outputs of the model after the algorithm has run 10,000 iterations. A map was produced for each of the ten transition sub-model for the forest cover change by implementing the model using the default parameters. The hidden layer is made up of a series of process and iterations that transforms the input data to produce the output layer.

4.5.2 Land cover distribution forecasting

The Markov Chain was applied to the transition potential maps produced by the 1999 and 2009 forest cover maps to predict the forest cover for 2019. In the process, the Markov chain calculated the extent of land cover that will be transitioning to the other for each land cover type over the specified future date. With the Markov chain being a stochastic process, the next stage of the system ($t+1$) is dependent on its current state (t) (Eastman, 2015). The extent of change that will be experienced was estimated by the Markov chain based on the two land cover maps used for the modelling. There was an estimation of the extent of land cover that would be expected to change from the later date to the specified future date. This is based on the projected transition potentials that are derived from the MLPNN (Aguejdad et al., 2017) and then creates the transition probability matrix. The transition probability matrix ($n*n$) obtained for the land cover maps is expressed by;

$$P = \begin{bmatrix} P_{11} & \dots & P_{1n} \\ \dots & \dots & \dots \\ P_{n1} & \dots & P_{nn} \end{bmatrix}, p_{ij} \geq 0, \sum_{j=1}^n p_{ij} = 1, i = 1 \quad (4.1)$$

where P denotes the probability matrix of n states,

p_{ij} denotes the transition probability of state i to j

The probability of the future state of a cell is also given as;

$$p(t+1) = p(t) * P \quad (4.2)$$

where $p(t+1)$ denotes the future land cover distribution,

$p(t)$ denotes the current land cover distribution,

P denotes the transition potential.

The process ended with the production of hard and soft prediction maps. The hard prediction map has the same land cover types as input maps. The model is based on the competitive land allocation model with similarity to a multi-objective decision process (Eastman, 2015). On the other hand, the soft prediction map has a similarity with the transition potential map. It expresses the change probability of every pixel for a transition in a continuous map format.

4.5.3 Validation and Projection

Validation is essential to assess accuracy of the prediction model. Once satisfactory performance is achieved, then it can be used for predicting the possible future land cover distribution. In this study, it was conducted by comparing the simulated 2019 map with the 2019 classification map (actual map) using kappa statistics (Kamusoko et al., 2009, Wang et al., 2012). The Kappa for no information (K_{no}) which is a variation of the standard kappa index is used to define the overall accuracy of the model. The ability of the simulation to predict the location is validated by kappa for grid-level location ($K_{location}$), while the quantity is predicted by Kappa for quantity [$K_{quantity}$] (Pontius Jr and Schneider, 2001). When each of the values is close to 1 then the model is defined as perfect, but if it is close to 0 then it is considered imperfect (Pontius Jr, 2002, Pontius Jr and Schneider, 2001).

The model produced a Kappa for no information (K_{no}) of 0.84, which was close to 1 and therefore considered satisfactory for the prediction. The prediction was subsequently conducted for the year 2029.

4.6 Results

4.6.1 Spatial distribution of cover types

Land cover maps were produced for 1989, 1999, 2009, and 2019 to ascertain the extent and spatial distribution of each land cover type of the forest reserve. Spatially, each land cover type is mostly distributed or clustered in certain parts of the reserve (Figure 4.2). The pattern of distribution for the cover types did not change markedly from that of 1989, as similar patterns were observed for each of the land cover maps produced for each decadal year.

The closed canopy forest is mostly distributed in the middle, northeastern, south, and southwestern parts of the reserve. The open canopy forest intersperses into the areas of the closed canopy and is also distributed in the southeastern, southwestern, and northeastern parts of the reserve with some patches found in the middle. The grassland could be found in the southeastern, northwestern, and northeastern boundaries of the reserve. Most of the bare sites are found within the grassland in the northeastern, southwestern, and southeastern boundaries of the forest.

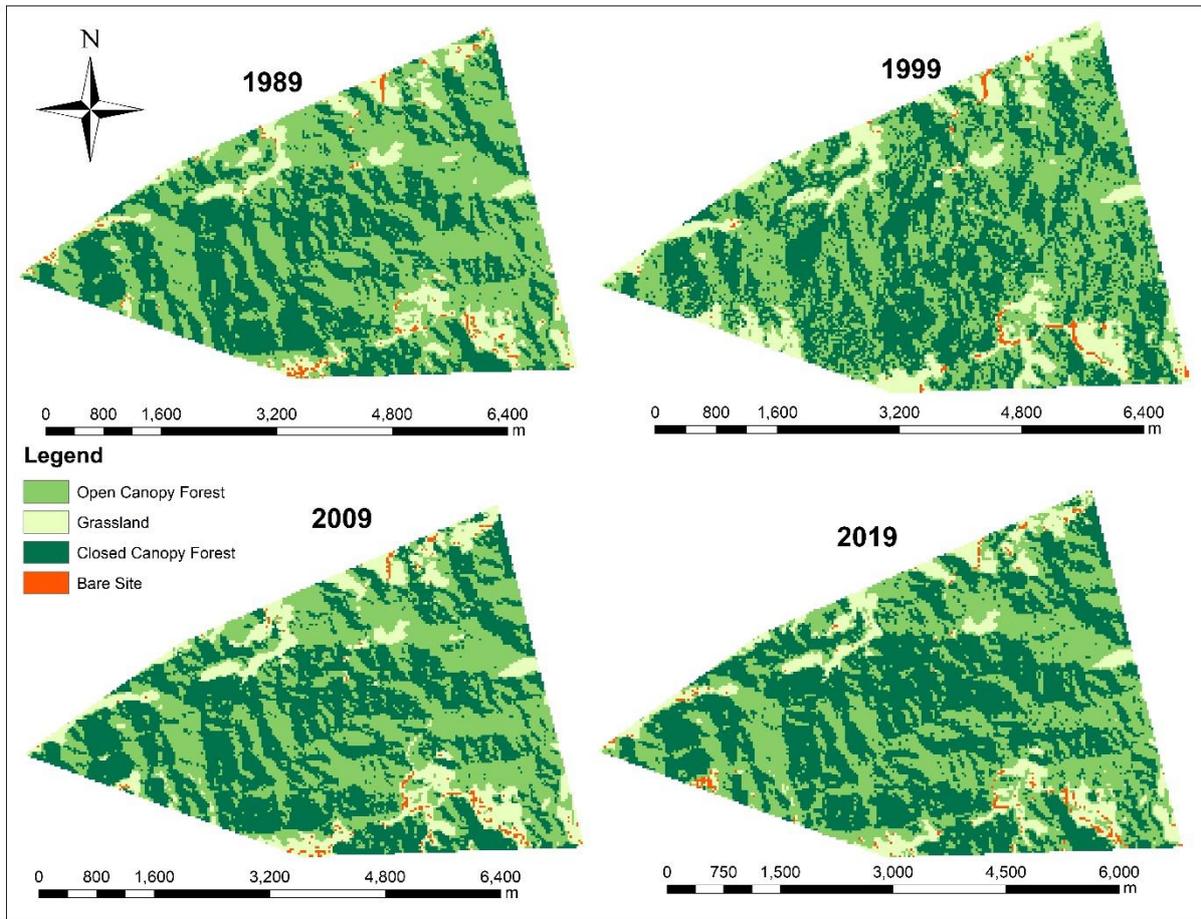


Figure 4.2: Land cover maps of the forest reserve produced from the classified images of 1989, 1999, 2009, and 2019.

A detailed accuracy assessment was conducted for the classification (Table 4.2). The overall accuracies obtained were 94.64%, 90.22%, 94.38% and 95.83% while the kappa coefficient was 0.92, 0.84, 0.91 and 0.93 for 1989, 1999, 2009 and 2019 images respectively. The high accuracies obtained reflect the robustness of the SVM algorithm in utilizing the selected Landsat bands to map each cover type. The overall accuracies for each classification were within acceptable limits of 80% and above, and hence the forest maps produced were satisfactory to be used for the modelling.

Table 4.2: Accuracy assessment of forest cover classifications for 1989, 1999, 2009, and 2019 images.

Accuracy	Forest cover class	1989	1999	2009	2019
User's Accuracy (%)	Closed Canopy forest	98.55	89.25	98.53	95.83
	Open Canopy forest	95.65	88.46	97.87	100
	Grassland	89.13	96.77	95.92	95.35
	Bare Sites	85.71	87.50	57.14	81.81
Producer's Accuracy (%)	Closed Canopy forest	98.55	93.26	98.53	100
	Open Canopy forest	97.78	82.14	97.87	93.33
	Grassland	95.35	96.77	88.68	95.35
	Bare Sites	54.54	87.50	80.00	81.81
Overall Accuracy (%)		94.64	90.22	94.38	95.83
Kappa Coefficient		0.92	0.84	0.91	0.93

4.7 Spatiotemporal Analysis

4.7.1 General forest cover changes

The land cover types of the forest reserve experienced considerable changes in their respective coverage between 1989 and 2019 (Table 4.3). In this period, the closed canopy forest had a change in extent from 883.46 ha in 1989 to 1059.23 ha in 2019. This accounts for a 7.93% increase representing a total area of 175.77 ha, occurring at a rate of 5.86 ha per year. The changes happened mostly in the middle and towards the northern parts of the forest (Figure 4.2).

On the other hand, the open canopy forest lost 8.18% of its coverage by decreasing in extent from 1091.99 ha in 1989 to 910.59 ha in 2019. The total area of change was 181.40, which happened at the rate of 6.04 ha per year (Table 4.3). The changes occurred mostly in the middle, towards the north and northeastern parts of the forest (Figure 4.2).

The grassland increased by 2.8% as a result of a change in extent from 218.97 ha as of 1989 to 226.55 ha as of 2019. This represented a total area of change by 7.60 ha which happened at a rate of 0.25 ha per year. These changes were concentrated around the southeastern and southwestern parts of the forest (Figure 4.2). The bare sites decreased from 22.94 ha in 1989 to 20.98 in 2019. This represents a change in area by 1.96 ha, which happened at a rate of 0.06 per year. The changes occurred mostly around the southern parts of the forest (Figure 4.2). The individual land cover type changes fed into the overall changes that were experienced between 1989 and 2019 (Table 4.3).

Table 4.3: Decadal land cover changes that occurred for the Nkandla Forest Reserve from 1989 to 2019.

Land Cover Class	1989		1999		2009		2019		Δ1989-1999	Δ1999-2009	Δ2009-2019
Land Cover Class	Area (ha)	%	Area (ha)	Area (ha)	Area (ha)						
Closed canopy	883.46	39.84	936.13	42.22	968.98	43.70	1059.23	47.77	52.66	32.86	90.55
Open canopy	1091.99	49.25	980.14	44.21	939.52	42.37	910.60	41.07	-111.85	-40.62	-28.91
Grassland	218.97	9.88	285.47	12.87	288.59	13.02	226.55	10.22	65.50	3.13	-62.04
Bare sites	22.94	1.03	15.62	0.70	20.26	0.91	20.98	0.94	-7.32	4.63	0.75
Total Area	2217.36	100	2217.36	100	2217.36	100	2217.36	100			

4.7.2 Decadal changes within forest cover types

The changes experienced in coverage between 1989 and 2019 were due to inter-transitions between the land cover types and were assessed at a decadal interval for 1989-1999, 1999-2009, and 2009-2019. The area of a land cover type that transitioned into another, served as a loss for it but as a gain for the other land cover type. Much of the transitions occurred between the closed and the open canopy forest types. The closed canopy consistently increased in coverage at each decadal interval by gaining more areas of the open canopy forest. This is a result of more trees in the open canopy forest growing (vertical and horizontal) to fill canopy gaps and form continuous layers. Much of these substantial changes occurred between 1989-1999, involving an area of 302.88 ha and between 1999-2009 involving an area of 319.22 (Details of this can be found in Table 4.5 of Appendix 4.1). This could be seen around the middle, eastern, and towards the northeastern parts of the forest (Figure 4.2). As a result of the consistent changes, the closed canopy forest became the dominant land cover from 2009 and remained as such as of 2019.

Despite that, the closed canopy also had a considerable area of its cover transitioning to open canopy forest consistently at each decadal interval. This is manifested in the increase in gap openings that break the continuous canopy layer of the closed canopy forest. Between 1999 and 2009, it experienced the largest loss to the open canopy forest involving an area of about 301.27 ha (Details can be found in Table 4.5 of Appendix 4.1). This is observed mostly around the eastern and southeastern parts of the forest (Figure 4.2). Some parts of the closed canopy forest ranging between 3 ha and 19 ha also transitioned to grassland. This happened due to the loss of tree cover in these parts making grassland take over. There was no complete turnover of the structural condition of the closed canopy forest. This is because about two-thirds of its cover experienced no change (persisted) at each decadal interval, which may explain the pattern of distribution remaining like that of 1989.

The open canopy similarly experienced a consistent and substantial change in each decade. Much of the gains it had were as a result of parts of the closed canopy forest experiencing gaps in its continuous canopy layer. The period it had such significant gain was 1989-1999 involving an area of 237.62 and 1999-2009 involving an area of 301.27 (Details of this can be found in Tables 4.6 and 4.7 of Appendix 4.2 and 4.3). In terms of the areas that transitioned from the open canopy forest to the closed canopy forest, the process is the same as explained under why the closed canopy increased in the extent of coverage. About two-thirds of its original extent experienced no change. Comparatively, the total area of tree cover of the open canopy growing into close canopy was more than that of the closed canopy losing tree cover to become open canopy. Hence, the open canopy lost its status as the dominant cover from 2009 onwards to the closed canopy forest.

Most of the area of the grassland had as of 1989 experienced no change, although some parts underwent a consistent change at each decade. The gains it had were as a result of taking over some parts of the closed and open canopy forests because of tree cover loss. Between 2009-2019, it had the highest gain of the open canopy forest, involving an area of 60.43 (Details of this can be found in Table 4.8 of Appendix 4.4). It took over some

areas of the bare sites at each decade to a varying extent ranging between 8 ha and 14 ha. This is as a result of grassland colonizing these areas of no vegetation cover. However, between 5 ha and 17 ha of its cover changed into the bare site.

The bare sites similarly experienced a change in each decade. Most of its increase in size was as a result of part of the grassland losing its vegetation cover to become bare. Such extent ranged between 4 ha and 17 ha, with the 1989-2009 period having the most change involving an area of about 16.70 ha.

4.7.3 Land cover predictions

The model was validated to ascertain its suitability for predicting the future land cover distribution of the forest. A satisfactory kappa accuracy of 0.84 was obtained which indicated its suitability. The prediction was conducted for the 2029 land cover distribution (Table 4.4 and Figure 4.3). Amongst all the land cover types, it is the closed canopy forest that may have the highest likely change in coverage. The prediction indicated that the closed canopy forest will experience a decline in coverage from 1059.23 (47.77%) in 2019 to 979.43 (44.17%) in 2029 (Table 4.4). This represents a loss of 79.80 ha at a rate of 7.98 ha per year, accounting for about 3.6% of its original area. These changes could likely happen around the eastern, southeastern, and northeastern parts of the reserve (Figure 4.3). The predicted loss of some areas of the closed canopy forest is a result of mostly the grassland taking over those places due to tree cover loss.

The open canopy was predicted to increase from 910.60 ha (41.07%) in 2019 to 920.78 ha (41.53%) in 2029. This involved an area of 10.18 ha increasing at a rate of 1.08 ha per year. The open canopy forest will not experience much change as compared to the closed canopy forest and the grassland as it will still cover about 41% of the forest between 2019 and 2029.

The grassland was predicted to gain about 68.02 ha (3%) more between 2019 and 2029 at a rate of 6.08ha per year. In 2029 it will increase from 226.55 ha (10.22%) to 294.57 ha (13.28%). Much of the gains of the grassland will be as a result of taking over some parts of the closed canopy forest. The bare sites were predicted to increase marginally from 20.98 ha (0.94%) to 22.58 ha (1.02%) in 2029.

Table 4.4: Actual land cover of 2019 and predicted land cover distribution of the forest for 2029.

Land cover types	2019		2029	
	(Actual)		(Predicted)	
	Area (ha)	%	Area (ha)	%
Closed Canopy	1059.23	47.77	979.43	44.17
Open Canopy	910.60	41.07	920.78	41.53
Grassland	226.55	10.22	294.57	13.28
Bare Site	20.98	0.94	22.58	1.02
Grand Total	2217.36	100	2217.36	100

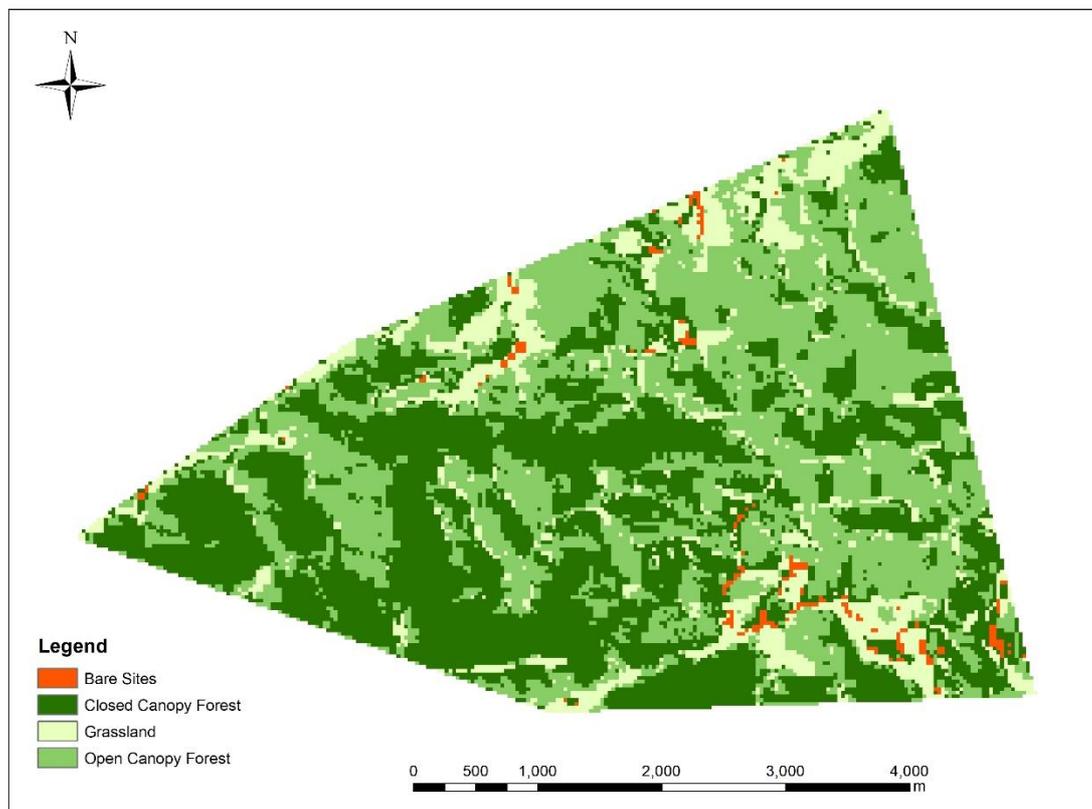


Figure 4.3: Predicted land cover distribution map of the forest for 2029.

4.8. Discussion

Forest cover change detection and future prediction analysis have many implications on decision making and policy directions both at the national and international levels, concerning the conservation of forest ecosystems, species diversity, and climate change. The information provided is vital for sustainable forest management in ascertaining the

condition of forests concerning trends and dynamics of changes. It ensures effective planning and prioritization of initiatives meant to sustain forest ecosystems as well as devise mitigation actions in case of loss of tree cover. The outcomes of this study are therefore essential at the local and national level considering the limited total area of natural forest reserves in South Africa. This was made possible by the freely available Remote Sensing satellite imagery and implementation of LULC Modelling.

The Landsat satellite imagery data proved useful in both the classification and change detection analysis. The historical data enables the research to assess the past and present changes for sustainable forest management. The sensor has a long history of providing data that has contributed to land cover analysis with much success (Phiri and Morgenroth, 2017). The Landsat data has equally proven useful in many other studies that have conducted LULC changes analysis in different forest ecosystems and landscapes (Addae and Oppelt, 2019, Da Ponte et al., 2017, Mihai et al., 2017, Vázquez-Quintero et al., 2016). The attributes of the Landsat images coupled with the methodological approach enhanced the spatial and temporal changes mapping and projections for the Nkandla Forest Reserve. The visible range bands, NIR, and SWIR bands which were used in the classification have good sensitivity for vegetation (Roy et al., 2014, Dube and Mutanga, 2015). This facilitated the delineation of each land cover type of the forest, thus enhancing the spatiotemporal analysis.

The SVM algorithm was also robust for the image classification as it was instrumental in enhancing the ability of the spectral bands in detecting each land cover type. The delineation was enhanced by the design and inherent capabilities of the SVM algorithm, as also reported in other studies (Cervantes et al., 2020, Wang et al., 2017, Thanh Noi and Kappas, 2018). As a non-parametric algorithm, it does not assume normality in distribution which makes it suitable to forest land cover classification as natural forest ecosystems are not normally distributed. The other capabilities include robust to high dimensional data and noise, fast prediction due to support vectors, and require less training data sets (Fassnacht et al., 2016). As a result of the effective classification carried out by the SVM, the MLPNN and MCM combination (Addae and Oppelt, 2019, Ranagalage et al., 2019), was able to spatially detect and map each marginal change and the inter-transitions among cover types at decadal intervals with good accuracy. It also enhanced the prediction of the future land cover distribution of the forest. Thus, the hybrid methodological approach was vital and can be replicated in similar studies.

The study found that the Nkandla Forest Reserve experienced changes through either increase or a decrease in the extent of each land cover type, resulting in the spatial and temporal changes that occurred over the 30 years (1989-2019). Spatially, the changes are not localized, but it spreads across the forest and mostly occurred in the middle, southern, southeastern, southwestern, and towards the northeastern part of the forest reserve. The direction of change reflected the inter-transitions between the land cover types. Generally, it has been indicated that forest cover changes are usually caused by natural (ecological) and anthropogenic factors (Deb et al., 2018, Fragal et al., 2016). This was evident for the forest reserve as both ecological and anthropogenic factors were observed to be driving the changes that occurred.

Ecological processes such as regeneration, tree stem growth, and mortality contributed to the high levels of gains and losses manifested in the inter-transitions among the cover types. The natural sub-tropical forest just like tropical forests undergoes complex ecological processes over time, which leads to changes within the tree cover. These forests are characterized by high diversity, high productive systems and net primary productivity [NPP] (Malhi et al., 2004), which lead to rapid tree growth. The NPP is a major moderator of carbon cycles and other ecological processes (Zhang et al., 2016). This may explain what was observed in most parts of the open canopy forest, where large areas transitioned into closed canopy forest consistently at each decadal interval. It is a confirmation of studies that suggest that increased growth and regeneration of tree species diversity occur in gaps within open canopy areas of forests (He et al., 2015).

Observations also indicate that some ecological processes are enhanced within the gap since sunlight can get to the forest floor (Muscolo et al., 2014, He et al., 2019). The process leads to high NPP resulting in increased stem recruitment rates and stand-level biomass growth (Zhang et al., 2016). It is, therefore, possible that most of these gap areas with young trees might have undergone such rapid growth to form continuous and interlocking canopy layers. More open canopy areas might have consistently experienced this over the 30 years. This accounts for the consistently increased coverage of the closed canopy forest. Such forest cover changes have positive impacts on ecosystem functioning and biodiversity interactions. The capacity of the forest is thus enhanced to provide ecosystem services such as carbon sequestration, enhanced tree species diversity, wildlife conservation, and habitat protection. If these processes are not interfered with through anthropogenic factors, then the forest will be positioned to continually perform its productive and protective functions effectively.

As the forest trees grow, mortality may increase through wind-throw, disease, and over maturity accounting for losses within the tree cover (Neumann et al., 2017). Tree mortality affects species diversity, forest structure, growing stock, yield, rate of growth, nutrient and carbon cycling (Fontes et al., 2018, MacLean, 2016). The transitioning of some parts of the closed canopy to open canopy could be attributed to these ecological processes. Mortality increases the level of gap opening, breaking the continuous canopy structure of the forest. This may disrupt the ecological cycles thereby affecting ecosystem services such as carbon sequestration and tree species diversity. The ecological processes, therefore, influenced the level of changes and are likely to further influence future forest cover distribution.

Human-induced factors also contributed to the transitioning of closed canopy forest to the open canopy forest or the open canopy to grassland. Uncontrolled cattle grazing and extraction of NTFPs by fringing communities of the Nkandla Forest Reserve could be contributing factors to these observed changes. Also, wildfires caused by fringe communities through crop and animal farms close to the boundaries affected areas of tree cover. It must be noted that not all human disturbances lead to fragmentation of forest or mass habitat loss, but they do have detrimental effects on the forest ecosystem and biodiversity (Peres et al., 2006). Subtle but permanent extraction of fractions of forest products is experienced through harvesting of firewood and NTFPs, hunting, and domestic grazing animals (Martorell and Peters, 2005, Boucher et al., 2011). The effects may include alteration in seed dispersal at forest sites that experience domestic animals,

firewood extraction, and hunting activities (Leal et al., 2014). These chronic disturbances cause forest degradation on smaller scales across forests affecting biological integrity. The disturbances interrupt and modify the ecological cycles leading to reduced regeneration and seedling growth. Human-induced modification of forest ecosystems has hence been identified to reshape and impact the systems significantly (Ribeiro et al., 2015, Boivin et al., 2016), which was evident for the forest.

The chronic human disturbances also account for the increased coverage of the grassland especially the taken over of parts of the closed and open canopy forests. Naturally, grasslands are prone to wildfires (Snyder et al., 2006, Davies et al., 2015), so fires caused by fringe communities around the boundaries spread and destroys tree cover. It is also possible that the felling of poles for domestic purposes near the boundaries might have opened these areas for grass to invade. Grassland areas are also susceptible to weeds and alien species (Chambers et al., 2007, Davies et al., 2011a), which can easily take over tree cover areas and lead to tree seedling and sapling mortality. Already some invasive species have been recorded in parts of the grassland (Ezemvelo KZN Wildlife, 2015b), and it could be a reason for the grassland gaining some more coverage in the closed and open canopy forests. These confirm that chronic human disturbances have detrimental effects on forest landscapes, and it will be important for forest managers to device measures and initiatives that could help curb such disturbances and ensure that functional abilities of forests are not compromised.

On the other hand, the extent of the bare sites remained low although it experienced some gain and losses in some areas. However, it took over some parts with tree cover so it will be important to check such bare sites and avoid its spread. The grassland took over some parts of bare sites, which is good in ecological terms as it will help to reduce erosion in such sites.

Historical mapping and spatiotemporal analysis have aided the simulation and prediction of possible future land cover (Addae and Oppelt, 2019, Vázquez-Quintero et al., 2016, Voight et al., 2019). The outcomes of these projections are important as they can inform forest management strategies and policymaking. The prediction of the future land cover distribution for the Nkandla Forest Reserve for 2029 revealed that the closed canopy forest will decline in coverage by about 79 ha of the total extent of 2019. The predicted loss is likely to happen in the northeastern, southeastern, and eastern parts of the forest. These areas are close to the boundaries where fringe communities are expanding with associated increased crop production, animal farming, and frequent fires. On the other hand, the open canopy areas will gain about 10 ha more by 2029, while the grassland will gain about 68 ha more by 2029. The increase in the extent of the grassland and open canopy increasing in extent will be due to the taking over parts of the closed canopy areas. The predicted cover loss for the closed and open canopy forests may be caused by the same process of ecological and anthropogenic factors. Chen et al. (2013) similarly found land cover types losing parts of their extent in predictions, although they might have experienced an increase in the earlier years. These predictions provide a scenario that should arouse the need for forest managers to put in measures like increased monitoring and surveillance in and around the forest.

Generally, forest landscapes may seem intact and unperturbed due to their tree cover. However, there can be subtle or small-scale changes that may go unnoticed and over time affect large areas. This leads to forest degradation and fragmentation and reducing its functionality. Forest managers must develop initiatives including financial commitments towards periodic monitoring. The monitoring will ensure early detection of changes and provide an understanding of the responses of the forest to the changes as well as potential impacts on its biodiversity (McGeoch et al., 2011, Daume et al., 2014). It will enable the management to initiate effective strategies for maintaining forest ecosystems for the continual supply of goods and services. The use of Remote Sensing technology and methods are recommended to be adopted to support forest monitoring.

There could also be the engagement of the fringing communities to promote collaborative forest management and conservation. Since the forests play a socio-ecological, cultural, and economic role at the local and national levels, engaging and including communities in conservation and monitoring initiatives could help reduce the immediate and possible future threat to forests. A reduction in chronic human disturbances will reduce the degradation and fragmentation of the forest reserve.

4.9. Conclusion

The spatial and temporal analysis revealed the land cover changes that happened in the Nkandla Forest Reserve between 1989 and 2019 at decadal intervals. These changes were manifested in the inter-transitions among the four land cover types especially between the closed and open canopy forests. However, the closed canopy had the most gains over the period which is positive as it improves the provision of ecosystem goods and services due to an increase in tree cover. These gains could be attributed to the ecological process and a non-interference of these processes will enhance the ability of the forest to continually provide ecosystem services. The chronic human disturbances are also drivers for such changes and have negative effects on the forest reserve and as such they must be addressed. The prediction of the 2029 land cover distribution of the forest revealed that there will be a decline in the coverage of the closed canopy forest, while the open canopy will increase marginally. The decline in the extent of the closed canopy forest is likely to be caused by grassland taking over some of its areas. Under this, the grassland is likely to gain some more areas amongst the cover types.

The quantitative and qualitative spatial and temporal insights provided through this study will be vital for sustainable forest management with regards to planning and developing effective forest monitoring and conservation strategies and initiatives. It is important to note that the effect of forest cover changes is not only felt locally or nationally, but it has a global impact ultimately. It is therefore recommended that approaches such as the use of Remote Sensing technology, which made this research possible, and non-technological strategies such as engaging and involving forest fringe communities could be adopted to improve monitoring and conservation of forest ecosystems. It should be done in a bid to ensuring that dwindling natural forests cover are curbed to secure and save these forest resources of local, national, and global significance for the continual provision of ecosystem goods and services.

Acknowledgment

The authors appreciate the support provided by Ms. Sharon Louw (Ecologist) of Ezemvelo KwaZulu-Natal Wildlife (EKZNW) in securing a permit to conduct the study in the Nkandla Forest Reserve. We thank Mr. Elliackim Zungu (Conservation Manager), Mr. Simon Makhaye, Mr. Sandebulawa Biyela and other supporting staff of EKZNW for assisting in field data collection. The authors also thank Mr. Charles Zungunde (University of KwaZulu-Natal) for driving the field team across the forest as well as assisting in data collection. We further thank the journal editors and reviewers for their efforts, time and insights.

Funding

This research and study received a block grant from the South African System Analysis Centre (SASAC) through the National Research Foundation (NRF) of South Africa, with grant number 118770. The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest

References

- ADDAE, B. & OPPELT, N. 2019. Land-Use/Land-Cover Change Analysis and Urban Growth Modelling in the Greater Accra Metropolitan Area (GAMA), Ghana. *Urban Science*, 3.
- AGRAWAL, A., CASHORE, B., HARDIN, R., SHEPHERD, G., BENSON, C. & MILLER, D. 2013. Economic contributions of forests. *Background Paper*, 1.
- AGUEJDAD, R., HOUET, T., HUBERT-MOY, L. & ASSESSMENT 2017. Spatial validation of land use change models using multiple assessment techniques: A case study of transition potential models. *Environmental Modeling*, 22, 591-606.
- AHAMMAD, R., STACEY, N. & SUNDERLAND, T. C. 2019. Use and perceived importance of forest ecosystem services in rural livelihoods of Chittagong Hill Tracts, Bangladesh. *Ecosystem services*, 35, 87-98.
- ARROYO-RODRIGUEZ, V., FAHRIG, L., TABARELLI, M., WATLING, J. I., TISCHENDORF, L., BENCHIMOL, M., CAZETTA, E., FARIA, D., LEAL, I. R., MELO, F. P. L., MORANTE-FILHO, J. C., SANTOS, B. A., ARASAGISBERT, R., ARCE-PENA, N., CERVANTES-LOPEZ, M. J., CUDNEY-VALENZUELA, S., GALAN-ACEDO, C., SAN-JOSE, M., VIEIRA, I. C. G., SLIK, J. W. F., NOWAKOWSKI, A. J. & TSCHARNTKE, T. 2020. Designing optimal human-modified landscapes for forest biodiversity conservation. *Ecol Lett*, 23, 1404-1420.
- BARLOW, J., LENNOX, G. D., FERREIRA, J., BERENGUER, E., LEES, A. C., MACNALLY, R., THOMSON, J. R., DE BARROS FERRAZ, S. F., LOUZADA, J. & OLIVEIRA, V. H. F. 2016. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature*, 535, 144-147.
- BOEDHIHARTONO, A. 2017. Can Community Forests Be Compatible With Biodiversity Conservation in Indonesia? *Land*, 6.

- BOIVIN, N. L., ZEDER, M. A., FULLER, D. Q., CROWTHER, A., LARSON, G., ERLANDSON, J. M., DENHAM, T. & PETRAGLIA, M. D. 2016. Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Sciences*, 113, 6388-6396.
- BOUCHER, D., ELIAS, P., LININGER, K., MAY-TOBIN, C., ROQUEMORE, S. & SAXON, E. 2011. The root of the problem: what's driving tropical deforestation today? *The root of the problem: what's driving tropical deforestation today?*
- CERVANTES, J., GARCIA-LAMONT, F., RODRÍGUEZ-MAZAHUA, L. & LOPEZ, A. 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
- CHAMBERS, J. C., ROUNDY, B. A., BLANK, R. R., MEYER, S. E. & WHITTAKER, A. 2007. What makes Great Basin sagebrush ecosystems invasible by *Bromus tectorum*? *Ecological Monographs*, 77, 117-145.
- CHAUDHARY, S., CHETTRI, N., UDDIN, K., KHATRI, T. B., DHAKAL, M., BAJRACHARYA, B. & NING, W. J. E. C. 2016. Implications of land cover change on ecosystems services and people's dependency: A case study from the Koshi Tappu Wildlife Reserve, Nepal. *Ecological Complexity*, 28, 200-211.
- CHEN, C.-F., SON, N.-T., CHANG, N.-B., CHEN, C.-R., CHANG, L.-Y., VALDEZ, M., CENTENO, G., THOMPSON, C. & ACEITUNO, J. 2013. Multi-Decadal Mangrove Forest Change Detection and Prediction in Honduras, Central America, with Landsat Imagery and a Markov Chain Model. *Remote Sensing*, 5, 6408-6426.
- DA PONTE, E., ROCH, M., LEINENKUGEL, P., DECH, S. & KUENZER, C. 2017. Paraguay's Atlantic Forest cover loss—Satellite-based change detection and fragmentation analysis between 2003 and 2013. *Applied Geography*, 79, 37-49.
- DAFF 2015. State of the forests report 2010-2012 report, Department Agriculture, Forestry and Fisheries, Republic of South Africa. *Unpublished*, 88.
- DAUME, S., ALBERT, M. & VON GADOW, K. 2014. Forest monitoring and social media – Complementary data sources for ecosystem surveillance? *Forest Ecology and Management*, 316, 9-20.
- DAVIES, K. W., BOYD, C. S., BATES, J. D. & HULET, A. 2015. Dormant season grazing may decrease wildfire probability by increasing fuel moisture and reducing fuel amount and continuity. *International Journal of Wildland Fire*, 24, 849-856.
- DAVIES, K. W., BOYD, C. S., BECK, J. L., BATES, J. D., SVEJCAR, T. J. & GREGG, M. A. 2011. Saving the sagebrush sea: an ecosystem conservation plan for big sagebrush plant communities. *Biological Conservation*, 144, 2573-2584.
- DEB, S., DEBNATH, M. K., CHAKRABORTY, S., WEINDORF, D. C., KUMAR, D., DEB, D. & CHOUDHURY, A. 2018. Anthropogenic impacts on forest land use and land cover change: modelling future possibilities in the Himalayan Terai. *Anthropocene*, 21, 32-41.
- DUBE, T. & MUTANGA, O. 2015. 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.
- EASTMAN, J. R. 2015. TerrSet tutorial. *J Clark Labs Clark University, Worcester, MA*.
- EZEMVELO KZN WILDLIFE 2015. Nkandla Forest Complex: Protected Area Management Plan. *Unpublished Version 1.0*, 173
- FASSNACHT, F. E., LATIFI, H., STEREŃCZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of

- studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- FOKENG, R. M., FORJE, W. G., MELI, V. M. & BODZEMO, B. N. 2020. Multi-Temporal Forest Cover Change Detection in the Metchie-Ngoum Protection Forest Reserve, West Region of Cameroon. *The Egyptian Journal of Remote Sensing Space Science*, 23, 113-124.
- FONTES, C. G., CHAMBERS, J. Q. & HIGUCHI, N. 2018. Revealing the causes and temporal distribution of tree mortality in Central Amazonia. *Forest Ecology and Management*, 424, 177-183.
- FRAGAL, E. H., SILVA, T. S. F. & NOVO, E. M. L. D. M. 2016. Reconstructing historical forest cover change in the Lower Amazon floodplains using the LandTrendr algorithm. *Acta Amazonica*, 46, 13-24.
- GILROY, J. J., WOODCOCK, P., EDWARDS, F. A., WHEELER, C., BAPTISTE, B. L., URIBE, C. A. M., HAUGAASEN, T. & EDWARDS, D. P. 2014. Cheap carbon and biodiversity co-benefits from forest regeneration in a hotspot of endemism. *Nature Climate Change*, 4, 503.
- GÓMEZ, C., WHITE, J. C. & WULDER, M. A. 2016. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 55-72.
- GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2020. Mapping natural forest cover using satellite imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, 18, 1-14.
- HANSEN, M. C. & LOVELAND, T. R. 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122, 66-74.
- HE, Z., LIU, J., SU, S., ZHENG, S., XU, D., WU, Z., HONG, W. & WANG, J. L.-M. 2015. Effects of forest gaps on soil properties in *Castanopsis kawakamii* nature forest. *PLoS one*, 10, e0141203.
- HE, Z., WANG, L., JIANG, L., WANG, Z., LIU, J., XU, D. & HONG, W. 2019. Effect of microenvironment on species distribution patterns in the regeneration layer of forest gaps and non-gaps in a subtropical natural forest, China. *Forests*, 10, 90.
- KAMUSOKO, C., ANIYA, M., ADI, B. & MANJORO, M. 2009. Rural sustainability under threat in Zimbabwe—simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. *Applied Geography*, 29, 435-447.
- KHATAMI, R., MOUNTRAKIS, G. & STEHMAN, S. V. 2016. A meta-analysis of Remote Sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, 177, 89-100.
- KUMAR, M., SINGH, M. P., SINGH, H., DHAKATE, P. M. & RAVINDRANATH, N. H. 2019. Forest working plan for the sustainable management of forest and biodiversity in India. *Journal of Sustainable Forestry*, 39, 1-22.
- LEAL, L. C., ANDERSEN, A. N. & LEAL, I. R. 2014. Anthropogenic disturbance reduces seed-dispersal services for myrmecochorous plants in the Brazilian Caatinga. *Oecologia*, 174, 173-181.
- LOHBECK, M., POORTER, L., MARTÍNEZ-RAMOS, M. & BONGERS, F. 2015. Biomass is the main driver of changes in ecosystem process rates during tropical forest succession. *Ecology*, 96, 1242-1252.
- LU, D., LI, G. & MORAN, E. 2014. Current situation and needs of change detection techniques. *International Journal of Image and Data Fusion*, 5, 13-38.

- MACLEAN, D. A. 2016. Impacts of insect outbreaks on tree mortality, productivity, and stand development. *The Canadian Entomologist*, 148, S138-S159.
- MAITHANI, S. 2015. Neural networks-based simulation of land cover scenarios in Doon valley, India. *Geocarto International*, 30, 163-185.
- MALHI, Y., BAKER, T. R., PHILLIPS, O. L., ALMEIDA, S., ALVAREZ, E., ARROYO, L., CHAVE, J., CZIMCZIK, C. I., FIORE, A. D. & HIGUCHI, N. 2004. The above-ground coarse wood productivity of 104 Neotropical forest plots. *Global Change Biology*, 10, 563-591.
- MARTORELL, C. & PETERS, E. M. 2005. The measurement of chronic disturbance and its effects on the threatened cactus *Mammillaria pectinifera*. *Biological Conservation*, 124, 199-207.
- MCGEOCH, M. A., DOPOLO, M., NOVELLIE, P., HENDRIKS, H., FREITAG, S., FERREIRA, S., GRANT, R., KRUGER, J., BEZUIDENHOUT, H. & RANDALL, R. M. 2011. A strategic framework for biodiversity monitoring in South African National Parks: essay. *Koedoe: African Protected Area Conservation and Science*, 53, 1-10.
- MCROBERTS, R. E. & WESTFALL, J. A. 2016. Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73, 625-633.
- MIHAI, B., SĂVULESCU, I., RUJOIU-MARE, M. & NISTOR, C. 2017. Recent forest cover changes (2002–2015) in the Southern Carpathians: A case study of the Iezer Mountains, Romania. *Science of the Total Environment*, 599, 2166-2174.
- MILLARD, K. & RICHARDSON, M. 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, 7, 8489-8515.
- MISHRA, V. N. & RAI, P. K. 2016. A Remote Sensing aided multi-layer perceptron-Markov chain analysis for land use and land cover change prediction in Patna district (Bihar), India. *Arabian Journal of Geosciences*, 9.
- MUSCOLO, A., BAGNATO, S., SIDARI, M. & MERCURIO, R. 2014. A review of the roles of forest canopy gaps. *Journal of Forestry Research*, 25, 725-736.
- NEUMANN, M., MUES, V., MORENO, A., HASENAUER, H. & SEIDL, R. 2017. Climate variability drives recent tree mortality in Europe. *Global Change Biology*, 23, 4788-4797.
- NGWIRA, S. & WATANABE, T. 2019. An analysis of the causes of deforestation in Malawi: a case of Mwazisi. *Land*, 8, 48.
- PAL, S. & ZIAUL, S. 2017. Detection of land use and land cover change and land surface temperature in English Bazar urban centre. *The Egyptian Journal of Remote Sensing and Space Science*, 20, 125-145.
- PERES, C. A., BARLOW, J. & LAURANCE, W. F. 2006. Detecting anthropogenic disturbance in tropical forests. *Trends Ecol Evol*, 21, 227-9.
- PHIRI, D. & MORGENROTH, J. 2017. Developments in Landsat land cover classification methods: A review. *Remote Sensing*, 9, 967.
- PONTIUS JR, R. G. 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, 68, 1041-1050.
- PONTIUS JR, R. G. & SCHNEIDER, L. C. 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, ecosystems & environment*, 85, 239-248.
- RANAGALAGE, M., WANG, R., GUNARATHNA, M. H. J. P., DISSANAYAKE, D. M. S. L. B., MURAYAMA, Y. & SIMWANDA, M. 2019. Spatial Forecasting of

- the Landscape in Rapidly Urbanizing Hill Stations of South Asia: A Case Study of Nuwara Eliya, Sri Lanka (1996–2037). *Remote Sensing*, 11.
- RAWAT, J. S. & KUMAR, M. 2015. Monitoring land use/cover change using Remote Sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science*, 18, 77-84.
- REDDY, C. S., PASHA, S. V., SATISH, K., SARANYA, K., JHA, C. & MURTHY, Y. K. 2018. Quantifying nationwide land cover and historical changes in forests of Nepal (1930–2014): implications on forest fragmentation. *Biodiversity and Conservation*, 27, 91-107.
- RIBEIRO, E. A. M., ARROYO-RODRÍGUEZ, V., SANTOS, B. A., TABARELLI, M. & LEAL, I. R. 2015. Chronic anthropogenic disturbance drives the biological impoverishment of the Brazilian Caatinga vegetation. *Journal of Applied Ecology*, 52, 611-620.
- RIMAL, B., ZHANG, L., KESHTKAR, H., HAACK, B. N., RIJAL, S. & ZHANG, P. 2018. Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information*, 7, 154.
- ROUSTA, I., SARIF, M. O., GUPTA, R. D., OLAFSSON, H., RANAGALAGE, M., MURAYAMA, Y., ZHANG, H. & MUSHORE, T. D. 2018. Spatiotemporal analysis of land use/land cover and its effects on surface urban heat island using Landsat data: A case study of Metropolitan City Tehran (1988–2018). *Sustainability*, 10, 4433.
- ROY, D. P., WULDER, M. A., LOVELAND, T. R., C.E, W., ALLEN, R. G., ANDERSON, M. C., HELDER, D., IRONS, J. R., JOHNSON, D. M., KENNEDY, R., SCAMBOS, T. A., SCHAAF, C. B., SCHOTT, J. R., SHENG, Y., VERMOTE, E. F., BELWARD, A. S., BINDSCHADLER, R., COHEN, W. B., GAO, F., HIPPLE, J. D., HOSTERT, P., HUNTINGTON, J., JUSTICE, C. O., KILIC, A., KOVALSKYY, V., LEE, Z. P., LYMBURNER, L., MASEK, J. G., MCCORKEL, J., SHUAI, Y., TREZZA, R., VOGELMANN, J., WYNNE, R. H. & ZHU, Z. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172.
- SANG, L., ZHANG, C., YANG, J., ZHU, D. & YUN, W. 2011. Simulation of land use spatial pattern of towns and villages based on CA–Markov model. *Mathematical and Computer Modelling*, 54, 938-943.
- SHARMA, R., RIMAL, B., BARAL, H., NEHREN, U., PAUDYAL, K., SHARMA, S., RIJAL, S., RANPAL, S., ACHARYA, R. P. & ALENAZY, A. A. 2019. Impact of land cover change on ecosystem services in a tropical forested landscape. *Resources*, 8, 18.
- SNYDER, R. L., SPANO, D., DUCE, P., BALDOCCHI, D. & XU, L. 2006. A fuel dryness index for grassland fire-danger assessment. *Agricultural and Forest Meteorology*, 139, 1-11.
- TEAM, R. D. C. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>.
- TEWKESBURY, A. P., COMBER, A. J., TATE, N. J., LAMB, A. & FISHER, P. F. 2015. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, 1-14.
- THANH NOI, P. & KAPPAS, M. J. S. 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. 18, 18.

- TUGUME, P., KAKUDIDI, E. K., BUYINZA, M., NAMAALWA, J., KAMATENESI, M., MUCUNGUZI, P. & KALEMA, J. 2016. Ethnobotanical survey of medicinal plant species used by communities around Mabira Central Forest Reserve, Uganda. *Journal of Ethnobiology Ethnomedicine*, 12, 1-28.
- VÁZQUEZ-QUINTERO, G., SOLÍS-MORENO, R., POMPA-GARCÍA, M., VILLARREAL-GUERRERO, F., PINEDO-ALVAREZ, C. & PINEDO-ALVAREZ, A. 2016. Detection and Projection of Forest Changes by Using the Markov Chain Model and Cellular Automata. *Sustainability*, 8.
- VITTEK, M., BRINK, A., DONNAY, F., SIMONETTI, D. & DESCLÉE, B. 2014. Land Cover Change Monitoring Using Landsat MSS/TM Satellite Image Data over West Africa between 1975 and 1990. *Remote Sensing*, 6, 658-676.
- VOIGHT, C., HERNANDEZ-AGUILAR, K., GARCIA, C. & GUTIERREZ, S. 2019. Predictive Modeling of Future Forest Cover Change Patterns in Southern Belize. *Remote Sensing*, 11.
- WANG, M., WAN, Y., YE, Z. & LAI, X. 2017. Remote Sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. *Information Sciences*, 402, 50-68.
- WANG, S., ZHENG, X. & ZANG, X. 2012. Accuracy assessments of land use change simulation based on Markov-cellular automata model. *Procedia Environmental Sciences*, 13, 1238-1245.
- WESSELS, K., VAN DEN BERGH, F., ROY, D., SALMON, B., STEENKAMP, K., MACALISTER, B., SWANEPOEL, D. & JEWITT, D. 2016. Rapid Land Cover Map Updates Using Change Detection and Robust Random Forest Classifiers. *Remote Sensing*, 8.
- WU, H., LI, Z., CLARKE, K. C., SHI, W., FANG, L., LIN, A. & ZHOU, J. 2019. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science*, 33, 1040-1061.
- YANG, X., ZHENG, X.-Q. & LV, L.-N. 2012. A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecological Modelling*, 233, 11-19.
- ZHANG, L., XIAO, J., ZHOU, Y., ZHENG, Y., LI, J. & XIAO, H. 2016. Drought events and their effects on vegetation productivity in China. *Ecosphere*, 7, e01591.
- ZHEN, Z., QUACKENBUSH, L. J., STEHMAN, S. V. & ZHANG, L. 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *International Journal of Remote Sensing*, 34, 6914-6930.
- ZHU, X., LIU, W., JIANG, X. J., WANG, P. & LI, W. 2018. Effects of land-use changes on runoff and sediment yield: Implications for soil conservation and forest management in Xishuangbanna, Southwest China. *Land Degradation and Development*, 29, 2962-2974.

Appendix 4.1

Table 4.5: General land cover changes of the forest reserve between 1989 and 2019.

2019					
1989	Closed Canopy	Open Canopy	Grassland	Bare Sites	Grand Total
Closed Canopy	795.71	87.48	0.27	0.00	883.46
Open Canopy	262.98	792.24	33.03	3.74	1091.99
Grassland	0.54	27.22	179.06	12.14	218.95
Bare Sites	0.00	3.66	14.19	5.09	22.92
Grand Total	1059.23	910.59	226.55	20.98	
Net Change	175.77	-181.40	7.6	-1.94	
Annual Change	5.86	-6.04	0.25	-0.06	
Annual Change rate (%)	0.61	-60	0.11	-0.29	

Note: The table presents the extent of each land cover type of the forest that persisted between the two periods as well as the area that was either gained from or lost to the other land cover types over the same period.

Appendix 4.2

Table 4.6: Land Cover Transition Matrix of the Forest Reserve for 1989-1999.

		1999			
1989	Closed Canopy	Open Canopy	Grassland	Bare Sites	Grand Total
Closed Canopy	632.80	237.62	13.03	0.00	883.46
Open Canopy	302.88	715.11	73.28	0.71	1091.99
Grassland	0.36	27.22	182.46	8.93	218.97
Bare Sites	0.09	0.18	16.70	5.98	22.94
Grand Total	936.13	980.14	285.47	15.62	2217.36
Net Change	52.66	-111.85	65.50	-7.32	
Annual Change	5.26	-11.19	6.65	-0.73	
Annual Change rate (%)	0.59	-1.07	2.69	-3.75	

Note: The table presents the extent of each land cover type of the forest that persisted between the two periods as well as the area that was either gained from or lost to the other land cover types over the same period.

Appendix 4.3

Table 4.7: Land Cover Transition Matrix of the Forest Reserve for 1999-2009

		2009			
1999	Closed Canopy	Open Canopy	Grassland	Bare Sites	Grand Total
Closed Canopy	630.93	301.27	3.84	0.09	936.13
Open Canopy	319.22	610.31	50.17	0.44	980.14
Grassland	18.84	27.76	225.93	12.94	285.47
Bare Sites	0.00	0.18	8.66	6.78	15.62
Grand Total	968.99	939.52	288.60	20.25	2217.36
Net Change	32.86	-40.62	3.13	4.63	
Annual Change	3.27	-4.06	0.13	0.46	
Annual Change rate (%)	0.35	-0.42	0.11	2.63	

Note: The table presents the extent of each land cover type of the forest that persisted between the two periods as well as the area that was either gained from or lost to the other land cover types over the same period.

Appendix 4.4

Table 4.8: Land Cover Transition Matrix of the Forest Reserve for 2009-2019

	2019				
2009	Closed Canopy	Open Canopy	Grassland	Bare Sites	Grand Total
Closed Canopy	863.65	104.62	0.71	0.00	968.98
Open Canopy	189.78	741.80	7.76	0.17	939.50
Grassland	5.71	60.43	209.15	13.30	288.59
Bare Sites	0.09	3.75	8.93	7.49	20.23
Grand Total	1059.23	910.59	226.55	20.98	
Net Change	90.55	-28.91	-62.04	0.75	
Annual Change	9.06	-2.89	-6.20	0.075	
Annual Change rate (%)	0.90	-0.31	-2.39	1.28	

Note: The table presents the extent of each land cover type of the forest that persisted between the two periods as well as the area that was either gained from or lost to the other land cover types over the same period.

CHAPTER 5: MAPPING THE ABOVEGROUND CARBON STOCK OF SUB-TROPICAL NATURAL FOREST USING SENTINEL 2 SATELLITE IMAGERY AND RANDOM FOREST ALGORITHM

Abstract

Accurate estimates of aboveground carbon (AGC) stocks are important for climate policies and forest management. However, such information is lacking for many sub-tropical forests in African countries. Therefore, this study utilized Sentinel 2 imagery and Random Forest (RF) regression models to predict and map the AGC of the Nkandla forest in KwaZulu-Natal, South Africa. The study compared the performance of four RF models and ranked the spectral variables to ascertain their significance in AGC modelling. The models were made up of an equal number of preselected sets of Spectral bands, Near Infrared (NIR) vegetation indices, Red-edge vegetation indices, and Combined Variables that were extracted from Sentinel 2 imagery as predicting variables. The results showed that all the models had a comparable coefficient of determination (R^2) but varied root mean squared error (RMSE). The Combined Variables model outperformed the other models ($R^2 = 0.949$, RMSE = 20.65 Mg/ha). The red-edge vegetation indices model was the second best with the spectral bands model emerging as the third. The NIR vegetation indices model was the least performing model due to its high RMSE. The study demonstrates the capabilities of the freely available fine-scale Sentinel 2 imagery and RF algorithm for AGC prediction and mapping.

Keywords: Carbon, Natural Forest, Management, Vegetation, Random Forest, Remote Sensing.

5.1 Introduction

Accurate estimation of aboveground carbon is vital for carbon accounting, carbon emission monitoring, climate change policies and, land use and forest management (Asner and Mascaro, 2014). The living biomass of forests accounts for about 40% of total terrestrial carbon (Pan et al., 2011), and serves as a proxy for estimating aboveground carbon stocks (Vicharnakorn et al., 2014). Aboveground carbon (AGC) information is essential due to fluctuations in carbon stock trends as a result of dwindling forest ecosystems caused by deforestation, degradation, and fragmentation of natural forests. Adequate and reliable AGC stock estimation at the local, regional and global levels is therefore essential for improving climate change policies (Chave et al., 2014). Periodic measurements and assessment of the trends and dynamics of carbon stocks are means by which such credible quantitative and qualitative information can be obtained.

Different methods have been used for forest AGC measurements, monitoring and assessments, including field inventory, Remote Sensing technology, and a combination of both (Jucker et al., 2017). Field inventory methods are identified to be time-consuming, expensive, and cover a very limited area (Thurner et al., 2014). Nonetheless, it is the most accurate (Kumar and Mutanga, 2017). On the other hand, the Remote Sensing technology has wide areas of coverage (White et al., 2016), serves as a source of cost-effective temporal data (Bozkaya et al., 2015), making it more advantageous to apply for AGC mapping. Some studies have adopted the approach of combining Remote Sensing methods and field inventory data as a much robust approach since the two methods complement each other and utilize their good attributes for improved results (Mohd Zaki et al., 2018).

Many studies have utilized various Remote Sensing sensors for the prediction of AGC across different forest ecosystems and landscapes. The sensor type is mostly determined by their availability, ecosystem type and objective of study. In regards to active sensors, the light detection and ranging [LiDAR] (Saatchi et al., 2011b, Asner et al., 2011, Brillì et al., 2019)

and synthetic aperture radar [SAR] (Sarker et al., 2013, Qazi et al., 2017, Mitchard et al., 2009) have been used and have proven useful and robust. The areas of the landscapes for these studies include forest, savanna, woodlands, and tree plantations.

Multispectral satellite imagery has also been used widely for AGC prediction. For instance, Fuchs et al. (2009) applied Quickbird and archived ASTER data using k-nearest neighbour (k-NN) methods and linear regression to estimate AGC and determine its spatial variability in the Siberian tundra forest. The Quickbird outperformed the ASTER data (RMSE = 6.42 t/ha, RMSEr = 44%). In the ranking of the variables, the Gray-level co-occurrence matrix (GLCM) contrast kernel of 25×25 pixels of the Quickbird performed better than all the other variables. The application of MODIS and regression models were used to assess the AGB and carbon stocks of different vegetation types in India (Devagiri et al., 2013). Carbon stock of 3 megatonnes (Mt) was produced by the regression models with a corresponding R^2 of 0.81. The prediction was done with NDVI values of specific months and that of December produced the best results. In another study, Wang et al. (2009) utilized spectral bands, principal components and band ratio variables of Landsat TM in spatial uncertainty analysis of forest vegetation carbon. Variations and uncertainties were observed due to different plot sizes with the coefficient of correlation (r) varying between 0.401 to 0.405. The least performing variables were band ratio [(TM2 + TM3 + TM5)/ TM7]. Landsat 8 data was also employed in estimating the living biomass and carbon for the lowland miombo landscape in Tanzania (Gizachew et al., 2016). The NDVI was selected and used among other vegetation due to ease of application and interpretation. The applied linear model had a root mean squared error (RMSE) of 44 t/ha. The total living biomass was estimated to be 140 Mt while the total living carbon was estimated to be 47% of the biomass.

The recently launched Sentinel 2 satellite imagery by the European Satellite Agency (ESA), Sentinel 2, has also provided satisfactory results in various vegetation studies. Castillo et al. (2017) used the Sentinel 1 SAR and the Sentinel 2 imagery in the estimation of AGB mangrove forest in the Philippines. The SAR based models were more accurate with a correlation of 0.82 and 0.83 and a corresponding RMSE of 27.8 – 28.5 Mg/ha. The red-edge bands (RE1, RE2 and RE3) combination with elevation data were the best predictors while the Inverted Red-edge Chlorophyll Index (IRECI) produced the best accuracy among other vegetation indices. Chen et al. (2019) similarly, evaluated the Sentinel 1 SAR, Sentinel 2 imagery and SRTM digital elevation model (DEM) in AGB prediction. The performance of Geographically Weighted Regression (GWR), Support Vector Machine regression (SVMR), Stepwise Regression (SWR), Random Forest (RF) and Artificial Neural Network (ANN) and were also evaluated. The Sentinel series and the STRM DEM were the best performers among the variables while the RF produced the best accuracy (mean error = 1.39, mean absolute error = 25.48, $r = 0.9769$ and RMSE = 61.11 Mg/ha). The capacity of the Sentinel 2 to estimate AGB was evaluated for a community forest in Nepal (Pandit et al., 2018a). The effect of the number of input variables used in predicting models was also investigated. The model that combined all the spectral bands and vegetation indices variables provided better AGB estimates with a corresponding R^2 of 0.81 and RMSE of 25.57 t/ha. The red-edge bands were found to be important variables that contributed to the accuracy.

South Africa has a limited area of its landmass of natural forest ecosystems (DAFF, 2015). Considering, the limited area of such natural forests, it is important to monitor and estimates its aboveground carbon to contribute to the national carbon database. However, information on the AGC stocks of most natural forests is non-existent which affects carbon management. Also, changes and dynamics in carbon trends cannot be accounted for as there is no baseline information to be used for temporal carbon analysis. The Nkandla Forest Reserve is one of the main forests in the KwaZulu-Natal province, South Africa which also lacks such carbon stocks

information. Only a few studies that have been done on AGC and AGB (Dube and Mutanga, 2015, Dube and Mutanga, 2016), but it is not representative and adequate on the available AGC stocks for the natural forests and contribute to the natural carbon database. Therefore, our study aims to predict and map the AGC stocks of the Nkandla Forest Reserve using Sentinel 2 imagery and Random Forest (RF) regression. It also seeks to assess compare and assess the predictive ability of different spectral products of the Sentinel 2. The output of this study will provide AGC stock distribution information which will be vital for local carbon accounting and management. The provision of the information will be beneficial to local and national carbon accounting and monitoring. It will also help to identify robust and informative Sentinel 2 spectral products that will contribute to the performance of the RF models and accuracies. This will be of much benefit in knowing which spectral products to select for carbon modelling in either similar subtropical forests or that of other climatic zones.

5.2 Materials and Methods

5.2.1 Study Area

The Nkandla Forest Reserve was established and gazetted as a forest reserve in 1918, and 1992 respectively. It is an Afromontane sub-tropical forest in the KwaZulu-Natal (KZN) province of South Africa with a total area of 2,218 ha. It is located on 28° 43' 50.88" S and 30° 7' 9.84" E (Figure 5.1). The forest experiences a peak average temperature of 27°C between December and January, and the lowest average temperature of 2°C in the winter months of June and July (Ezemvelo KZN Wildlife, 2015b). It has a generally steep and undulating topography and an altitude of a minimum level of 500 m and exceeding 1,300m. It has four main land covers made up of closed canopy forest (1,059.23 ha), open canopy forest (910.60 ha), grassland (226.55 ha) and bare sites [20.97 ha] (Gyamfi-Ampadu et al., 2020). The grasslands are found on hilltops and downhills with some patches interspersing the areas with tree cover. Some areas around the forest boundaries experience frequent fires due to the activities of the forest fringe communities. The fringe communities graze their domestic animals such as cattle in the grasslands of the forest and obtain some non-timber forest products (NTFPs). An increase in commercial plantations and expanding communities possess some threat to the forest.

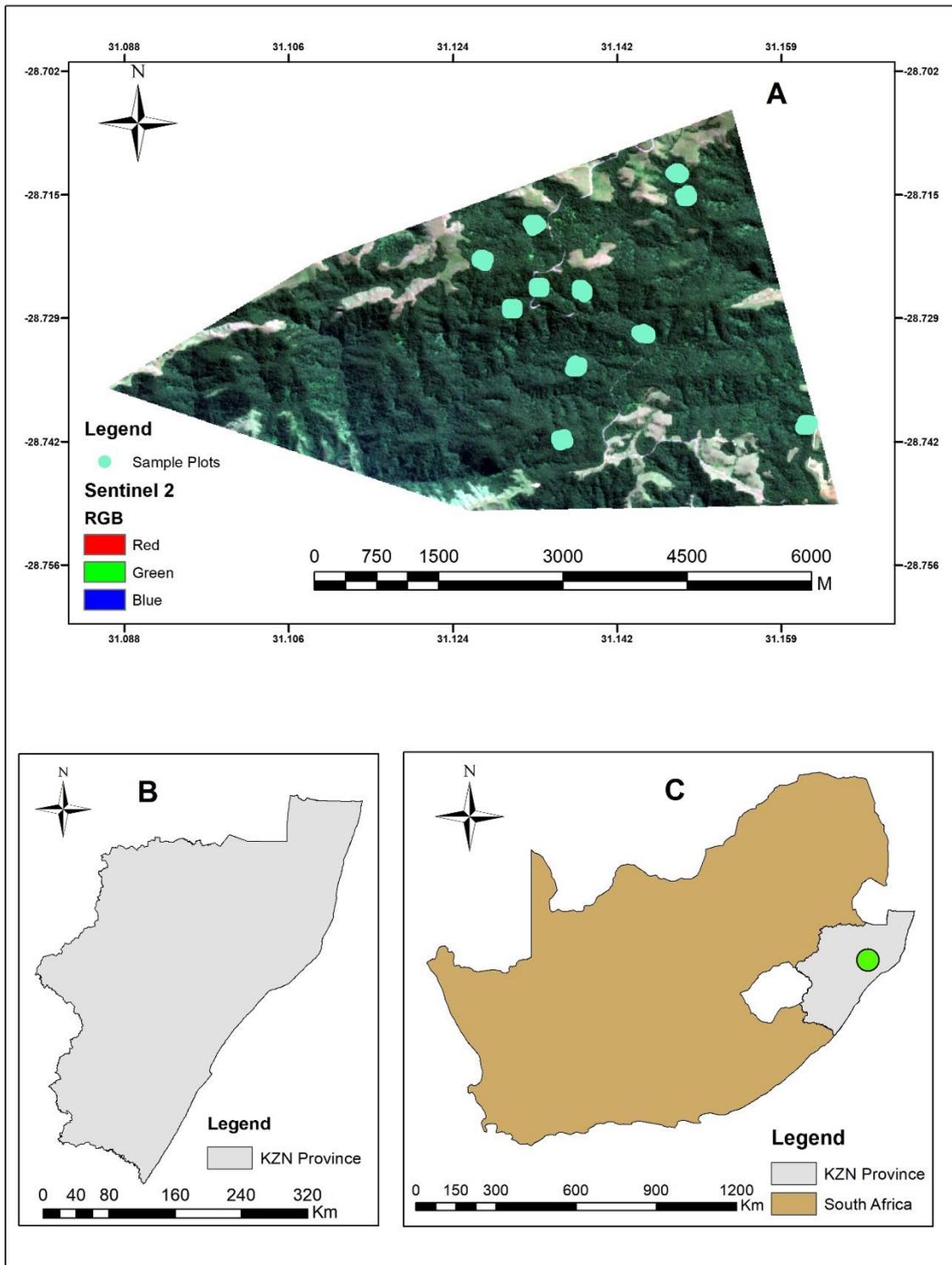


Figure 5.1: Map of the study area. Note: A is a map of the Nkandla Forest Reserve, B is a map of the KwaZulu-Natal (KZN) province and C is the map of South Africa indicating the location of KZN province and the Nkandla Forest Reserve.

5.3 Field inventory and AGC estimation

Field data were collected from 275 sampling plots (20 m x 20 m) that were set in the forest, between 24 April 2019 and 7 May 2019. We selected this plot size to enhance alignment with the spatial resolution of the Sentinel 2 imagery. Since AGC is derived from AGB estimations, the dimensions of living trees were measured and recorded with a diameter at breast height (DBH) criteria of ≥ 5 cm. The information recorded for each tree was the DBH, height, species name (local and scientific), and the GPS coordinates. The DBH was measured with a standard diameter tape (in cm), and the heights with a Vertex Hypsometer.

The AGB of every single tree was calculated by using a generic allometric equation developed by Chave et al. (2014) as presented below:

$$AGB_{tree} = 0.0673 \times (WD \times DBH^2 \times H)^{0.973} \quad (5.1)$$

Where; AGB_{tree} is the AGB of the individual tree species, $WD(g/cm^3)$ is the wood density of the tree, DBH (cm) is the diameter at breast height of the tree and H is the total height of the tree.

The wood densities of trees were obtained from the databases for pantropical global wood densities (Chave et al., 2009, Zanne et al., 2009) and the African wood density (Carsan et al., 2012), based on their availability. This standard equation and species wood densities were used because it was developed for common trees and has worked well in many studies.

After determining the AGB, the standard carbon fraction of 0.50 was applied to derive the AGC (IPCC, 2006). The computed individual tree AGC was then aggregated to generate the plot level AGC and used as the dependent variable for the Random Forest regression models.

5.3.1 Field data statistics

The descriptive statistics of the field plot level AGC presented in Table 5.1, illustrate that about 63% of the trees are with DBH of between 5 cm to 20 cm, and 27% between 20 cm to 40 cm. While the rest are big trees (above 40 cm) represented by 10%.

Table 5.1: Descriptive statistics of field measured data.

Parameter	DBH (cm)	Height (m)	AGC (Mg/ha)
Mean	19.24	12.11	2.20
Minimum	5	5.5	0.11
Maximum	113	32.5	206.45
Standard Deviation	13.51	5.07	6.54

5.4 Remote Sensing imagery data

The Remote Sensing imagery data used in the study were spectral bands and spectral indices generated from the near infrared (NIR) and red-edge regions of the electromagnetic spectrum.

5.4.1 Spectral bands variables

A cloud-free Sentinel 2 imagery which was captured on the 14th of April 2019 was downloaded from the USGS website (<https://earthexplorer.usgs.gov>). The Sentinel 2 imagery has 13 spectral bands ranging from the visible to the shortwave infrared regions of the electromagnetic spectrum. The image was preprocessed to optimal parameters to facilitates its use for the required analysis. The imagery was atmospherically corrected with the semi-automatic classification plugin (SCP) of the QGIS 3.10 software. The dark object subtraction (DOS1) SCP plugin was also applied to transform the image radiance to spectral reflectance. For this AGC prediction, ten spectral bands were selected (Table 5.2). The bands 1, 9, and 10 were not used as they contain aerosols, water vapour, and cloud information. The selected bands are sensitive to living biomass, carbon, and their respective resolutions are optimal for quantifying these elements (Wang et al., 2018a, Chen et al., 2018, Pandit et al., 2018a). The use of spectral bands that are sensitive to vegetation is important in vegetation modelling and predictions. Furthermore, the sensitivity of the spectral bands to vegetation enhances their correlation with the dependent variable that is used to extract their spectral reflectance values for the analysis. All the bands were resampled to 10 m spatial resolution using the nearest neighbourhood analysis in the SNAP toolbox to enhance stacking for improved analysis. Related AGB studies suggest that improving the spatial resolution of Sentinel 2 imagery to finer scales reduces uncertainties and improves accuracies (Attarchi and Gloaguen, 2014, luVaglio Laurin et al., 2017).

Table 5.2: Details of spectral bands

Band	Name/Code	Central Wavelength	Spatial Resolution (m)
B2	Blue (490 nm)	0.490	10
B3	Green (560 nm)	0.560	10
B4	Red (665 nm)	0.665	10
B5	RE1 (705 nm)	0.705	20
B6	RE2 (740 nm)	0.740	20
B7	RE3 (783 nm)	0.783	20
B8	NIR (842 nm)	0.842	10
B8A	NIR2 (865 nm)	0.865	20
B11	SWIR1 (1610 nm)	1.610	20
B12	SWIR 2 (2190 nm)	2.190	20

RE: Rede edge; NIR: near-infrared; SWIR: shortwave infrared

5.4.2 Spectral Indices variables

In this study, we made use of eight conventional NIR vegetation indices that have mostly been adopted in AGC and AGB studies. They were generated from the resampled and atmospherically corrected Sentinel 2 imagery with the appropriate mathematical expressions (Table 5.2). The use of preprocessed imagery is necessary to improve outputs of vegetation indices and to relate them to the image reflectance (Byrd et al., 2014, Zhang et al., 2014). These vegetation indices used in our study were useful in vegetation related studies (Wicaksono, 2017, Zhang et al., 2019, Askar et al., 2018) but yet to have a wide application in direct AGC prediction and mapping making it necessary to test their capabilities for AGC estimations.

Seven red-edge region vegetation indices were also generated from the Sentinel 2 imagery with the application of suitable band mathematical expressions (Table 5.2). They were also generated from the atmospherically corrected and resampled image. The Sentinel 2 generated red-edge vegetation indices have recently been applied and tested in vegetation studies (Lin et al., 2019, Xie et al., 2018, Zhang et al., 2018) but are yet to be specifically used as a predicting set of variables in AGC estimation. Therefore, it is worth applying them directly to the AGC prediction and ascertain their capabilities for the modelling as well.

Table 5.1: Details of NIR and Red-edge region vegetation indices variables

Predictor variable	Relevant band/index	Sentinel 2 Band Mathematical expressions	Resolution /Reference
NIR Vegetation Indices	EVI	$2.5 * (B8 - B4) / (B8 + 6 * B4 - 7.5 * B3 + 1)$	(Huete et al., 2002)
	DVI	$B8 - B4$	(Tucker, 1979)
	NDVI	$(B8 - B4) / (B8 + B4)$	(Rouse et al., 1974)
	GNDVI	$(B8 - B3) / (B8 - B3)$	(Buschmann and Nagel, 1993)
	RDVI	$(B8 - B4) / \sqrt{(B4 + B4)}$	(Roujean and Breon, 1995)
	GRVI	$B8 / B3$	(Sripada et al., 2005)
	SR	$B8 / B4$	(Birth and McVey, 1968)
Red-edge Vegetation Indices	MSR	$(B8 / B4) - 1 / \sqrt{(B8 / B4) + 1}$	(Chen, 1996)
	NDVIre1	$(B8 - B5) / (B8 + B5)$	(Gitelson and Merzlyak, 1994)
	NDVIre2	$(B8a - B5) / (B8a + B5)$	
	REI	$(RE2 - RE1) / (RE2 + RE1)$	(Zarco-Tejada et al., 2001)
	NDre1	$(B6 - B5) / (B6 + B5)$	(Gitelson and Merzlyak, 1994)
	NDre2	$(B7 - B5) / (B7 + B5)$	(Barnes et al., 2000)
	MSRre	$(B8 / B5) - 1 / \sqrt{(B8 / B5) + 1}$	(Chen, 1996)
MSRren	$(B8a / B5) - 1 / \sqrt{(B8a / B5) + 1}$	(Fernández-Manso et al., 2016)	

RE: Rede edge; NIR: near-infrared; SWIR: shortwave infrared, EVI: Enhance Vegetation Index, DVI: Difference Vegetation Index, NDVI: Normalised Difference Vegetation Index, RDVI: Renormalised Difference Vegetation Index, SR: Simple Ratio, GNDVI: Green Normalized Difference Vegetation Index, NDVIre1: Normalized Difference Vegetation Index red-edge 1, REI: Red-edge Index, NDre1: Normalized Difference red-edge 1, NDre2: Normalized Difference red-edge 2, MSRre: Modified Simple Ratio red-edge, MSRren: Modified Simple Ratio red-edge narrow.

5.4.3 Variable selection

The prediction of AGC requires the use of independent variables for estimating the AGC stocks. Inferences from some studies suggest that the number of predicting variables in a regression influences the output (Pandit et al., 2018a, Millard and Richardson, 2015, Mutowo et al., 2018b). Models with a larger number of variables usually have an advantage over those with a smaller set of variables. We compared the predictive abilities of four RF models with each having an equal number of variables. The variables were made of a set of spectral bands, NIR and red-edge vegetation indices. This was done to provide an equal ground for comparing the performance of each model. The best and significant variables were preselected from each variable set to reduce the possibility of redundancy and multicollinearity as a result of a high number of predictor variables.

We applied the Recursive Feature Elimination (RFE) algorithm to execute the selection process. The RFE algorithm was applied separately to each set of spectral variables using the RF function with 10-fold cross-validation in the R statistical package. The RFE algorithm selects variables that are important to significantly contribute to the model accuracy. This is done through four main steps which are 1) training of the RF model, 2) computation of the permutation importance measure, 3) elimination of the less relevant variables (features) and 4) repetition of the first 3 steps till no further variables (features) remains. The process is done by recomputing the permutation importance at each step of variable performance assessment and elimination. It selects a smaller size and more efficient variable subset since, the most informative variables are well ranked in the last stages of the steps of the backward procedure (Gregorutti et al., 2017). As a result of different training sets, it is likely to obtain slightly different variables at every iteration (Wang et al., 2018a). Therefore, the algorithm was implemented in 20 iterations for each set of variables to ensure the selection of credible, robust and informative ones (Li et al., 2016). At each iteration, the least number of variables with minimum was selected as the optimal value. The most frequently occurring variables were selected as the optimal ones.

At the end of the RFE implementation, the top five variables each were selected from the spectral bands ($n=10$), NIR vegetation indices ($n=8$) and Red-edge vegetation indices variables ($n=7$) to form four separate RF models. Each set of variables selected was used in a model corresponding to the spectral product type from which they were selected. The Blue, RE1, RE2, RE3 and NIR bands were selected from the set of spectral bands, the DVI, EVI, NDVI, MSR, SR were selected from the set of NIR vegetation indices while the REI, NDre1, MSRren, NDre2, NDVIre1 were selected from the red-edge vegetation indices. Afterwards, the original set of variables of the spectral bands and the vegetation indices were combined ($n=25$) and the top five were also selected to form the fourth model. This model was subsequently called the Combined Variables model. The variables selected from this model were RE1, RE2, SWIR2, NDVI and MSRren. Each set of the five selected variables were used in the separate RF models.

The breakdown of the spectral products used for the RF models into a specific set of spectral bands only, NIR vegetation indices only and red-edge vegetation indices only is important as it enhances a full assessment of the predictive abilities of each set when they are used alone in an AGC predicting model. There are many categories of spectral products with the advancement of Remote Sensing modelling of vegetation and some might likely be much more robust and informative than others. Hence, our approach to break them down into a specific set of products and used for the AGC prediction. This further enhanced the analysis of the contribution of each set of products to the accuracies of the predicting models. It also enhanced the testing of their performances and analyze their performance and contribution to modelling accuracies.

5.5 Random Forest Regression Modelling

We used Random Forest regression models to predict and map the AGC stock. The Random Forest (RF) is a non-parametric machine learning algorithm that is capable of undertaking both classification and regression. The algorithm uses a bagging approach to split the data; a part of the data is for training and building the decision tree. The remaining part is for estimating the out-of-bag (OOB) error for each tree. This result is an unbiased estimation of the generalization error. An advantage of the RF is that it does not overfit data, handle complex data and also able to deal with the problem of multicollinearity (Ramoelo et al., 2015, Abdel-Rahman et al., 2013a, Rodríguez-Galiano et al., 2011, Breiman, 2001). The RF *n*tree and *m*try are two main parameters that determine the outputs of the RF. The *n*tree is the total number of decision trees grown in the model with a default value of 500, while the *m*try is the number of predictor variables selected that performs the splitting at each node. The default values of these parameters produce optimum results (Belgiu and Drăguț, 2016, Duro et al., 2012). Besides, the RF allows the assessment of the statistical significance of each predicting variable in the model.

Four separate RF regression models were developed for the selected top five variables of the (a) Spectral bands, (b) NIR vegetation indices, (c) Red-edge vegetation indices and (d) Combined Variables to relate them with the field observed data. These predicting models were executed in the R statistical software environment (Team, 2013) through the “randomForest” package (Liaw and Wiener, 2002). The field observed data were extracted separately with each spectral variable set and subsequently partitioned into 70% training data (192 plots), and 30% independent validation data (83 plots) in a random selection approach. Each RF model was calibrated with the training data and then applied the bootstrapping of 500 iterations to predict the AGC using the independent 83 validation set for each of them. Bootstrapping is a resampling approach that selects samples randomly in a replacement manner and allows conclusions to be drawn from the samples instead of making assumptions about the estimator (Mooney et al., 1993). In the process, 192 plot data were drawn at each bootstrap iteration with replacement from 192 samples while the remaining 83 samples were used to validate the model outputs (Fassnacht et al., 2014c).

5.6 Models evaluation and variables importance

The predictive performance of the four RF regression models were compared and evaluated based on two main statistical parameters. These were the coefficient of determination (R^2) and the root mean squared error (RMSE) values. These accuracy parameters were calculated from the means of the 500 bootstrapped samples. The model that produced the highest R^2 and least RMSE was considered the most accurate. We produce a final AGC stock map of the studied forest from the best RF model.

The variable importance feature in the RF algorithm allowed us to evaluate the significance of each of the variables in the model accuracy. The significance of the variables relates to the increasing percentage in mean squared error (%IncMSE), which denotes the effect of a variable in a model when it is removed. We evaluated this, to determine which Sentinel 2 variables (products) could be considered as best in AGC prediction for natural forests

5.7 Results

5.7.1 Random Forest models performance

The performance of the four RF models was evaluated and compared based on the defined accuracy parameters (Table 5.4). Comparable result (R^2) has been observed between all the RF models, which may be due to the strength of the RF and the use of an equal number of predictor variables in the models. However, there was a marked difference in the RMSE. The Combined Variables model outperformed the other RF models with R^2 of 0.949 and RMSE of 20.65 Mg/ha. The Red-edge vegetation indices variables model was the second-best with R^2 of 0.945 and RMSE of 21.60 Mg/ha, the Spectral bands' variables model was the third-best by returning R^2 of 0.948 and RMSE of 21.83 Mg/ha. Lastly, the NIR vegetation indices variables model returned R^2 of 0.948 and RMSE of 22.94 Mg/ha, making it the least performing model.

Table 5.2: Model accuracies.

Predictor Variable	R^2	RMSE (Mg/h)
Combined Variables ($n=5$)	0.949	20.65
Red-edge vegetation indices only ($n=5$)	0.945	21.60
Spectral bands ($n=5$)	0.948	21.83
NIR Vegetation indices ($n=5$)	0.946	22.94

Scatter plots showing the correlation between the RF model predictions and the field observed (measured) outputs for each model are presented in Figure 5.2. The scatter plots were produced with a randomly selected independent validation set for each RF model. It could be observed that a good relationship exists between the field measured and RF predicted AGC.

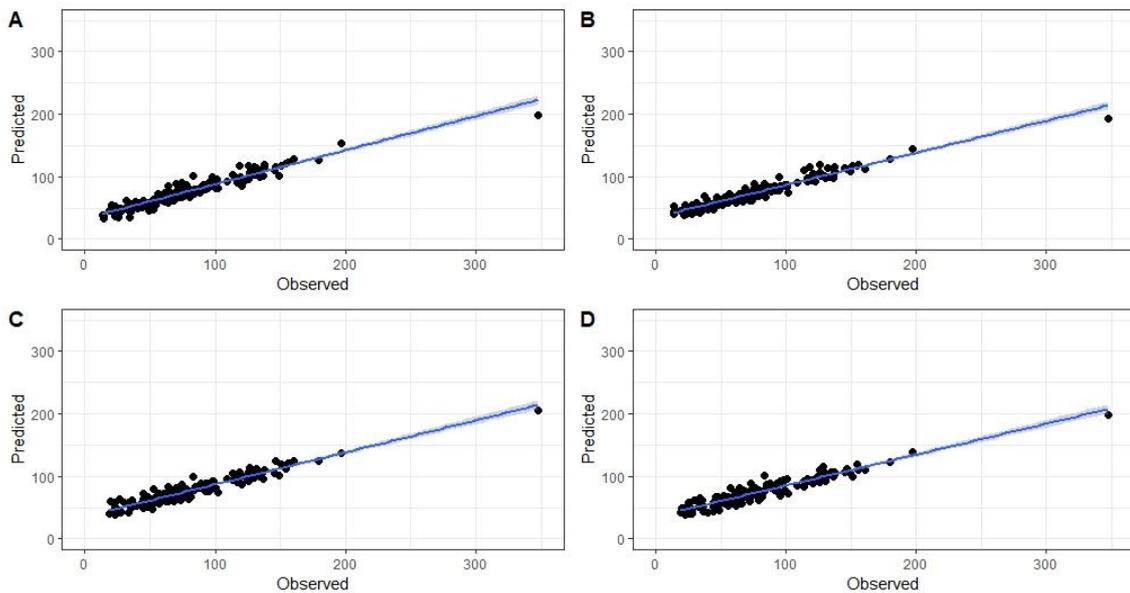


Figure 5.2: Scatter plots indicating the relationship between the observed and predicted AGC. The line of best fit is represented by the blue line with the shaded region representing the 95%

confidence interval. Note: A represents the Combined Variables model (A), B represents the Red-edge vegetation indices model, C represents the Spectral bands model and D represents the NIR vegetation indices.

5.7.2 AGC map

The Combined Variables RF model emerged as the best among the others. Hence, it was utilized to produce the AGC map to show the spatial distribution and concentration across the Nkandla Forest Reserve (Figure 5.3). The carbon levels are classified into ranges to show the different levels of concentration. A high amount of the variance in the AGC was predictable by the model variables. The closed and open canopy areas of the forest which form about 90% of the forest have the highest concentration of carbon ranging from 50 mg/ha and above due to its tree cover. The grasslands which constitute about 10% of the land cover of the forest ranged below 50 Mg/ha.

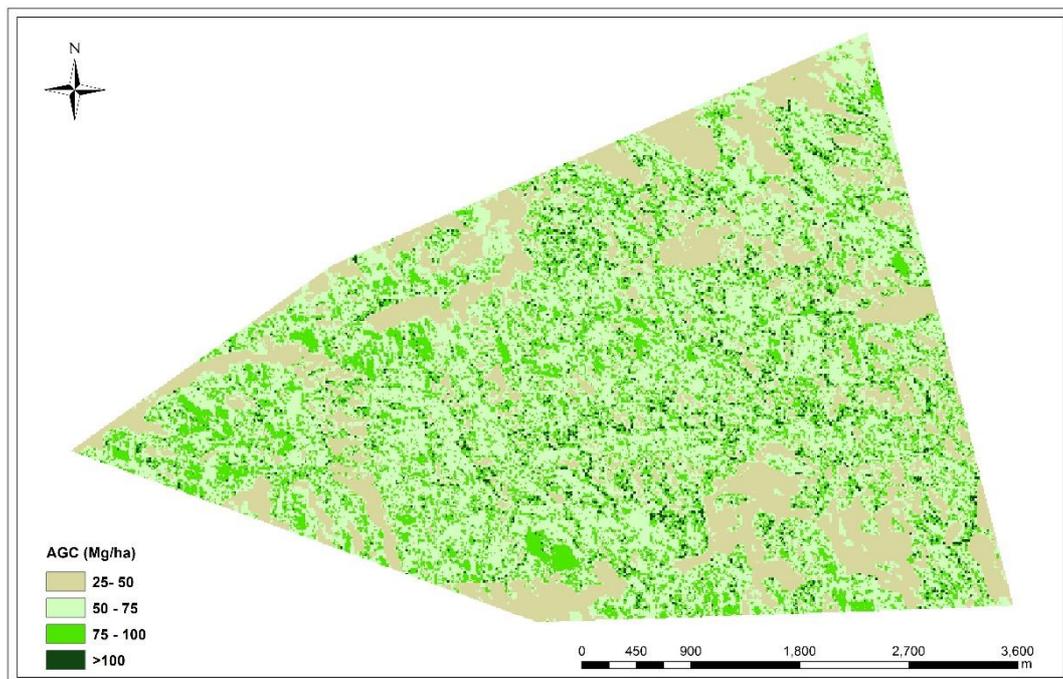


Figure 5.3: Final AGC distribution map produced for the Nkandla Forest Reserve.

5.7.3 AGC prediction variable importance

The variable importance feature of the RF algorithm was used to determine and rank the importance of every variable based on their significance in each model. The variable importance is a measure of the significance of each variable to the prediction accuracy of the model. The criteria used in this study was the percentage increasing mean squared error (% IncMSE) for each variable in each RF model.

The Combined Variables RF model had the red-edge bands such as the RE2 and the RE1 emerged as the first and third important variables (Figure 5.4A). The NDVI was the second most important variable while the MSRren was the least important. Concerning the Red-edge vegetation indices RF model, the REI was the most important variable while the least variable was the NDVIre1 (Figure 5.4B). The MSRen and the NDre1 had comparable %IncMSE showing their equal importance to the model and prediction of the AGC stock. The three Sentinel 2 red-edge bands, RE1, RE2 and RE2 emerged as the top three most important variables for the spectral band's RF model (Figure 5.4C). The least ranked variable was the Blue band. The DVI was the most important variable, followed by the MSR and NDVI in respective order as the top three variables for the NIR vegetation indices RF model (Figure 5.4D). The least important variable was the SR.

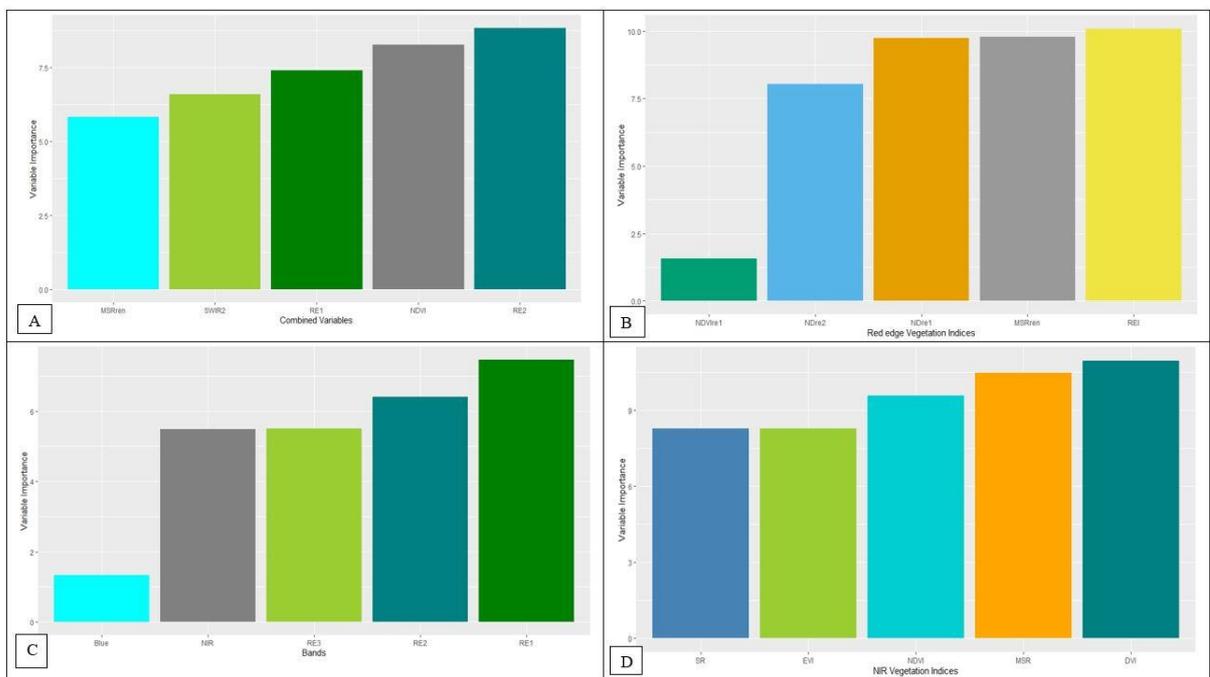


Figure 5.4: Variable importance for the variables of the four RF model of models. A is for the Combined Variables model, B is for the Red-edge vegetation indices, C is for the Spectral bands model and D is for the NIR vegetation indices model. The variables have been ranked based on the least to the most important variable from left to right.

5.8 Discussion

Our study found that all four RF models had a comparable coefficient of determination (R^2). This may be due to the use of an equal number of variables used in the models. Inferences drawn from AGC and AGB studies suggest that models with a large number of predicting variables produced a high R^2 and lower RMSE values as compared to models with a smaller number of variables (Millard and Richardson, 2015, Mutowo et al., 2018a, Pandit et al., 2018a). This reveals the critical role the number of predicting variables has on model accuracies. As a result of the satisfactory findings of our study, it may be important to consider using the same

number of predicting variables for models especially in comparative studies to provide possible equal grounds for comparison.

A second reason for the comparative accuracy could be due to the capabilities of the RF algorithm. The RF algorithm can handle different complex Remote Sensing data and variables types (Fassnacht et al., 2016). Furthermore, it works effectively with the different types of data sets without overfitting and as well deals with multicollinearity (Fassnacht et al., 2016, Ramoelo et al., 2015). These attributes make the RF robust for the prediction of AGC and the production of satisfactory accuracies. The finding of our study is similar to other studies that found the RF algorithm contributing to predicting model accuracies (Dube and Mutanga, 2015, Ghosh and Behera, 2018). Hence, it has been recommended by many studies to be adopted for vegetation monitoring and assessments.

Thirdly, the pre-selection of robust and informative variables for the four RF models could also account for the comparable accuracies. The selection of significant predictors is necessary to prevent the challenge of saturation, redundancy, computational complexities and low accuracies (Millard and Richardson, 2013, Nitze et al., 2015). Furthermore, it makes models simple and easy to interpret, reduces the variance of the model and therefore overfitting and also fast training of the model. These might have enhanced the predictive ability of the RF models. The Recursive Feature Elimination (RFE) algorithm could be adopted for the selection of important and robust variables when developing predicting models. The use of the RFE algorithm was important for developing the RF regression models as the selected variables were robust in the prediction. The usefulness of the RFE algorithm has also been found by other studies (Ghosh and Joshi, 2014, Wang et al., 2018a). Hence, the prior selection of the best and statistically significant variables could be adopted as a major component of predictive studies involving AGC and other vegetation studies.

As a result of the comparable R^2 obtained by all the four models, the determination of the best model was based on the one that had a lower RMSE than the other three. Thus, the Combined Variables model performed better than the other models due to its slightly lower RMSE. The Sentinel-2 products that contributed significantly to the output of the combined variables RF model are the RE1, RE2, SWIR2, NDVI and MSRren as the predicting variables. This is explained by the combined effect of the unique capabilities and information of the variables. The output of the model is reflective of the sensitivity of these variables to vegetation attributes such as chlorophyll, canopy structure and phenology. Rajah et al. (2019) stated that a strong relationship exists between the red-edge and the SWIR with vegetation parameters, thus further explaining the findings of our studies. These findings share similarities with other vegetation studies that also found the combination of Sentinel 2 spectral bands and vegetation indices variables performing better than single set variables (Mutowo et al., 2018a). Spectral bands in the red-edge region are sensitive to green vegetation biomass variations in contrast to senescence stage vegetation and are less susceptible to the problem of saturation (Todd et al., 1998, Mutanga and Skidmore, 2004). It explains the correlative ability of the red-edge bands with biomass and AGC (Sibanda et al., 2015, Dube and Mutanga, 2016). The MSRren shares in the capabilities of the red-edge bands as it is derived from those regions. Moreover, MSRren may have lessened the problem of data saturation in the model as the red-edge vegetation indices are less prone to the problem of saturation (Mutanga and Skidmore, 2004). The inclusion of the NDVI in the model might have also paid off in improving the predictive ability of the model. The NDVI is among the commonly used NIR vegetation indices applied in vegetation studies which is able to characterize canopy growth and vigor (Sripada et al., 2005, Xue and Su, 2017). It also relates to photosynthesis and canopy structure (Grace et al., 2007, Gamon et al., 1995). These attributes of the NDVI complimented that of the red-edge bands and the MSRren to make the combined variable much more robust. Similarly, the SWIR2 also proved

significant as it has also been found to be sensitive to species composition (Aklilu Tesfaye and Gessesse Awoke, 2020). These characteristics of the variables proved their importance in the modelling and affirm that a careful combination of key variables of spectral bands and vegetation indices could be a robust approach to predicting AGC.

Although the merger of the capabilities RE1, RE2, SWIR2, NDVI and MSRren contributed to the Combined Variables model accuracy, they were ranked with the RF importance variable feature to ascertain which of them contributed much to the performance and accuracy of the model. The RE1 and RE2 were the first and third variables respectively, while the NDVI was the second. This ranking output further affirms the superior robustness of the red-edge bands over the other spectral variables.

Despite the selected Spectral bands, NIR and Red-edge vegetation indices variables RF model had slightly higher RMSE margins than the Combined variables model, it is not indicative that they are not appropriate for the prediction of AGC of natural forests. This is because each of these models produced some level of satisfactory results with the comparable high R^2 and varying RMSE values, indicating their correlation with the field measured AGC. The Red-edge vegetation RF variables model was the second as it performed better than the spectral bands and the NIR vegetation models because it also had a slightly lower RMSE than them. The RE1, NDre1, MSRren, NDre2, NDVIre1 formed the variables in this model. The red-edge vegetation indices are sensitive to leaf and chlorophyll content and phenological stages (Aklilu Tesfaye and Gessesse Awoke, 2020, Xie et al., 2018), and hence reflecting the robustness they share with the bands from which they are derived. Emerging studies have been exploring the use of more Sentinel 2 red-edge derived vegetation indices through existing and new modelling approaches to ascertain their usefulness for vegetation related research (Fernández-Manso et al., 2016, Xie et al., 2018, Zhang et al., 2018). However, these red-edge vegetation indices appear to have been applied in AGC studies on a limited scale. Therefore, our study outcomes could contribute to ascertaining their usefulness for AGC prediction and mapping for natural and heterogenous forest ecosystems. Similarly, Fernández-Manso et al. (2016) found the red-edge vegetation indices outperforming the NIR vegetation indices. Additionally, Forkuor et al. (2020) also indicated that the red-edge vegetation indices performed better than other vegetation indices including those derived from the SWIR region. The calculation of red-edge vegetation indices was not possible owing to the unavailability of red-edge bands in old generation satellite sensors such as Landsat, MODIS and ASTER (Mutowo et al., 2018b). However, it is now possible in new generation sensors such as Sentinel 2, thus providing a means of further using them for AGC quantification to provide more insights on improving their utilization. In order to evaluate the performance of five red-edge vegetation indices, they were also ranked with the RF variable importance feature. The RE1, MSRren and NDre1 were the top three variables of the red-edge vegetation indices set and were significant contributors to its statistical output.

The Sentinel 2 bands RF model was the third-best performing model as it had the third lower error margin. It had the RE1, RE2, RE3, NIR and the Blue bands. The findings of other studies have also affirmed the capabilities of the Sentinel 2 spectral bands (Wicaksono, 2017, Zhang et al., 2019, Ghosh and Behera, 2018). The robustness of the red-edge bands has been seen in all the models and needs not be overemphasized. The NIR band displays strong reflectance for sites with high vegetation, hence its use in most vegetation related studies (Gizachew et al., 2016). On the other hand, the blue band has low reflectance and might have contributed to the error margin of the model. It was the least important variable when it was ranked alongside the other four bands by the RF algorithm. The RE1, RE2 and RE3 (red-edge) bands were the top three most important variables in the spectral band's RF model. This finding is similar to that of other studies, including Castillo et al. (2017) and Pandit et al. (2018a) who also found the

red-edge bands to be the most important variables and had a better prediction than the other spectral bands. Comparatively, Sentinel 2 may have an advantage over the other multispectral imagery (optical sensors) since it is the only one with three red-edge bands (Han et al., 2017). These red-edge bands are found to be more informative (Kira et al., 2016), making them effective for vegetation attributes monitoring and measurements including AGC.

The NIR vegetation indices RF model was the least performing model as it had the highest RMSE. Its variables were the DVI, NDVI, MSR, EVI and the SR. In terms of their characteristics, the DVI can distinguish between vegetation and soil (Tucker, 1979), while the NDVI can detect vegetation growth and vigour. The MSR has an increased sensitivity to the biophysical parameters of vegetation (Chen, 1996) whereas the EVI optimizes vegetation signals. Furthermore, it corrects soil background signals and reduces the influences of the atmosphere through the use of the blue reflectance region (Huete, 1988). Lastly, the SR is effective to use under various conditions and it is created by the band with the strongest vegetation reflectance and the band with the deepest absorption of chlorophyll (Birth and McVey, 1968). Generally, vegetation indices derived from Sentinel 2 are confirmed to correlate well with vegetation (Adan, 2017). It has been found that the effects of environmental conditions and shadows on spectral reflectance, especially in complex vegetation stands, can be minimized by vegetation indices (Adam et al., 2014a). Although all these attributes are related to the vegetation, it could not effectively deal with the error. That notwithstanding, they may be used for use in AGC studies as they had a high correlation with the vegetation based on their high R^2 . The DVI, NDVI and MSR emerged as the top three most important variables while the SR was the least important.

It may be worth indicating that the different modelling approaches coupled with data types, can potentially influence the AGC stock quantification as each research works with and settles on what finally produces optimal results. The use of Sentinel 2 imagery and the RF regression algorithm with the application of these identified important and statistically significant variables could be adopted for AGC studies in other natural forest ecosystems. It will lead to the identification of the best variables and robust models. This can deepen the understanding surrounding the utilization and application of the imagery and algorithms in explaining AGC stocks and dynamics.

The AGC map that was produced helps to identify the ranges, concentration and spatial distribution of AGC across the forest and it will be useful for carbon management. A high amount of the AGC is concentrated in the closed canopy and open canopy forests of the reserve due to the high tree cover. This is found almost across the entire forest except for areas that have dense grasslands. This map will be beneficial to the management of the forest to know where to set priorities with regards to AGC management. Furthermore, the map could also help to validate future AGC spatiotemporal analysis, monitoring and mapping.

The field data indicated that the forest is not old, and it has the potential to sequester more carbon as it continues to grow. This could be guaranteed if there are no human modifications, like land use land cover changes, deforestation and degradation which would interfere with the ecological and biological processes of the forest reserve. Effective management will be required for this forest to ensure the conservation of trees and ensure the realization of the potential forest capacity as a carbon sink.

The outputs of our study could have implications on natural forest management, carbon accounting and climate change adaptation and mitigation programmes both locally and internationally. It provides information on the range and distribution of AGC, useful to national and international databases on carbon stocks for informed decision making and policy

directions. An example is an assessment of including certain forests in the Reducing Emissions from Deforestation and forest Degradation-plus (REDD+) initiative of the United Nations (Avtar et al., 2020).

Finally, the results of this study demonstrate the potential of Sentinel 2 satellite imagery and RF regression algorithm in predicting and mapping AGC stocks in sub-tropical natural forests. This is in line with related studies that similarly observed the ability of the imagery and the algorithm suitable for AGB, AGC and vegetation monitoring (Chen et al., 2019, Ghosh and Behera, 2018, Forkuor et al., 2020). The findings could contribute to further deepening the understanding of the dynamics and approaches in the application of Sentinel 2 imagery and RF machine learning algorithm for AGC prediction and mapping in different forest ecosystems, landscapes and ecological zone.

5.9 Conclusion

Climate change, forest cover changes, carbon accounting and forest management makes AGC estimations important at the local and international levels. Also, due to the dynamic nature of forest ecosystems and increasing human interventions, it will be worth quantifying the local and national AGC stocks to support management decisions and policies. Several approaches involving the combination of many Remote Sensing data, field measurements and machine learning algorithms are being explored on different scales to provide the quantitative and qualitative information needed for decision making. One of those is the application of the Sentinel 2 and RF regression algorithm in the AGC prediction, estimation and mapping used in this study.

The results showed that the selection and combination of their key variables could improve prediction accuracies. The red-edge bands and their derived vegetation indices were observed to be informative and robust for AGC modelling as compared to the other spectral bands and the NIR vegetation indices of the other models. Also, the use of an equal number of variables in predicting models could be considered for comparative studies as it provides a potential equal basis for comparison. The modelling approach and the identified statistically significant variables may be adopted and utilized in the estimation and mapping of AGC stocks.

Furthermore, this study demonstrates the capabilities and robustness of Sentinel 2 data and RF regression algorithm in the direct AGC prediction and spatial mapping in natural sub-tropical forests. The AGC estimation is important as it is the first for the study area and will support forest management. Also, since this research is among the studies that have tested the direct prediction and mapping of AGC in natural forests, it may serve the larger Remote Sensing community and forest managers because of the vital information it provides. The freely available Sentinel 2 imagery and RF machine learning algorithm could be further used in studies especially in tropical and subtropical zones such as Africa to enhance contribution to national and global AGC carbon stock estimates.

Acknowledgment

The authors appreciate the support provided by Ms. Sharon Louw (Ecologist) of Ezemvelo KwaZulu-Natal Wildlife (EKZNW) in securing a permit to conduct the study in the Nkandla

Forest Reserve. We thank Mr. Elliackim Zungu (Conservation Manager), Mr. Simon Makhaye, Mr. Sandebulawa Biyela and other supporting staff of EKZNW for assisting in field data collection. The authors also thank Mr. Charles Zungunde (University of KwaZulu-Natal) for driving the field team across the forest as well as assisting in data collection. We further thank the journal editors and reviewers for their efforts, time and insights.

Funding

This research and study received a block grant from the South African System Analysis Centre (SASAC) through the National Research Foundation (NRF) of South Africa, with grant number 118770. The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest

References

- ABDEL-RAHMAN, E., M, AHMED, F. B. & ISMAIL, R. 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, 34, 712-728.
- ADAM, E., MUTANGA, O., ABDEL-RAHMAN, E. M. & ISMAIL, R. 2014. Estimating standing biomass in papyrus (*Cyperus papyrus* L.) swamp: exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing*, 35, 693-714.
- ADAN, M. 2017. *Integrating Sentinel-2 derived vegetation indices and terrestrial laser scanner to estimate above-ground biomass/Carbon in Ayer Hitam tropical forest Malaysia*. Master's Thesis, The University of Twente, Enschede, The Netherlands.
- AKLILU TESFAYE, A. & GESSESSE AWOKE, B. 2020. Evaluation of the saturation property of vegetation indices derived from sentinel-2 in mixed crop-forest ecosystem. *Spatial Information Research*.
- ASKAR, NUTHAMMACHOT, N., PHAIRUANG, W., WICAKSONO, P. & SAYEKTININGSIH, T. 2018. Estimating Aboveground Biomass on Private Forest Using Sentinel-2 Imagery. *Journal of Sensors*, 2018, 1-11.
- ASNER, G. P., HUGHES, R. F., MASCARO, J., UOWOLO, A. L., KNAPP, D. E., JACOBSON, J., KENNEDY-BOWDOIN, T. & CLARK, J. K. 2011. High-resolution carbon mapping on the million-hectare Island of Hawaii. *Frontiers in Ecology and the Environment*, 9, 434-439.
- ASNER, G. P. & MASCARO, J. 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, 140, 614-624.
- ATTARCHI, S. & GLOAGUEN, R. 2014. Improving the estimation of above ground biomass using dual polarimetric PALSAR and ETM+ data in the Hyrcanian mountain forest (Iran). *Remote Sensing*, 6, 3693-3715.
- AVTAR, R., TSUSAKA, K. & HERATH, S. 2020. Assessment of forest carbon stocks for REDD+ implementation in the muyong forest system of Ifugao, Philippines. *Environmental Monitoring Assessment*, 192, 1-18.

- BARNES, E., CLARKE, T., RICHARDS, S., COLAIZZI, P., HABERLAND, J., KOSTRZEWSKI, M., WALLER, P., CHOI, C., RILEY, E. & THOMPSON, T. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, 2000.
- BELGIU, M. & DRĂGUȚ, L. 2016. Random forest in Remote Sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- BIRTH, G. S. & MCVEY, G. R. 1968. Measuring the color of growing turf with a reflectance spectrophotometer 1. *Agronomy Journal*, 60, 640-643.
- BOZKAYA, A. G., BALCIK, F. B., GOKSEL, C. & ESBAH, H. 2015. Forecasting land-cover growth using remotely sensed data: a case study of the Igneada protection area in Turkey. *Environ Monit Assess*, 187, 18.
- BREIMAN, L. 2001. Random forests. *Machine learning*, 45, 5-32.
- BRILLI, L., CHIESI, M., BROGI, C., MAGNO, R., ARCIDIACO, L., BOTTAI, L., TAGLIAFERRI, G., BINDI, M. & MASELLI, F. 2019. Combination of ground and Remote Sensing data to assess carbon stock changes in the main urban park of Florence. *Urban Forestry & Urban Greening*, 43.
- BUSCHMANN, C. & NAGEL, E. 1993. In vivo spectroscopy and internal optics of leaves as basis for Remote Sensing of vegetation. *International Journal of Remote Sensing*, 14, 711-722.
- BYRD, K. B., O'CONNELL, J. L., DI TOMMASO, S. & KELLY, M. 2014. Evaluation of sensor types and environmental controls on mapping biomass of coastal marsh emergent vegetation. *Remote Sensing of Environment*, 149, 166-180.
- CARSAN, S., ORWA, C., HARWOOD, C., KINDT, R., STROEBEL, A., NEUFELDT, H. & JAMNADASS, R. 2012. African wood density database.
- CASTILLO, J. A. A., APAN, A. A., MARASENI, T. N. & SALMO, S. G. 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 70-85.
- CHAVE, J., COOMES, D., JANSEN, S., LEWIS, S. L., SWENSON, N. G. & ZANNE, A. E. 2009. Towards a worldwide wood economics spectrum. *Ecology letters*, 12, 351-366.
- CHAVE, J., REJOU-MECHAIN, M., BURQUEZ, A., CHIDUMAYO, E., COLGAN, M. S., DELITTI, W. B., DUQUE, A., EID, T., FEARNESIDE, P. M., GOODMAN, R. C., HENRY, M., MARTINEZ-YRIZAR, A., MUGASHA, W. A., MULLER-LANDAU, H. C., MENCUCCINI, M., NELSON, B. W., NGOMANDA, A., NOGUEIRA, E. M., ORTIZ-MALAVASSI, E., PELISSIER, R., PLOTON, P., RYAN, C. M., SALDARRIAGA, J. G. & VIEILLEDENT, G. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob Chang Biol*, 20, 3177-90.
- CHEN, J. M. 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22, 229-242.
- CHEN, L., REN, C., ZHANG, B., WANG, Z. & XI, Y. 2018. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. *Forests*, 9.
- CHEN, L., WANG, Y., REN, C., ZHANG, B. & WANG, Z. 2019. Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data. *Remote Sensing*, 11.
- DAFF 2015. State of the forests report 2010-2012 report, Department Agriculture, Forestry and Fisheries, Republic of South Africa. *Unpublished*, 88.

- DEVAGIRI, G. M., MONEY, S., SINGH, S., DADHAWAL, V. K., PATIL, P., KHAPLE, A., DEVAKUMAR, A. S. & HUBBALLI, S. 2013. Assessment of above ground biomass and carbon pool in different vegetation types of south western part of Karnataka, India using spectral modeling. *Tropical Ecology* 54, 149-165.
- DUBE, T. & MUTANGA, O. 2015. 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. 2016. The impact of integrating WorldView-2 sensor and environmental variables in estimating plantation forest species aboveground biomass and carbon stocks in uMgeni Catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119, 415-425.
- DURO, D. C., FRANKLIN, S. E. & DUBÉ, M. G. 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of environment*, 118, 259-272.
- EZEMVELO KZN WILDLIFE 2015. Nkandla Forest Complex: Protected Area Management Plan. *Unpublished* Version 1.0, 173
- FASSNACHT, F. E., LATIFI, H., STEREŃCZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- FASSNACHT, F. E., NEUMANN, C., FÖRSTER, M., BUDDENBAUM, H., GHOSH, A., CLASEN, A., JOSHI, P. K., KOCH, B. & SENSING, R. 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE Journal of Selected Topics in Applied Earth Observations*, 7, 2547-2561.
- FERNÁNDEZ-MANSO, A., FERNÁNDEZ-MANSO, O. & QUINTANO, C. 2016. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, 50, 170-175.
- FORKUOR, G., BENEWINDE ZOUNGRANA, J.-B., DIMOBE, K., OUATTARA, B., VADREVU, K. P. & TONDOH, J. E. 2020. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study. *Remote Sensing of Environment*, 236.
- FUCHS, H., MAGDON, P., KLEINN, C. & FLESSA, H. 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sensing of Environment*, 113, 518-531.
- GAMON, J. A., FIELD, C. B., GOULDEN, M. L., GRIFFIN, K. L., HARTLEY, A. E., JOEL, G., PEÑUELAS, J. & VALENTINI, R. 1995. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5, 28-41.
- GHOSH, A. & JOSHI, P. K. 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26, 298-311.
- GHOSH, S. M. & BEHERA, M. D. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*, 96, 29-40.
- GITELSON, A. & MERZLYAK, M. N. 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves.

- Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143, 286-292.
- GIZACHEW, B., SOLBERG, S., NAESSET, E., GOBAKKEN, T., BOLLANDSAS, O. M., BREIDENBACH, J., ZAHABU, E. & MAUYA, E. W. 2016. Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. *Carbon Balance Manag*, 11, 13.
- GRACE, J., NICHOL, C., DISNEY, M., LEWIS, P., QUAIFFE, T. & BOWYER, P. 2007. Can we measure terrestrial photosynthesis from space directly, using spectral reflectance and fluorescence? *Global Change Biology*, 13, 1484-1497.
- GREGORUTTI, B., MICHEL, B. & SAINT-PIERRE, P. 2017. Correlation and variable importance in random forests. *Statistics Computing*, 27, 659-678.
- GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2020. Mapping natural forest cover using satellite imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, 18, 1-14.
- HAN, J., WEI, C., CHEN, Y., LIU, W., SONG, P., ZHANG, D., WANG, A., SONG, X., WANG, X. & HUANG, J. 2017. Mapping above-ground biomass of winter oilseed rape using high spatial resolution satellite data at parcel scale under waterlogging conditions. *Remote Sensing*, 9, 238.
- HUETE, A. 1988. Huete, AR A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment. Remote Sensing of environment*, 25, 295-309.
- HUETE, A., DIDAN, K., MIURA, T., RODRIGUEZ, E. P., GAO, X. & FERREIRA, L. G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195-213.
- IPCC 2006. *2006 IPCC guidelines for national greenhouse gas inventories*, Intergovernmental Panel on Climate Change.
- JUCKER, T., CASPERSEN, J., CHAVE, J., ANTIN, C., BARBIER, N., BONGERS, F., DALPONTE, M., EWJIK, K. Y., FORRESTER, D. I. & HAENI, M. 2017. Allometric equations for integrating Remote Sensing imagery into forest monitoring programmes. *Global change biology*, 23, 177-190.
- KIRA, O., NGUY-ROBERTSON, A. L., ARKEBAUER, T. J., LINKER, R. & GITELSON, A. A. 2016. Informative spectral bands for remote green LAI estimation in C3 and C4 crops. *Agricultural Forest Meteorology*, 218, 243-249.
- KUMAR, L. & MUTANGA, O. 2017. Remote Sensing of Above-Ground Biomass. *Remote Sensing*, 9.
- LI, X., CHEN, W., CHENG, X. & WANG, L. 2016. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. *Remote Sensing*, 8, 514.
- LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. *R news*, 2, 18-22.
- LIN, S., LI, J., LIU, Q., LI, L., ZHAO, J. & YU, W. 2019. Evaluating the Effectiveness of Using Vegetation Indices Based on Red-Edge Reflectance from Sentinel-2 to Estimate Gross Primary Productivity. *Remote Sensing*, 11.
- MILLARD, K. & RICHARDSON, M. 2013. Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR-SAR fusion using a random forest classifier. *Canadian Journal of Remote Sensing*, 39, 290-307.
- MILLARD, K. & RICHARDSON, M. 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, 7, 8489-8515.
- MITCHARD, E. T., SAATCHI, S. S., WOODHOUSE, I. H., NANGENDO, G., RIBEIRO, N., WILLIAMS, M., RYAN, C. M., LEWIS, S. L., FELDPAUSCH, T. & MEIR, P.

2009. Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, 36.
- MOHD ZAKI, N. A., LATIF, Z. A. & SURATMAN, M. N. 2018. Modelling above-ground live trees biomass and carbon stock estimation of tropical lowland Dipterocarp forest: integration of field-based and remotely sensed estimates. *International Journal of Remote Sensing*, 39, 2312-2340.
- MOONEY, C. F., MOONEY, C. L., MOONEY, C. Z., DUVAL, R. D. & DUVALL, R. 1993. *Bootstrapping: A nonparametric approach to statistical inference*, sage.
- MUTANGA, O. & SKIDMORE, A. K. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25, 3999-4014.
- MUTOWO, G., MUTANGA, O. & MASOCHA, M. 2018a. Evaluating the Applications of the Near-Infrared Region in Mapping Foliar N in the Miombo Woodlands. *Remote Sensing*, 10.
- MUTOWO, G., MUTANGA, O. & MASOCHA, M. 2018b. Mapping foliar N in miombo woodlands using sentinel-2 derived chlorophyll and structural indices. *Journal of Applied Remote Sensing*, 12, 046028.
- NITZE, I., BARRETT, B. & CAWKWELL, F. 2015. Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation*, 34, 136-146.
- PAN, Y., BIRDSEY, R. A., FANG, J., HOUGHTON, R., KAUPPI, P. E., KURZ, W. A., PHILLIPS, O. L., SHVIDENKO, A., LEWIS, S. L. & CANADELL, J. G. 2011. A large and persistent carbon sink in the world's forests. *Science*, 1201609.
- PANDIT, S., TSUYUKI, S. & DUBE, T. 2018. Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sensing*, 10.
- QAZI, W. A., BAIG, S., GILANI, H., WAQAR, M. M., DHAKAL, A. & AMMAR, A. 2017. Comparison of forest aboveground biomass estimates from passive and active Remote Sensing sensors over Kayar Khola watershed, Chitwan district, Nepal. *Journal of Applied Remote Sensing*, 11.
- RAJAH, P., ODINDI, J., MUTANGA, O. & KIALA, Z. 2019. The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. *Nature Conservation*, 35, 41-61.
- RAMOELO, A., CHO, M. A., MATHIEU, R., MADONSELA, S., VAN DE KERCHOVE, R., KASZTA, Z. & WOLFF, E. 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 43, 43-54.
- RODRÍGUEZ-GALIANO, V. F., ABARCA-HERNÁNDEZ, F., GHIMIRE, B., CHICA-OLMO, M., ATKINSON, P. M. & JEGANATHAN, C. 2011. Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environmental Sciences*, 3, 44-49.
- ROUJEAN, J.-L. & BREON, F.-M. 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51, 375-384.
- ROUSE, J., HAAS, R., SCHELL, J. & DEERING, D. Monitoring vegetation systems in the Great Plains with ERTS. Third earth resources technology satellite-1 symposium, Vol. 1, 10-14 December, 1973, 1974. NASA Scientific and Technical Information Office Washington, DC, 309-317.

- SAATCHI, S. S., HARRIS, N. L., BROWN, S., LEFSKY, M., MITCHARD, E. T., SALAS, W., ZUTTA, B. R., BUERMANN, W., LEWIS, S. L., HAGEN, S., PETROVA, S., WHITE, L., SILMAN, M. & MOREL, A. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc Natl Acad Sci U S A*, 108, 9899-904.
- SARKER, M. L. R., NICHOL, J., IZ, H. B., AHMAD, B. B. & RAHMAN, A. A. 2013. Forest Biomass Estimation Using Texture Measurements of High-Resolution Dual-Polarization C-Band SAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 3371-3384.
- SIBANDA, M., MUTANGA, O. & ROUGET, M. 2015. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110, 55-65.
- SRIPADA, R. P., HEINIGER, R. W., WHITE, J. G. & WEISZ, R. 2005. Aerial color infrared photography for determining late-season nitrogen requirements in corn. *Agronomy Journal*, 97, 1443-1451.
- TEAM, R. D. C. 2013. R: A language and environment for statistical computing.
- THURNER, M., BEER, C., SANTORO, M., CARVALHAIS, N., WUTZLER, T., SCHEPASCHENKO, D., SHVIDENKO, A., KOMPTER, E., AHRENS, B., LEVICK, S. R. & SCHMULLIUS, C. 2014. Carbon stock and density of northern boreal and temperate forests. *Global Ecology and Biogeography*, 23, 297-310.
- TODD, S., HOFFER, R. & MILCHUNAS, D. 1998. Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing*, 19, 427-438.
- TUCKER, C. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150.
- VAGLIO LAURIN, G., PIROTTI, F., CALLEGARI, M., CHEN, Q., CUOZZO, G., LINGUA, E., NOTARNICOLA, C. & PAPAIE, D. 2017. Potential of ALOS2 and NDVI to estimate forest above-ground biomass, and comparison with lidar-derived estimates. *Remote Sensing*, 9, 18.
- VICHARNAKORN, P., SHRESTHA, R., NAGAI, M., SALAM, A. & KIRATIPRAYOON, S. 2014. Carbon Stock Assessment Using Remote Sensing and Forest Inventory Data in Savannakhet, Lao PDR. *Remote Sensing*, 6, 5452-5479.
- WANG, D., WAN, B., QIU, P., SU, Y., GUO, Q., WANG, R., SUN, F. & WU, X. 2018. Evaluating the Performance of Sentinel-2, Landsat 8 and Pléiades-1 in Mapping Mangrove Extent and Species. *Remote Sensing*, 10.
- WANG, G., OYANA, T., ZHANG, M., ADU-PRAH, S., ZENG, S., LIN, H. & SE, J. 2009. Mapping and spatial uncertainty analysis of forest vegetation carbon by combining national forest inventory data and satellite images. *Forest Ecology and Management*, 258, 1275-1283.
- WHITE, J. C., COOPS, N. C., WULDER, M. A., VASTARANTA, M., HILKER, T. & TOMPALSKI, P. 2016. Remote Sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42, 619-641.
- WICAKSONO, P. 2017. Mangrove above-ground carbon stock mapping of multi-resolution passive remote-sensing systems. *International Journal of Remote Sensing*, 38, 1551-1578.
- XIE, Q., DASH, J., HUANG, W., PENG, D., QIN, Q., MORTIMER, H., CASA, R., PIGNATTI, S., LANEVE, G., PASCUCCI, S., DONG, Y. & YE, H. 2018. Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index

- Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11, 1482-1493.
- XUE, J. & SU, B. 2017. Significant Remote Sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017, 17.
- ZANNE, A., LOPEZ-GONZALEZ, G., COOMES, D., ILIC, J., JANSEN, S., LEWIS, S., MILLER, R., SWENSON, N., WEIEMANN, M. & CHAVE, J. 2009. Global wood density database. Data from: Towards a worldwide wood economics spectrum. Dryad Digital Repository.
- ZARCO-TEJADA, P. J., MILLER, J. R., NOLAND, T. L., MOHAMMED, G. H. & SAMPSON, P. H. 2001. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience Remote Sensing*, 39, 1491-1507.
- ZHANG, G., GANGULY, S., NEMANI, R. R., WHITE, M. A., MILESI, C., HASHIMOTO, H., WANG, W., SAATCHI, S., YU, Y. & MYNENI, R. B. 2014. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sensing of Environment*, 151, 44-56.
- ZHANG, M., DU, H., ZHOU, G., LI, X., MAO, F., DONG, L., ZHENG, J., LIU, H., HUANG, Z. & HE, S. 2019. Estimating Forest Aboveground Carbon Storage in Hang-Jia-Hu Using Landsat TM/OLI Data and Random Forest Model. *Forests*, 10.
- ZHANG, Z., LIU, M., LIU, X. & ZHOU, G. 2018. A New Vegetation Index Based on Multitemporal Sentinel-2 Images for Discriminating Heavy Metal Stress Levels in Rice. *Sensors*, 18.

CHAPTER 6: EVALUATING MULTI-SENSORS SPECTRAL AND SPATIAL RESOLUTIONS FOR TREE SPECIES DIVERSITY PREDICTION



Article

Evaluating Multi-Sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction

Enoch Gyamfi-Ampadu ^{1,*} , Michael Gebreslasie ¹  and Alma Mendoza-Ponce ²

¹ School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban 4000, South Africa; gebreslasie@ukzn.ac.za

² Centro de Ciencias de la Atmósfera, Ciudad Universitaria, Universidad Nacional Autónoma de México Investigación Científica s/n, C.U., Coyoacán, Mexico City 04510, Mexico; amendozap@atmosfera.unam.mx

* Correspondence: 217082006@stu.ukzn.ac.za

Abstract

Forests contribute significantly to terrestrial biodiversity conservation. Monitoring of tree species diversity is vital due to climate change factors. Remote Sensing imagery is a means of data collection for predicting diversity of tree species. Since various sensors have different spectral and spatial resolutions, it is worth comparing them to ascertain which could influence the accuracy of prediction of tree species diversity. Hence, this study evaluated the influence of the spectral and spatial resolutions of PlanetScope, RapidEye, Sentinel 2 and Landsat 8 images in diversity prediction based on the Shannon diversity index (H'), Simpson diversity Index (D_1) and Species richness (S). The Random Forest regression was applied for the prediction using the spectral bands of the sensors as variables. The Sentinel 2 was the best image, producing the highest coefficient of determination (R^2) under both the Shannon Index ($R^2 = 0.926$) and the Species richness ($R^2 = 0.923$). Both the Sentinel and RapidEye produced comparable higher accuracy for the Simpson Index ($R^2 = 0.917$ and $R^2 = 0.915$, respectively). The PlanetScope was the second-accurate for the Species richness ($R^2 = 0.90$), while the Landsat 8 was the least accurate for the three diversity indices. The outcomes of this study suggest that both the spectral and spatial resolutions influence prediction accuracies of satellite imagery.

Keywords: natural forests; diversity; prediction; sensors; random forest; conservation

6.1 Introduction

Forests cover about one-third of the earth's total landmass and contain a large amount of terrestrial biodiversity (Gamfeldt et al., 2013, Aerts and Honnay, 2011). Forest biodiversity is an expression of the differences among the living organisms present in the ecosystem and it is considered as one of the means of measuring forest health and stability (Wang and Gamon, 2019). The interdependence and interaction among the species influence and facilitate the provision of ecosystem goods and services (Iverson and McKenzie, 2013). These ecosystem goods and services include carbon sequestration and storage, provision of habitats for wildlife, production of non-timber forest products (NTFPs), regulation of water and biogeochemical cycles (Millennium Ecosystem Assessment, 2005). Although forest biodiversity includes trees, animal species and other life forms, trees seem to be the most essential elements as without them there will be no forest and most ecosystem goods and services provision will be hindered.

The prediction and estimation of tree species diversity provide forest managers, ecologists and conservationists information for forest management decisions. The spatial information obtained through the estimation of the tree species is vital for effective forest management and biodiversity conservation (Turner et al., 2003); and it provides better understanding of forest ecological processes such as tree growth rates, species recruitment, and net productivity (Luo et al., 2019). In recent years, remote sensors have provided data that help predict, estimate and map forests at various levels (Fundisi et al., 2020, Grabska et al., 2019b). This is due to its large spatial coverage, less time consumption, and cost-effectiveness as compared to traditional inventories and assessments (Adelabu et al., 2013, Ustuner et al., 2016, Kavzoglu and Mather, 2003). That notwithstanding, the methodological approach that establishes the relationship between Remote Sensing imagery and field data is identified as a robust means of predicting tree species diversity.

The advances in remote sensors, data characteristics, and processing systems have increased the potential of satellite imagery in providing accurate and robust spatially explicit estimates of tree species diversity. Satellite imagery from sensors has been employed for tree species diversity assessments for various types of forests in many areas. The output of these assessments has demonstrated the ability of Remote Sensing satellite imagery to predict species diversity based on field derived measured data. Furthermore, many of these studies prescribe images and approaches that could be adopted in the modelling process.

Different types of sensors, including multispectral (Chrysafis et al., 2020, Madonsela et al., 2018, Mutowo and Murwira, 2012a, Arekhi et al., 2017, Foody and Cutler, 2006, Gillespie et al., 2016, Gould, 2000, Grabska et al., 2019b), hyperspectral (Carlson et al., 2007, Colgan et al., 2012, Oldeland et al., 2010, Kalacska et al., 2007, Laurin et al., 2014, Schäfer et al., 2016) and active ones like the Light detection and ranging (LiDAR) have been used over the years for the prediction of tree species diversity in different forest types, and climatic zones and scales. Hyperspectral images can predict tree diversity with better accuracy due to their numerous narrow bands (Nagendra et al., 2013, Ghosh et al., 2014). On some occasions, the hyperspectral data are fused or combined with LiDAR (Zhao et al., 2018, Sun et al., 2019). This data fusion approach helps to take advantage of the ability of the hyperspectral data to detect different vegetation communities and the ability of the LiDAR data to measure the structural attributes of tree species. Furthermore, the LiDAR data can pass through clouds, which allows incident rays to reach the target feature and the reflected rays to get to the sensor. However, the high cost of acquiring hyperspectral and LiDAR data has hindered the mass application in diversity prediction (Mutowo and Murwira, 2012a).

Multispectral images could be said to have been used much in the prediction of tree species diversity. One of the most used is the Landsat images which have proven useful in the forest zones within which they were applied (Arekhi et al., 2017, Madonsela et al., 2017, Mohammadi and Shataee, 2010). Over the years there has been an improvement in its spectral bands and how they sense vegetation, especially with the Landsat 8 (Dube and Mutanga, 2015). Since the success of diversity prediction across different forest zones depends on the ability of the spectral bands to correlate with tree species characteristics, it is important to adopt images that have a high sensitivity to forests. Another satellite imagery that has been used is the Advanced Spaceborne Thermal Emission and Reflection Radiometer [ASTER] (Feilhauer and Schmidlein, 2009a, Mutowo and Murwira, 2012a). It seems to have not had much application in the prediction of diversity across many forest types as compared to the Landsat satellite imagery. However, the studies that have used it have found its spatial and spectral product capable of producing good prediction accuracies (Mutowo and Murwira, 2012b). Another Remote Sensing imagery that has also proven robust and informative and is also freely available is Sentinel 2 imagery. It has a medium resolution and a large number of spectral bands that enhance its accuracy outputs (Grabska et al., 2019b, Persson et al., 2018, Chrysafis et al., 2020). It is also one of the images that have been used extensively for many vegetation studies. It is the only Remote Sensing imagery that has three red edge bands which give it some level of advantage over the other satellite imageries (Mutowo et al., 2018a). This is because of the chlorophyll information it contains which contributes to the high sensitivity it has for vegetation. These reasons could be the basis for the high accuracies it produces in diversity studies.

The spectral and spatial products of images including the bands, vegetation indices and texture variables are normally used as the predicting variables. Apart from the spectral bands, one of the most used predictors is the normalized difference vegetation index (NDVI) (Arekhi et al., 2017, Gould, 2000, Pau et al., 2012, Oindo and Skidmore, 2010, Krishnaswamy et al., 2009). It is derived from the bands with the highest absorption (Red) and reflectance (near infrared), which makes it useful under various conditions. However, one of the drawbacks of the NDVI is data saturation in areas with a high leaf area index (LAI). This could also likely affect diversity estimation under certain circumstances. Texture variables such as the Gray Level Co-occurrence Matrix (GLCM) have also contributed to the prediction of diversity (George-Chacon et al., 2019, Ozdemir et al., 2008, Fundisi et al., 2020). The spectral bands such as the near infrared (NIR), red edge (RE) and the shortwave infrared (SWIR) have been found to be important in many diversity studies (Otunga et al., 2018, Tigges et al., 2013, van Deventer et al., 2017, Tuominen et al., 2018). These are located in the regions in the electromagnetic spectrum that contains vegetation information and including any of them in predictive models could improve performance outputs. The use of any of these variables could be dependent on the forest type and the tree cover density. For instance, most forests in temperate and boreal zones may not be of high density and heterogenous as compared to tropical and subtropical forests. As such forests in the tropical and subtropical forest are likely to require robust predictors as compared to temperate and boreal forests. The capabilities of the predictors may be sensor-dependent and the advancement in their design over the years has made diversity studies much more successful.

The methods and modelling techniques are also one of the main factors that contribute to diversity prediction outputs. Most studies have resorted to the use of regression which is carried out by either parametric or non-parametric machine learning algorithms. The Random Forest (RF) which is a non-parametric algorithm is one of the main algorithms that have been extensively used in predictions (Pedro et al., 2015, Chrysafis et al., 2020, Mallinis et al., 2020). As a non-parametric algorithm, it does not assume a normal distribution of data and it is optimal to be used for diversity modelling of natural forests due to these characteristics. The linear regression which is a parametric algorithm has as well been used in many studies (Mutowo and Murwira, 2012b, John et al., 2008, Madonsela et al., 2018). The modelling technique in the use of these algorithms is an important factor to consider as one of the things that affect accuracies.

The prediction of tree species diversity in many forests and climatic zones have become necessary with time due to factors such as increasing climate change that are negatively affecting species. The availability and advancement of different sensors are continually being tested for their suitability for diversity modelling as well as increasing knowledge in their application. However, none of these studies has been carried out for subtropical natural forests in South Africa, which creates a gap in tree diversity management. It must be noted that subtropical natural forests are characterised by high tree species diversity and density (Ouyang et al., 2016, Huang et al., 2017, Sun et al., 2017). As such, it will require informative and robust imagery to predict and map their tree species diversity. Thus, evaluating multi-sensors performance and identifying the best based on their spectral and spatial resolutions is beneficial for the application of imagery in diversity prediction and mapping. Hence, our study aimed to assess how the performance and accuracies of PlanetScope, RapidEye, Sentinel 2 and Landsat 8 images could be influenced by their spectral and spatial resolution in the prediction of tree species diversity for a subtropical natural forest in KwaZulu-Natal (KZN) province, South

Africa. The Shannon Index (H'), Simpson Index (D_1) and the Species richness (S) together with RF regression modelling, are utilised to identify which image has a good relationship with them and produce good accuracy. The outcomes of our study will provide information on how spectral and spatial resolution could influence image model accuracies, which can provide a guide in the decision making on the imagery to select for predicting tree species diversity of subtropical natural forests. It will also contribute to existing knowledge and approach to modelling of diversity for forest management and conservation. Furthermore, it could assist forest managers in devising measures that can enhance the conservation and protection of forest diversity.

6.2. Materials and Methods

6.2.1. Study Area

The Nkandla Forest Reserve is an Afromontane sub-tropical forest type, and it was established in 1918. It is found in the north of KwaZulu-Natal province, South Africa. It has a total area of 2217 ha and located on $28^{\circ} 43' 50.88''$ S and $30^{\circ} 7' 9.84''$ E (Figure 6.1). A peak average temperature of 27°C is experienced between December and January, and the lowest average temperature of 2°C in the winter months of June and July (Ezemvelo KZN Wildlife, 2015b). It has an undulating and steep topography with an altitude of a minimum level of 500 m and exceeding 1300 m. It has four land cover types made up of closed canopy forest (1,059.23 ha), open canopy forest (910.60 ha), grassland (226.55 ha) and bare sites [20.97 ha] (Gyamfi-Ampadu et al., 2020). It has common tree species such as *Cryptocarya myrtifolia*, *Trichilia dregeana*, *Bridelia micrantha*, *Elaeodendron croceum*, *Podocarpus henkelii* and *Olea capensis*.

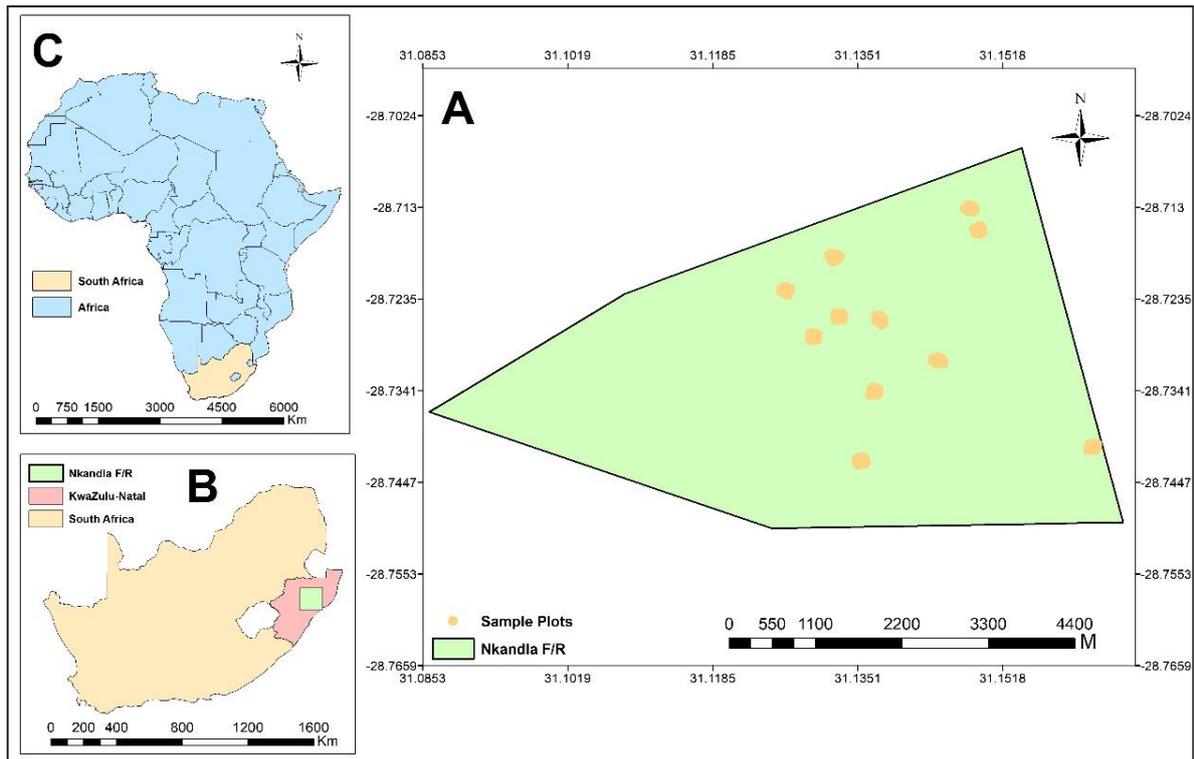


Figure 6.1: Map of the study area. Note: A is the Nkandla Forest Reserve, B is the map of South Africa indicating the location of KZN province and the forest and C is the map of Africa indicating the location of South Africa.

6.2.2. Field Inventory and Diversity Indices Estimation

Tree data information was collected from the accessible parts of the Nkandla Forest Reserve between 24 April 2019 and 7 May 2019. This is because it was observed from a reconnaissance survey that some parts of the forest, especially the western portions were inaccessible due to the presence of high elevation and deep slopes. Therefore, the inventory was restricted to the middle and northeastern parts which have gentle slopes. Existing transects were followed and a systematic approach was used in setting up the sampling plots in the gentle slope and relatively flat terrain. Eleven 100 m x 100 m plots (1 ha) were randomly set up in areas with gentle slopes and flat terrain. Each of the 1 ha plots was subdivided into 25 subplots of 20 m x 20 m sizes each to facilitate the data collection. Thus, the tree data was obtained from a total of 275 subplots. In each subplot, the diameter at breast height (DBH) of tree species ≥ 5 cm was measured with a diameter tape. Other information recorded for the trees was the species name (local and scientific) and the GPS coordinates of the trees. The individual number of species was summed up for each sampling plot. This approach of tree inventory did not compromise on the data collected because similar numbers and types of tree species were measured and recording in most of the sample plots. The tree data were further compared and confirmed for similarity with the tree list in the management plan of the forest obtained from the Ezemvelo KZN Wildlife.

The relative number of each tree species was used to compute the Shannon Index (H') (Shannon, 1948), Simpson Index (D_1) (Simpson, 1949) and Species richness (S) (Morris et al., 2014) for each species. There was done by using the mathematical functions in equations 5.1, 5.2 and 5.3 for the three diversity indices respectively. These indices have been well established and they allow for comparison of tree species diversity levels at different scales (Daly et al., 2018) and they as well help to account for the evenness and richness of diversity for each site. The Species richness takes into consideration the absolute number of species in a particular ecosystem, while the evenness takes into consideration the relative abundance of each species (Mallinis et al., 2020). The Shannon Diversity Index (H') accounts for both the species richness, and species abundance (Ifo et al., 2016). The original Simpson Index (D) emphasizes on the evenness component of diversity (Simpson, 1949). The Shannon Index is sensitive towards species rarity and abundance, while the Simpson Index is sensitive towards abundance in species distribution (Morris et al., 2014). These indices have been used widely and are confirmed to have a relationship with the spectral reflectance of Remote Sensing sensors (Oldeland et al., 2010).

$$H' = - \sum_{i=1}^S p_i \times \ln(p_i) \quad (6.1)$$

$$D_1 = 1 - \sum_{i=1}^S p_i \quad (6.2)$$

$$S = N \quad (6.3)$$

where p_i is the proportionate abundance of the i th species in the sampling plot, S is the total number of all species in a sampling plot, and \ln is the natural logarithm of the proportionate abundance of species in the sampling plot.

6.2.3. Field Inventory Data Analysis

The descriptive statistics for the Shannon Index (H') and Simpson Index (D_1) and Species richness (S) that were computed for the field inventory are presented in Table 6.1.

Table 6.1: Descriptive statistics of Shannon Index (H'), Simpson Index (D_1) and Species richness (S) produced from the field inventory data.

Parameter	Shannon Index	Simpson Index	Species richness
Mean	2.055	0.891	9
Minimum	0.949	0.155	4
Maximum	2.718	0.993	15
Standard Deviation	0.290	0.068	2.47

6.2.4. Remote Sensing Data

We used four different sensors of different spectral and spatial resolutions because the study's focus was to compare and assess multi-sensor spectral and spatial resolution effects on accuracies in tree species diversity prediction. The satellite imageries used in the study were Landsat 8, Sentinel 2, RapidEye and PlanetScope (Table 6.2). All the images were cloud-free. The Landsat 8 has 10 spectral bands covering the visible to the shortwave infrared (SWIR) region of the electromagnetic (EM) spectrum with a spatial resolution of 30 m. The Sentinel 2 has a spectral resolution of 13 also ranging from the visible range to the SWIR region of the spectrum with varying spatial resolution. The blue, green, red and near infrared (NIR) spectral bands have a spatial resolution of 10 m, while the three red edges bands, narrow near infrared (NNIR) and the two shortwave infrared bands (SWIR 1 and SWIR2) have a spatial resolution of 20 m. The coastal aerosol (Band 1), water vapour (Band 9) and cirrus bands (Band 10) have a spatial resolution of 60 m. The Landsat 8 and the Sentinel 2 are both freely available imagery that has been used extensively for vegetation related studies. The Landsat 8 is provided by the United States Geological Service (USGS) while Sentinel 2 is provided by the European Space Agency (ESA).

The RapidEye has a spatial resolution of 5 m and five spectral bands which ranges from the visible to the NIR region of the EM spectrum. It is also among the sensors that have been used extensively for vegetation studies. On the other hand, the PlanetScope is a relatively new sensor and it is yet to be much used in diversity prediction. It has four spectral bands ranging from the visible to the NIR of the EM spectrum with a spatial resolution of 3 m. Both the RapidEye and PlanetScope are commercial sensors provided by the Planet Team.

A Landsat 8 image captured on 8 May 2019 was downloaded from the Earth Explorer website (www.usgs.gov) of the USGS. The Landsat 8 image was atmospherically corrected from Top-of-Atmosphere to surface reflectance using the apparent reflection function in ArcGIS 10.6.1. The coastal aerosol band (Band 1), the panchromatic band (Band 8), Cirrus (Band 9) and thermal infrared bands (Bands 11 and 12) were not included in the bands considered for the analysis. They were excluded because the band 1 contains aerosols, band 8 is panchromatic, band 9 contains cloud information, while bands 11 and 12 contains thermal information. The Sentinel 2 image was captured on 14 April 2019 and was similarly downloaded from the Earth Explorer website (www.usgs.gov) of the USGS. It was atmospherically corrected using the semi-automatic classification plugin (SCP) of the QGIS 3.10 software. The image radiance was

transformed into spectral reflectance with the dark object subtraction (DOS1) SCP plugin of the QGIS 3.10 software. The image was further resampled to 10 m spatial resolution using the SNAP toolbox for the spectral bands to have a uniform resolution, as they are varied. This operation was done to enhance the analysis. Bands 1, 9 and 10 were excluded because they contain aerosols, water vapour and cloud information respectively. The PlanetScope and the RapidEye images were downloaded from the Planet Explorer website (www.planet.com / www.api.planet.com). The PlanetScope was captured on 30 April 2019 while the RapidEye was captured on 18 June 2019. The two images were atmospherically corrected by the suppliers (Planet Team) and subsequently provided to be downloaded for the analysis. The characteristics of each of the four images have been detailed in Table 6.2.

Table 6.2: Details of the spectral and spatial resolution of satellite imageries.

Sentinel 2			Landsat 8			RapidEye			PlanetScope		
Bands	Bandwidth (nm)	Spatial resolution (m)	Bands	Bandwidth	Spatial resolution (m)	Bands	Bandwidth (nm)	Spatial resolution (m)	Bands	Bandwidth (nm)	Spatial resolution (m)
Blue	458-523	10	Blue	452-512	30	Blue	440-510	5	Blue	455-515	3
Green	543-573	10	Green	533-590	30	Green	520-590	5	Green	500-590	3
Red	650-680	10	Red	636-673	30	Red	630-685	5	Red	590-670	3
RE1	698-713	20	NIR	851-879	30	RE	690-730	5	NIR	780-860	3
RE2	733-748	20	SWIR1	1566-1651	30	NIR	760-850	5			
RE3	773-793	20	SWIR2	2107-2294	30						
NIR	785-899	10									
NNIR	855-875	20									
SWIR2	1565-1655	20									
SWIR2	2100-2280	20									

6.2.5. Important Variables Selection

The Recursive Feature Elimination (RFE) algorithm was subsequently used to select important variables to be used as input variables for the Random Forest regression model for each of the four images. This process is very important as it helps to eliminate noisy variables and reduce redundancy and computational complexities (Ghosh and Joshi, 2014, Wang et al., 2018a). The RFE process of elimination is carried out in a stepwise approach involving; 1) the training of the RF model, 2) computing the permutation importance measure, 3) eliminating of the less relevant variables (features) and 4) repeating the first 3 steps until no further variables remain (Wang et al., 2018a). The most informative variables are ranked in the last stage of the steps of the backward procedure and the algorithm selects a smaller size and more efficient variable subset.

The SWIR1, SWIR2, RE2, NIR and NNIR bands were selected for Sentinel 2, whereas the Red, NIR and RE bands were selected for the RapidEye. The VNIR bands were maintained by the algorithm for the PlanetScope after the running of several iterations. Lastly, the Green, Red, NIR and SWIR1 bands were selected for the Landsat 8.

6.2.6. Random Forest Regression Modelling

Random Forest (RF) (Breiman, 2001) regression models were used to predict the tree species diversity based on the Shannon diversity (H') and Simpson diversity ($D1$) and Species richness (S) derived from the field measured data. The prediction established the relationship between the diversity indices and the spectral characteristics of the image data. The RF is a non-parametric machine learning algorithm that can undertake both classification and regression (Breiman, 2001). A bagging system is used to split the data by the algorithm where a part of the data is used for training and building the decision tree. The remaining set is used for estimating the out-of-bag (OOB) error for each tree. The RF algorithm has an advantage of not overfitting data because there is a convergence of the generalization error when the number of trees increases (Rodríguez-Galiano et al., 2011, Breiman, 2001). It is also able to deal with the problem of multicollinearity (Abdel-Rahman et al., 2013b, Ramoelo et al., 2015). The RF has two main parameters that contribute to the accuracies of models. These are the *n_{tree}* and the *m_{try}* and they may be tuned or left in defaults values. The *n_{tree}* has a default value of 500 and it is the total number of decision trees grown in the model. The default value of the *m_{try}* is the total number of predictor variables divided by 3 ($N/3$) when used in regression models. Studies that have used the default values of both parameters have obtained satisfactory results (Duro et al., 2012, Nitze et al., 2015). Aside from these characteristics, the RF enables the assessment and ranking of statistical significance of each predicting variable in the model with the use of its variable importance feature.

The four models were implemented with the “randomForest” package (Liaw and Wiener, 2002) in the R statistical software environment (Team, 2013). The spectral pixel values of each of the four images were extracted and used in the models. The 275 sample plot values of each of the Shannon (H') and Simpson ($D1$) diversity indices and Species richness (S) values computed from tree species data were partitioned into 70% training data (192), and 30% independent validation data (83) in a random selection approach. We calibrated each RF regression model

with the training data and then applied the bootstrapping of 500 iterations to predict the diversity.

A parameter optimization process was carried out to find the best *n*tree and *m*try values for the RF model of each of the four satellite imageries. The “tuneRF” function in the “randomForest” package was used to find the optimal *m*try value for the models. The value obtained after the process was 1 for all the models. On the other hand, the optimal *n*tree values obtained for the Sentinel 2, RapidEye, PlanetScope and the Landsat 8 models were 600, 500, 900 and 400 respectively. The *n*tree and the *m*try values were then used in models for predicting the tree species diversity. The independent 83 validation set of each image was subsequently used for the validation of prediction accuracies.

6.2.7. Models Evaluation

The four RF regression models’ predictive abilities were compared and assessed based on two main statistical parameters. These parameters were the coefficient of determination (R^2), and the root mean squared error (RMSE). The means of the 500 bootstrapped samples were used to calculate the accuracy parameters values. The RF regression model with the highest R^2 and lowest RMSE values was determined as the most accurate.

6.2.8. Variable Importance

The variable importance feature of the RF algorithm was applied to evaluate and rank the predicting variables according to their statistical importance in contributing to the accuracy of each model. The importance of each variable is determined by the percentage increase in mean squared error (%IncMSE). The %IncMSE denotes the effect of a predicting variable in a model when it is removed from it. This was assessed to determine the spectral bands that play an important role in the prediction and correlated well with the Shannon Diversity Index (H') and Simpson Diversity Index (D1) and Species richness (S) for the subtropical natural forest.

6.3. Results

6.3.2. Sensor Performance Evaluation

The RF model was utilized to evaluate the performance of the four sensors for the prediction of the tree species diversity for Shannon Index, Simpson Index, and the Species richness. Their performances were evaluated based on the R^2 and the RMSE. The model with the highest R^2 and lowest RMSE was considered more accurate and robust.

As illustrated in Table 6.3, the Sentinel 2 image model was the most accurate ($R^2 = 0.926$, RMSE = 0.148) for the prediction of tree species diversity derived using Shannon Index while the RapidEye emerged as the second accurate ($R^2 = 0.902$, RMSE = 0.147) for the same diversity index. The PlanetScope model was the third accurate ($R^2 = 0.898$, RMSE = 0.156) with the Landsat 8 model being the least accurate ($R^2 = 0.529$, RMSE = 1.748).

The Sentinel 2 and the RapidEye were the most accurate with a comparable accuracy output ($R^2 = 0.917$, $RMSE = 0.043$ and $R^2 = 0.915$, $RMSE = 0.044$ respectively) for the tree species prediction tree with the Simpson Diversity Index [D_1] (Table 6.3). Whereas the PlanetScope produced the second-best accuracy ($R^2 = 0.899$, $RMSE = 0.045$), and Landsat 8 was the least accurate ($R^2 = 0.410$, $RMSE = 0.063$).

The Sentinel 2 was once more the most accurate ($R^2 = 0.923$, $RMSE = 1.983$), under the Species richness (S), while the PlanetScope was the second accurate ($R^2 = 0.900$, $RMSE = 1.293$) (Table 6.3). The RapidEye was the third accurate ($R^2 = 0.833$, $RMSE = 1.287$), and the Landsat 8 was the least accurate model ($R^2 = 0.532$, $RMSE = 1.746$).

Table 6.1: RF sensor model accuracies for the Sentinel 2, RapidEye, PlanetScope and the Landsat 8 for Shannon Index, Simpson Index and the Species Richness.

Image	Shannon Index			Simpson Index			Species Richness		
	R^2	RMSE	p value	R^2	RMSE	p value	R^2	RMSE	p value
Sentinel 2	0.926	0.148	$<2.2 \times 10^{16}$	0.917	0.043	$<2.2 \times 10^{16}$	0.923	1.183	$<2.2 \times 10^{16}$
RapidEye	0.902	0.147	$<2.2 \times 10^{16}$	0.915	0.044	$<2.2 \times 10^{16}$	0.833	1.287	$<2.2 \times 10^{16}$
PlanetScope	0.898	0.156	$<2.2 \times 10^{16}$	0.899	0.045	$<2.2 \times 10^{16}$	0.900	1.293	$<2.2 \times 10^{16}$
Landsat 8	0.529	1.748	$<2.2 \times 10^{16}$	0.410	0.063	$<2.2 \times 10^{16}$	0.532	1.746	$<2.2 \times 10^{16}$

The statistical evaluation conducted for the prediction has been presented in Table 6.4. It was observed that there was a slight underestimation for the prediction under the Shannon Index and the Species richness by all four images. On the other hand, the prediction for the Simpson Index had the field measured values and the predicted values correlated much better as they were within ranges of each other. Scatter plots produced by each RF model of the imageries which establishes the relationship between the field measured and predicted diversity under the Shannon Index, Simpson Index and Species richness are presented in Figures 6.2–6.4.

Table 6.2: The statistical analysis of the prediction made with the RF for each of the images under the Shannon Index, Simpson Index and the Species richness.

Satellite Image	Parameter	Shannon Index	Simpson Index	Species richness
Sentinel 2	Mean	2.05	0.89	9.24
	Minimum	1.51	0.55	6.58
	Maximum	2.34	0.95	12.7
	Standard deviation	0.15	0.04	1.33
RapidEye	Mean	2.05	0.89	9.22
	Minimum	1.40	0.53	5.73
	Maximum	2.44	0.94	13.48
	Standard deviation	0.18	0.04	1.41
PlanetScope	Mean	2.05	0.89	9.26
	Minimum	1.61	0.56	6.25
	Maximum	2.44	0.95	12.67
	Standard deviation	0.15	0.04	1.30
Landsat 8	Mean	2.06	0.89	9.04
	Minimum	1.73	0.73	6.13
	Maximum	2.42	0.95	13.37
	Standard deviation	0.15	0.04	1.41

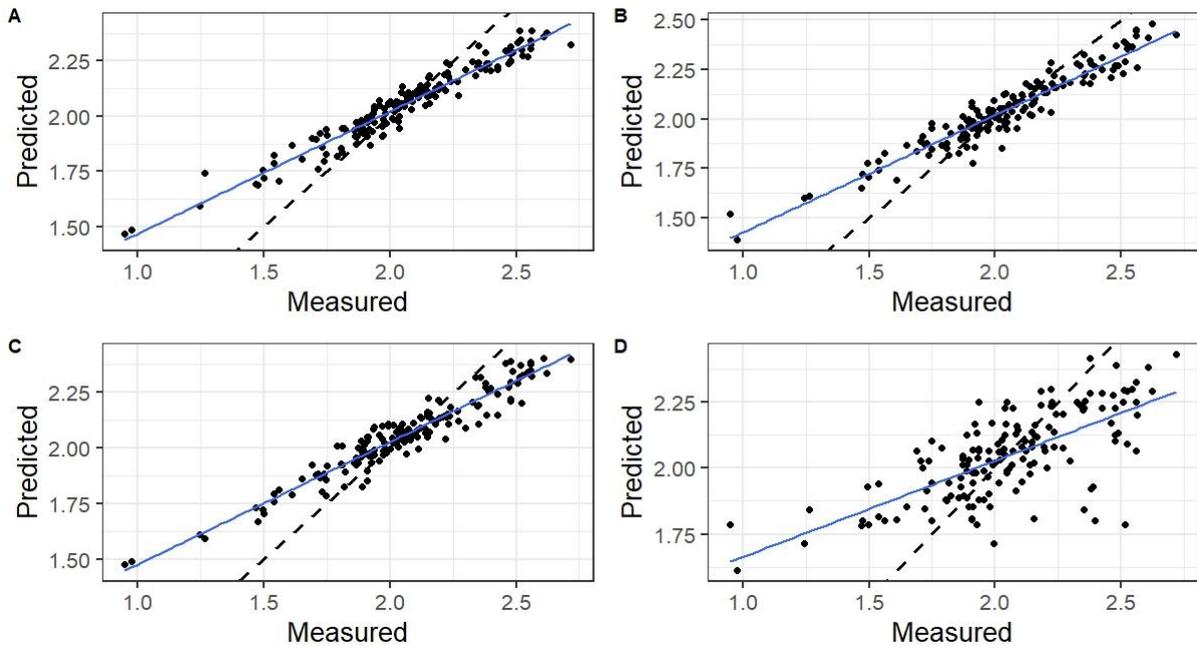


Figure 6.2: Scatter plot for the Shannon Index prediction. (A) is for Sentinel 2, (B) is for RapidEye, (C) is for PlanetScope, and (D) is for Landsat 8. The blue line is the line of best fit and the dashed line is the 1:1 line as shown on the individual plots.

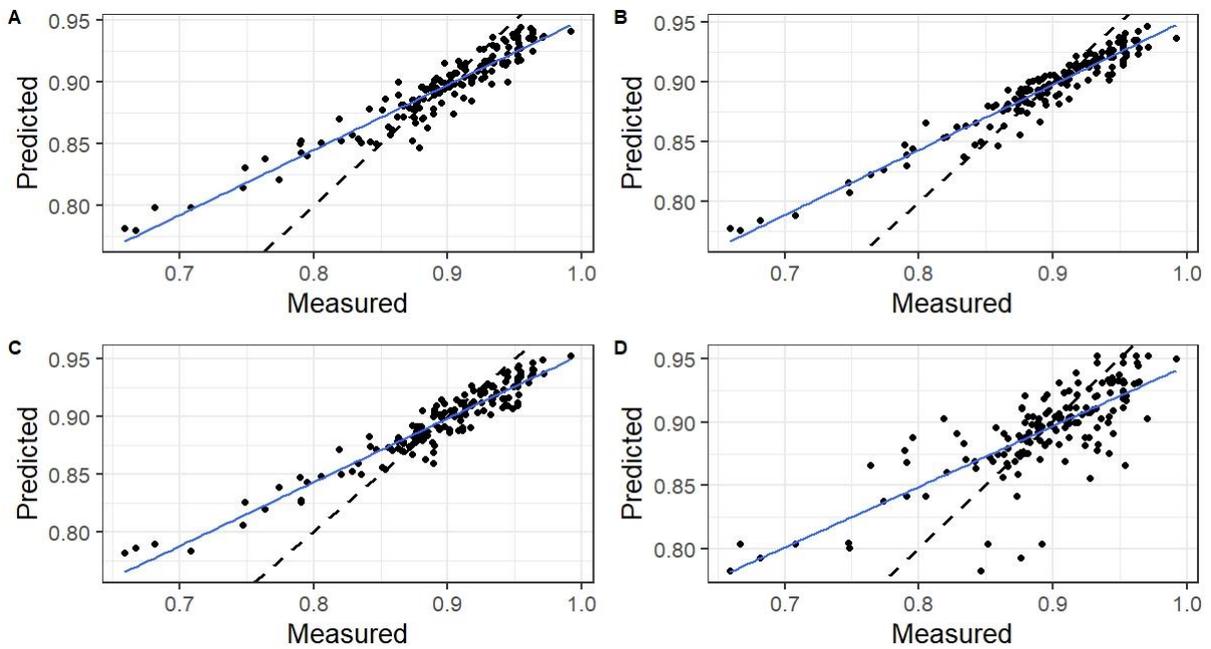


Figure 6.3: Scatter plot for the Simpson Index prediction. (A) is for Sentinel 2, (B) is for RapidEye, (C) is for PlanetScope, and (D) is for Landsat 8. The blue line is the line of best fit and the dashed line is the 1:1 line as shown on the individual plots.

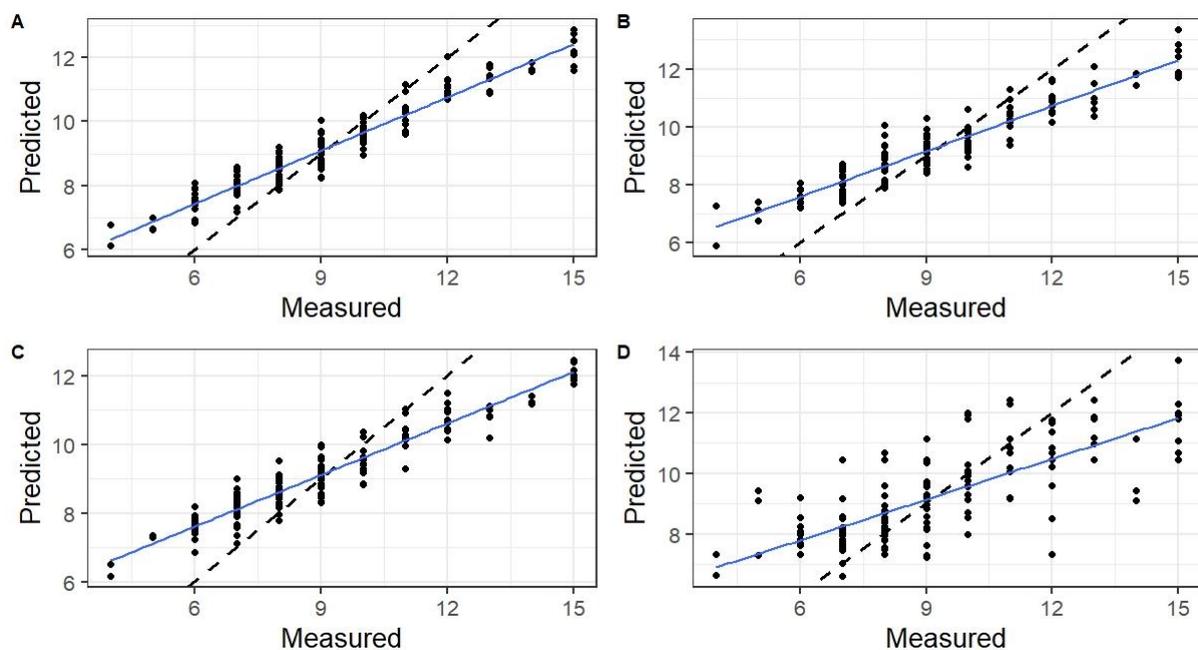


Figure 6.4: Scatter plot for the Species richness predictions. (A) is for Sentinel 2, (B) is for RapidEye, (C) is for PlanetScope, and (D) is for Landsat 8. The blue line is the line of best fit and the dashed line is the 1:1 line as shown on the individual plots.

6.3.3. Predicting Important Variables

The Variable Importance feature of the RF was utilized to rank the importance of each Remote Sensing variable for the prediction of the tree species diversity. RF regression algorithm provides the percentage increase mean square error (%IncMSE), which was used to rank the variables. The variables for each of the sensors under the Shannon Index, Simpson Index and the Species richness were ranked in decreasing order of importance for each RF model.

Table 6.5 illustrates the ranking of the important variables of the Sentinel-2 spectral bands used for the prediction of tree species diversity under the Shannon, Simpson, and Species richness indices. The SWIR1 band was the most important variable for the tree species diversity predicted using the Shannon index. The second-best to the least important variables were the SWIR 2, RE2, NNIR and the NIR respectively.

For the Simpson Index (D_1) predictions, the RE2 was the most important variable, while the SWIR1 was the second important variable. The third and fourth positions were occupied by the SWIR2 and NNIR. The NIR was again least important in the prediction. The %IncMSE values indicated that it played a very minimal role for this diversity index. Regarding the Species richness, the SWIR1 was once more the most important variable, while the NNIR, SWIR2, RE2 and NIR followed as second to the least, respectively.

Table 6.3. Variable importance ranking for the Sentinel 2 image model under the Shannon Index, Simpson Index and the Species richness.

Shannon Index		Simpson Index		Species richness	
Band	%IncMSE	Band	%IncMSE	Band	%IncMSE
SWIR1	17.75	RE2	6.71	SWIR1	18.57
SWIR2	14.11	SWIR1	6.05	NNIR	16.19
RE2	13.52	SWIR2	5.43	SWIR2	13.27
NNIR	8.97	NNIR	4.50	RE2	10.88
NIR	4.89	NIR	-0.03	NIR	6.99

The important variable ranking for the RapidEye spectral bands used in the RF model is illustrated in Table 6.6 for the Shannon Index, Simpson Index and the Species Richness. The most important variable under the Shannon Index was the Red band. The second was the NIR whereas the RE was last. The ranking of the most important variables for the Simpson Index and the Species richness was the same. The NIR was the most important band, while the RE and Red bands were the second and third respectively.

Table 6.4. Variable importance ranking for the RapidEye image model under the Shannon Index, Simpson Index and the Species richness.

Shannon Index		Simpson Index		Species richness	
Band	%IncMSE	Band	%IncMSE	Band	%IncMSE
Red	19.64	NIR	6.28	NIR	12.58
NIR	14.24	RE	4.79	RE	9.25
RE	14.01	Red	1.91	Red	8.39

Table 6.7 displays the important variables for the PlanetScope image under the three diversity indices. The Green and Red bands had the same level of significance under the Shannon Index in their contribution to the accuracy of the image’s model. They shared the first position while the Blue band was third and the NIR was the least significant. For the Simpson Index, the Green band was the important variable for the prediction done under the Simpson Index. The NIR was the second-best contributor to the accuracy with the Blue and Red being third and last. With the Species richness, the NIR emerged as the best variable and the Green band was the second best. The Red and Blue bands shared the third and fourth positions respectively.

Table 6.5: Variable importance ranking for the PlanetScope image model under the Shannon Index, Simpson Index and the Species richness.

Shannon Index		Simpson Index		Species Richness	
Band	%IncMSE	Band	%IncMSE	Band	%IncMSE
Green	9.93	Green	6.03	NIR	14.54
Red	9.91	NIR	5.83	Green	13.46
Blue	7.78	Blue	5.44	Red	12.78
NIR	7.04	Red	2.01	Blue	10.11

The variable importance of the Landsat 8 was not much different from that of the Sentinel 2, RapidEye and the PlanetScope as presented in Table 6.8. The SWIR1 which is one of the key spectral bands of the Landsat 8 was the most important variable for the Shannon Index. The NIR occupied the second position, followed by the Green and Red bands as third and last respectively. The reverse was the situation under the Simpson Index, where the NIR was the most important and the SWIR1 was the second. The Red band emerged as the third and the Green was the last. In the case of the Species richness, the SWIR1 was the best variable and the NIR following as the second-best variable. The Red and Green bands occupied the third and last position respectively.

Table 6.6. Variable importance ranking for the Landsat 8 image model under the Shannon Index, Simpson Index and the Species Richness.

Shannon Index		Simpson Index		Species Richness	
Band	%IncMSE	Band	%IncMSE	Band	%IncMSE
SWIR1	18.01	NIR	8.36	SWIR1	18.96
NIR	15.38	SWIR1	7.87	NIR	16.71
Green	14.16	Red	5.11	Red	13.91
Red	11.96	Green	3.77	Green	13.80

6.4. Discussion

In recent years there has been the launch and availability of free and commercial remote sensors that produce imageries which are adopted for forest vegetation-related research. The spectral and spatial attributes are vital for Remote Sensing imagery, and these could influence their suitability, and robustness for the characterization and prediction of forest attributes such as tree species diversity (Ganivet et al., 2019). The sensor type influences and contributes much to accuracy (Fassnacht et al., 2014a). Therefore, the assessment of different sensors based on their spectral and spatial resolution in the prediction and mapping of species diversity is

beneficial to ecologists and Remote Sensing experts. It is worth noting, that each sensor does have its strength and limitation (Lu, 2006), as a result of their spectral and spatial resolutions. This was displayed in the accuracy produced under each of the three diversity indices. Furthermore, it indicates the relationship between the predicting spectral variables and the indices.

The Sentinel 2 imagery was the most accurate and performed better than the RapidEye, PlanetScope and Landsat 8 for the prediction using the Shannon Index and the Species richness. It was also the best image together with the RapidEye under the Simpson Index as both had a comparable high R^2 and low RMSE. Several factors could account for the higher performance of the Sentinel 2 than the other images. Firstly, the five important spectral bands (RE2, NIR, NNIR, SWIR1 and SWIR 2) selected through the application of the Recursive Feature Elimination (RFE) may have been robust than that of the other three images. The availability of the red edge and the SWIR bands for the Sentinel 2 might have also contributed significantly to its accuracy. The red edge and SWIR bands, which are also positioned in Sentinel 2, have a higher sensitivity to healthy vegetation and minimum susceptibility to saturation (Todd et al., 1998, Mutanga and Skidmore, 2004). These attributes of the bands make them effective for diversity prediction in high density natural forests. It is important to note that the sensitivity of the red edge and the SWIR bands enhance their correlation with vegetation (Rajah et al., 2019). This sensitivity may be attributed to the narrow bandwidth and their location in the electromagnetic spectrum. It is also likely that the spectral bands of the Sentinel 2 are more informative than that of the RapidEye, PlanetScope and the Landsat 8. This may explain the better correlation of the Sentinel 2 with the field measured Shannon Index, Simpson Index and the Species richness that led to its high accuracy. Also, the larger number of spectral bands used for the RF regression model of the Sentinel 2 could have enhanced its capability and influenced the high accuracy. Findings of Rocchini et al. (2007) indicated that a large number of spectral bands increased diversity prediction accuracy, thus, suggesting the preference of large number spectral bands to a small number (Wang et al., 2018b). The spatial resolution of the Sentinel 2 could have also been a key factor because the pixels of the image are likely to have more tree species falling within it. As a result, more information on vegetation might have been preserved for the image. Since the three diversity indices rely on the types and number of species, the ability of the image to have more trees falling well within its pixels is vital for higher accuracy in predictions. Our study shares a similarity with Mallinis et al. (2020), who also found the Sentinel 2 performing better than the RapidEye in species diversity prediction in the Mediterranean region. Among other reasons, the study indicated that the absence of SWIR bands for the RapidEye could be a contributing factor, which has some relation to our findings. The inherent capability of the Sentinel 2 that enhances the detection and characterizing of vegetation have been confirmed in other research (Chrysafis et al., 2020, Immitzer et al., 2019, Martin-Gallego et al., 2020), which further validates our findings.

A knowledge of the variables that contributed most to the accuracy of models is important in modelling. It helps to select key variables that are robust, reduces redundancy and noise in the prediction and characterisation of vegetation attributes (Millard and Richardson, 2015, Ghosh and Joshi, 2014). With regards to the Sentinel 2, the RE2, SWIR1 and the SWIR2 contributed significantly to accuracy outputs both under the Shannon index and the Species richness, mainly due to their high sensitivity to vegetation. Immitzer et al. (2019) also observed that the red edge and the SWIR bands were useful and produced better accuracy for broadleaf species

classification. In addition, the importance of the red edge band is emphasized by Grabska et al. (2019b), while Persson et al. (2018) and (Bolgen et al., 2018) highlight the significance of the SWIR vegetation variability classification and separation. It is worth stating that, although the NIR had a higher reflectance for healthy vegetation, it was the least contributor to the higher accuracy of the Sentinel 2. It was not robust enough for the prediction as it could not enhance the capabilities of the image. With the advancement and increase in Remote Sensing imagery and their application to vegetation and forests attribute characterisation and mapping, the identification of these key bands is vital.

In the prediction with the Simpson Index, the RapidEye performed better than the PlanetScope and the Landsat 8 as it produced a comparable higher accuracy together with the Sentinel 2. This could have been due to the availability of the red edge and the NIR band for the RapidEye (Parson, 2013, Immitzer et al., 2019, Bolyn et al., 2018), which may have significantly contributed to the higher accuracy it produced under this diversity index. Although it has been suggested that having a larger number of variables are important (Rocchini et al., 2007, Wang et al., 2018b), it is also possible that selecting few but robust and informative bands as inputs variables for a model could help reduce noise and produce higher accuracies. That might have worked for the RapidEye under the Simpson Index. On the other hand, its finer spatial resolution could have had an effect on accuracies under the Shannon Index and the Species richness. It is indicated that higher spatial resolution of satellite imageries usually contain the structural attributes of vegetation community, but some information on the species type and the relative abundance is lost (Nagendra, 2001). This may further account for why it placed second to the Sentinel 2 under the Shannon Index and Species richness. Taking individually, its coefficient of determination for the three diversity indices ranged between 0.83 and 0.92, accounting for its good explanation of the variance and suitability for diversity modelling. The RapidEye has been found useful in vegetation studies such as intra and inter-species biomass prediction (Dube et al., 2014), forest structural information (Wallner et al., 2014), tree species classification (Adelabu et al., 2013) and urban vegetation classification (Tigges et al., 2013). Hence, it could further be evaluated in similar studies to ascertain its suitability for diversity prediction.

The PlanetScope is a relatively new image as compared to the RapidEye, Sentinel 2 and Landsat 8. It was the second-best image for the Simpson Index, but the third-best for the Shannon and the Species Richness in the prediction. Although four spectral bands were used for its RF regression model, its spectral bands are likely less informative and sensitive to vegetation as compared to the RapidEye and the Sentinel 2. Its bands are made up of only the visible and near infrared (VNIR) and lacks bands such as the red edge and the SWIR. This might have also accounted for the low accuracies it had as compared to the Sentinel 2 and the RapidEye. As identified by our study findings and other vegetation related studies (Mallinis et al., 2020), the red edge and SWIR are very useful and contribute to model accuracies. Similarly, to the RapidEye, the fine spatial resolution of the PlanetScope might have also reduced its ability to have a high number of species, thereby producing lower accuracies for the Shannon Index and the Species richness. On the positive side, it has a very good temporal resolution (revisit time) of one day, which makes it suitable for time series species diversity studies. It could also be accessed for vegetation phenological and seasonal variation studies because of the daily revisit time that could capture seasonal changes observed in vegetation. In the variable importance assessment, the Green and NIR bands were much more accurate respectively for

the Shannon Index, Simpson Index and the Species richness. Generally, the VNIR bands are common to most satellite images and are sensitive and correlate well with vegetation (Dube and Mutanga, 2015). Among the VNIR bands, the Red, Green and the NIR have high reflectance for healthy vegetation and could be considered as part of the spectral bands employed for diversity prediction in high density natural subtropical forests.

The low performance of the Landsat may be directly related to the low spatial resolution as compared to the other images. Its accuracy for the Shannon and Simpson Indices were just about half of that of the other images. Contrary to the findings of our studies, it has provided satisfactory accuracies in studies, (Madonsela et al., 2017, Maeda et al., 2014), although it was not compared with other images. On a more general basis, it is among the images that have been used for vegetation studies including diversity (Madonsela et al., 2017, Gillespie et al., 2016, Gould, 2000). Furthermore, its bands have been designed and improved for detecting and mapping vegetation (El-Askary et al., 2014, Pahlevan et al., 2014), and it has proven to be useful for those vegetation studies. Similarly, the most important variables among the spectral bands used for the prediction under the three indices were the SWIR 1 and the NIR. The importance of these bands needs not be overemphasized as their capabilities have already been indicated for the other images. On an individual basis, it may be useful for diversity prediction as has been found in vegetation related studies. Its high amount of historical data could be explored for multitemporal and time series diversity studies.

Generally, the spectral bands had a high relationship with the Shannon index, Simpson index and the Species richness with most of the accuracies for the Sentinel 2, RapidEye and PlanetScope. Successful diversity estimation with the utilisation of Remote Sensing data would be dependent on the spectral variables that could suitably capture the species diversity for the landscape in question (Madonsela et al., 2017). Therefore, spectral bands in the VNIR up to the SWIR region could be used to further ascertain their suitability for diversity prediction and mapping in natural subtropical forests.

Concerning the diversity indices, the use of either one of them could be dependent on the objective of the study, the forest type and the image. Spectral bands respond differently to them in their application to diversity prediction. However, little attention has been given to finding out much about their sensitivity to the species distribution patterns (Madonsela et al., 2017), with the use of spectral variables. Since species abundance, richness and evenness are likely to change with time, it may be important to determine the indices that best correlate with spectral variables through seasonal and temporal studies.

The Random Forest regression algorithm was very beneficial in the prediction by each image model. It demonstrated the capability to handle different types of complex Remote Sensing image data (Belgiu and Drăguț, 2016). Since it is a non-parametric machine learning algorithm, it does not assume normality (Fassnacht et al., 2016). This attribute is useful for natural forests since they are mostly heterogeneous and do not have a normal distribution. Furthermore, it can handle redundancy, reduce noise and deal with multicollinearity (Breiman, 2001, Abdel-Rahman et al., 2013b). All these might have probably influenced the functioning of the models to produce satisfactory accuracies. It could explain why the RF is mostly adopted for most vegetation related studies including diversity prediction.

The findings of our study have shown the capability of the images and important spectral bands most especially for the Sentinel 2, RapidEye and PlanetScope that are optimal for the prediction and mapping of tree species diversity. The output of our study is important for forest managers and ecologists in the modelling and prediction of tree species diversity. This could assist forest managers and ecologists in the selection of images and spectral bands for the prediction of diversity in natural subtropical forests. Generally, it could assist in the application of Remote Sensing technology and modelling in the estimation of diversity.

6.5. Conclusions

Our study assessed how spectral and spatial resolutions influence the accuracy of Remote Sensing imagery models based on the Shannon index, Simpson index and Species richness for the Nkandla natural forest in South Africa. Since various sensors perceive vegetation differently based on their spatial and spectral resolutions, finding a suitable one for the prediction of the tree species diversity in high density natural forests is important. It has been demonstrated in our studies and others that both the spectral and spatial resolutions of satellite imagery have much influence on the accuracies of images. The medium spatial resolution of Sentinel 2 and its spectral resolution makes it more capable in the prediction of diversity. Although the RapidEye, PlanetScope and Landsat 8 had lower performances than the Sentinel 2, it is not indicative that they may not be used for diversity prediction in natural subtropical forests. Since their abilities have been demonstrated in our study, they may be used to further ascertain the condition under which they could work better. On an individual basis, each of the imageries may be applied as they produced satisfactory accuracies. Also, since there are no generic spectral and spatial resolutions for diversity prediction currently, more studies could be carried out to test different sensors in various forest types to ascertain which could work much better.

Acknowledgment

The authors appreciate the support provided by Ms. Sharon Louw (Ecologist) of Ezemvelo KwaZulu-Natal Wildlife (EKZNW) in securing a permit to conduct the study in the Nkandla Forest Reserve. We thank Mr. Elliackim Zungu (Conservation Manager), Mr. Simon Makhaye, Mr. Sandeulawa Biyela and other supporting staff of EKZNW for assisting in field data collection. The authors also thank Mr. Charles Zungunde (University of KwaZulu-Natal) for driving the field team across the forest as well as assisting in data collection. We further thank the journal editors and reviewers for their efforts, time and insights.

Funding

This research and study received a block grant from the South African System Analysis Centre (SASAC) through the National Research Foundation (NRF) of South Africa, with grant number 118770. The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest

References

- ABDEL-RAHMAN, E. M., AHMED, F. B. & ISMAIL, R. 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *34*, 712-728.
- ADELABU, S., MUTANGA, O., ADAM, E. & CHO, M. A. 2013. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. *Journal of Applied Remote Sensing*, *7*.
- AERTS, R. & HONNAY, O. 2011. Forest restoration, biodiversity and ecosystem functioning. *J BMC ecology*, *11*, 29.
- AREKHI, M., YILMAZ, O. Y., YILMAZ, H. & AKYUZ, Y. F. 2017. Can tree species diversity be assessed with Landsat data in a temperate forest? *Environ Monit Assess*, *189*, 586.
- BELGIU, M. & DRĂGUȚ, L. 2016. Random forest in Remote Sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24-31.
- BOLYN, C., MICHEZ, A., GAUCHER, P., LEJEUNE, P. & BONNET, S. 2018. Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *Biotechnologie, Agronomie, Société et Environnement*, *22*, 16.
- BREIMAN, L. 2001. Random forests. *Machine learning*, *45*, 5-32.
- CARLSON, K. M., ASNER, G. P., HUGHES, R. F., OSTERTAG, R. & MARTIN, R. E. 2007. Hyperspectral Remote Sensing of Canopy Biodiversity in Hawaiian Lowland Rainforests. *Ecosystems*, *10*, 536-549.
- CHRYSAFIS, I., KORAKIS, G., KYRIAZOPOULOS, A. P. & MALLINIS, G. 2020. Predicting Tree Species Diversity Using Geodiversity and Sentinel-2 Multi-Seasonal Spectral Information. *Sustainability*, *12*.
- COLGAN, M., BALDECK, C., FÉRET, J.-B. & ASNER, G. 2012. Mapping Savanna Tree Species at Ecosystem Scales Using Support Vector Machine Classification and BRDF Correction on Airborne Hyperspectral and LiDAR Data. *Remote Sensing*, *4*, 3462-3480.
- DALY, A. J., BAETENS, J. M. & DE BAETS, B. J. M. 2018. Ecological diversity: measuring the unmeasurable. *6*, 119.
- DUBE, T. & MUTANGA, O. 2015. 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., 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 (Basel)*, *14*, 15348-70.
- DURO, D. C., FRANKLIN, S. E. & DUBÉ, M. G. 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of environment*, *118*, 259-272.

- EL-ASKARY, H., ABD EL-MAWLA, S., LI, J., EL-HATTAB, M. & EL-RAEY, M. 2014. Change detection of coral reef habitat using Landsat-5 TM, Landsat 7 ETM+ and Landsat 8 OLI data in the Red Sea (Hurghada, Egypt). *International journal of Remote Sensing*, 35, 2327-2346.
- EZEMVELO KZN WILDLIFE 2015. Nkandla Forest Complex: Protected Area Management Plan. *Unpublished* Version 1.0, 173
- FASSNACHT, F. E., HARTIG, F., LATIFI, H., BERGER, C., HERNÁNDEZ, J., CORVALÁN, P. & KOCH, B. 2014. Importance of sample size, data type and prediction method for Remote Sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment*, 154, 102-114.
- FASSNACHT, F. E., LATIFI, H., STEREŃCZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- FEILHAUER, H. & SCHMIDTLEIN, S. 2009. Mapping continuous fields of forest alpha and beta diversity. *Applied Vegetation Science* 12, 429–439.
- FOODY, G. M. & CUTLER, M. E. J. 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological Modelling*, 195, 37-42.
- FUNDISI, E., MUSAKWA, W., AHMED, F. B. & TEFAMICHAEL, S. G. 2020. Estimation of woody plant species diversity during a dry season in a savanna environment using the spectral and textural information derived from WorldView-2 imagery. *PLoS One*, 15, e0234158.
- GAMFELDT, L., SNÄLL, T., BAGCHI, R., JONSSON, M., GUSTAFSSON, L., KJELLANDER, P., RUIZ-JAEN, M. C., FRÖBERG, M., STENDAHL, J. & PHILIPSON, C. D. 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature communications*, 4, 1340.
- GANIVET, E., BLOOMBERG, M. J. F. E. & MANAGEMENT 2019. Towards rapid assessments of tree species diversity and structure in fragmented tropical forests: A review of perspectives offered by remotely-sensed and field-based data. 432, 40-53.
- GEORGE-CHACON, S. P., DUPUY, J. M., PEDUZZI, A. & HERNANDEZ-STEFANONI, J. L. 2019. Combining high resolution satellite imagery and lidar data to model woody species diversity of tropical dry forests. *Ecological Indicators*, 101, 975-984.
- GHOSH, A., FASSNACHT, F. E., JOSHI, P. K. & KOCH, B. 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, 26, 49-63.
- GHOSH, A. & JOSHI, P. K. 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26, 298-311.
- GILLESPIE, T. W., DE GOEDE, J., AGUILAR, L., JENERETTE, G. D., FRICKER, G. A., AVOLIO, M. L., PINCETL, S., JOHNSTON, T., CLARKE, L. W. & PATAKI, D. E. 2016. Predicting tree species richness in urban forests. *Urban Ecosystems*, 20, 839-849.
- GOULD, W. 2000. Remote Sensing of Vegetation, Plant Species Richness, and Regional Biodiversity Hotspots. *Ecological Applications*, 10, 1861-1870.
- GRABSKA, E., HOSTERT, P., PFLUGMACHER, D. & OSTAPOWICZ, K. 2019. Forest stand species mapping using the Sentinel-2 time series. *J Remote Sensing*, 11, 1197.
- GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2020. Mapping natural forest cover using satellite imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, 18, 1-14.

- HUANG, Y., MA, Y., ZHAO, K., NIKLAUS, P. A., SCHMID, B. & HE, J.-S. 2017. Positive effects of tree species diversity on litterfall quantity and quality along a secondary successional chronosequence in a subtropical forest. *Journal of Plant Ecology*, 10, 28-35.
- IFO, S. A., MOUTSAMBOTE, J.-M., KOUBOUANA, F., YOKA, J., NDZAI, S. F., BOUETOU-KADILAMIO, L. N. O., MAMPOUYA, H., JOURDAIN, C., BOCKO, Y. & MANTOTA, A. B. 2016. Tree species diversity, richness, and similarity in intact and degraded forest in the tropical rainforest of the Congo Basin: case of the forest of Likouala in the Republic of Congo. *International Journal of Forestry Research*, 2016.
- IMMITZER, M., NEUWIRTH, M., BÖCK, S., BRENNER, H., VUOLO, F. & ATZBERGER, C. 2019. Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *J Remote Sensing*, 11, 2599.
- IVERSON, L. R. & MCKENZIE, D. J. L. E. 2013. Tree-species range shifts in a changing climate: detecting, modeling, assisting. 28, 879-889.
- JOHN, R., CHEN, J., LU, N., GUO, K., LIANG, C., WEI, Y., NOORMETS, A., MA, K. & HAN, X. 2008. Predicting plant diversity based on Remote Sensing products in the semi-arid region of Inner Mongolia. *Remote Sensing of Environment*, 112, 2018-2032.
- KALACSKA, M., SANCHEZ-AZOFEIFA, G. A., RIVARD, B., CAELLI, T., WHITE, H. P. & CALVO-ALVARADO, J. C. 2007. Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy. *Remote Sensing of Environment*, 108, 82-96.
- KAVZOGLU, T. & MATHER, P. 2003. The use of backpropagating artificial neural networks in land cover classification. *J International journal of Remote Sensing*, 24, 4907-4938.
- KRISHNASWAMY, J., BAWA, K. S., GANESHAIAH, K. & KIRAN, M. 2009. Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. *Remote Sensing of Environment*, 113, 857-867.
- LAURIN, V. G., CHEUNG-WAI CHAN, J., CHEN, Q., LINDSELL, J. A., COOMES, D. A., GUERRIERO, L., DEL FRATE, F., MIGLIETTA, F. & VALENTINI, R. 2014. Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. *PLoS One*, 9, e97910.
- LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. *R news*, 2, 18-22.
- LU, D. 2006. The potential and challenge of Remote Sensing-based biomass estimation. *International Journal of Remote Sensing*, 27, 1297-1328.
- LUO, W., LIANG, J., CAZZOLLA GATTI, R., ZHAO, X. & ZHANG, C. J. J. O. E. 2019. Parameterization of biodiversity–productivity relationship and its scale dependency using georeferenced tree-level data. 107, 1106-1119.
- MADONSELA, S., CHO, M. A., RAMOELO, A. & MUTANGA, O. 2017. Remote Sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, 133, 116-127.
- MADONSELA, S., CHO, M. A., RAMOELO, A., MUTANGA, O. & NAIDOO, L. 2018. Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation*, 66, 106-115.
- MAEDA, E. E., HEISKANEN, J., THIJS, K. W. & PELLIKKA, P. K. E. 2014. Season-dependence of Remote Sensing indicators of tree species diversity. *Remote Sensing Letters*, 5, 404-412.
- MALLINIS, G., CHRYSAFIS, I., KORAKIS, G., PANA, E. & KYRIAZOPOULOS, A. P. 2020. A Random Forest Modelling Procedure for a Multi-Sensor Assessment of Tree Species Diversity. *Remote Sensing*, 12.

- MARTIN-GALLEGO, P., APLIN, P., MARSTON, C., ALTAMIRANO, A. & PAUCHARD, A. 2020. Detecting and modelling alien tree presence using Sentinel-2 satellite imagery in Chile's temperate forests. *Forest Ecology Management*, 474, 118353.
- MILLARD, K. & RICHARDSON, M. 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, 7, 8489-8515.
- MILLENNIUM ECOSYSTEM ASSESSMENT, M. J. S. 2005. Ecosystems and human well-being.
- MOHAMMADI, J. & SHATAEE, S. 2010. Possibility investigation of tree diversity mapping using Landsat ETM+ data in the Hyrcanian forests of Iran. *Remote Sensing of Environment*, 114, 1504-1512.
- MORRIS, E. K., CARUSO, T., BUSCOT, F., FISCHER, M., HANCOCK, C., MAIER, T. S., MEINERS, T., MÜLLER, C., OBERMAIER, E. & PRATI, D. 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology evolution*, 4, 3514-3524.
- MUTANGA, O. & SKIDMORE, A. K. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25, 3999-4014.
- MUTOWO, G. & MURWIRA, A. 2012a. Relationship between remotely sensed variables and tree species diversity in savanna woodlands of Southern Africa. *International Journal of Remote Sensing*, 33, 6378-6402.
- MUTOWO, G. & MURWIRA, A. 2012b. The spatial prediction of tree species diversity in savanna woodlands of Southern Africa. *Geocarto International*, 27, 627-645.
- MUTOWO, G., MUTANGA, O. & MASOCHA, M. 2018. Evaluating the Applications of the Near-Infrared Region in Mapping Foliar N in the Miombo Woodlands. *Remote Sensing*, 10.
- NAGENDRA, H. 2001. Using Remote Sensing to assess biodiversity. *International Journal of Remote Sensing*, 22, 2377-2400.
- NAGENDRA, H., LUCAS, R., HONRADO, J. P., JONGMAN, R. H. G., TARANTINO, C., ADAMO, M. & MAIROTA, P. 2013. Remote Sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, 33, 45-59.
- NITZE, I., BARRETT, B. & CAWKWELL, F. 2015. Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation*, 34, 136-146.
- OINDO, B. O. & SKIDMORE, A. K. 2010. Interannual variability of NDVI and species richness in Kenya. *International Journal of Remote Sensing*, 23, 285-298.
- OLDELAND, J., WESULS, D., ROCCHINI, D., SCHMIDT, M. & JÜRGENS, N. 2010. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecological Indicators*, 10, 390-396.
- OTUNGA, C., ODINDI, J., MUTANGA, O. & ADJORLOLO, C. 2018. Evaluating the potential of the red edge channel for C3 (*Festuca* spp.) grass discrimination using Sentinel-2 and Rapid Eye satellite image data. *Geocarto International*, 34, 1123-1143.
- OUYANG, S., XIANG, W., WANG, X., ZENG, Y., LEI, P., DENG, X. & PENG, C. 2016. Significant effects of biodiversity on forest biomass during the succession of subtropical forest in south China. *Forest Ecology and Management*, 372, 291-302.
- OZDEMIR, I., NORTON, D. A., OZKAN, U. Y., MERT, A. & SENTURK, O. 2008. Estimation of tree size diversity using object oriented texture analysis and aster imagery. *Sensors*, 8, 4709-4724.

- PAHLEVAN, N., LEE, Z., WEI, J., SCHAAF, C. B., SCHOTT, J. R. & BERK, A. 2014. On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic Remote Sensing. *Remote Sensing of Environment*, 154, 272-284.
- PARSON, E. A. 2013. Climate Engineering in Global Climate Governance: Implications for Participation and Linkage. *Transnational Environmental Law*, 3, 89-110.
- PAU, S., GILLESPIE, T. W. & WOLKOVICH, E. M. 2012. Dissecting NDVI–species richness relationships in Hawaiian dry forests. *Journal of Biogeography*, 39, 1678-1686.
- PEDRO, M. S., RAMMER, W. & SEIDL, R. 2015. Tree species diversity mitigates disturbance impacts on the forest carbon cycle. *Oecologia*, 177, 619-630.
- PERSSON, M., LINDBERG, E. & REESE, H. 2018. Tree species classification with multi-temporal Sentinel-2 data. *J Remote Sensing*, 10, 1794.
- RAJAH, P., ODINDI, J., MUTANGA, O. & KIALA, Z. 2019. The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. *Nature Conservation*, 35, 41-61.
- RAMOELO, A., CHO, M. A., MATHIEU, R., MADONSELA, S., VAN DE KERCHOVE, R., KASZTA, Z. & WOLFF, E. 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation Geoinformation*, 43, 43-54.
- ROCCHINI, D., RICOTTA, C. & CHIARUCCI, A. 2007. Using satellite imagery to assess plant species richness: The role of multispectral systems. *J Applied Vegetation Science*, 10, 325-331.
- RODRÍGUEZ-GALIANO, V. F., ABARCA-HERNÁNDEZ, F., GHIMIRE, B., CHICA-OLMO, M., ATKINSON, P. M. & JEGANATHAN, C. 2011. Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environmental Sciences*, 3, 44-49.
- SCHÄFER, E., HEISKANEN, J., HEIKINHEIMO, V. & PELLIKKA, P. 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. *Ecological Indicators*, 64, 49-58.
- SHANNON, C. 1948. A mathematical theory of communication. *J The Bell system technical journal*, 27, 379-423.
- SIMPSON, E. 1949. Measurement of diversity. *J nature*, 163, 688-688.
- SUN, Y., HUANG, J., AO, Z., LAO, D. & XIN, Q. 2019. Deep Learning Approaches for the Mapping of Tree Species Diversity in a Tropical Wetland Using Airborne LiDAR and High-Spatial-Resolution Remote Sensing Images. *Forests*, 10.
- SUN, Z., LIU, X., SCHMID, B., BRUELHEIDE, H., BU, W. & MA, K. 2017. Positive effects of tree species richness on fine-root production in a subtropical forest in SE-China. *Journal of Plant Ecology*, 10, 146-157.
- TEAM, R. D. C. 2013. R: A language and environment for statistical computing.
- TIGGES, J., LAKES, T. & HOSTERT, P. 2013. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. *Remote Sensing of Environment*, 136, 66-75.
- TODD, S., HOFFER, R. & MILCHUNAS, D. 1998. Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing*, 19, 427-438.
- TUOMINEN, S., NÄSI, R., HONKAVAARA, E., BALAZS, A., HAKALA, T., VILJANEN, N., PÖLÖNEN, I., SAARI, H. & OJANEN, H. 2018. Assessment of classifiers and Remote Sensing features of hyperspectral imagery and stereo-photogrammetric point clouds for recognition of tree species in a forest area of high species diversity. *Remote Sensing*, 10, 714.

- TURNER, W., SPECTOR, S., GARDINER, N., FLADELAND, M., STERLING, E. & STEININGER, M. 2003. Remote Sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18, 306-314.
- USTUNER, M., SANLI, F. B. & ABDIKAN, S. 2016. Balanced Vs Imbalanced Training Data: Classifying Rapideye Data with Support Vector Machines. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B7, 379-384.
- VAN DEVENTER, H., CHO, M. A. & MUTANGA, O. 2017. Improving the classification of six evergreen subtropical tree species with multi-season data from leaf spectra simulated to WorldView-2 and RapidEye. *International Journal of Remote Sensing*, 38, 4804-4830.
- WALLNER, A., ELATAWNEH, A., SCHNEIDER, T. & KNOKE, T. 2014. Estimation of forest structural information using RapidEye satellite data. *Forestry*, 88, 96-107.
- WANG, D., WAN, B., QIU, P., SU, Y., GUO, Q., WANG, R., SUN, F. & WU, X. 2018a. Evaluating the Performance of Sentinel-2, Landsat 8 and Pléiades-1 in Mapping Mangrove Extent and Species. *Remote Sensing*, 10.
- WANG, R. & GAMON, J. A. 2019. Remote Sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, 231, 111218.
- WANG, R., GAMON, J. A., SCHWEIGER, A. K., CAVENDER-BARES, J., TOWNSEND, P. A., ZYGIELBAUM, A. I. & KOTHARI, S. 2018b. Influence of species richness, evenness, and composition on optical diversity: A simulation study. *J Remote Sensing of Environment*, 211, 218-228.
- ZHAO, Y., ZENG, Y., ZHENG, Z., DONG, W., ZHAO, D., WU, B. & ZHAO, Q. 2018. Forest species diversity mapping using airborne LiDAR and hyperspectral data in a subtropical forest in China. *Remote Sensing of Environment*, 213, 104-114.

**CHAPTER 7: IDENTIFYING THE BEST SEASON FOR PREDICTING TREE
SPECIES DIVERSITY USING SENTINEL 2 SATELLITE IMAGERY AND
RANDOM FOREST ALGORITHM**

Abstract

Tree species diversity contributes to the functional traits of forest ecosystems and is an important measure of forest health. Therefore, Remote Sensing prediction of tree species diversity is essential because it provides useful information for sustainable forest tree species diversity management. Remote sensors record different spectral signatures of tree species within the various seasons of the year thereby affecting the performance of satellite imagery models used to predict tree species diversity. Information on the best season is very important for tree species diversity prediction. However, there is currently no information on the best season for predicting tree species diversity for the sub-tropical natural forests in South Africa. Thus, our study aimed at identifying the best season for tree species diversity prediction by evaluating the performance of Sentinel 2 imageries captured in summer, spring, autumn and winter. The Random Forest (RF) regression algorithm was utilized to model and identify the best season using texture variables derived from Sentinel-2 imageries of each of the four seasons. The texture variables derived from the Sentinel-2 imagery were used as independent variables, and the Shannon Diversity Index (H') values derived from the field survey was used as the dependent variable for the models. The summer imagery outperformed the spring, autumn, and winter imageries by a small margin. Its coefficient of determination (R^2) was 0.94, and its Root Mean Squared Error (RMSE) was 0.130. The spring imagery emerged as the second-best while the autumn imagery followed as the third best. The winter imagery was the last best imagery. The result indicates that the best season for tree species diversity prediction of sub-tropical natural forests using Remote Sensing imagery could be the summer. The findings have vital implications for forest managers and researchers when estimating tree species diversity with satellite imageries.

Keyword: Species, Diversity, Imagery, Prediction, Random Forest, Accuracy, Sentinel 2, Management

7.1 Introduction

Tree species diversity is one of the means of measuring forest health (Arekhi et al., 2017) and it contributes significantly to the ecological process that enhances the net growth of a forest. The functional traits of forests are enhanced by tree species diversity and the intactness and level of diversity facilitate the provision of multiple ecological benefits (Gamfeldt et al., 2013). Among these ecological benefits includes the regulation of water, carbon and biogeochemical cycles (Millennium Ecosystem Assessment, 2005). Although there are natural threats to tree species diversity, human-induced factors are the major factors of decline (Zhao et al., 2018). They are observed through deforestation, homogenization of ecosystems through the planting of single and mostly exotic species and agricultural land expansion (Laurin et al., 2014). Reliable information is needed in both the local management and conservation efforts and as well as for the international community in the global policy direction for conserving biodiversity (Sinton, 2017). Therefore, estimation and monitoring of tree species diversity is a requirement for mitigation measures to biodiversity losses and effective management, conservation and protection of forest ecosystems (Chrysafis et al., 2020). The estimation of

tree diversity is a worthy operation in enhancing the availability of spatial distribution information in identifying hotspots and the level of diversity to support decision making.

Effective methods and technology are essential for the assessment of diversity at all levels in providing adequate information for tree species management and conservation (Schäfer et al., 2016). It is especially true and vital in the tropical and subtropical zones that are highly diverse (Gaston, 2000) and where there is a poor understanding of tree species diversity and its importance to the forests (Milliken et al., 2010). Remote Sensing has proven to have the capability to map the spatial and temporal distribution diversity in various forest types (Carlson et al., 2007, John et al., 2008, Wallis et al., 2017, Zhao et al., 2018).

Promising results have been reported in Remote Sensing modelling and prediction of tree species diversity in various ecological zones and forest types. For instance; Feilhauer and Schmidtlein (2009a) employed the ASTER imagery in Kyrgyzstan to predict the alpha diversity among tree species. The Partial Least Squares Regression (PLSR) was used for the modelling with the ASTER spectral products and digital elevation model features (elevation, slope, aspect) used as input variables. The model was able to explain 61% of the variance. Mutowo and Murwira (2012a) similarly utilized the ASTER spectral variables in assessing the relationship between them and the tree species diversity for the savanna woodland in South Africa. The standard deviation of the near-infrared band (stdev NIR) and the soil adjusted vegetation index (SAVI) were used as predicting variables in a linear regression model for the prediction. The combined stdev NIR and the SAVI could explain between 60% and 64% of the variance in the diversity, which was better than when they were used on a singular basis. Madonsela et al. (2018) evaluated the capabilities of the Normalized Difference Vegetation Index (NDVI) and the woody canopy cover (WCC) for predicting tree species diversity in the savanna zone of South Africa using Landsat 8. The relationship between the tree species diversity and the WCC was evaluated in the first model while the second model evaluated the relationship between the combined WCC and the NDVI for diversity prediction in a factorial model. The findings of the study indicated that there was a significant relationship between WCC and diversity. The relationship between the combined WCC and the NDVI and the tree species diversity was found to be insignificant.

The high-resolution imaging spectroscopy data and light detection and ranging (LiDAR) have also been used for tree species diversity prediction. The Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) was employed to map species richness in Hawaii (Carlson et al., 2007). The spatial and spectral derivatives of the portions of the electromagnetic spectrum that relates to biochemical properties such as water nitrogen content were assessed. The linear regression and the combined wavelengths linked with biochemical analysis were used to establish the relationship between spectral reflectance and species richness. Lastly, Sun et al. (2019) utilized the LiDAR and VHR-RGB data together with modified deep learning methods (AlexNet, VGG16, and ResNet50) for estimation of the diversity in a tropical wetland in Southern China. The prediction was based on four diversity indices (the Shannon diversity index, the Simpson diversity index, the Margalef richness index and the Pielou evenness index) calculated from field data. The highest overall accuracy was produced by the VGG16, displaying the potential of deep learning for tree species diversity prediction.

Some studies also considered the effects of seasons on the prediction of tree species diversity. Maeda et al. (2014) conducted in the Afromontane forest of Taita hills in Kenya used 15

Landsat time-series data to ascertain the extent to which the relation between Remote Sensing indicators and the tree species diversity depends on the season of the year. Through the simple and multivariate regression modelling, the authors established that the season of the year influences the relationship between Remote Sensing metrics and tree species diversity. The recently launched Sentinel 2 imagery was applied together with Geodiversity data (aspect and elevation) to assess multi-seasonal and single-season diversity prediction in the Mediterranean region (Chrysafis et al., 2020). The findings indicated that multi-seasonal models produced the highest accuracies, while single-season models of mid-summer and mid-autumn produced the second-best accuracies. Furthermore, the study reported that Geodiversity data had very little influence on the accuracies.

Each season of the year exhibits climatic conditions that cause changes in the phenology and morphology of tree species. Tree's phenology denotes the biological processes in trees, which include flowering, budburst, seed set, seed dispersal and leaf fall that relates to prevailing climatic conditions (Davi et al., 2011, Yang et al., 2017). On the other hand, tree morphology relates to the physical structures such as roots, stems, branches, twigs as well as leaves, flowers and fruits. The changes in phenological and morphological traits in trees due to prevailing conditions are likely to influence how they are perceived by sensors and affect the spectral signatures that are recorded. Thus, the performance and accuracy produced by the predicting model of a sensor could also be influenced by the phenological and morphological cycles, depending on the season of the year the imagery is captured. For instance, Sentinel 2 imageries captured in each of the four seasons, that is summer, winter, spring and autumn may produce high or low accuracy depending on the season in which they were captured. Although there is a possibility of the performance being affected, many studies that predict tree species diversity do not take the season imageries are captured into consideration. This is because there is currently no information on the best season for which tree species diversity prediction could be undertaken. Some limitations in the performance of the predicting models of imageries could be observed as a result. Therefore, our study focuses on identifying the best season for predicting tree species diversity in a sub-tropical natural forest in South Africa. It is expected that high accuracies could be produced by the models of satellite imageries captured in the best season. Also, it could be a piece of useful information for ecologists, forest managers and Remote Sensing scientists in conducting such studies in natural forests.

7.2 Materials and Methods

7.2.1 Study Area

The subtropical Nkandla natural forest reserve was established in 1918 and it is found in the KwaZulu-Natal (KZN) province of South Africa. It has a total area of about 2,217 ha and is located on 28° 43' 50.88" S and 30° 7' 9.84" E (Figure 7.1). It experiences the highest average temperature of 27°C between the summer months of December and January and records the lowest average temperature of 2°C in the winter months of June and July (Ezemvelo KZN Wildlife, 2015b). It has a steep and undulating topography across the entire forest. Its altitude is at the lowest level of 500 m and exceeds 1300 m at the highest range. The forest has been classified into four land cover types, comprising of closed canopy forest, open canopy forest, grassland and bare sites (Gyamfi-Ampadu et al., 2020).

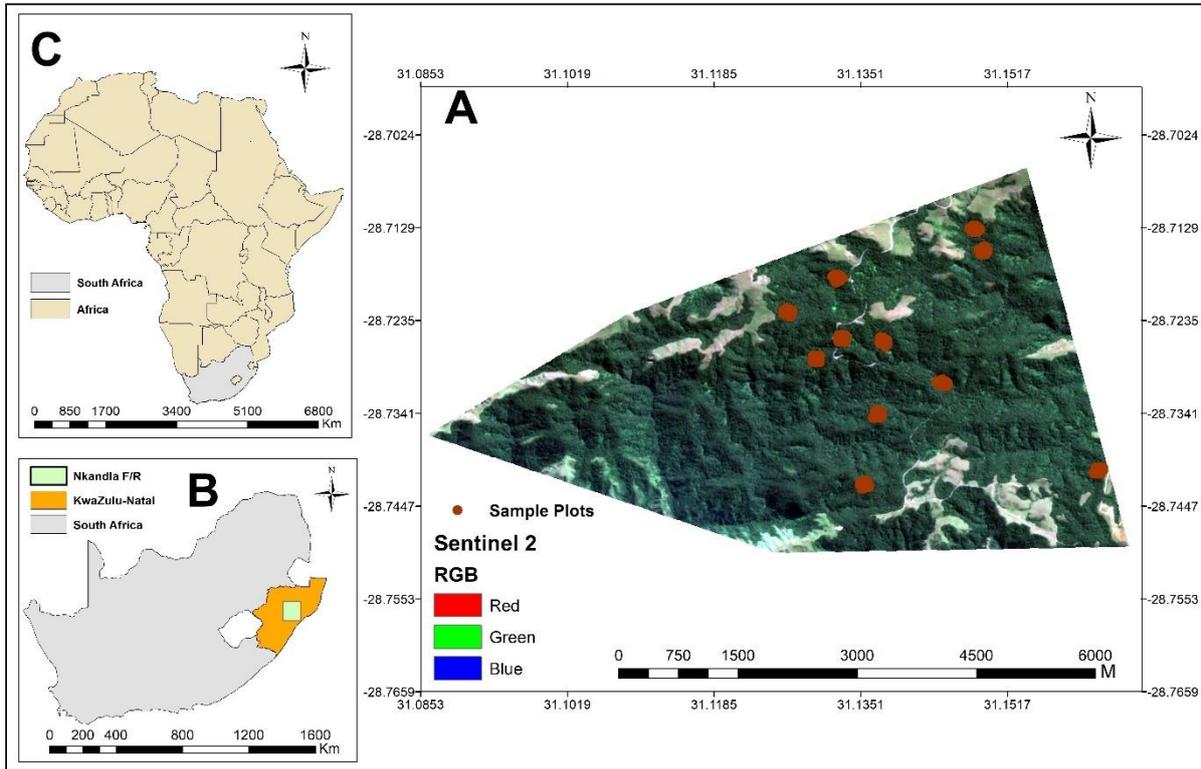


Figure 7.1: Map of the study area. Note: A represents the Nkandla Forest Reserve, B is a map of South Africa indicating the location of KZN province and the forest reserve and C is an Africa indicating the location of South Africa.

7.3 Tree Data Collection and Diversity Estimation

The tree inventory survey was undertaken between 24 April 2019 and 07 May 2019. Parts of the middle and western section of the forest were found to be inaccessible mainly due to deep valleys. As such, the tree information was collected from the accessible parts that have gentle slopes (Figure 7.1). A systematic approach was used in setting up the plots in gentle slope after following existing transects. Firstly, eleven 1 ha (100 m x 100 m) plots were set up in the gentle slope areas. After that, each of the plots was divided into 25 subplots of 20 m x 20 m. A total of 275 plots were obtained from the subplots and the tree information was subsequently collected from them. The diameter at breast height (DBH) of all trees ≥ 5 cm was measured and recorded in each of the plots. The local and scientific names as well as the coordinate of each tree was also recorded. The sampling design and the collection of the data in the accessible parts of the forest were not a compromise on the data as similar tree species were recorded in the majority of the sampling plots. The tree data was used to compute the Shannon Diversity Index [H'] (Shannon, 1948) and used as the dependent variable for the prediction. The Shannon Diversity Index accounts for species evenness, abundance and rarity (Ifo et al., 2016, Morris et al., 2014). The mathematical function used for the diversity indices have been expressed in equations 1:

$$H' = - \sum_{i=1}^S p_i \times \ln(p_i) \quad (7.1)$$

Where p_i is the proportionate abundance of the i th species in the sampling plot, S is the total abundance of all species in a sampling plot, and \ln is the natural logarithm of the proportionate abundance of species in the sampling plot.

7.4 Remote Sensing imagery data

We used the Sentinel 2 imagery for our study because it has provided good accuracies in many other vegetation studies (Immitzer et al., 2019, Mutowo et al., 2018a, Pandit et al., 2018a, Grabska et al., 2019a). Additionally, Sentinel 2 imagery was used because it produced the best accuracy among four satellite imageries that were used to assess the effects of spatial and spectral resolution on tree species diversity prediction (Gyamfi-Ampadu et al., 2021). It has 13 spectral bands that range from the visible to shortwave infrared (SWIR) regions of the electromagnetic spectrum and the spectral bands have varying spatial resolutions. Since our study sought to identify the best season for predicting tree species diversity, four imageries of Sentinel 2 were downloaded from the United State Geological Services [USGS] website (www.usgs.gov). The four imageries were made up of an imagery each for summer, autumn, winter and spring yearly seasons. The summer imagery was captured on 29th January 2020, autumn on 14th April 2019, winter on 18th July 2019 and spring on 6th October 2019. It is important to note that the date of capture of each of the imageries was in the middle of the season. The seasons are at their peak around that period. The imageries were atmospherically corrected with the semi-automatic plugin of the QGIS 3.6.1 software. The bands 1, 9 and 10 were excluded from the bands used for the analysis as they contain aerosols, water vapour and cloud information. The details of the 10 spectral bands used in the study are found in Table 7.1.

Table 7.1: Details of Sentinel 2 spectral bands used in the analysis

Band	Name/Code	Central Wavelength	Spatial Resolution (m)
B2	Blue	0.490	10
B3	Green	0.560	10
B4	Red	0.665	10
B5	Red-edge 1	0.705	20
B6	Red-edge 2	0.740	20
B7	Red-edge 3	0.783	20
B8	Near Infrared	0.842	10
B8a	Near Infrared 2	0.865	20
B11	Shortwave Infrared 1	1.610	20
B12	Shortwave Infrared 2	2.190	20

7.5 Deriving of texture variables

The variables used as inputs for predicting models is an important factor that determines the accuracy of tree species diversity studies. One of such is texture variables that are derived from texture analysis. Texture analysis refers to mathematical functions, procedures and models applied to extracting spatial information on earth features from imageries (Srinivasan and Shobha, 2008). It defines the local spatial organization of spectral values varying spatially that are repeated in a region of bigger spatial scales (Srinivasan and Shobha, 2008). Hence, the image texture contains vital information on spatial and structural information about earth surface features (Coburn and Roberts, 2004, Kayitakire et al., 2006). In tree species diversity prediction, the spatial and structural information could be of value as the pixel size may determine the level and number of species that could be captured within. The spectral and spatial resolution of satellite imageries influences the accuracies of tree species diversity predictions (Gyamfi-Ampadu et al., 2021). Therefore, variables that are derived for spatial information of spectral bands could enhance the accuracy output.

Texture variables were derived from each set of 10 spectral bands of Sentinel 2 imagery of each of the four seasons and used as the independent (predicting) variables. Seven texture variables (Contrast, Dissimilarity, Homogeneity, Entropy, Grey-level Cooccurrence Matrix (GLCM) Mean, GLCM Variance, and GLCM Correlation) were derived using the 5 x 5, 7 x 7, and 9 x 9 window size for each season. The mathematical functions of each of the variables and their description can be found in Table 7.2.

Table 7.2: Mathematical functions of the texture variables and their descriptions

Texture variable	Mathematical function	Description
Contrast	$\sum_{ij=0}^{N-1} P_{ij} (i - j)^2$	It measures the overall amount of a window's local variation. (Yuan et al., 1991)
Dissimilarity	$\sum_{ij=0}^{N-1} P_{ij i-j }$	It is similar to contrast. However, the difference is that where there is an exponential increase in contrast weights in a diagonal movement, dissimilarity weight increases linearly.
Homogeneity	$\sum_{ij=1}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$	It is a measure of the smoothness of image texture. When there is a large change in spectral values, it results in small homogeneity values, while a small change results in larger homogeneity values. (Tuttle et al., 2006)
Entropy	$-\sum_{ij=1}^{N-1} P_{ij} \ln(P_{ij})$	The disorder in the imagery is measured by Entropy. Many GLCM features have small values when the image is not uniform, implying that Entropy is very large (Baraldi and Pannigiani, 1995).
GLCM Variance	$\sigma^2 = \sum_{ij=1}^{N-1} P_{ij} (i - \mu_i)^2$ $\sigma^2 = \sum_{ij=1}^{N-1} P_{ij} (i - \mu_j)^2$	The variability of the spectral response of the pixel is accounted for by Variance (Tuttle et al., 2006). It takes into consideration the pairwise combination of the variability
GLCM Mean	$\frac{1}{2} \left[\sum_{ij=1}^{N-1} iP(i-j) + jP_{ij} \right]$	The average grey level in the local window represented by the Mean (Pacifici et al., 2009).

GLCM Correlation	$\sum_{ij=1}^{N-1} P_{ij} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$	It measures the grey level linear dependency within an imagery (Kayitakire et al., 2006).
------------------	--	---

Note: P is the texture index while i and j refer to the texture pixels. σ and μ represent the mean and standard deviation, respectively. N is the dimension of the co-occurrence matrix.

After the derivation process, the summer, spring, autumn and winter imageries had a total of 70 texture variables each for every window (5x5, 7 x 7 and 9 x 9). In order to ascertain the best window to use for the prediction, a pre-testing modelling was carried for all the four imageries through RF regression model analysis using the texture variables derived for each of the 5 x 5, 7 x 7, and 9 x9 windows. The 7 x 7 window emerge as the best after the pre-testing modelling. Hence, the 7 x 7 window was subsequently used for each of the four imageries' models.

7.5.1 Predicting Variables Selection

Important predicting variables were selected from each set of 70 texture variables of the summer, autumn, winter and spring imageries using the Recursive Feature Elimination (RFE) algorithm. The benefits of this operation are that it helps to reduce computational complexities, reduce redundancy, and eliminate noisy variables (Ghosh and Joshi, 2014, Wang et al., 2018a). Out of the 70 predicting variables of a single season, the top five predicting variables were selected. The same process was executed for all the remaining seasons.

The Band-6-Entropy, Band-12-Entropy, Band-8A-Entropy, Band-8-Homogeneity and Band-5-GLCM variance were selected for the summer imagery. The Band-5-Contrast, Band-5-GLCM correlation, Band-5-Dissimilarity, Band-6-Homogeneity and Band-4-Contrast were selected for the autumn imagery. The Band-6-GLCM correlation, Band-8A-Homogeneity, Band-6-Entropy, Band-7-Contrast and Band8A-Entropy were selected for the winter imagery, while the Band-11-Entropy, Band-2-Entropy, Band-11-GLCM correlation, Band-7-Entropy and Band-3-Homogeneity were selected for the spring imagery.

7.5.2 Random Forest regression models

The Random Forest is an ensemble-based non-parametric algorithm that operates through the production of many decision trees by using a subset of training variables that are randomly selected (Breiman, 2001, Belgiu and Drăguț, 2016). It can be used both for classification and regression analysis. It uses a bagging method by utilizing the randomly selected samples to generate a forest, built on Classification and Regression Trees [CART] (Abdel-Rahman et al., 2013a). Generally, two-thirds of the samples are used in training the trees, while the remaining one-third (out of the bag [OOB]) is used for internal cross-validation. This mostly determines the performance of the RF model (Breiman, 2001, Rodríguez-Galiano et al., 2011). The RF algorithm can handle a complex number of variables, it does not overfit data and able to deal with multicollinearity in modelling (Abdel-Rahman et al., 2013a, Ramoelo et al., 2015). Some internal parameters that contribute to the performance are the $mtry$ and $nree$. These are user-

defined parameters that could be optimized (tuned) or left in default values. It has been found that the model accuracies are more sensitive to the *mtry* than the *ntree* (Ghosh and Joshi, 2014).

The RF model was implemented and analyzed for each of the four imageries using the “randomForest” package (Liaw and Wiener, 2002). Extraction of the pixel values was carried out for each imagery model. In a random selection approach, 275 sample values for Shannon Index (H') were subsequently partitioned into training samples of 70% (192) and validation samples of 30% (83) for each of the four imageries. A tuning of the *mtry* and *ntree* parameters was carried out to determine the best values that could be used in each model. The *tuneRF* function was used to obtain the best *mtry* value in the R statistical software environment (Team, 2013). The *mtry* values that were obtained for the summer, autumn, winter and spring were 2, 1, 1 and 2 respectively. The *ntree* values of 100 to 1000 at an interval of 100 was tested to obtain the best value for each imagery model. After the testing, the best values obtained for summer, autumn, winter and spring imageries were 900, 700, 600, and 700 respectively. A 500 bootstrapping iterations were used to predict the tree species diversity based on the derived Shannon Index (H') after training the model with the 70% sample. The 30% validation set of each of the four seasons was then used to validate the prediction accuracies for each of them.

7.5.3 Model evaluation and tree species diversity map production

The performance of the four season models was evaluated with the coefficient of determination (R^2) and the root mean squared error (RMSE). The model that produced the highest R^2 and lowest RMSE was the most accurate among the others.

A tree species diversity map was produced by the imagery that produced the best accuracy. This map is meant to provide a spatial distribution of tree species diversity across the forest by showing areas of high and low diversity.

7.5.4 Importance Variables Evaluation and Diversity Map Production

The importance of each of the predicting variables of each of the four seasons was evaluated and ranked for their significance and contribution to the accuracies obtained for each imagery. This was done by using the variable importance feature in the RF algorithm. The determination of the importance of each variable was based on the percentage increase in mean squared error (%IncMSE). The %IncMSE expresses the effect of a variable used in a model when it is eliminated from it.

7.6 Results

7.6.1 Field Measured Data Analysis

Table 7.3 presents the descriptive statistics of the Shannon Diversity Index (H') values derived from the field measurement data of each sample plot. These values were used as the dependent variable of the RF models.

Table 7.3: Descriptive statistics of the field inventory data.

Parameter value	Statistical values
Mean	2.055
Minimum	0.949
Maximum	2.718
Standard deviation	0.290

7.6.2 Image Model Performance

The best imagery was determined by the highest R^2 and the lowest RMSE. The accuracies of all the imageries have been presented in Table 7.4. The summer imagery model produced the overall best accuracy with a slightly higher accuracy value ($R^2 = 0.94$, RMSE = 0.130) while the spring imagery produced the second-best accuracy ($R^2 = 0.93$, RMSE = 0.138). The winter and autumn imageries had a comparable R^2 of 0.92. Among the two, the autumn imagery produced a lower RMSE of 0.138 making it the third-best imagery. The winter imagery was hence the last performing imagery model due to a higher RMSE of 0.144.

Table 7.4: RF model accuracies of the four seasonal imageries.

Seasonal Image	Accuracies	
	R^2	RMSE
Summer	0.94	0.130
Spring	0.93	0.138
Autumn	0.92	0.138
Winter	0.92	0.144

A scatter plot was produced to illustrate the relationship between the field measured values and the predictions made by RF models of the four seasons imageries (Figure 7.2).

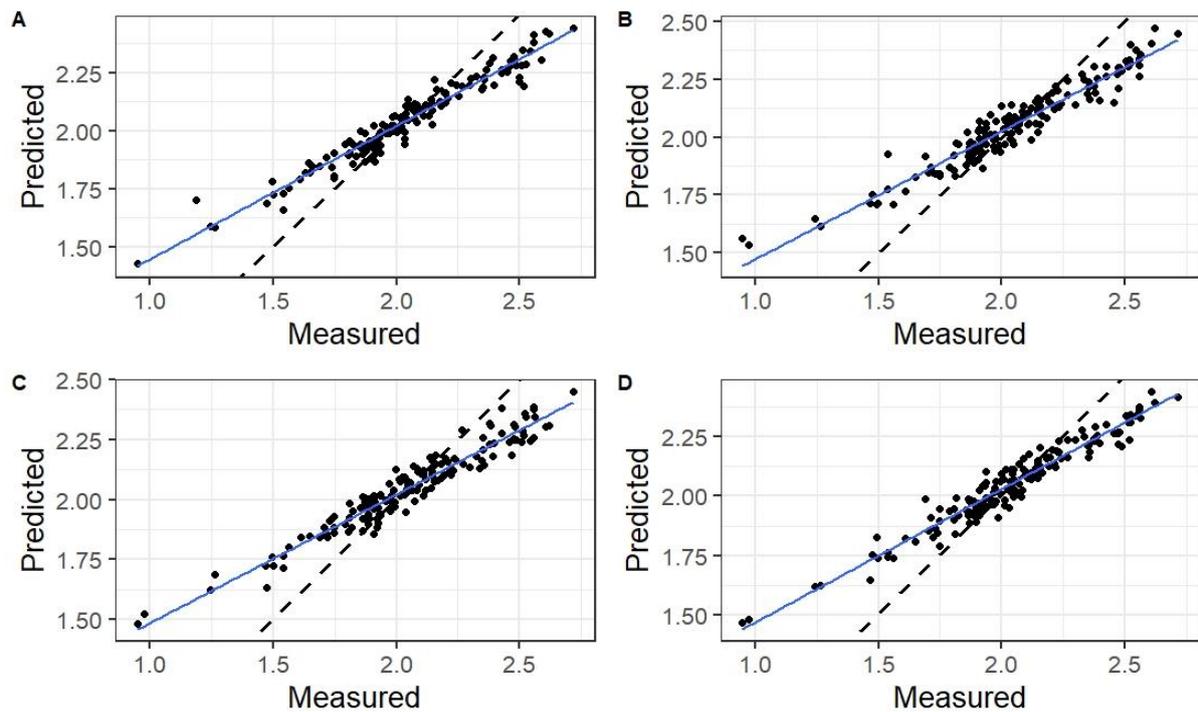


Figure 7.2: Scatter plots for the tree species diversity prediction. (A) is for the Summer imagery, (B) is for Spring imagery, (C) is for Autumn imagery, and (D) is for Winter imagery. The blue line is the line of best-fit while the dashed line is the 1:1 as illustrated in the individual plots.

A map that illustrates the tree species diversity distribution for the Nkandla Forest Reserve was produced with the summer imagery since it produced the best accuracy (Figure 7.3).

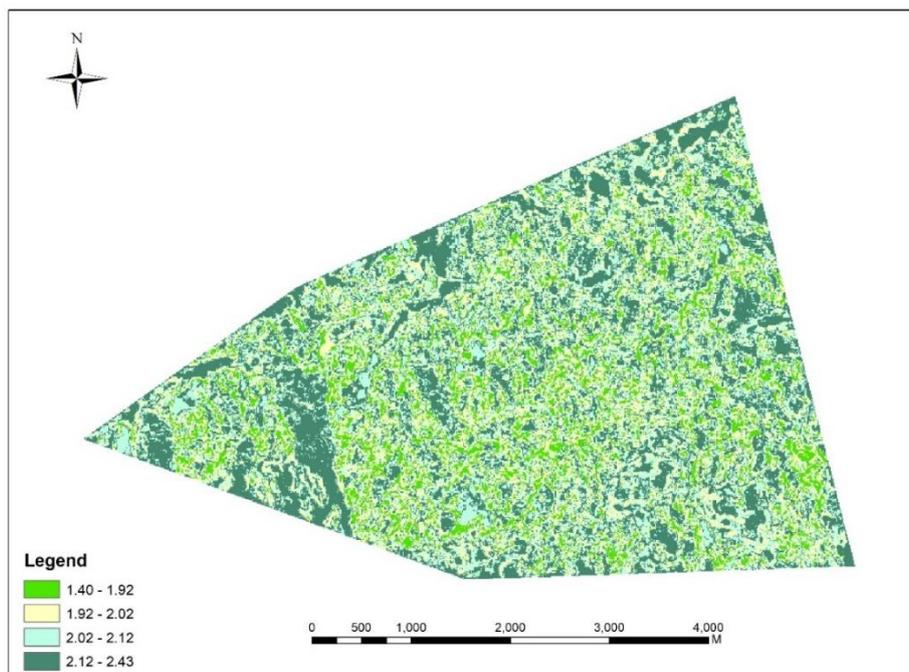


Figure 7.3: The diversity map produced for the Nkandla Forest Reserve.

7.6.3 Important variables

With the help of the variable importance feature in the RF algorithm, the predicting variables were ranked based on their contribution to the accuracy of each imagery model. The ranking of the significance of each variable was determined by the percentage increase in mean square error (%IncMSE). The Band-8-Homogeneity was the most important variable for the summer imagery while the Band-6-Entropy, Band-12-Entropy, Band-8A-Entropy and Band-5-GLCMV (Variance) followed from second to the least important respectively (Table 7.5)

Table 7.5: The variable importance ranking of the Summer imagery.

Predicting texture variable	%IncMSE
Band-8-Homogeneity	27.93
Band-6-Entropy	24.74
Band-12-Entropy	22.58
Band-8A-Entropy	21.75
Band-5-GLCMV	12.20

Concerning the spring imagery, the Band-11-Entropy was ranked as the most important variable. The Band-11-GLCM Correlation was second-best whereas the was Band-2-Entropy third-best variable (Table 7.6). The Band-3-Homogeneity was fourth, while the Band-7-Entropy was least important.

Table 7.6: The variable importance ranking of the Spring imagery.

Predicting texture variable	%IncMSE
Band-11-Entropy	17.70
Band-11-GLCMC (Correlation)	12.94
Band-2-Entropy	10.93
Band-3-Homogeneity	10.16
Band-7-Entropy	5.57

The Autumn imagery had the Band-5-Contrast to be the most important variable and the Band-5-Dissimilarity was second-best. The third to least were Band-5-GLCM Correlation, Band-4-Contrast and Band-6-Homogeneity respectively (Table 7.7).

Table 7.7: Variable importance ranking for Autumn imagery.

Predicting texture variable	%IncMSE
Band-5-Contrast	20.10
Band-5-Dissimilarity	19.94
Band-5-GLCMC (Correlation)	19.62
Band-7-Contrast	21.75
Band-5-GLCM variance	12.20

With regards to the winter imagery, the Band-7-Contrast emerged as the most important variable and the Band-6-GLCM correlation was the second-best. The Band-8A-Homogeneity was the third-best while the Band-8A-Homogeneity was the fourth and the Band-6-Entropy was the least important (Table 7.8).

Table 7.8: Variable importance ranking for Winter imagery.

Predicting texture variable	%IncMSE
Band-7-Contrast	25.56
Band-5-GLCMC (Correlation)	14.99
Band-8A-Entropy	14.39
Band-8A-Homogeneity	14.39
Band-6-Entropy	13.81

7.7 Discussion

The four seasons of the year relate to changes in climatic conditions and are important in determining the phenological cycles, morphology and senescence period of trees and vegetation in general. The leaf and canopy density is related to the phenology and morphology of trees and is likely to be affected, depending on the season of the year. The phenological process involves leaf budding, leaf flushing, fruiting and leaf fall in relation to the seasons of the year due to climatic variations (Kikim and Yadava, 2001, Davies et al., 2011b). The phenological pattern of tree species exhibits a strong seasonality in subtropical forest ecosystems (Kikim and Yadava, 2001), as a result of the prevailing climatic conditions. It is worth noting that sensors specific phenological stages and processes are not recorded by satellite sensors but rather general measurements of activities and growth of vegetation (Atkinson et al., 2012). Therefore, the spectral reflectance may be influenced by the seasonal changes experienced by tree species, thus determining how vegetation is perceived and recorded by the sensor. The spectral signature can be significantly changed and influenced by the condition of the vegetation at certain seasons of the year (Maeda et al., 2014). The morphology of natural forests trees species may either remain the same or change at different seasons and it may as well affect the spectral reflectance of imageries (White et al., 1997).

Therefore, identifying the best season for tree species diversity prediction is of importance to forest managers, ecologists, Remote Sensing scientists. Tree species diversity prediction done in the best season could help to improve accuracies.

In our study, the summer imagery performed slightly better than the spring, winter and autumn imageries. The summer image had a better correlation with the Shannon Diversity Index resulting in the higher accuracy produced. The findings may be an indication that the best season for tree species diversity prediction could be the summer and imageries captured within this season are likely to perform better by producing high accuracy as observed for the sub-tropical natural forest of South Africa. The reason behind this finding could be related to the condition of tree species (phenology and morphology) during the summer months.

The capture date of the summer imagery was in the middle of the season when the season is at its peak (Bond et al., 2003, Davis and Joubert, 2011), and the availability of summer rains guarantees continual photosynthesis and growth during these months. Although there is high temperature experienced most of the day, the composition of vegetation and the canopy are likely to remain unchanged. Rainfall which contributes immensely to the growth rates of vegetation is still present during the summer months. The rainfall that intersperses the dry period is enough to reduce the potential threats to phenology and morphology of tree species during the summer months. Therefore, it is possible that the phenology of the tree species is not stalled and the morphology of tree species may also not be affected as their condition may not change. This is mainly because the forests do not experience “no growth temperature” due to lack of rainfall, but experience “growth temperatures” with the availability of rains that maintain growth (Ellery et al., 1991). As such, the leaf density and tree characteristics are not reduced by the seasonal variation during the summer months (Bond et al., 2003). Furthermore, since the peak of the growing season is mid-summer, the rate of evapotranspiration and transpiration do not show marked differences (Dye and Versfeld, 2007). As such, the functional traits of the forests are probably maintained throughout the year. Dye and Versfeld (2007) found the plantation forest to retain a high green leaf index throughout the year, and the sub-tropical natural forest may share similarities with the plantation forest trees. As a result, the canopy of trees does not enter the senescence and deciduous stage and can maintain their growth rate and physical characteristics. Thus, the spectral signature recorded by the sensor at this time of the year may be from a community of healthy tree species and resulting in the slightly higher performance of the summer imagery.

A tree species diversity map of the Nkandla Forest Reserve was produced with the summer imagery since it was the most accurate. The map will be of importance at the local level for the forest managers and will assist them to know areas of low, moderate, and high tree species diversity. This information will be vital for planning involving initiatives such as enrichment planting in less diverse areas.

The textural variables selected as predicting variables for the RF model by the Recursive Feature Elimination algorithm was also useful and contributed to the accuracies obtained. This is because some texture variables can be redundant and highly correlated (Pacifci et al., 2009), which may affect models performance and reduce accuracies. Also, a large number of features can lead to high computational complexities (Ghosh and Joshi, 2014). Therefore, the selection of a subset of the predicting variables is useful as it can improve the accuracies of models because the noisy and redundant variables are eliminated. In the application of the RF model,

the main variables that contributed to the higher accuracy of the summer imagery were the Band-8-Homogeneity. The entropy was also important to model as three entropy variables derived for bands 6, 12 and 8A made good contribution to the imagery's accuracy. These variables could be adopted as predicting variables when using a summer imagery for tree species diversity prediction in a subtropical natural forest.

The spring imagery was the second-best by also producing a slightly better accuracy than the autumn and winter imageries. The spring season follows the winter period when the growth of some species might have ceased or lost their leaves, especially for deciduous species. Most tree species seem to have their phenological period occurring during the transition of the winter and spring seasons (Kikim and Yadava, 2001). Therefore, the spring period is a time of flowering and growth of new leaves for some species in response to longer days and available warm air and soil temperatures. However, the process does not happen at the same time for all species causing a spatial variation across the forest landscape (Liang and Schwartz, 2009). The phenomenon is likely to affect the spectral reflectance and signature of imageries captured in the spring season in the sub-tropical zone like our study area. The spectral reflectance may be low due to the new growth of tree leaves as compared to the time when they have fully grown leaves in summer. This could be the reason for the slightly lower accuracy produced by the spring imagery. In terms of the performance of the predicting variables, the entropy variable also had a significant contribution to the accuracy of the spring imagery. The Entropy variable derived for the bands 11 and 2 were first and third respectively, while the Band-11-GLCMC (Correlation) was ranked second. These texture variables could also be adopted for the spring imageries in case they are used for diversity predictions.

With the summer and the spring imageries having emerged as the first and second-best imageries respectively, the autumn imagery followed with the third-best accuracy. The autumn season is associated with a fall of leaves (Soudani et al., 2012) and deciduous species are likely to be much more affected than evergreen species. The phenological process is likely to reduce or stalled within the autumn season as the chlorophyll content in leaves decomposes, the nitrogen content is withdrawn and leaves eventually fall (Kodani et al., 2002). Hence, lower rates of the leaf area index (LAI) is observed among tree species, especially for deciduous tree species. The spectral reflectance and signatures of satellite imagery may be affected as a low reflectance may be recorded from the tree species due to the loss of leaves. Such a situation might have had some level of effect on the autumn imagery as it will have fewer tree canopy leaves reflecting incidence radiation for recording by the sensor. The lower accuracy recorded for the autumn imagery may be related to this condition of tree species prevailing in the autumn season. With regards to the variable importance assessment, the texture variables derived from band 5 of the Sentinel 2 contributed significantly to the accuracy of the autumn imagery. The Band-5-Contrast, Band-5-Dissimilarity and Band-5-GLCMC (Correlation) were the first three variables among the five predicting variables of the RF model of the imagery.

The Winter imagery had a comparable R^2 with the autumn image, but it was the least accurate due to its higher RMSE as compared to the summer, spring and autumn imageries. It may be related to the natural phenomenon of the tree leaves having lesser growth rates and photosynthetic activities and some also undergoing senescence. Breunig et al. (2015) indicated that in the winter months, spectral reflectance is much affected because of lowered sun illumination. Furthermore, it has been found that the LAI decreases (Li et al., 2017) and to some extent shadow could cast on the tree species canopy by the higher local area topography

which modifies the level of incident solar radiation and the reflectance reaching the sensor (Song and Woodcock, 2003). The dry and cold winter season is usually associated with the maximum dropping of leaves (Kikim and Yadava, 2001) and emergent trees and in the second upper stratum may be much vulnerable. Furthermore, Breunig et al. (2015) observed reduced plant area index (PAI) and increase gap openings in the winter months. Therefore, the phenological and morphological characteristics of species could be much more affected as the winter months are the harsh season of the year for the tree species. These reasons may explain the slightly less accuracy the winter imagery produced, and the higher level of error as compared to the summer, spring and autumn imagery. Concerning the importance variable analysis, the Band-7-Contrast, Band-5-GLCMC (Correlation) and Band-8A-Entropy were the top three predicting variables. In situations where it is only the winter season imagery that is available, these predicting variables could be utilized as they proved useful for the model of the imagery for prediction in the subtropical natural forest.

7.9 Conclusion

Different types of forest, satellite imageries, and algorithms used for modelling and prediction are likely to provide different prediction accuracies of forest attributes such as tree species diversity depending on the season. The findings of our study are vital for ecologists, conservationists, and other vegetation Remote Sensing experts. Since the season of the year could influence the spectral reflectance of imageries and how vegetation is perceived and recorded by the sensor, it is important to identify the best season for tree species diversity prediction to enhance the production of good accuracies. The summer season emerged as the best for predicting tree species diversity for sub-tropical natural forest which may be different from what other sites in literature produced. It is an indication that the summer season could probably be the best season for the prediction of tree species diversity of the natural sub-tropical forests. This information is vital for tree species diversity prediction, especially at the local level. Other studies may be recommended for similar sub-tropical forest zone to ascertain the best season for predictions. Our study is unique based on the predicting variables (texture variables) and the modelling approach employed. Key predicting variables identified for the prediction could also be adopted for tree species diversity in similar climatic and ecological zones.

Acknowledgment

The authors appreciate the support provided by Ms. Sharon Louw (Ecologist) of Ezemvelo KwaZulu-Natal Wildlife (EKZNW) in securing a permit to conduct the study in the Nkandla Forest Reserve. We thank Mr. Elliackim Zungu (Conservation Manager), Mr. Simon Makhaye, Mr. Sandebulawa Biyela and other supporting staff of EKZNW for assisting in field data collection. The authors also thank Mr. Charles Zungunde (University of KwaZulu-Natal) for driving the field team across the forest as well as assisting in data collection. We further thank the journal editors and reviewers for their efforts, time and insights.

Funding

This work is based on the research supported wholly by the National Research Foundation (NRF) of South Africa (Grant Number: 132924). The funders had no role in the study design, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest

References

- ABDEL-RAHMAN, E., M, AHMED, F. B. & ISMAIL, R. 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, 34, 712-728.
- AREKHI, M., YILMAZ, O. Y., YILMAZ, H. & AKYUZ, Y. F. 2017. Can tree species diversity be assessed with Landsat data in a temperate forest? *Environ Monit Assess*, 189, 586.
- ATKINSON, P. M., JEGANATHAN, C., DASH, J. & ATZBERGER, C. 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, 123, 400-417.
- BARALDI, A. & PANNIGIANI, F. 1995. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience Remote Sensing*, 33, 293-304.
- BELGIU, M. & DRĂGUȚ, L. 2016. Random forest in Remote Sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- BOND, W. J., MIDGLEY, G. F. & WOODWARD, F. I. 2003. What controls South African vegetation — climate or fire? *South African Journal of Botany*, 69, 79-91.
- BREIMAN, L. 2001. Random forests. *Machine learning*, 45, 5-32.
- BREUNIG, F. M., GALVÃO, L. S., DOS SANTOS, J. R., GITELSON, A. A., DE MOURA, Y. M., TELES, T. S. & GAIDA, W. 2015. Spectral anisotropy of subtropical deciduous forest using MISR and MODIS data acquired under large seasonal variation in solar zenith angle. *International Journal of Applied Earth Observation and Geoinformation*, 35, 294-304.
- CARLSON, K. M., ASNER, G. P., HUGHES, R. F., OSTERTAG, R. & MARTIN, R. E. 2007. Hyperspectral Remote Sensing of Canopy Biodiversity in Hawaiian Lowland Rainforests. *Ecosystems*, 10, 536-549.
- CHRYSAFIS, I., KORAKIS, G., KYRIAZOPOULOS, A. P. & MALLINIS, G. 2020. Predicting Tree Species Diversity Using Geodiversity and Sentinel-2 Multi-Seasonal Spectral Information. *Sustainability*, 12.
- COBURN, C. & ROBERTS, A. C. 2004. A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing*, 25, 4287-4308.
- DAVI, H., GILLMANN, M., IBANEZ, T., CAILLERET, M., BONTEMPS, A., FADY, B. & LEFEVRE, F. 2011. Diversity of leaf unfolding dynamics among tree species: new insights from a study along an altitudinal gradient. *Agricultural Forest Meteorology*, 151, 1504-1513.

- DAVIES, Z. G., EDMONDSON, J. L., HEINEMEYER, A., LEAKE, J. R. & GASTON, K. J. 2011. Mapping an urban ecosystem service: quantifying above-ground carbon storage at a city-wide scale. *Journal of Applied Ecology*, 48, 1125-1134.
- DAVIS, C. & JOUBERT, A. 2011. Southern Africa's climate: Current state and recent historical changes. CSIR.
- DYE, P. & VERSFELD, D. 2007. Managing the hydrological impacts of South African plantation forests: An overview. *Forest Ecology and Management*, 251, 121-128.
- ELLERY, W. N., SCHOLES, R. J. & MENTIS, M. T. 1991. An initial approach. to predicting the sensitivity of the South African grassland biome to climate change. *South African Journal of Science*, Vol. 87, 499-503.
- EZEMVELO KZN WILDLIFE 2015. Nkandla Forest Complex: Protected Area Management Plan. *Unpublished Version 1.0*, 173
- FEILHAUER, H. & SCHMIDTLEIN, S. 2009. Mapping continuous fields of forest alpha and beta diversity. *Applied Vegetation Science* 12, 429-439.
- GAMFELDT, L., SNÄLL, T., BAGCHI, R., JONSSON, M., GUSTAFSSON, L., KJELLANDER, P., RUIZ-JAEN, M. C., FRÖBERG, M., STENDAHL, J. & PHILIPSON, C. D. 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature communications*, 4, 1340.
- GASTON, K. J. 2000. Global patterns in biodiversity. *Nature*, 405, 220-227.
- GHOSH, A. & JOSHI, P. K. 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26, 298-311.
- GRABSKA, E., HOSTERT, P., PFLUGMACHER, D. & OSTAPOWICZ, K. 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, 11.
- GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2020. Mapping natural forest cover using satellite imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, 18, 1-14.
- GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2021. Evaluating Multi-Sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction. *Remote Sensing*, 13.
- IFO, S. A., MOUTSAMBOTE, J.-M., KOUBOUANA, F., YOKA, J., NDZAI, S. F., BOUETOU-KADILAMIO, L. N. O., MAMPOUYA, H., JOURDAIN, C., BOCKO, Y. & MANTOTA, A. B. 2016. Tree species diversity, richness, and similarity in intact and degraded forest in the tropical rainforest of the Congo Basin: case of the forest of Likouala in the Republic of Congo. *International Journal of Forestry Research*, 2016.
- IMMITZER, M., NEUWIRTH, M., BÖCK, S., BRENNER, H., VUOLO, F. & ATZBERGER, C. 2019. Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *J Remote Sensing*, 11, 2599.
- JOHN, R., CHEN, J., LU, N., GUO, K., LIANG, C., WEI, Y., NOORMETS, A., MA, K. & HAN, X. 2008. Predicting plant diversity based on Remote Sensing products in the semi-arid region of Inner Mongolia. *Remote Sensing of Environment*, 112, 2018-2032.
- 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, 390-401.
- KIKIM, A. & YADAVA, P. S. 2001. Phenology of tree species in subtropical forests of Manipur in north eastern India. *Tropical Ecology*, 42, 269-276.
- KODANI, E., AWAYA, Y., TANAKA, K. & MATSUMURA, N. 2002. Seasonal patterns of canopy structure, biochemistry and spectral reflectance in a broad-leaved deciduous *Fagus crenata* canopy. *Forest Ecology Management*, 167, 233-249.

- LAURIN, V. G., CHEUNG-WAI CHAN, J., CHEN, Q., LINDSELL, J. A., COOMES, D. A., GUERRIERO, L., DEL FRATE, F., MIGLIETTA, F. & VALENTINI, R. 2014. Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. *PLoS One*, 9, e97910.
- LI, X., MAO, F., DU, H., ZHOU, G., XU, X., HAN, N., SUN, S., GAO, G. & CHEN, L. 2017. Assimilating leaf area index of three typical types of subtropical forest in China from MODIS time series data based on the integrated ensemble Kalman filter and PROSAIL model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126, 68-78.
- LIANG, L. & SCHWARTZ, M. D. 2009. Landscape phenology: an integrative approach to seasonal vegetation dynamics. *Landscape Ecology*, 24, 465-472.
- LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. *R news*, 2, 18-22.
- MADONSELA, S., CHO, M. A., RAMOELO, A., MUTANGA, O. & NAIDOO, L. 2018. Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation*, 66, 106-115.
- MAEDA, E. E., HEISKANEN, J., THIJIS, K. W. & PELLIKKA, P. K. E. 2014. Season-dependence of Remote Sensing indicators of tree species diversity. *Remote Sensing Letters*, 5, 404-412.
- MILLENNIUM ECOSYSTEM ASSESSMENT, M. J. S. 2005. Ecosystems and human well-being.
- MILLIKEN, W., ZAPPI, D., SASAKI, D., HOPKINS, M. & PENNINGTON, R. T. 2010. Amazon vegetation: how much don't we know and how much does it matter? *Kew Bulletin*, 65, 691-709.
- MORRIS, E. K., CARUSO, T., BUSCOT, F., FISCHER, M., HANCOCK, C., MAIER, T. S., MEINERS, T., MÜLLER, C., OBERMAIER, E. & PRATI, D. 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology evolution*, 4, 3514-3524.
- MUTOWO, G. & MURWIRA, A. 2012. Relationship between remotely sensed variables and tree species diversity in savanna woodlands of Southern Africa. *International Journal of Remote Sensing*, 33, 6378-6402.
- MUTOWO, G., MUTANGA, O. & MASOCHA, M. 2018. Evaluating the Applications of the Near-Infrared Region in Mapping Foliar N in the Miombo Woodlands. *Remote Sensing*, 10.
- PACIFICI, F., CHINI, M. & EMERY, W. J. 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, 113, 1276-1292.
- PANDIT, S., TSUYUKI, S. & DUBE, T. 2018. Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sensing*, 10.
- RAMOELO, A., CHO, M. A., MATHIEU, R., MADONSELA, S., VAN DE KERCHOVE, R., KASZTA, Z. & WOLFF, E. 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 43, 43-54.
- RODRÍGUEZ-GALIANO, V. F., ABARCA-HERNÁNDEZ, F., GHIMIRE, B., CHICA-OLMO, M., ATKINSON, P. M. & JEGANATHAN, C. 2011. Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environmental Sciences*, 3, 44-49.

- SCHÄFER, E., HEISKANEN, J., HEIKINHEIMO, V. & PELLIKKA, P. 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. *Ecological Indicators*, 64, 49-58.
- SHANNON, C. 1948. A mathematical theory of communication. *J The Bell system technical journal*, 27, 379-423.
- SINTON, J. 2017. Review of Jill Jonnes. Urban forests: a natural history of trees and people in the American Cityscape. Springer.
- SONG, C. & WOODCOCK, C. E. 2003. Monitoring forest succession with multitemporal Landsat images: Factors of uncertainty. *IEEE Transactions on Geoscience Remote Sensing*, 41, 2557-2567.
- SOUDANI, K., HMIMINA, G., DELPIERRE, N., PONTAILLER, J.-Y., AUBINET, M., BONAL, D., CAQUET, B., DE GRANDCOURT, A., BURBAN, B. & FLECHARD, C. 2012. Ground-based Network of NDVI measurements for tracking temporal dynamics of canopy structure and vegetation phenology in different biomes. *Remote Sensing of Environment*, 123, 234-245.
- SRINIVASAN, G. & SHOBHA, G. Statistical texture analysis. Proceedings of world academy of science, engineering and technology, 2008. 1264-1269.
- SUN, Y., HUANG, J., AO, Z., LAO, D. & XIN, Q. 2019. Deep Learning Approaches for the Mapping of Tree Species Diversity in a Tropical Wetland Using Airborne LiDAR and High-Spatial-Resolution Remote Sensing Images. *Forests*, 10.
- TEAM, R. D. C. 2013. R: A language and environment for statistical computing.
- TUTTLE, E. M., JENSEN, R. R., FORMICA, V. A. & GONSER, R. 2006. Using Remote Sensing image texture to study habitat use patterns: a case study using the polymorphic white-throated sparrow (*Zonotrichia albicollis*). *Global Ecology Biogeography*, 15, 349-357.
- WALLIS, C. I. B., BREHM, G., DONOSO, D. A., FIEDLER, K., HOMEIER, J., PAULSCH, D., SÜßENBACH, D., TIEDE, Y., BRANDL, R., FARWIG, N. & BENDIX, J. 2017. Remote Sensing improves prediction of tropical montane species diversity but performance differs among taxa. *Ecological Indicators*, 83, 538-549.
- WANG, D., WAN, B., QIU, P., SU, Y., GUO, Q., WANG, R., SUN, F. & WU, X. 2018. Evaluating the Performance of Sentinel-2, Landsat 8 and Pléiades-1 in Mapping Mangrove Extent and Species. *Remote Sensing*, 10.
- WHITE, M. A., THORNTON, P. E. & RUNNING, S. W. J. G. B. C. 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. 11, 217-234.
- YANG, H., YANG, X., HESKEL, M., SUN, S. & TANG, J. 2017. Seasonal variations of leaf and canopy properties tracked by ground-based NDVI imagery in a temperate forest. *Sci Rep*, 7, 1267.
- 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, 47-54.
- ZHAO, Y., ZENG, Y., ZHENG, Z., DONG, W., ZHAO, D., WU, B. & ZHAO, Q. 2018. Forest species diversity mapping using airborne LiDAR and hyperspectral data in a subtropical forest in China. *Remote Sensing of Environment*, 213, 104-114.

CHAPTER 8: SYNTHESIS REPORT OF THE THESIS

8.1 Synthesis of the chapters

All the specific objectives that were defined for this study were achieved. From the six manuscripts, it could be realised that Remote Sensing has a good potential in its application for the mapping of natural forests. In most specific terms the mapping of forest cover, detection of forest cover change and forecasting of future land cover distribution, the carbon stock estimation, tree species diversity prediction and the identifying of the best season for tree species predictions were all carried out successfully. Generally, more such studies would be recommended to enhance sustainable forest management based on the information provided through Remote Sensing mapping of sub-tropical natural forests such as the Nkandla Forest Reserve. A synthesis of each chapter has been provided below for each of the manuscripts.

8.2 Two Decades Progress on The Application of Remote Sensing for Monitoring Tropical and Sub-Tropical Natural Forests: A Review

Interesting observations were made in our review of relevant literature on the progress made in Remote Sensing application to tropical and sub-tropical natural forests. The tropical and sub-tropical climatic zones are found in Southern America, Asia and Sub-Saharan Africa. The predefined thematic areas of which were, AGB and AGC estimations, tree species identification, tree species diversity prediction and forest cover mapping and change detection were identified to be key areas of Remote Sensing mapping of natural forests. Africa was observed to be lagging behind Southern America and Asia in the number of studies carried over the period of assessment. More research is recommended in African countries for the natural forest as it is faced with a high rate of deforestation and climate change

It was observed from the finding that sensors are being improved to map complex and highly diverse natural tropical and sub-tropical forests and also to meet the demands of science and practice in Remote Sensing applications. Freely available Landsat and Sentinel 2 remains highly used especially in African researchers. Commercial very high resolution, hyperspectral and active sensors mostly have less application in Africa and more research funding may be made provided for African research. High research budget needs to be made available for African researchers to be able to carry out highly advanced research that will be of much use to sustainable forest management and conservation.

The application of machine learning algorithms has also increased over time and has led to improved accuracies. Since accuracy is a means of measuring the credibility of Remote Sensing mapping and monitoring, more advanced forms of modelling using machine learning algorithms are employed by researchers.

This review outcomes are of importance to Remote Sensing researchers undertaking key studies for tropical and sub-tropical natural forests. The research outputs will be a good guide for the selection of Remote Sensing data and machine learning algorithms that can facilitate modelling and provide good research outputs. More research is recommended in these thematic areas and other relevant ones to provide adequate and credible information to forest managers and ecologists towards efficient conservation and protection initiatives.

8.3 Mapping Natural Forest Cover Using Satellite Imagery of Nkandla Forest Reserve, Kwazulu-Natal, South Africa.

Multiple benefits are obtained from natural forest ecosystems in their provision of ecosystem goods and services which are of importance to direct and indirect forest dependants. Research has shown that Remote Sensing forest cover mapping is a major means of obtaining information that contributes immensely to making informed and evidence-based decisions that enhance sustainable forest management. The Landsat 8 imagery could successfully delineate the land cover classes of the Nkandla Forest Reserve. These land cover classes are closed canopy forest, open canopy forest, grassland and bare sites. The RF and SVM were used for the land cover classification and their performances were evaluated. They both proved robust for the classification as each of them produced accuracies that were above 95%.

The finding is confirmed by other studies that also found the SVM and RF producing higher accuracies in land use/cover classification (Paneque-Gálvez et al., 2013, Petropoulos et al., 2012, Pelletier et al., 2016, Yin et al., 2017). The SVM, however, produced performed better than the RF with slightly higher accuracy. Adam et al. (2014) similarly found the SVM performing better than the RF. Both algorithms are recommended for mapping of forest cover due to their inherent capabilities in handling complex natural forests.

The visible and the SWIR bands were the most important variables that significantly contributed to the accuracy due to their sensitivity to vegetation. Many other studies have similarly found these bands useful for vegetation mapping. Thus, they are recommended to be included in spectral that are used for forest cover mapping. Knowledge of land cover classes is a key part of forest management and so, this mapping is a worthy operation.

8.4 Multi-Decadal Spatial and Temporal Forest Cover Change Analysis of Nkandla Natural Reserve, South Africa.

The focus of this study which was to detect changes in the land cover of the Nkandla Forest Reserve revealed a persistent change for all the four land cover classes (closed canopy forest, open canopy forest, grassland and bare sites). Inter-transitions were observed among the cover classes which led to the changes. The closed canopy forest has the most gain over the period and it became the dominant land cover type. This is an indication of tree cover taking over canopy gaps within the open canopy forest. Such changes are positive as more tree cover enhances the provision of ecosystem goods and services. A negative type that involves the taking over of closed canopy areas by the open canopy forest, grassland and bare sites is not a good sign as it lowers the level of ecosystem services provision. Natural and human causes were found to be the major causes of the changes.

The prediction of future land cover classes is important as it helps to put in mitigation measures for negative forms of changes. The forecasting in spatial distribution carried out for 2029 revealed that the closed canopy forest will have the most changes as compared to the open canopy forest, grassland and bare sites. It will experience a decline while the open canopy

forest and bare sites will have a slight increase. The grassland was predicted to have the most increase by 2029.

The mapping was made possible with the Landsat data due to the availability of historical data which dates back to the 1970s. The Landsat imageries could delineate the four land use/cover classes of the Nkandla Forest Reserve with high accuracies. Other studies were also successful in the application of Landsat data in land use/cover classification (Addae and Oppelt, 2019, Da Ponte et al., 2017, Mihai et al., 2017, Vázquez-Quintero et al., 2016). The MLPNN and the MCM applied in similar studies were found to facilitate the spatial and temporal change detection (Addae and Oppelt, 2019, Ranagalage et al., 2019).

The spatio-temporal information provided through our study will be of importance to the forest managers with regard to planning. It is recommended that pragmatic measures should be put in place to prevent the human form of disturbances to the forest to maintain its integrity. This will ensure the continual provision of ecosystem goods and services.

8.5 Mapping of Aboveground Carbon Stock in Sub-Tropical Natural Forest using Sentinel 2 Satellite Imagery and Random Forest Algorithm.

The prediction of AGC with Remote Sensing data has been more necessary due to increasing climate change and deforestation. It is also important in carbon accounting and forest management. The combination of Sentinel 2 and RF for the modelling of the AGC of the Nkandla Forest Reserve was also carried out successfully. The prediction abilities of the spectral bands, NIR vegetation indices and red-edge vegetation were compared for the modelling of the AGC. This helped to identify the sentinel 2 spectral products that will be much robust for the AGC prediction.

The red-edge bands and their derived vegetation indices were observed to be informative and robust for AGC modelling as compared to the other spectral bands and the NIR vegetation indices of the other models. The modelling methods and the identified important variables may be adopted and utilized in the estimation and mapping of AGC stocks in other climatic zones to ascertain their usefulness.

The outcomes of our study shares similarities with some studies that found the RF algorithm contributing to predicting model accuracies (Dube and Mutanga, 2015, Ghosh and Behera, 2018). Furthermore, the red edge-edge bands of the Sentinel 2 were identified as being informative and much robust in vegetation studies due to their high sensitivity. The finding of our study is similar to other studies that found the RF algorithm contributing to predicting model accuracies (Dube and Mutanga, 2015, Ghosh and Behera, 2018).

8.6 Evaluating Multi-Sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction.

Many studies are carrying out tree species diversity prediction in many climatic zones and Remote Sensing imageries due to its importance in forest management. Spatially explicit information provided through such predictions has been demonstrated to be vital as it helps to

identify tree species diversity hot spots and less diverse areas in natural forests. Although many studies have used various Remote Sensing imageries for tree species diversity predictions across different forest zones, the influence of the spectral and spatial resolution of sensors on the prediction accuracies had not been fully assessed. Hence, our study evaluated the influence of the spectral and spatial resolution on the prediction accuracies in tree species diversity by comparing the performance of Landsat 8, Sentinel 2, RapidEye and the PlanetScope. The Shannon Index (H'), Simpson Index (D_1) and the Species richness (S) together with RF regression modelling were utilised to identify which imageries had a good relationship with them and produce good accuracy.

The Sentinel 2 outperformed the RapidEye, PlanetScope and Landsat 8. The RapidEye was the second-best imagery while the PlanetScope was the third-best imagery. The Landsat 8 performed the least. Our study shares a similarity with Mallinis et al. (2020), who also found the Sentinel 2 performing better than the RapidEye in species diversity prediction in the Mediterranean region. This observation is due to inherent capability of the Sentinel 2 that have been confirmed in vegetation related studies (Chrysafis et al., 2020, Immitzer et al., 2019, Martin-Gallego et al., 2020).

The findings indicated that both spatial and spectral resolution influence the predicting accuracies of imageries in tree species diversity predictions for natural sub-tropical forests. More such studies are recommended for other forest zones to ascertain how the accuracies of different imageries could be influenced by the spectral and spatial resolutions. Such information would be vital for forest managers and ecologists.

8.7 Identifying the Best Season for Predicting Tree Species Diversity using Sentinel 2

Satellite Imagery and Random Forest Algorithm

The seasons of the year determine phenology and morphological changes and characteristics of tree species. Changes are realised in the tree species phenology and morphology based on the seasons of the year and are likely to affect how sensors perceive vegetation and may further affect the accuracies of imageries. It is, therefore, important to identify the best season for the prediction of tree species diversity due to these changes in tree characteristics that affect spectral signatures and reflectance of sensors. A combination of RF regression algorithm and texture variables were used for the modelling. From the results of our study, the summer season was observed to be the best for tree species diversity due to the higher performance of the summer imagery as compared to the spring, autumn and winter imageries. Similar studies are recommended for other forest zones to ascertain the best season for tree species diversity predictions.

In a similarly study, Chrysafis et al. (2020) found the multi-seasonal models producing the highest accuracies than single-season models of mid-summer. Although the approach used for the referenced study was not used in our study, it is worth mentioning the output. Subsequent studies could adopt this modelling approach as it may enhance the accuracy. An inconsistency was observed with in relation to Chrysafis et al. (2020) as the mid-summer season was found to the second best season while it was found to be the first in our study.

References

- ADAM, E., MUTANGA, O., ODINDI, J. & ABDEL-RAHMAN, E. M. 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35, 3440-3458.
- ADDAE, B. & OPPELT, N. 2019. Land-Use/Land-Cover Change Analysis and Urban Growth Modelling in the Greater Accra Metropolitan Area (GAMA), Ghana. *Urban Science*, 3.
- CHRYSAFIS, I., KORAKIS, G., KYRIAZOPOULOS, A. P. & MALLINIS, G. 2020. Predicting Tree Species Diversity Using Geodiversity and Sentinel-2 Multi-Seasonal Spectral Information. *Sustainability*, 12.
- DA PONTE, E., ROCH, M., LEINENKUGEL, P., DECH, S. & KUENZER, C. 2017. Paraguay's Atlantic Forest cover loss—Satellite-based change detection and fragmentation analysis between 2003 and 2013. *Applied Geography*, 79, 37-49.
- DUBE, T. & MUTANGA, O. 2015. 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.
- GHOSH, S. M. & BEHERA, M. D. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*, 96, 29-40.
- IMMITZER, M., NEUWIRTH, M., BÖCK, S., BRENNER, H., VUOLO, F. & ATZBERGER, C. 2019. Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *J Remote Sensing*, 11, 2599.
- MALLINIS, G., CHRYSAFIS, I., KORAKIS, G., PANA, E. & KYRIAZOPOULOS, A. P. 2020. A Random Forest Modelling Procedure for a Multi-Sensor Assessment of Tree Species Diversity. *Remote Sensing*, 12.
- MARTIN-GALLEGO, P., APLIN, P., MARSTON, C., ALTAMIRANO, A. & PAUCHARD, A. 2020. Detecting and modelling alien tree presence using Sentinel-2 satellite imagery in Chile's temperate forests. *Forest Ecology Management*, 474, 118353.
- MIHAI, B., SĂVULESCU, I., RUJOIU-MARE, M. & NISTOR, C. 2017. Recent forest cover changes (2002–2015) in the Southern Carpathians: A case study of the Iezer Mountains, Romania. *Science of the Total Environment*, 599, 2166-2174.
- PANEQUE-GÁLVEZ, J., MAS, J.-F., MORÉ, G., CRISTÓBAL, J., ORTA-MARTÍNEZ, M., LUZ, A. C., GUÈZE, M., MACÍA, M. J. & REYES-GARCÍA, V. 2013. Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *International Journal of Applied Earth Observation and Geoinformation*, 23, 372-383.
- PELLETIER, C., VALERO, S., INGLADA, J., CHAMPION, N. & DEDIEU, G. 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156-168.
- PETROPOULOS, G. P., KALAITZIDIS, C. & PRASAD VADREVVU, K. 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences*, 41, 99-107.
- RANAGALAGE, M., WANG, R., GUNARATHNA, M. H. J. P., DISSANAYAKE, D. M. S. L. B., MURAYAMA, Y. & SIMWANDA, M. 2019. Spatial Forecasting of the Landscape in Rapidly Urbanizing Hill Stations of South Asia: A Case Study of Nuwara Eliya, Sri Lanka (1996–2037). *Remote Sensing*, 11.

- VÁZQUEZ-QUINTERO, G., SOLÍS-MORENO, R., POMPA-GARCÍA, M., VILLARREAL-GUERRERO, F., PINEDO-ALVAREZ, C. & PINEDO-ALVAREZ, A. 2016. Detection and Projection of Forest Changes by Using the Markov Chain Model and Cellular Automata. *Sustainability*, 8.
- YIN, H., KHAMZINA, A., PFLUGMACHER, D. & MARTIUS, C. 2017. Forest cover mapping in post-Soviet Central Asia using multi-resolution remote sensing imagery. *Sci Rep*, 7, 1375.