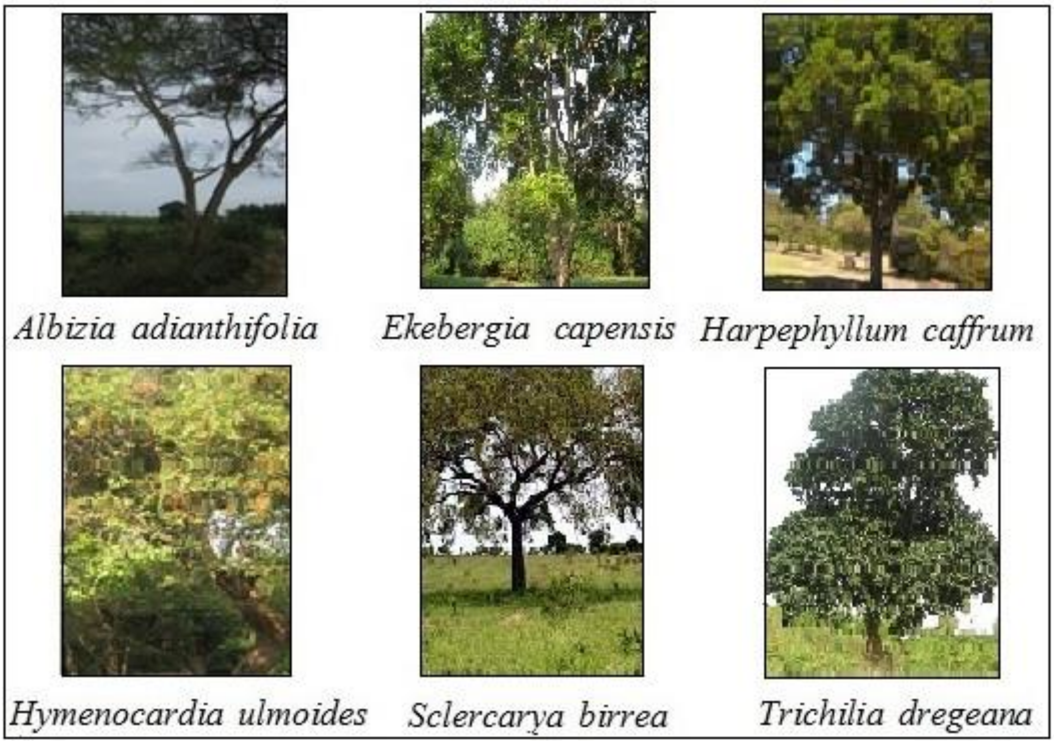
Remote Sensing of Endangered Tree Species in the Fragmented Dukuduku Indigenous Forest of KwaZulu-Natal, South Africa

Galal Elawad Khaled Omer



A thesis submitted to the College of Agriculture, Engineering and Science, at the University of KwaZulu-Natal, in fulfilment of the academic requirements for the degree of Doctor of Philosophy in Environmental Sciences

July 2016

Pietermaritzburg

South Africa

ABSTRACT

Dukuduku indigenous forest in South Africa provides varied products and usable materials for human needs that include construction and fence poles, raw material for craft work, livestock browse and medicine to the poor rural communities. The Dukuduku indigenous forest is dominated by many rare and endangered tree species, for example *Syzygium cordatum*, *Cussonia zuluensis*, *Ficus natalensis*, *Canthium inerme*, *Strychnos madagascariensis*, *Strychnos spinosa*, *Albizia adianthifolia*, *Ekebergia capensis*, *Harpephyllum caffrum*, *Hymenocardia ulmoides*, *Sclercarya birrea* and *Trichilia dregeana*. These tree species play a significant role in ecosystem functioning and services, land use dynamics and other socio-economic aspects. Such aspects include ecological, economic, livelihood, security-based and well-being benefits. However, some tree species in the Dukuduku indigenous forest have become endangered and threatened by a number of factors such as the rapid harvesting rate and climate. Conventional approaches of monitoring and mapping endangered tree species are complex and require intensive fieldwork which is a costly, time-consuming and subjective protocol, particularly when carried out in highly fragmented ecosystems. In this regard, remotely-sensed data offer a practical and economical means of quantifying indigenous forest fragmentation over large areas. Remote sensing is capable of providing rapid, relatively inexpensive and near-real-time data that could be used for monitoring endangered tree species especially in indigenous forest ecosystems where data collection may be difficult. The advent of advanced imaging systems and supervised learning algorithms has made it increasingly practical to facilitate the development of models that adequately map endangered tree species and estimate their biophysical and biochemical properties in the fragmented forest ecosystems. Recently, vegetation maps have been produced using the advanced imaging systems such as WorldView-2 and robust classification algorithms like support vectors machines (SVM) and artificial neural networks (ANN). However, delineation of endangered tree species from other land use/cover and estimating their biophysical (leaf area index: LAI) and biochemical (leaf nitrogen: N and carbon: C_N) traits in a fragmented ecosystem using high spatial resolution imagery has largely remained elusive due to the complexity of the species structure and their distribution.

Therefore, in the current research, WorldView-2 data with SVM and ANN classification algorithms were firstly used to map the spatial distribution of detailed land use/cover classes in the Dukuduku indigenous forest. The study successfully mapped eight land use/cover patterns, achieving an overall classification accuracy of 78% with an allocation disagreement (AD) of 17% and a quantity disagreement (QD) of 5%. The results demonstrated the ability of remotely-sensed data for mapping land use/cover at high accuracy within an indigenous forest ecosystem.

Secondly, the study examined the utility of the advanced WorldView-2 data for distinguishing endangered tree species from other land use/cover using SVM and ANN classification algorithms. The results showed the robustness of the two machine learning algorithms for classifying the endangered tree species. Overall accuracy of 77% (AD of 12% and QD of 11%) for SVM and 75% (AD of 16% and QD of 9%) for ANN were achieved. Furthermore, the advent of the additional WorldView-2 bands for mapping the six endangered tree species was also tested. Results showed that the additional WorldView-2 bands produced almost the same overall accuracy of 70% for both SVM (AD = 14% and QD = 16%) and ANN (AD = 19% and QD = 11%).

Thirdly, the relationship between LAI of the six endangered tree species and spectral vegetation indices (SVIs) derived from WorldView-2 data in fragmented and intact indigenous forest ecosystems were tested using SVM and ANN regression algorithms. The results showed that LAI at tree species level could be accurately estimated using the fragmented stratum data compared with the intact stratum data. Specifically, the study showed that the accurate LAI predictions were achieved for *Hymenocardia ulmoides* tree using the fragmented stratum data and SVM regression model based on a validation data set (R^2_{Val} = 0.75, $RMSE_{Val}$ = 1.37% of the mean).

Finally, multispectral WorldView-2 spectral variables with the SVM and ANN regression methods were explored for estimating and mapping forest leaf N and C_N concentrations of fragmented and intact indigenous forests ecosystems. The results showed that the accurate forest foliar N predictions were achieved for the fragmented data using the SVM (R^2_{Val} = 0.77, $RMSE_{Val}$ = 1.07% of the mean) and ANN (R^2_{Val} = 0.70 and $RMSE_{Val}$ = 5.40% of the mean) regression methods. The study also showed that the accurate forest foliar C_N predictions were achieved for the fragmented data using SVM (R^2_{Val} = 0.67, $RMSE_{Val}$ = 1.64% of the mean) and ANN (R^2_{Val} = 0.51, $RMSE_{Val}$ = 2.21% of the mean). It was observed that that SVM regression approach achieved relatively more accurate models for estimating the forest leaf N and C_N concentrations in the fragmented and intact indigenous forest ecosystems compared to the ANN regression method.

Overall, the results provide accurate information that is important for forest managers and researchers for making informed decisions regarding the conservation and management of the land use/cover patterns and endangered tree species in Dukuduku indigenous forest in South Africa.

PREFACE

The research work described in this thesis was carried out in the School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, from February 2012 to March 2016, under the supervision of Prof. Onisimo Mutanga and Dr. Elfatih M. Abdel-Rahman (School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, South Africa).

I would like to declare that the research work reported in this thesis has never been submitted in any form to any other university. It therefore represents my original work except where due acknowledgments are made.

Galal Elawad Khaled Omer   Signed: ____________________Date_______________________

As the candidate`s supervisors, we certify the above statement and have approved this thesis for submission.

1. Prof. Onisimo Mutanga   Signed: ______________________ Date ______________________

2. Dr. Elfatih M. Abdel-Rahman   Signed: _________________ Date ______________________

DECLARATION 1- PLAGIARISM

I, Galal Elawad Khaled Omer, hereby declare that this thesis submitted for the Doctor of Philosophy degree at the University of KwaZulu-Natal is my original work and has not been previously submitted to any other institution of higher education. I further declare that:

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

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

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

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written but the general information attributed to them has been referenced.

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

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

Signed ______________________

DECLARATION 2 - PUBLICATIONS AND MANUSCRIPTS

1. **G. Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, “Potential utility of the WorldView-2 multispectral data and support vector machines algorithm to classifying land use/cover in Dukuduku landscape, Kwazulu-Natal, South Africa” *The 10th African Association of Remote Sensing of the Environment* (*AARSE*) *conference*, Johannesburg, South Africa, pp. 309-318, 2014.

2. **G. Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, “Exploring the utility of the additional WorldView-2 bands and support vector machines in mapping land use/cover in a fragmented ecosystem, South Africa” *South African Journal of Geomatics*, 4 (4), 414-433, 2015.

3. **G. Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, “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*, PP (99), 1-16, 2015.

4. **G. Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, “Empirical prediction of leaf area index (LAI) of endangered tree species in intact and fragmented indigenous forests ecosystems using WorldView-2 data and two robust machine learning algorithms” *Remote Sensing*, 8 (4), 1-26, 2016.

5. **G. Omer**, O. Mutanga, E.M. Abdel-Rahman, K. Peerbhay and E. Adam, “Mapping leaf nitrogen and carbon concentrations of intact and fragmented indigenous forest ecosystems using empirical modeling techniques and WorldView-2 data”. Under preparation.

Signed_______________________________

DEDICATION

To my beloved father, Elawad Khaled Omer, much-loved mother Hussna El Hassan Abdalla, brothers and sisters for their constant encouragement and patience for my success. I also dedicate this project to my lovely wife, Fatima Abdelkarim Abdelrahman, and precious son, Ahmed Galal Elawad.

ACKNOWLEDGEMENTS

First of all, all praise goes to the Almighty Allah for the gift of my life, granting me health, power, desire, dreams and the ability to complete my Ph.D. thesis.

It is a great pleasure to convey my deepest gratitude and immense admiration to my academic advisor and eventually my thesis supervisor, Prof. Onisimo Mutanga, for his constant advice, confidence, and scientific guidance in structuring this work, commitment, critical comments, and moral support which enabled me to complete my research. This would not have been possible without his invaluable inspiration, continued encouragement, and continuous support for taking the challenge of this research idea and giving the thesis a final shape. Prof., you have taught me to be a researcher. I learnt how to critically write and review research work that saw our papers being readily accepted by international journals.

I would like to express heartfelt gratitude to Dr. Elfatih M. Abdel-Rahman, who patiently and tirelessly supervised this study to completion. His help in the field of remote sensing and GIS has been constructive. Thank you for accepting the task of being my co-supervisor, even though the study had already begun, and for your generous time and resources. Thank you for your much appreciated inspiration and dedication. I really enjoyed working with you.

I would like to thank Professor Fethi B. Ahmed, who was the first supervisor, until he resigned from his position at the University of KwaZulu-Natal. Thank you for giving me this great opportunity to read for my Ph.D. abroad.

My sincere appreciation is also extended to Dr. Elhadi Adam who provided me with support, advice, guidance and helpful suggestions especially in the sections related to remote sensing techniques during the research work. I also express my thanks to Dr. Khalid Mansour who provided a lot of information related to my research and for his guidance in preparing the research proposal.

Sampling in indigenous forest fields is not an easy task and is considered by field experts as risky. I am grateful to Mr. Charles Otunga, Mr. Lucky Nkomo (University of KwaZulu-Natal) and the Inkanyamba Development Trust and Manukelana Arts and Indigenous Nursery in KwaZulu-Natal for the assistance you provided during the field work. This helped me in achieving the objectives of the research. My appreciation extends to the R development core team and Tanagra team for their very powerful open source packages for the statistical analysis.

My thanks are extended to Dr. Riyad Ismail (Sappi forests, South Africa), Dr. Samuel Adelabu (University of the Free State), Dr. John Odindi, Dr. Mercy Ojoyi, Mr. Romano Lottering and Dr. Nafisa Sobratee (University of KwaZulu-Natal) for their helpful comments and great assistance during data analysis. I am also indebted to my colleagues at the University of KwaZulu-Natal, particularly Mr. Charles Otunga, Mr. Elkhatab Abdalla, Ms. Sithabile Hlahla, Dr. Kabir Peerbhay and Dr. Timothy Dube for their helpful comments and assistance during data analysis and write-up of the thesis.

My deepest thanks extends to the University of KwaZulu-Natal administration for accepting me as a Ph.D. candidate and for providing me with essential Ph.D. resources, their financial support for field work, satellite imageries and laboratory analyses was also appreciated. Special thanks extends to the staff of Geography Department at the University of KwaZulu-Natal for their support and friendship. My special gratitude goes to Mrs. Shanita Ramroop for arranging safe and comfortable transport and accommodation as well as organizing other logistics related to my study. Mr. Donavan DeVos is thanked for his assistance for being a true friend during my study. He was always ready to help me. Similar appreciation goes to the cartography unit staff specially Mr. Brice Gijsbertsen and Mr. Victor Bangamwabo for their help in facilitating the field work, for assisting me in ArcMap, and for configuring the GPS (Geographical Positioning System).

I am gratefully indebted to the University of Khartoum and Ministry of High Education, Sudan for financial support, assistance and follow-up during my study. My thanks are extended to the staff of Faculty of Forestry, University of Khartoum, Sudan especially to Dr. Abdallah Mirgany, the Dean of Faculty of Forestry; Dr. Omer Said Musa, the former Head of Forest Conservation and Protection Department, who provided a lot of information related to my research, encouraged and advised me to undertake my Ph.D. degree study in South Africa. Special thanks are also extended to Dr. El Amin Sanjak, Dr. Mudawi Elobeid, Dr. Abdelnasir Hanno, Mr. Hassan Hasab Elrasoul and Shaikh Omer El Obeid, who helped in various ways including follow-up of my progress through frequent contacts during the course of this research work.

The faces which always shine in the deepest part in my heart are my mother, father, brothers and sisters. My thanks are due to their endless encouragement, inspiration, and committed support. Here, I would like to thank my great brother, Omer Elawad who has kindled the lighthouse in my way towards achieving success in my life. My sister, Hadia Elawad, is greatly appreciated for suspending her job to travel to South Africa to take care of my first born, Ahmed while my beloved wife and I were busy studying. Hadia has demonstrated the meanings of sacrifice and selflessness.

Finally, I am grateful to my wife Fatima Mohammed (Om Abuha). I sincerely say thank you for your patience and for carrying the burden of looking after your study and our son. It has helped me to focus on my Ph.D. study. My achievement as well as the son’s excellence is your reward.

Galal Elawad Khaled Omer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

General Introduction

1.1 Background

Indigenous forests stretch across different parts of Africa and are particularly concentrated in the southern and eastern parts of the continent (Eeley *et al.*, 2001). Indigenous forests are a source of valued resources that highly contribute to rural communities in southern Africa (Shackleton and Shackleton, 2004; Eldeen and van Staden, 2007; van Wyk, 2008). They play a vital role in nutrient and carbon cycling, provide habitats for fauna and flora, and reduces soil erosion (Eldeen, 2005; Brendler *et al.*, 2010). The indigenous forest biome also supports a high proportion of South Africa flora and fauna diversity, e.g. forest mammals and forest birds represent more than 13% of the total terrestrial of these taxa in southern Africa (Eeley *et al.*, 1999).

In South Africa, indigenous forests cover about 0.2% of the country's land surface, and consist of many small, fragmented and largely scattered patches in relatively dry landscapes (Ndlovu, 2013). The country has a strong history of utilizing tree species for traditional healing (Louw *et al.*, 2002; Eldeen, 2005). The country also has a high rate of plant diversity, with some 30000 species of flowering plants and 80% of these species are endemic (Goldblati, 1978; Fennell *et al.*, 2004). This incorporates a large diversity of plants including, among others, trees, shrubs, and herbs (Louw *et al.*, 2002). An example of such fragmented forests in KwaZulu Natal, South Africa is the Dukuduku forest. Dukuduku indigenous forest is one of the best preserved remnants of South African coastal forests (Ntombela, 2003). Dukuduku indigenous forest provides variety of benefits and products that include traditional medicine, resin, and livestock browsing (Eldeen, 2005; Brendler *et al.*, 2010; Cho *et al.*, 2012; Mlambo, 2013). The medicinal tree species are a very important component provided by indigenous forests and play a vital role in providing subsistence and income (Hamilton, 2004). However, the coastal forest in the Dukuduku area is highly threatened by the rapid growth of informal human settlements and agricultural systems (Ndlovu, 2013). It is interesting to note that local communities in Zululand use different tree species to treat human diseases, such as fever, stomach ache dysentery, snake and scorpion bites (Hutchings, 1996; Sewram *et al.*, 2000; Eldeen and van Staden, 2007; Brendler *et al.*, 2010). In addition, the Dukuduku indigenous forest has become home to an increasing number of mainly subsistence farmers, some of whom form part of the group of land claimants (Sundnes, 2013). Therefore, these growing illegal squatters lead to large scale and rapid destruction of the forest ecosystem. For the aforementioned reasons, some tree species have become endangered, threatened and rare in different forests in South Africa. In particular, the Dukuduku forest has experienced intense harvesting which has led to the disappearance of some of these tree species (van Wyk *et al.*, 2006; Mlambo, 2013). The intensive harvesting of these tree species for different use has caused over utilization of resources and has subsequently affected the forest ecosystem in the area (Baillie *et al.*, 2004; Wiersum *et al.*, 2006).

Moreover, studies have shown that the distribution of some rare and threatened tree species has decreased due to over utilization of resources for wood carving and traditional medicine in Dukuduku area (Eldeen, 2005; Mlambo, 2013). The Dukuduku area is dominated by several rare tree species, for example *Syzygium cordatum*, *Cussonia zuluensis*, *Ficus natalensis*, *Canthium inerme*, *Strychnos madagascariensis*, *Strychnos spinosa*, *Albizia adianthifolia*, *Ekebergia capensis*, *Harpephyllum caffrum*, *Hymenocardia ulmoides*, *Sclercarya birrea* and *Trichilia dregeana* (Watt and Breyer-Brandwijk, 1962; Jäger *et al.*, 1996). Among these tree species, six tree species (Figure 1.1) were observed to be under severe threat and endangered (Mlambo, 2013). Therefore, this research focuses on classifying these tree species amongst other land use/cover and estimates their biophysical and biochemical parameters. Endangered tree species require sound management and protection protocols to assess the ecosystems' services and resilience in the value chain (Eldeen, 2005; Lyons *et al.*, 2005; Pouteau *et al.*, 2012). This requires intensive fieldwork to geo-locate and identify endangered tree species from other land use/cover classes and estimate their biophysical and biochemical properties (Rushton *et al.*, 2004; Pouteau *et al.*, 2012).

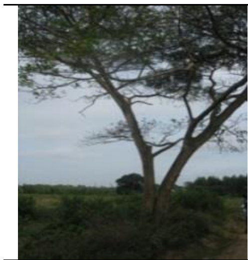
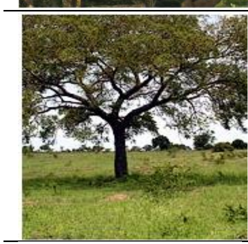
Field Photograph	Scientific name	Isizulu name	English name
	*Albizia adianthifolia*	IGowani	Flat crown
	*Ekebergia capensis*	UmYamathi	Cape ash
	*Harpephyllum caffrum*	UmGwenya	Wild Plum
	*Hymenocardia ulmoides*	UmBambahlangu	Red- heart tree
	*Sclercarya birrea*	UmGanu	Marula
	*Trichilia dregeana*	UmKhuhlu	Forest mahogany

Figure 1.1: The endangered tree species selected in the present study

1.2 Endangered Tree Species in Indigenous Forest

In the past few decades, traditional methods have been used to quantify, map and monitor vegetation and tree species. These methods provide significantly better results in mapping endangered tree species over small areas. However, the traditional ground-based forest surveys are complex and require intensive and difficult fieldwork that involves interrogating local expert knowledge in terms of species identification and to support accurate information on the spatial and temporal distribution of endangered tree species, dynamics and characteristics. This exercise is too expensive, and time-consuming, particularly in a large fragmented ecosystem. In order to understand the spatial distribution of endangered tree species and develop sustainable forest management practices to monitor their functions in the indigenous forest ecosystem, it is important to (i) improve the understanding of the dynamics of indigenous forest ecosystems, and (ii) develop an early warning system for forest fragmentation and loss of forest species diversity. Furthermore, there is more precise information available from forest surveys and there is a critical requirement to develop real-time spatially-explicit data for modeling the spatial distribution and predicting the biophysical (e.g. LAI: leaf area index) and biochemical (e.g. N: nitrogen and C_N: carbon) indicators needed for the rapid assessment and proactive management of the endangered tree species. One of the best ways to improve the management and monitoring of indigenous forest ecosystems is to estimate the biophysical (e.g. LAI) and biochemical (e.g. foliar N and C_N concentrations) attributes. These tree characteristics are proxies for ecosystem resilience, conservation and forest health. The forest LAI is an important biophysical attribute for modeling the energy and mass exchange between the land surface and the atmosphere of terrestrial ecosystems as it is one of the most useful indicators of forest growth, biomass and net primary production (Asner *et al.*, 2003). Therefore, predicting LAI of some tree species that play a vital role in ecosystem services is a necessary and valuable information.

Forest leaf N and C_N are also among the most important biochemical components of tree organic matter, and the estimation of their concentrations can help to monitor the nutrient uptake processes and forest health. Estimating forest foliar N and C_N concentrations of different forest ecosystems such as the fragmented and intact Dukuduku forests could help resource managers to understand the impact of various socio-ecological mechanisms on indigenous forest species and the vulnerability of these ecosystems to external and internal perturbations. In this regard, a complementary remotely-sensed data has successfully been used to provide a fairly accurate, repetitive and unbiased means for mapping and monitoring vegetation and tree species. Therefore, techniques that make use of the advantages of remote sensing are needed for mapping endangered tree species in order to determine the condition of indigenous forest ecosystems.

1.3 Application of Remote Sensing for Characterizing Endangered Tree Species in an Indigenous Forest Ecosystem

Obtaining accurate information for the sake of mapping and monitoring endangered tree species distribution is an important technical task for sustainable indigenous forest management. Remote sensing is a powerful tool that can obtain accurate information for mapping and monitoring vegetation species in different indigenous forest ecosystems (Vogelmann, 1995; Fuller, 2001; Cho *et al.*, 2013). However, remotely-sensed mapping and monitoring of changes in the spatial distribution of vegetation species in fragmented ecosystems faces some challenges that are associated with the characteristics of indigenous vegetation species as well as with the pixel size of some sensors. The challenges facing scientists in terms of the application of remote sensing for distinguishing between vegetation species in fragmented ecosystem are as follows: (i) vegetation species phenology changes as a result of climate change, particularly precipitation, which leads to the spectral variability of the same species (Ray, 1995); (ii) the likelihood of nonlinear mixing due to the multiple scattering of light from the species canopies and/or leaves, which leads to an overestimation of green vegetation species; and (iii) vegetation species adaptations to harsh environmental factors, which make the spectral reflectance of these species different (Ray and Murray, 1996).

Multispectral imagery (i.e. Landsat, IKONOS and Système Pour l'Observation de la Terre (SPOT)) are affordable, relatively available, and provide accurate data for discriminating among endangered tree species in fragmented ecosystems. In general, multispectral data have brought great opportunities for classifying land use/cover and tree species in homogenous and intact landscapes (Pu and Landry, 2012). Multispectral remote sensing has largely been used for mapping and modeling forest species, their biochemical and biophysical parameters (Ferwerda *et al.*, 2005; Davi *et al.*, 2006; Pilger, 2008). However, multispectral remotely-sensed data have high spatial resolution and provide more bands with lower spatial resolution. The lower spatial resolution of multispectral remotely-sensed data might not accurately classify land use/cover and tree species in a large landscape (Cho *et al.*, 2013). Multiple objects within a pixel can lead to spectral confusion and poor discriminating among discrete and continuous cover classes causing ambiguous land use/cover and tree species classes (Aplin, 2003). Conversely, these challenges hinder the classification of land use/cover classes and tree species when multispectral remotely-sensed data are captured from heterogeneous and fragmented forest ecosystems (Foody, 2002).

The advent of hyperspectral remotely-sensed data can overcome the limitations of multispectral data by offering spectral data of many and contiguous wavebands for more reliable and accurate land use/cover and tree species mapping (Vaiphasa *et al.*, 2007; Petropoulos *et al.*, 2012). Hyperspectral remotely-sensed data are considered one of the most advanced techniques for studying species dynamics, because it has many narrow wavelengths of less than 10 nm (Vaiphasa *et al.*, 2007). Many studies that have investigated the use of hyperspectral remote sensing in characterizing vegetation and have demonstrated the usefulness of the red edge region of the electromagnetic spectrum for mapping vegetation species and for estimating their biochemical and biophysical characteristics (Mutanga, 2005; Cho and Skidmore, 2006; Adam *et al.*, 2010; Adam *et al.*, 2012). The red edge region offers accurate details on the variations in species structure and condition (Cho *et al.*, 2008a). However, the uses of hyperspectral data have their own limitations in terms of cost, availability and high dimensionality.

Recently, the newly launched multispectral satellites like Sentinel-2, WorldView-2, WorldView-3, and RapidEye, have brought great opportunities for mapping land use/cover and tree species. Among these satellites, WorldView-2 which provides relatively better spectral resolution of eight wavebands with a pixel size of 2m (DigitalGlobe, 2010; Omar, 2010). WorldView-2 offers key spectral bands such as red edge and yellow, depicting tree spectral characteristics more accurately as compared to the other conventional wavebands like green and red (Dlamini, 2010; Ozdemir and Karnieli, 2011; Pu and Landry, 2012). For instance, the usefulness of WorldView-2 data for mapping individual tree species in indigenous and plantation forest ecosystems and characterizing coastal landscapes, has been demonstrated in many studies, among others, Navulur (2009), Chen *et al.* (2011) and Peerbhay *et al.* (2014). On the other hand, literature demonstrates that WorldView-2 satellite offers a better alternative data source for quantifying and providing timely spatial variations of forest biophysical (e.g. LAI and biomass) and biochemical (e.g. N) attributes (Cho *et al.*, 2013; Cho *et al.*, 2014). However, there is a lack of knowledge on the performance of WorldView-2 spectral subsets and variables for mapping land use/cover and distinguishing between endangered tree species as well as predicting their LAI, N and C_N concentrations in fragmented landscapes. Consequently, the challenges would be to assess and monitor both the distribution and quantity (e.g. LAI) of endangered tree species using remotely-sensed data in order to provide the most suitable level of detail and accuracy for mapping purposes. This facilitates a better understanding of the species quantity interaction in a spatial context. Therefore, this study focuses on the use of WorldView-2 data to accomplish the task of mapping land use/cover and to distinguish among six endangered tree species. Specifically, the study explored the advent of the WorldView-2 spectral subsets (WorldView-2 eight bands, WorldView-2 standard bands, and WorldView-2 additional bands) for mapping land use/cover and endangered tree species, as well as for estimating some tree and forest foliar biophysical (LAI) and biochemical (N and C_N) traits. The results of this study would therefore fill the gap in scientific research that requires further investigations necessary for reliable and accurate monitoring of indigenous forest especially in a fragmented ecosystem. The strength of the WorldView-2 multispectral data, however, needs to be further investigated for estimating and mapping other biophysical (e.g. biomass) biochemicals (e.g. P and K) traits in indigenous and tropical forest ecosystems within a fragmented ecosystem.

1.4 Aim and Objectives

The aim of the study was to investigate the utility of multispectral WorldView-2 data for mapping six endangered tree species and other land use/cover types in the fragmented Dukuduku indigenous forest ecosystem in South Africa. The study further explored the possibility to estimate biophysical and biochemical traits of the six endangered tree species.

The specific objectives of the current study were to:

1. Investigate the utility of high spatial resolution multispectral WorldView-2 data and advanced machine learning classification algorithms for mapping the land use/cover classes in a fragmented Dukuduku indigenous forest ecosystem;

2. Examine the utility of the advanced multispectral WorldView-2 data for mapping endangered tree species in the fragmented Dukuduku indigenous forest ecosystem using machine learning classification algorithms;

3. Test the utility of spectral vegetation indices (SVIs) calculated from the multispectral WorldView-2 data for predicting endangered tree species LAI in the fragmented and intact indigenous forest ecosystems using machine learning regression algorithms; and

4. Map fragmented and intact indigenous forest leaf N and C_N concentrations using multispectral WorldView-2 spectral variables and machine learning regression algorithms.

1.5 Scope of the Study

This study successfully mapped the spatial distribution of eight land use/cover classes in the Dukuduku area of northern KwaZulu-Natal province, South Africa using the very high spatial resolution multispectral WorldView-2 data. The utility of multispectral WorldView-2 datasets for mapping six endangered tree species in a fragmented Dukuduku forest ecosystem was also investigated using support vector machines (SVM) and artificial neural networks (ANN) classification algorithms. The study further demonstrates the utility of multispectral WorldView-2 derived SVIs and the SVM and ANN regression algorithms for estimating LAI of six endangered tree species under different management practices. The strength and potential of multispectral WorldView-2 dataset for estimating the concentrations of fragmented and intact indigenous forests foliar N and C_N concentrations was subsequently evaluated using 24 SVIs derived from WorldView-2 data and two robust and effective machine learning SVM and ANN regression algorithms. In this context, relatively more emphasis was placed on the prediction of forest and endangered tree species biophysical and biochemical characteristics because it is considered as the most limiting factor for the ecological, hydrological, and economic roles of tree species in indigenous forest ecosystem (Eldeen, 2005; Brendler *et al.*, 2010; Pouteau *et al.*, 2012).

1.6 Description of the Study Area

1.6.1 General

The study area includes two forest management protocols which are commonly practiced in the southern African indigenous forest ecosystems. These include fragmented and intact indigenous forests ecosystems. Fragmented indigenous forests are managed by the local communities and the traditional leaders. These are the forests and/or woodlands where individual tree crowns do not overlay to form a continuous canopy layer and are largely spaced. Intact indigenous forests are managed by the officials (e.g. Department of Forestry) and are defined as areas of land that are occupied by trees of continuous canopy layer. The Dukuduku region consists of fragmented and intact indigenous forest ecosystems which are dominated by rare tree species.

1.6.2 Dukuduku Indigenous Forest Ecosystem

The study was conducted in the Dukuduku indegenous forest which is located on the northern bank of Umfolozi River floodplain, South Africa. The study area is located between latitute 28°52'25''S and longitude 32°17'23''E (Figure 1.2). The Dukuduku forest area covers more than 6000 hectares (ha) of indigenous coastal forest on the rolling savannahs of the inland across the dune line along the KwaZulu-Natal coast, from southern KwaZulu-Natal province to beyond the Mozambican boundary (Ndlovu, 2013). The subtropical climate dominating the study area has warm moist summers and mild dry winters. The mean daily maximum temperatures are 26°C in January and 21°C in July, while mean daily minimum temperatures are 19°C in January and 9°C in July (Von Maltitz *et al.*, 2003). The rainy season falls between November and March with a mean annual rainfall of 1250 mm (Von Maltitz *et al.*, 2003). In South Africa, the Dukuduku forest is also considered to be one of the largest remaining stretches of coastal forest. However, due to a high number of individual illegal settlement and intensive agricultural systems, the natural vegetation surrounding the forest has been extensively removed (Cho *et al.*, 2012; Ndlovu, 2013). Similarly, increasing human activities and settlement in the area have led to an increase in ecosystem fragmentation (van Wyk *et al.*, 2006; Ndlovu, 2013). Therefore, the Dukuduku forest is facing many threats presented by the destruction of indigenous vegetation, forest plantation and agricultural systems. The majority of forests in the area are dominated by several indigenous vegetation including different age groups and other types land use/cover. The most dominant tree species in the area include *Syzygium cordatum* and C*ussonia zuluensis*. However, it is observed that six other tree species (Figure 1.1) in the Dukuduku forest are under severe threat and endangered in both the fragmented and intact forest ecosystems as they face rapid harvesting for woodcarving and other uses (Watt and Breyer-Brandwijk, 1962; Jäger *et al.*, 1996; Mlambo, 2013).

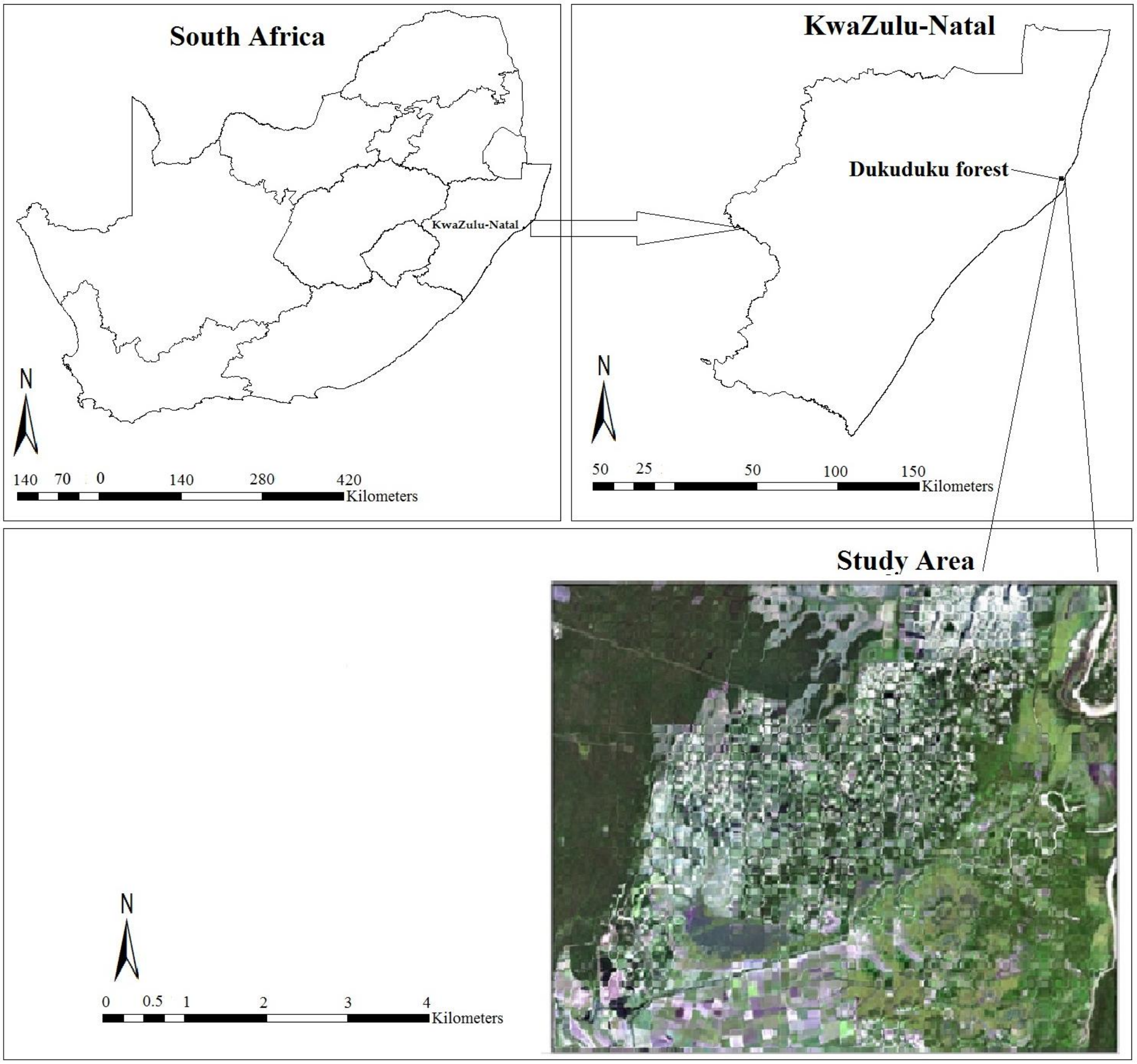


Figure 1.2: The location of the Dukuduku indigenous forest in KwaZulu-Natal province, South Africa

1.7 Outline of the Thesis

This thesis consists of a set of research papers addressing each of the objectives listed in Section 1.4. The papers have been either published or under preparation. Three papers have already been published, and one is still in preparation. Each paper has been written as a stand-alone article that can be read separately from the rest of the thesis but that draws separate conclusions that link to the overall research objective. As a result, a number of overlaps and replications occur in the sections “Introduction” and “Methods” in the different chapters. This problem is deemed to be of little significance when one considers the critical peer-review process and the fact that the different chapters are papers that can be read separately without losing the overall context. This thesis consists of six chapters from the introduction to the synthesis.

Chapter One: This chapter serves as an introduction to the study.

Chapter Two: In this chapter a new very high spatial resolution multispectral WorldView-2 imagery with additional band sets, is evaluated for classifying the land use/cover classes. Specifically, the study explored the utility of the three WorldView-2 spectral datasets including (8B: all bands, SB: four standard bands, and AB: four additional bands) to better improve the classification accuracy and distinguishing amongst various land use/cover classes using the SVM algorithm within a Dukuduku fragmented ecosystem.

Chapter Three: This chapter focuses on examining the utility of the WorldView-2 data for mapping six endangered tree species and other land use/cover classes in the fragmented Dukuduku indigenous forest using the advanced SVM and ANN classification algorithms. Furthermore, the chapter looks at utilization of the advent of the additional bands of WorldView-2 for mapping endangered tree species.

Chapter Four: This chapter investigates the potential of SVIs calculated from high spatial resolution WorldView-2 imagery to better improve prediction of six endangered tree species LAI in fragmented and intact indigenous forest ecosystems. The study also examines whether there are significant differences between the trees LAI of intact and fragmented indigenous forest ecosystems.

Chapter Five: This chapter provides an evaluation of the utility of different SVIs derived from WorldView-2 data for estimating and mapping intact and fragmented indigenous forests leaf N and C_N concentrations. Two robust machine learning SVM and ANN regression methods were used for deriving the predictive models. Furthermore, the chapter tests if there are significant differences in foliar N and C_N concentrations between the intact and fragmented indigenous forest ecosystems.

Chapter Six: This chapter highlights the summary and the main study outputs. Conclusions are also derived based on the findings of the preceding chapter. Some relevant recommendations for future study on the application of remote sensing in indigenous tree species mapping are outlined. A special emphasis is directed to the operational use of remotely-sensed data for mapping and monitoring endangered tree species.

At the end, a list of references is provided.

CHAPTER TWO

Exploring the Utility of the Additional WorldView-2 Bands and Support Vector Machines for Mapping Land Use/Cover in a Fragmented Ecosystem, South Africa

This chapter is based on:

1. G. **Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, "Exploring the utility of the additional WorldView-2 bands and support vector machines in mapping land use/cover in a fragmented ecosystem, South Africa" *South African Journal of Geomatics*, 4 (4), pp. 414-433, 2015.

2. G. **Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, "Potential utility of the WorldView-2 multispectral data and support vector machines algorithm to classifying land use/cover in Dukuduku landscape, Kwazulu-Natal, South Africa" Presented at the *10th African Association of Remote Sensing of the Environment* (*AARSE*) *conference*, Johannesburg, South Africa, pp. 309-318, 2014.

ABSTRACT

Land use/cover classification is a key research field in environmental applications of remote sensing on the earth's surface. The advent of new high resolution multispectral sensors with unique bands has provided an opportunity to map the spatial distribution of detailed land use/cover classes over a large fragmented area. The objectives of the present study were to: (1) map land use/cover classes using multispectral WorldView-2 data and support vector machines (SVM) in a fragmented ecosystem; and (2) compare the accuracy of three WorldView-2 spectral data sets for distinguishing amongst various land use/cover classes in a fragmented ecosystem. WorldView-2 spectral subsets comprising four standard bands (SB: blue, green, red and near infrared-1), and four additional bands (AB: coastal blue, yellow, red edge and near infrared-2) as well as all eight multispectral bands (8bands: 8B) were used for land use/cover classification. The overall classification accuracies of 78% for the land use/cover classification based on all eight multispectral bands. However, overall accuracy was 64% and 51% for the AB and SB land use/cover classifications, respectively. This indicates that additional bands such as red edge improve land use/cover classification. There were significant differences between the performance of all WorldView-2 subset pair comparisons (8B versus SB, 8B versus AB and SB versus AB) as demonstrated by the results of McNemar's test (Z score $\geq$1.96). This study concludes that WorldView-2 multispectral data and the SVM classifier have the potential to map land use/cover classes in a fragmented ecosystem. The study also offers relatively accurate information that is important for the indigenous forest managers in KwaZulu-Natal, South Africa for making informed decisions regarding conservation and management of land use/cover patterns.

Keywords: Land use/cover classification, fragmented ecosystem, WorldView-2, support vector machines

2.1 Introduction

Land use/cover is a fundamental variable that influences and links with many parts of human and physical systems and is a vital data component for many aspects of environmental change (Foody, 2002; Otukei and Blaschke, 2010). The changes in land use/cover have significant effects on basic ecosystem processes including biogeochemical cycling and land degradation (Penner, 1994; Foley *et al.*, 2005; Otukei and Blaschke, 2010). Similarly, the land use/cover maps are be used for environmental monitoring, management as well as modeling (Otukei and Blaschke, 2010). Despite this important role, land use/cover mapping is still faces a complex challenge in relation to ambiguous classes used (Cingolani *et al.*, 2004; Otukei and Blaschke, 2010). Additionally, fragmented ecosystems in many parts of Africa are characterized by the removal and clearing of the forest for pasture, agriculture and settlements leading to vegetation species loss (van Wyk *et al.*, 2006; Cho *et al.*, 2013). In most cases, indigenous forests are fragmented into patches of various sizes and shapes surrounded by a matrix of different land use/cover classes (Benitez-Malvido, 1998; Cho *et al.*, 2013). In this context, information relating to the dynamics, distribution and productivity of land use/cover is not only beneficial to the source of economic security but is also needed for fragmented ecosystems inventory, management and monitoring (Cingolani *et al.*, 2004; Pignatti *et al.*, 2009; Cho *et al.*, 2013). In order to meet the management and monitoring requirements of fragmented ecosystems, more specific information from land use/cover surveys and inventories is needed. However, it is quite difficult and challenging to produce land use/cover maps using traditional field survey approaches.

Traditional approaches are complex and require intensive fieldwork which is a costly and time-consuming, particularly in highly heterogeneous and fragmented ecosystems. In this regard, remote sensing is a particularly useful tool, as it has successfully been used for tree species and land cover classification (Clark *et al.*, 2005; Larsen, 2007). Moreover, multispectral sensors such as Landsat and SPOT cover large areas of the earth’s surface at repeated time intervals, making remote sensing a perfect alternative to traditional approaches for land use/cover. Recently, the developments of high spatial resolution multispectral sensors such as IKONOS have brought unique opportunities for classifying and monitoring land use/cover (Pu and Landry, 2012). Multispectral data either have high spatial resolution but offer only a few bands like blue, green, red, and near infrared (NIR), or they offer relatively more bands but with lower spatial resolution. The low spatial resolution multispectral sensors might not accurately map land use/cover classes in a heterogeneous and fragmented ecosystem (Foody, 2002; Cho *et al.*, 2012). Multiple objects within a pixel in such a case can lead to spectral confusion and poor distinction amongst discrete and continuous cover types, resulting in ambiguous land use/cover classes used (Cingolani *et al.*, 2004). On the other hand, multispectral data of very fine spatial resolution may not capture in the intra-class variability accurately when course land use/cover classes are mapped. As far as the spectral resolution is concerned, the advent of hyperspectral data can overcome the limitations of multispectral data by providing spectral data of many and contiguous wavebands (Vaiphasa *et al.*, 2007) for more accurate and reliable land use/cover maps (Pal, 2006; Petropoulos *et al.*, 2012). However, the use of hyperspectral data has its own limitations in terms of cost, availability, processing, and high dimensionality (Vaiphasa *et al.*, 2007; Dalponte *et al.*, 2009).

Recently, high spatial resolution multispectral sensors such as RapidEye, WorldView-2 and Sentinel-2 have been designed with relatively fewer additional bands to overcome the limitations of their spectral capabilities over other high spatial resolution multispectral sensors of conventional bands (standard bands) such as QuickBird. The potential of WorldView-2 data, for instance, has been demonstrated in a number of diverse studies that include, predicting and mapping forest structural parameters (Ozdemir and Karnieli, 2011), urban land cover mapping (Zhou *et al.*, 2012), and discriminating commercial forest species (Peerbhay *et al.*, 2014). These studies have demonstrated the utility of the eight available spectral bands of WorldView-2 imagery for mapping and predicting a feature of interest and concluded that WorldView-2 data have considerably improved the classification and prediction accuracies compared to conventional sensors. For instance, Pu and Landry (2012) explored the use of WorldView-2 and IKONOS data sets for mapping tree species and found that the bands available in WorldView-2 significantly increased the classification accuracy compared to the bands available in IKONOS sensor. These studies have limitations related to paucity of knowledge on the performance of WorldView-2 spectral subsets in vegetation and land use/cover types with advanced classification methods.

Various classification methods have been implemented in order to map vegetation species and land use/cover classes using WorldView-2 data. These methods include discriminant analysis

(Pu and Landry, 2012), decision trees (Heumann, 2011), and maximum likelihood and minimum distance to the mean classifiers (Cho *et al.*, 2011; McCarthy and Halls, 2014). All these classifiers have used supervised classification methods with conventional multispectral data. Amongst these classifiers, the maximum likelihood and minimum distance to the mean have been the most widely used (Kavzoglu and Mather, 2003; Otukei and Blaschke, 2010). These two classifiers have the ability to generate acceptable accuracy, simplicity and availability in most image processing packages (Zhang *et al.*, 2007; Cho *et al.*, 2011). However, all these classifiers have their own limitations that are related particularly to distributional assumptions and to mapping areas with limited training samples (Kavzoglu and Mather, 2003; Cho *et al.*, 2012).

To tackle these problems, powerful classification methods are essentially used for mapping land use/cover (Lu and Weng, 2007). These classification methods include, ANN, random forest (RF) and SVM (Civco, 1993; Pal, 2003; Pal, 2006; Heumann, 2011). SVM, ANN and RF offer a precise way to map land use/cover and tree species from remote sensing images without depending on any assumptions (Breiman, 2001; Dixon and Candade, 2008; Xiong *et al.*, 2010). RF is widely used for mining and classifying hyperspectral data for plant species identification and classification (Lawrence *et al.*, 2006; Verikas *et al.*, 2011; Naidoo *et al.*, 2012) while the application of SVM and ANN classifiers have been mainly explored in forest species classification using multispectral imagery (Xiong *et al.*, 2010; Yoon *et al.*, 2011; Nitze *et al.*, 2012). Amongst these methods, attention has been paid to the use of SVM classifier due to its superior image-handling ability (Vapnik, 1998). Numerous studies have used SVM classifier and multispectral imagery for land use/cover mapping (Kavzoglu and Colkesen, 2009; Otukei and Blaschke, 2010; Petropoulos *et al.*, 2012; Adam *et al.*, 2014). These researchers have found relatively better or similar performances obtained by this classifier as compared to other classifiers when multispectral and hyperspectral data were used. The exploration of the utility of WorldView-2 additional bands for improving the accuracy of land use/cover maps in a fragmented ecosystem is needed. From the available literature and to the best of the researcher's knowledge, no study utilized different WorldView-2 spectral subsets and SVM classifier to map land use/cover class in a fragmented ecosystem. Therefore, the objectives of the present study were to: (1) map different land use/cover classes using WorldView-2 data and SVM classifier in a fragmented ecosystem; and (2) compare the accuracy of three WorldView-2 spectral subsets in distinguishing amongst various land use/cover classes in a fragmented ecosystem.

2.2 Methodology

2.2.1 Image Acquisition and Pre-processing

A cloud-free WorldView-2 multispectral image covering the study area was acquired on 1st December 2013. WorldView-2 image (Figure 2.1) consists of eight multispectral bands in the 400-1040 nm spectral range with a spatial resolution of 2 m and swath width of 16.4 km at nadir. The spectral bands of WorldView-2 are coastal blue (400–450 nm), blue (450–510 nm), green (510–580 nm), yellow (585–625 nm), red (630–690 nm), red edge (705–745 nm), NIR-1 (770–895 nm), and NIR-2 (860–1040 nm). The image was atmospherically corrected and transformed to canopy reflectance using the Quick Atmospheric Correction (QUAC) extension in Environment for Visualizing Images (ENVI) 4.7 software (ENVI, 2009). QUAC determines atmospheric compensation parameters directly from the information contained within the image (pixel spectra) thus allowing for the retrieval of accurate reflectance spectra (Shen *et al.*, 2005; Agrawal and Sarup, 2011). The image was then referenced to the Universal Transverse Mercator (UTM zone 36 South) projection using WGS-84 Geodetic datum. The acquired image was geometrically corrected by DigitalGlobe™. After the geometric and atmospheric correction, the WorldView-2 image was spectral subsets to four SB and four new additional bands (AB). These subsets together with all eight 8B of WorldView-2 were compared for mapping land use/cover classes using SVM supervised classifier.

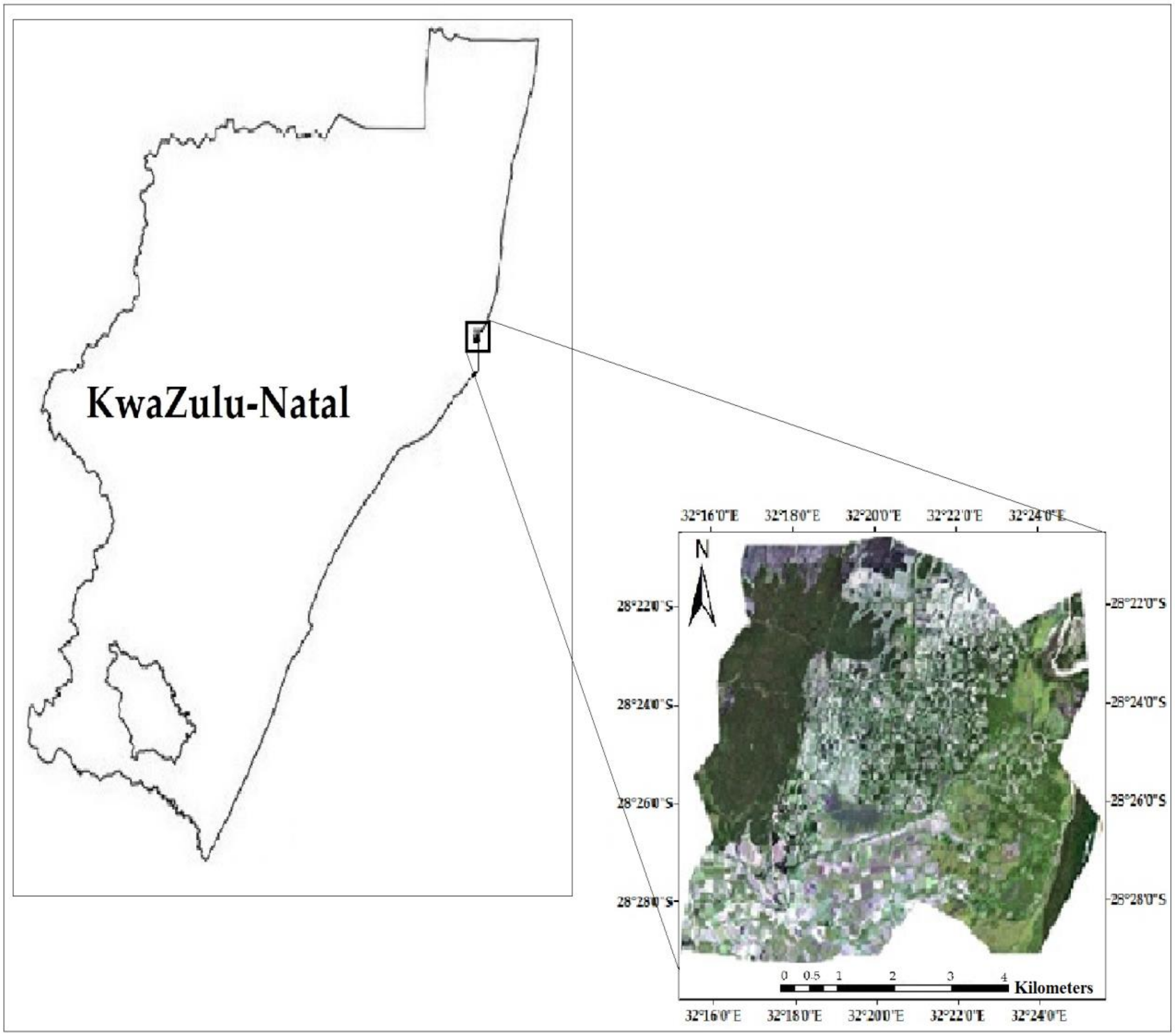


Figure 2.1: Location and a true-color WorldView-2 image of the study area

2.2.2 Field Data Collection

The field campaign was carried out on 7th December 2013, within a week of the WorldView-2 imagery acquisition. This was done in order to collect ground reference data of eight land use/cover classes, namely dune forest (DF), indigenous forest (IF), fragmented forest (FF), *Eucalyptus spp* (EP), *Pinus spp* (PN), mature sugarcane (MS), young sugarcane (YS), and grassland (GL) using a handheld Leica GS20 GPS with sub-meter accuracy. During the field visit, a total of 75 sample data points were collected for each class. The ground reference data were collected using random sampling protocol to adequately sample land use/cover classes based on their representative sizes within the study area. The reference data were then divided randomly into training (70%) and test (30%) dataset using Hawth's Analysis tool in ArcGIS 9.3. The SVM classifier was trained on 70% of a randomly selected holdout sample and final accuracy was assessed using the remaining 30% samples.

2.3 Statistical Analysis

The effectiveness of SVM classifier to map land use/cover classes was investigated in this study. The classifier was trained on 70% (n = 53) of a randomly selected holdout sample and final accuracy assessments were evaluated using the remaining 30% (n = 22) of the dataset. When the training positions and classes were allocated, classification signatures were created for the eight land use/cover classes in the study area. After assessing and adjusting the signatures, SVM supervised classification method was then employed to classify the WorldView-2 image. SVM parameters were optimized and then input into the ENVI software to map the classes on WorldView-2 image. The e1071 library version 2.15.2 in R statistical packages (R Development Core, 2012) was employed for SVM parameters optimization.

2.3.1 Support Vector Machines (SVM) Classifier

SVM (Cortes and Vapnik, 1995) is a learning technique that analyzes data and recognizes patterns. The algorithm was successfully used for classifying multispectral images (Cihlar, 2000; Muñoz-Villers and López-Blanco, 2008; Kavzoglu and Colkesen, 2009; Mountrakis *et al.*, 2011; Petropoulos *et al.*, 2012) for various purposes. SVM was originally introduced as a binary classifier (Cortes and Vapnik, 1995). However, real remote sensing problems usually include identification of multiple classes. Amendments are made to the simple SVM binary classifier to run as a multi-class classifier using methods such as one-against-one (OAO) and one-against-all (OAA) procedures. The algorithm is then assigned to the correct class by using a voting mechanism (Mathur and Foody, 2008; Krahwinkler and Rossman, 2011). SVM is a distribution-free algorithm that requires few training data points, and does not encounter any overfitting problem (Cortes and Vapnik, 1995; Burges, 1998; Brown *et al.*, 1999; Everingham *et al.*, 2007). SVM attempts to maximize the margin; that is the distance between the data points of each class to the optimal separating linear hyperplane axes created from each variable (Petropoulos *et al.*, 2011). There are two supporting hyperplanes in the boundaries of the data distribution and the data points on the margin of these hyperplanes are the support vectors of the algorithm and the optimal hyperplaneis in the middle of the margin. Many classes are not linearly separable, hence SVM uses kernel trick to adjust for finding a nonlinear (e.g. polynomial) separating hyperplane in a high-dimensional feature space using optimization function (Hornik *et al.*, 2006; Yang, 2011). For detailed description on SVM theory and principles see Cortes and Vapnik (1995), Burges (1998), Hornik *et al.* (2006) and Mathur and Foody (2008).

In the present study, WorldView-2 subsets (8B, SB and AB) were used for defining the space feature of SVM. A radial basis function was used to find an optimal hyperplane that can differentiate amongst land use/cover classes in the Dukuduku landscape. Radial basis function performed relatively better for classifying remotely-sensed data when compared with the polynomial kernel (Huang *et al.*, 2002; Kavzoglu and Colkesen, 2009; Pal, 2009; Yang, 2011) and requires optimization of only two parameters. These are the cost function (C) which controls the tradeoff between maximization of the margin width and minimizing the number of misclassified data points in the training dataset samples, and gamma (λ) which is the width parameter of the radial basis function kernel (Hornik *et al.*, 2006). The OAO procedure is used to implement a multiclass-based SVM model. The regularization of the C and λ parameters was performed using a 10-fold cross validation method (Hsu *et al.*, 2009; Yang, 2011). The dataset was divided into 10 subsets of equal size, SVM models were then trained on nine subset samples, and tested on the removed one and the process was repeated ten times until all subset samples had served as test samples. The pair parameter that minimizes the classification error was then considered as the optimal values for final classification.

2.3.2 Accuracy Assessment

A confusion matrix was constructed to compare the true class with the class assigned by SVM and to calculate the overall accuracy, producer's accuracy (PA) and user's accuracy (UA). Overall accuracy has the advantage of being directly interpretable as the percentage (%) between the number of correctly classified samples and the number of test samples. PA refers to the probability of a certain class being correctly classified, while UA represents the likelihood that a sample belongs to a specific class and the classifier accurately assigns it such a class. In addition, two parameters were calculated from the cross-tabulation matrix to evaluate the reliability of SVM classifier. These include quantity disagreement (QD) and allocation disagreement (AD) which were developed by Pontius and Millones (2011). The QD is the amount of the contrast between the number of test data and predicted data, while the AD describes the number of expected classes that have less than optimal spatial location in comparison to the test data. Depending on the accuracy metrics achieved for each WorldView-2 dataset in each accuracy assessment method, a statistical analysis can be performed to test if there was any significant difference between the classification results of three WorldView-2 spectral subsets. Hence, McNemar's test was performed to test whether there were any significant differences amongst the confusion matrices of the three WorldView-2 spectral datasets. McNemar's test is a nonparametric test based upon standardized normal test statistic calculated from error matrices of SVM classifier given as follows (Foody, 2004a; Leeuw *et al.*, 2006):

$$Z = \frac{f12 - f21}{\sqrt{f12 + f21}} \qquad \text{Equation 2. 1}$$

where $f12$ denotes the number of samples that are misclassified on the first confusion matrix but correctly classified on the second confusion matrix. $f21$ denotes the number of samples that are misclassified on the second confusion matrix but correctly classified on the first confusion matrix. A difference in accuracy between the confusion matrices of different WorldView-2 spectral subsets is statistically significant ($p \leq 0.05$) if Z score is more than 1.96 (Foody, 2004a; Leeuw *et al.*, 2006).

2.4 Results

2.4.1 Optimization of Support Vector Machines

The results of grid search and 10-fold cross validation method indicated optimal values of λ and *C* for SVM, respectively, of 0.1 and 10 for 8B, 0.1 and 1000 for SB, and 0.1 and 100 for AB. When these optimal values were input into SVM classifier, minimum classification error of 32.00%, 35.30%, and 36.30% for 8B, SB and AB, respectively were obtained. The optimal values of λ and *C* were input into SVM classification algorithm to map different land use/cover classes in the study area using the three WorldView-2 spectral data sets.

2.4.2 Accuracy Assessment

Figure 2.2 shows land use/cover maps obtained using SVM classifier. The main visual difference between the maps is that a relatively homogeneous map was produced when the WorldView-2

8B was used as compared with other spectral subsets. The maps also show that the Dukuduku indigenous forest was mainly surrounded by grassland and commercial forest plantation, while the grassland on the northeastern part of the study area was fragmented. Most of sugarcane farms in the study area were at a mature growth stage.

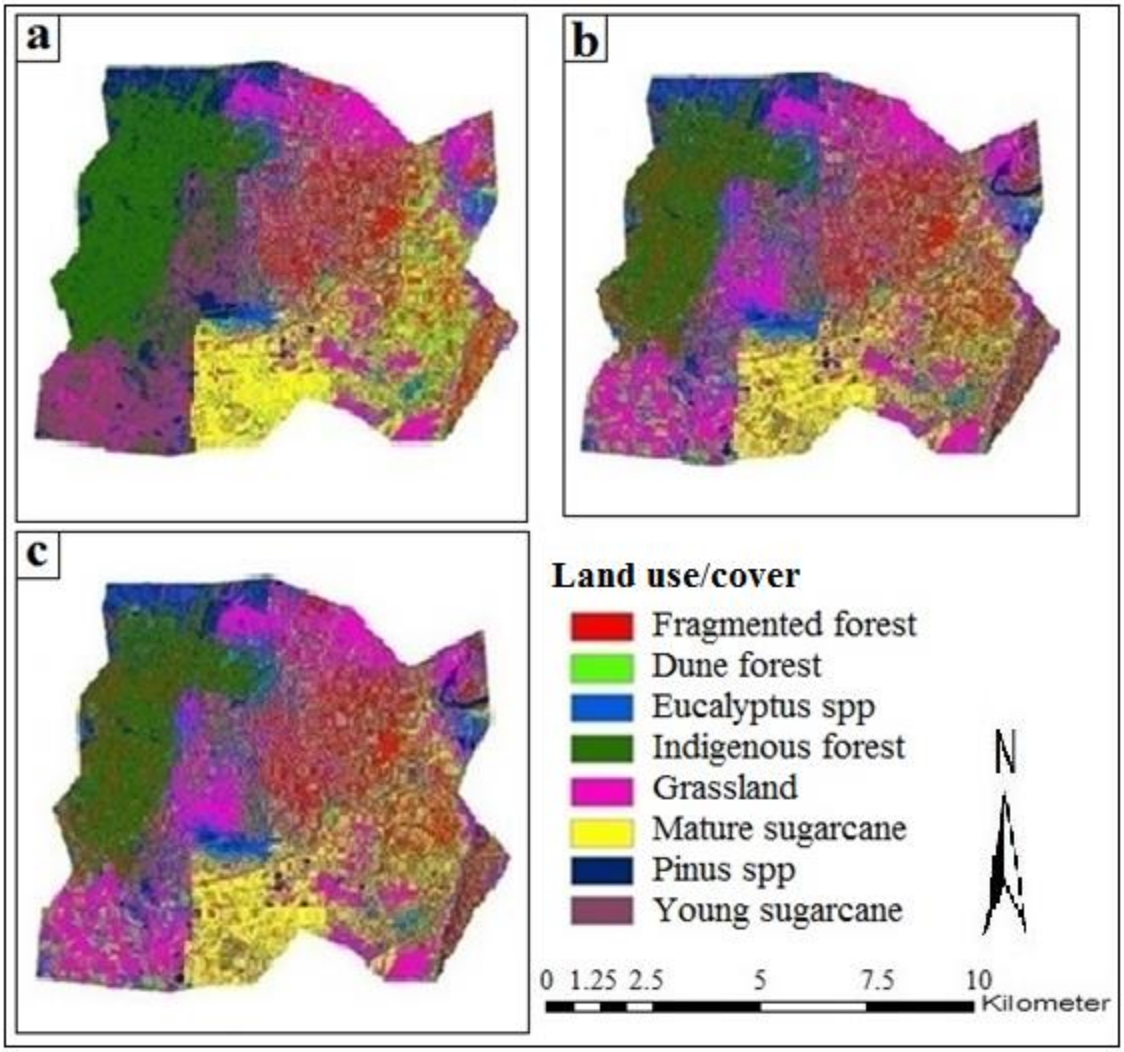


Figure 2.2: Land use/cover classification maps obtained using support vector machines classifier: (a) all eight WorldView-2 bands, (b) four standard WorldView-2 bands and (c) four additional WorldView-2 bands

Furthermore, the overall accuracy assessment for mapping land use/cover class was 78% (total disagreement = 22%), 51% (total disagreement = 49%) and 64% (total disagreement = 36%) using 8B, SB and AB respectively (Tables 2.1, 2.2 and 2.3). SVM classifier obtained QD values of 5%, 13% and 14% for 8B, SB and AB, respectively (Tables 2.1, 2.2 and 2.3). The tables also show relatively high AD values of 17%, 36% and 22%.

Table 2.1: Classification confusion matrix of support vector machines (SVM) classifier using WorldView-2 8B for the 30% test data sets. The confusion matrix includes overall accuracy (OA), quantity disagreements (QD) and allocation disagreements (AD)

Class	**FF**	**EP**	**IF**	**GL**	**MS**	**DF**	**PN**	**YS**	**Total**
FF	**15**	0	0	3	1	5	0	0	24
EP	1	**20**	5	0	0	1	0	0	27
IF	0	1	**13**	2	1	0	1	1	19
GL	1	0	2	**17**	1	0	1	0	22
MS	3	0	1	0	**16**	0	0	0	20
DF	2	1	0	0	1	**16**	0	0	20
PN	0	0	1	0	2	0	**20**	1	24
YS	0	0	0	0	0	0	0	**20**	20
Total	22	22	22	22	22	22	22	22	**176**
OA (%)	78								
QD (%)	5								
AD (%)	17								

FF = Fragmented forest, EP = *Eucalyptus spp*, IF = Indigenous forest, GL = Grassland, MS = Mature sugarcane, DF = Dune forest, PN = *Pinus spp*, YS = Young sugarcane

Table 2.2: Classification confusion matrix of support vector machines (SVM) classifier using WorldView-2 SB for the 30% test data sets. The confusion matrix includes overall accuracy (OA), quantity disagreements (QD) and allocation disagreements (AD)

Class	**FF**	**EP**	**IF**	**GL**	**MS**	**DF**	**PN**	**YS**	**Total**
FF	**13**	0	3	5	4	4	0	0	29
EP	0	**13**	3	0	0	4	0	0	20
IF	0	4	**12**	6	0	1	1	0	24
GL	0	0	0	**2**	0	0	0	0	2
MS	1	0	3	5	**8**	4	3	2	26
DF	6	4	1	1	0	**9**	0	2	23
PN	0	1	0	0	6	0	**17**	2	26
YS	2	0	0	3	4	0	1	**16**	26
Total	22	22	22	22	22	22	22	22	**176**
OA (%)	51								
QD (%)	13								
AD (%)	36								

FF = Fragmented forest, EP = *Eucalyptus spp*, IF = Indigenous forest, GL = Grassland, MS = Mature sugarcane, DF = Dune forest, PN = *Pinus spp*, YS = Young sugarcane

Table 2.3: Classification confusion matrix of support vector machines (SVM) classifier using WorldView-2 AB for the 30% test data sets. The confusion matrix includes overall accuracy (OA), quantity disagreements (QD) and allocation disagreements (AD)

Class	**FF**	**EP**	**IF**	**GL**	**MS**	**DF**	**PN**	**YS**	**Total**
FF	**19**	0	2	4	4	0	0	0	29
EP	0	**20**	2	0	0	1	0	0	23
IF	0	0	**11**	3	0	3	0	0	17
GL	0	0	1	**8**	1	0	0	0	10
MS	1	1	0	3	**9**	1	2	2	19
DF	0	1	1	0	0	**13**	0	3	18
PN	0	0	0	2	8	1	**19**	2	32
YS	2	0	5	2	0	3	1	**15**	28
Total	22	22	22	22	22	22	22	22	**176**
OA (%)	64								
QD (%)	14								
AD (%)	22								

FF = Fragmented forest, EP = *Eucalyptus spp*, IF = Indigenous forest, GL = Grassland, MS = Mature sugarcane, DF = Dune forest, PN = *Pinus spp*, YS = Young sugarcane

Generally, all land use/cover classes achieved over 70% producer's and user's accuracies, with exception of FF (PA = 63% and UA = 68%) and IF (PA of 68% and UA of 59%) when 8B subset was used (Figure 2.3). The results indicate that the values of the UA were less than 50% for dune forest, grassland, and mature sugarcane classes on SB subset (Figure 2.4) and for grassland and mature sugarcane on AB subset (Figure 2.5). According to McNemar's test, there was significant difference ($Z \geq 1.96$) at 95% confidence level amongst the confusion matrices of SVM classifier using WorldView-2 8B, SB and AB subsets (Table 2.4). Table 2.5 shows areas under each land use/cover class obtained from WorldView-2 8B, SB and AB subsets using SVM classification algorithm. The incomparable areas obtained by WorldView-2 subsets also confirm the dissimilar performance of SVM classification algorithm. The study area is dominated by indigenous forest, commercial plantation and grassland.

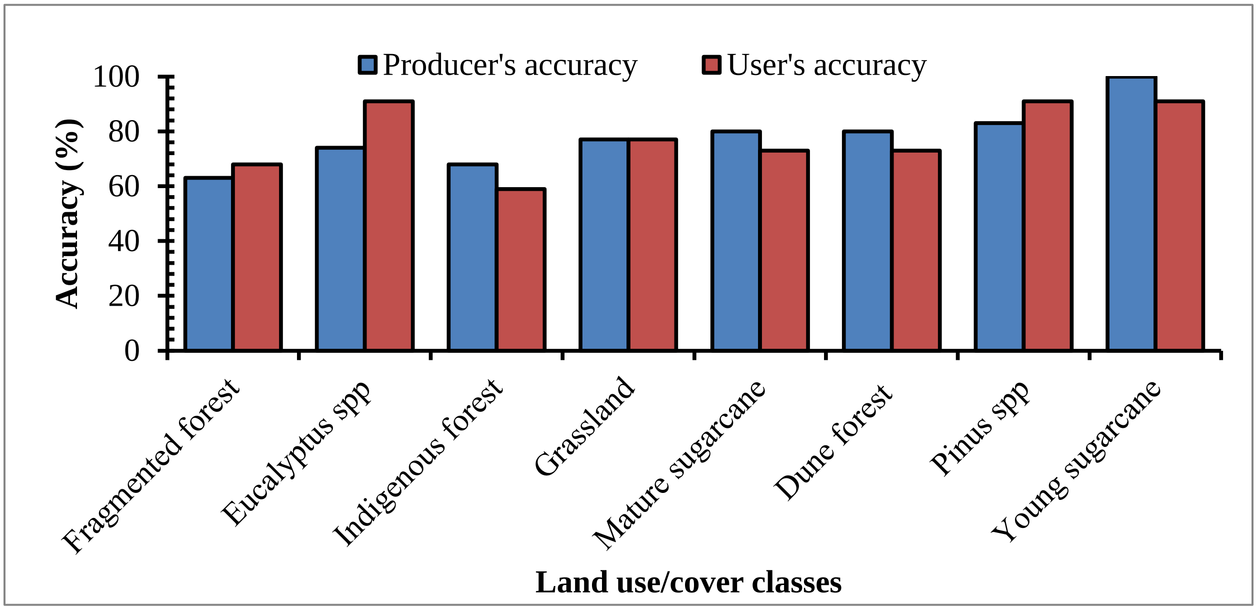


Figure 2.3: Producer's accuracy (%) and user's accuracy (%) of the studied eight land use/cover classes using all eight bands subset (8B) and support vector machines classifier for the 30% test data sets

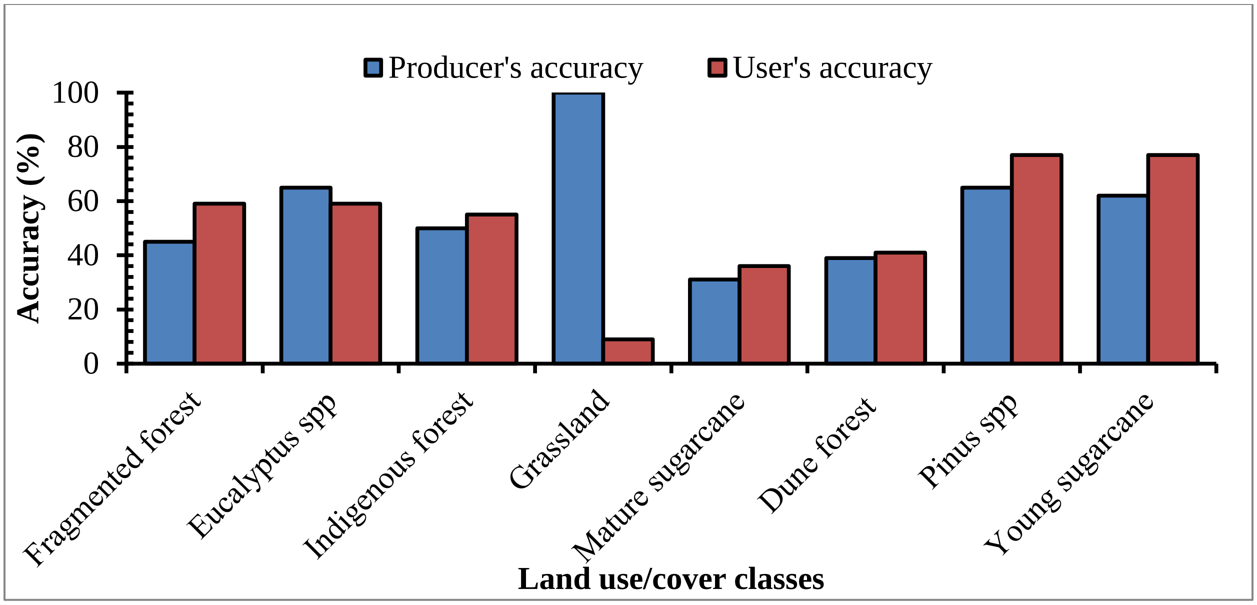


Figure 2.4: Producer's accuracy (%) and user's accuracy (%) of the studied eight land use/cover classes using standard bands subset (SB) and support vector machines classifier for the 30% test data sets

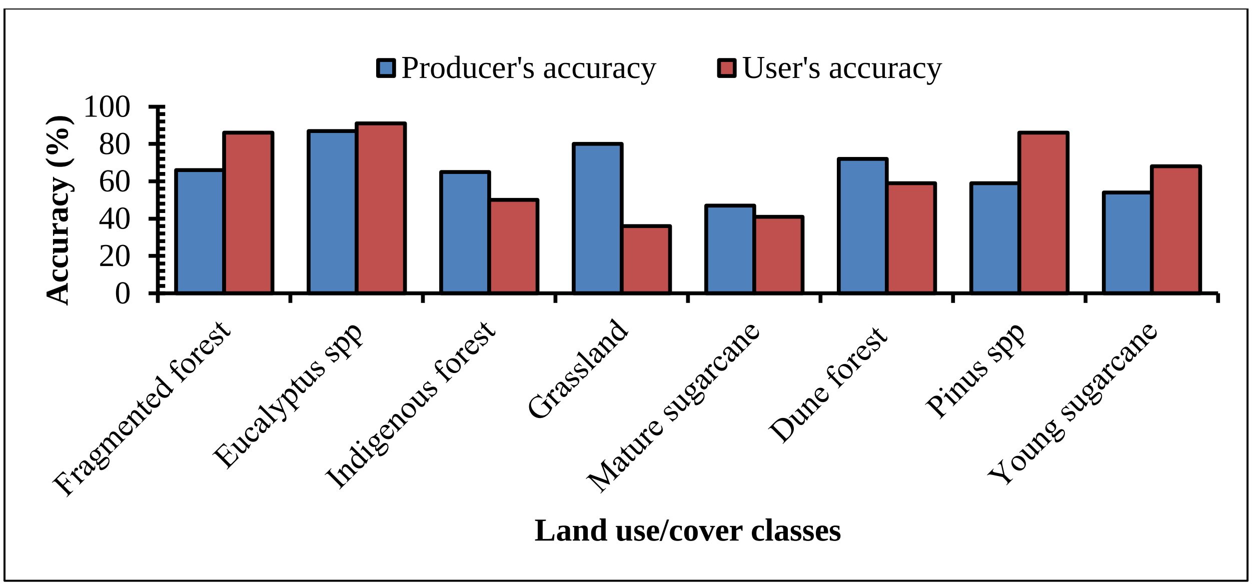


Figure 2.5: Producer's accuracy (%) and user's accuracy (%) of the studied eight land use/cover classes using additional bands subset (AB) and support vector machines classifier for the 30% test data sets

Table 2.4: McNemar's test result for comparing classification confusion matrices obtained using WorldView-2 8B, SB, and AB

		SB		
		Correctly Classified	Misclassified	Total
8B	Correctly Classified	98	12	110
	Misclassified	11	55	66
	Total	109	67	176
	Z score	5.42		
		AB		
		Correctly Classified	Misclassified	Total
8B	Correctly Classified	96	14	110
	Misclassified	12	54	66
	Total	108	68	176
	Z score	5.17		
		AB		
		Correctly Classified	Misclassified	Total
SB	Correctly Classified	93	16	109
	Misclassified	15	52	67
	Total	108	68	176
	Z score	4.52		

8B = all WorldView-2 eight bands subset, SB = four WorldView-2 standard bands subset and AB = four WorldView-2 additional bands subset

Table 2.5: Area of each land use/cover class in the study area obtained from all WorldView-2 eight bands, four WorldView-2 standard bands and four WorldView-2 additional bands subsets based on support vector machines classification algorithm

Land use/cover class	**Area (ha)**		
	8B	**SB**	**AB**
Fragmented forest	1918.52	1931.66	2131.63
Dune forest	2146.92	1945.23	1572.41
Eucalyptus spp	1579.09	2636.70	1520.79
Indigenous forest	4292.60	2814.09	4569.94
Grassland	2762.34	4388.52	2697.19
Mature sugarcane	2348.33	2106.46	2444.29
Pinus spp	1848.63	1818.57	1700.74

Table 2.5 (Continued)

Land use/cover class	Area (ha)		
	8B	**SB**	**AB**
Young sugarcane	2991.09	2246.29	3250.52

8B = all WorldView-2 eight bands subset, SB = four WorldView-2 standard bands subset and AB = four WorldView-2 additional bands subset

2.5 Discussion

The use of different types of spectral and spatial resolutions of optical sensors have been studied extensively in land use/cover mapping. These resolutions however obtained different degrees of classification accuracies. Nonetheless, concerns on the accuracy levels still continue (Lu and Weng, 2007). Moreover, cost and availability of remotely-sensed data with suitable spatial, spectral resolutions continue to be the main constraints facing large applications in land use/cover mapping. Land use/cover maps obtained from commonly used medium spatial-resolution multispectral sensors have often been arbitrated to be inadequate for operational application (Foody, 2002). On the other hand, the use of imagery from fine spatial resolution sensors has its own limitations in terms of cost, availability, processing, and high dimensionality. The limitations that characterize medium scale remotely-sensed data prevent the application and combination of field data with remotely-sensed data (Lu and Weng, 2007). Against this background, this study demonstrates that WorldView-2 data is effective for mapping land use/cover classes using SVM classifier in a fragmented ecosystem. This result is in conformity with Cho *et al.* (2013), and Ghosh and Joshi (2014) who concluded that using WorldView-2 data with advanced classification methods in a fragmented ecosystem leads to improved classification accuracy.

The main finding of the present study was that WorldView-2 8B significantly outperformed SB and AB subsets in mapping land use/cover classes in a fragmented ecosystem. The finding is in conformity with other studies (Elsharkawy *et al.*, 2012; Pu and Landry, 2012; Peerbhay *et al.*, 2014) that demonstrated the utility of WorldView-2 8B for mapping land use/cover class in intact landscapes. There are two reasons that may have led to high accuracy. Firstly, the land use/cover class in the Dukuduku forest consists of vegetation and WorldView-2 additional bands are effective in differentiating vegetated surfaces (Marchisio *et al.*, 2010; Yang, 2011; Alsubaie, 2012). Secondly, the SVM classification algorithm is useful for land use/cover mapping of fragmented ecosystems because SVM reduces classification error on test data points without a prior assumption about their distribution (Mountrakis *et al.*, 2011; Ghosh and Joshi, 2014).

Moreover, SVM is a known versatile classifier that constructs models based on a small data from different classes (Cortes and Vapnik, 1995) maximizing the margin between the support vectors and the hyperplane. The classification error is therefore significantly minimized. In the present study, a nonlinear kernel function was used to perform SVM classification. A nonlinear kernel is an efficient method to solve inseparability problems that may be found in the land use/cover classes. The relatively good performance of SVM classifier obtained in this study is consistent with the findings of Huang *et al.* (2002), Kavzoglu and Colkesen (2009) and Petropoulos *et al.* (2012) who utilized a kernel functions analysis of SVM for classifying remotely-sensed data and concluded that the classifier leads to improved classification accuracy. A number of researchers have found that SVM was the best classification technique for mapping land use/cover using high spatial resolution imagery such as WorldView-2 (Pal, 2006; Chen, 2011; Pu and Landry, 2012). From previous studies and available literature, WorldView-2 subsets (8B, SB and AB) have never been compared for mapping land use/cover classes in areas with small dataset samples such as the fragmented ecosystem in the Dukuduku area of KwaZulu-Natal, South Africa. The study showed that SVM classifier was unable to fully deal with the high spectral variation inherent in some land use/cover classes like mature sugarcane and grassland which obtained relatively lower UA and PA (see Figures 2.4 and 2.5). This is a common problem when classifying heterogeneous landscapes using high spatial resolution (WorldView-2 image) based on per-pixel classification techniques (Lu and Weng, 2007).

Although the eight land use/cover classes could be separated accurately using only the four standard bands, the use of the WorldView-2 additional bands (coastal blue, yellow, red edge and NIR-2) led to a considerable improvement in the classification accuracy. That is expected when advanced machine learning algorithm is used with WorldView-2 data. The additional wavebands are expected to provide an increase of up to 30% in classification accuracy (Zhou *et al.*, 2012). The low UA for the GL and MS classes indicate that there is a probability that pixels classified as GL and MS may not actually exist on the ground. That is expected since the physiological age of a mature sugarcane crop could be similar to densely vegetated grassland as a result of some confounding factors such as weeds and abiotic stressors (Abdel-Rahman *et al.*, 2013) and hence show similar spectral characteristics. The relatively high allocation disagreements shown in tables 2.1, 2.2 and 2.3 of the confusion matrices were expected since pixels covered by multi-classes could probably be mismatched in terms of spatial patterns between test ground truth instances and predicated test samples. However, these classification results are of a good practical application as the QD ranged between 5% and 14% for the different classification results.

In summary, the findings of the present study are promising for accurate mapping of land use/cover in fragmented areas as it demonstrates the possibility of mapping land use/cover classes using WorldView-2 data and SVM classifier. Moreover, the results provide reliable information on land use/cover classes in the Dukuduku area that could be used for the design of management plans policies as a basis for assessing and monitoring natural resources, ecological fragmentation and the ecosystem function. This information is therefore critical in the management of one of the most valuable landscapes in South Africa. Furthermore, mapping sugarcane ages is useful for the southern African sugar industry for making informed decisions with regard to sugarcane harvesting and milling. Further research is needed to widen the use of WorldView-2 imagery in identifying the rare forest species within the indigenous forest in the north-western part of the study area. This study mapped eight course land use/cover classes using remotely-sensed data with a fine pixel size. The intra-classes variability could have exceeded the fine pixel size of the WorldView-2 image. Hence, remotely-sensed data of medium spatial resolution (e.g., 10 or 20 m) could yield relatively better classification results.

2.6 Conclusions

The present study shows a successful application of multispectral WorldView-2 data and the machine learning SVM classifier for mapping eight land use/cover classes in a fragmented ecosystem. The results showed that WorldView-2 8B significantly outperformed both SB and AB subsets in mapping land use/cover classes, achieving an overall accuracy of 78%. On the other hand, WorldView-2 AB subset yielded significantly higher classification accuracy than SB subset. The results further demonstrated that the classification error for mature sugarcane and grassland was relatively higher. The study provides land use/cover maps that could be used as essential information for decision-making regarding land management and policy strategies in the fragmented Dukuduku area. It is recommended that further studies should look at identifying threatened (rare) tree species within the indigenous Dukuduku forest.

2.7 Acknowledgments

I would like to thank the University of KwaZulu-Natal, South Africa and the University of Khartoum, Sudan for funding the research. My appreciation also extends to the R development core team for their open source packages for the statistical analysis. My gratitude further extends to Inkanyamba Development Trust and Manukelana Arts and Indigenous Nursery for facilitating the field data collection. I would also like to thank Dr. Samuel Adelabu for his helpful comments and assistance during data analysis.

CHAPTER THREE

Performance of Support Vector Machines and Artificial Neural Networks for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa

This chapter is based on:

G. **Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, "Performance of support vector machines and artificial neural networks 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,* PP (99), pp. 1-16, 2015.

ABSTRACT

Endangered tree species play a significant role in ecosystem functioning and services, land use dynamics, and other socio-economic aspects. Such aspects include ecological, economic, livelihood, security-based and well-being benefits. The development of techniques for mapping and monitoring endangered tree species is thus critical for understanding the functioning of ecosystems. The advent of advanced imaging systems and supervised learning algorithms has provided an opportunity to map endangered tree species over fragmented areas. Recently, vegetation maps have been produced using advanced imaging systems such as WorldView-2 and robust classification algorithms such as support vectors machines (SVM) and artificial neural networks (ANN). However, delineation of endangered tree species in a fragmented ecosystem using high resolution imagery has largely remained elusive due to the complexity of the species structure and their distribution. Therefore, the aim of the present study was to examine the utility of the advanced WorldView-2 data for mapping endangered tree species in the fragmented Dukuduku indigenous forest of South Africa using SVM and ANN classification algorithms. Specifically, additional WorldView-2 bands were tested for mapping six endangered tree species. WorldView-2 spectral subsets comprising four standard bands (SB) and four additional bands (AB) as well as all eight multispectral bands (8bands: 8B) were classified using SVM and ANN methods. The results showed the robustness of the two machine learning algorithms for mapping the endangered tree species with an overall accuracy of 77% for SVM and 75% for ANN using 8B. The SB produced overall accuracy of 65% for SVM and 64% for ANN. The AB produced almost the same overall accuracy of 70% for both SVM and ANN. There were significant differences between the performances of the two classification algorithms as demonstrated by the results of McNemar's test (Z score $\geq$1.96). This study concludes that SVM and ANN classification algorithms with WorldView-2 8B have the potential to map endangered tree species in the Dukuduku indigenous forest. This study offers relatively accurate information that is important for forest managers to make informed decisions regarding management, and conservation protocols of endangered tree species.

Keywords: Endangered tree species, indigenous forest, Dukuduku, WorldView-2, support vector machines, artificial neural networks

3.1 Introduction

Indigenous forests span across different parts of Africa with relatively more existence in the southern and eastern parts of the continent (Eeley *et al.*, 2001). In South Africa, indigenous forests consist of many small, fragmented and largely scattered patches and cover approximately 0.2% of the country's land surface (Benitez-Malvido, 1998; Cho *et al.*, 2012; Cho *et al.*, 2013). Apart from their ecological, economic, livelihood security and well-being, indigenous forests in the country provide some medicinal products to the communities in the rural areas and contribute to the concept of "ecosystem services" (Shackleton and Shackleton, 2004; Eldeen and van Staden, 2007; van Wyk, 2008). One such indigenous forest in South Africa is the Dukuduku forest that is located in KwaZulu-Natal province. Dukuduku forest provides varied products and usable materials for human needs that include construction and fence poles, raw material for craft work, livestock browse and medicine to the poor rural communities (Eldeen, 2005; Eldeen and van Staden, 2007; van Wyk, 2008; Brendler *et al.*, 2010; Cho *et al.*, 2012). Different tree species in the forest play a vital role in providing such useful needs. It is interesting to note that local communities in Zululand, South Africa use some of these tree species to treat human diseases such as fever, stomachache, dysentery, snake and scorpion bites, malaria, inflammations, backache and facilitating childbirth (Hutchings, 1996; Jäger *et al.*, 1996; Sewram *et al.*, 2000; Eldeen and van Staden, 2007; Brendler *et al.*, 2010). Therefore, some tree species in the Dukuduku indigenous forest have become endangered and threatened because of the rapid harvesting rate and removal (van Wyk *et al.*, 2006; Cho *et al.*, 2013). These activities have resulted in over-exploitation of natural resources and caused severe forest fragmentation and serious threats to the conservation of tree species diversity in the indigenous Dukuduku forest ecosystem (Baillie *et al.*, 2004; Kätsch, 2006; Wiersum *et al.*, 2006; Cho *et al.*, 2012).

Endangered tree species need specific management and conservation protocols in order to play significant roles in ecosystem functioning, land use dynamics, and other socio-economic aspects in the value chain (Eldeen, 2005; Lyons *et al.*, 2005; Pouteau *et al.*, 2012). This requires intensive fieldwork to geo-locate and identify endangered tree species and characterize as well as estimate their coverage and distribution (Rushton *et al.*, 2004; Pouteau *et al.*, 2012). In this context, more precise information from forest survey is needed for mapping and monitoring endangered tree species in order to develop sustainable forest management practices. However, traditional field survey protocols are costly, time-consuming and often lack the necessary geospatial accuracy. Remotely-sensed data have been regarded as valuable source of information over the past decades for classifying and monitoring forest species and vegetation communities (Clark *et al.*, 2005; Quackenbush *et al.*, 2006; Van Aardt and Wynne, 2007; Pignatti *et al.*, 2009). However, mapping of tree species (e.g. endangered tree species) still faces complex challenges in relation to ambiguous classes used. Multiple objects within a pixel can lead to spectral confusion and poor distinction amongst different cover types (Aplin, 2003; Cingolani *et al.*, 2004; Herold *et al.*, 2008). In particular, these challenges hinder the classification of tree species when multispectral data are captured in fragmented forests (Foody, 2002; Cho *et al.*, 2012). This is due to broader and fewer spectral measurements collected by some multispectral sensors like Landsat and SPOT5 which might lead to spectral overlap between tree species.

Recently, the advent of new generation satellites with high spectral and spatial resolution such as Sentinel-2, WorldView-3, WorldView-2, RapidEye and Pleiades has brought unique opportunities for mapping trees at species level. Among these satellites, WorldView-2 and WorldView-3 offer key spectral bands like yellow, coastal blue and red edge that help in depicting tree characteristics. The utility of WorldView-2 image, for instance, has been demonstrated in various studies that include, predicting and mapping forest structural parameters (Ozdemir and Karnieli, 2011), mapping of tree species (Navulur, 2009), monitoring plantation forest (Omar, 2010), mapping increaser and decreaser grass species in degraded rangelands (Mansour and Mutanga, 2012), and the detection of invasive alien plants (Dlamini, 2010). These studies discussed the utility of the eight available spectral bands of WorldView-2 imagery and concluded that the WorldView-2 data have considerably improved the classification and prediction accuracies of features of interest compared to other multispectral data. These studies, however, have some limitations related to lack of knowledge on the performance of WorldView-2 spectral subsets for mapping tree species using advanced and robust classification algorithms.

Various advanced classification algorithms such as classification trees, RF, ANN, and SVM have been used to extract tree species information from multi-sensor and multispectral remote sensing images (Atkinson and Tatnall, 1997; Breiman, 2001; Petropoulos *et al.*, 2012; Adelabu *et al.*, 2013). Amongst these classification methods, attention has been accorded to the use of RF, SVM and ANN due to their superior image-handling abilities (Dixon and Candade, 2008; Adelabu *et al.*, 2013). RF, SVM and ANN offer a precise way to map vegetation cover and tree species from remote sensing images without to depend on any assumptions (Breiman, 2001; Mutanga and Skidmore, 2004a; Dixon and Candade, 2008; Xiong *et al.*, 2010; Yoon *et al.*, 2011). RF is widely used for mining and classifying hyperspectral data for plant species identification and classification (Lawrence *et al.*, 2006; Verikas *et al.*, 2011; Naidoo *et al.*, 2012) whilst the application of SVM and ANN classification algorithms have been mainly explored in forest species classification using multispectral imagery (Xiong *et al.*, 2010; Yoon *et al.*, 2011; Nitze *et al.*, 2012). In addition, SVM is a well-known machine learning algorithm which has frequently been used to locate multiple linear or potentially nonlinear class samples by a variety of kernel approaches (Bennett and Campbell, 2000). The kernel approach takes on several methods such as polynomial and a radial basis function that have revealed accurate results for vegetation classification (Pal and Mather, 2005). Radial basis function has many advantages which include its effectiveness as it works in an infinite dimensional feature space and having a single parameter conversely to the other well working kernels (Cortes and Vapnik, 1995; Hsu *et al.*, 2009; Krahwinkler and Rossman, 2011).

On the other hand, ANN is machine learning systems comprising of inter-connected linkages of modest processing elements. The algorithm is characterized by robust pattern recognition power, allowing it to represent complex multi-variate data forms (Atkinson and Tatnall, 1997). ANN has many advantages over the statistical methods which include easy adaptation to different kinds of data and input structures, and the ability to generalize for technique with multiple images (Bishop, 1995; Paola and Schowengerdt, 1995; Mathers, 1999; Lu and Weng, 2007), as well as the ability to categorize data with limited training data compared with traditional classifiers (Mathers, 1999). Previous studies demonstrated that SVM and ANN perform better than the conventional classification methods like maximum likelihood, minimum distance to the mean (Dixon and Candade, 2008; Shafri *et al.*, 2009; Otukei and Blaschke, 2010; Xiong *et al.*, 2010; Nitze *et al.*, 2012). These conventional classification methods depend on assumptions that may limit their utilities for many datasets and for mapping areas with limited training samples (Kavzoglu and Mather, 2003; Dixon and Candade, 2008; Otukei and Blaschke, 2010). Moreover, ANN is often referred to as a black-box technique that could encounter an over-fitting problem on the test dataset (Kimes *et al.*, 2000; Qiu and Jensen, 2004). Perfectly describing processes that interpret input data into output classes could, however, be challenging due to the combined use of multiple nonlinear activation functions at different layers (Kavzoglu and Mather, 2003). To the best of the researcher's knowledge, no study utilized very high resolution WorldView-2 imagery and SVM and ANN classification algorithms to map endangered tree species in a fragmented ecosystem. Therefore, the objective of the present study was to examine the utility of the advanced WorldView-2 satellite data for mapping endangered tree species in the fragmented Dukuduku indigenous forest of South Africa using SVM and ANN classification algorithms. Specifically, additional WorldView-2 bands were tested for mapping six endangered tree species.

3.2 Methodology

3.2.1 Field Data Collection

Intensive field work was conducted to identify endangered tree species associated with indigenous forest in the study area within one week of the WorldView-2 imagery acquisition. The purposive sampling approach was employed to geo-locate six endangered tree species grown in fragmented and intact Dukuduku indigenous forest ecosystems in the area using a handheld Leica GS20 GPS with sub-meter accuracy. The six tree species were purposively selected for analysis in the study. These trees are regarded as rare and endangered species because of their rapid harvesting and removal (van Wyk *et al.*, 2006). The networks of road and open paths were used to assist in selecting the endangered tree species by walking in various directions in the study area. Figure 3.1 shows the morphological and spectral characteristics of the six tree species. The target species were identified with the aid of expert knowledge. In total, 827 sample points were collected from the six endangered tree species and land use/cover classes in the study area. The sample points for each class were 101 (*Albizia adianthifolia*), 71 (*Ekebergia capensis*), 62 (*Harpephyllum caffrum*), 70 (*Hymenocardia ulmoides*), 80 (*Sclercarya birrea*), 68 (*Trichilia dregeana*), 75 (sugarcane), 75 (grassland), 75 (plantation forest), and 150 (other forest species that include coastal and dune forest species) (Figure 3.2). These points were then used as ground-truth data to classify the different WorldView-2 spectral subsets based on the pixel spectral signatures of the classes. For the tree classes, one pixel was used per individual tree crown.

3.2.2 Image Acquisition and Pre-Processing

In this study, a cloud-free WorldView-2 satellite imagery was acquired from the study area on 1st December 2013. WorldView-2 image is the first very high resolution satellite imagery that has the ability to acquire data of eight spectral bands in the 400 –1040 nm spectral range with spatial resolution of two metres and swath width of 16.4 Kilometers. The eight spectral bands of WorldView-2 consist of four standard bands situated in the blue (450 – 510 nm), green (510 – 580 nm), red (630 – 690 nm), and NIR-1 (770 – 895 nm), and four additional bands which are coastal blue (400 – 450 nm), yellow (585 – 625 nm), red edge (705 – 745 nm) and NIR-2 (860 – 1040 nm). The WorldView-2 image was atmospherically corrected using the QUAC procedure in Interactive Data Language (IDL) ENVI 4.7 software (ENVI, 2009). QUAC determines atmospheric compensation parameters directly from the information contained within the image (pixel spectra), thus allowing for the retrieval of accurate reflectance spectra (Shen *et al.*, 2005). No geometric correction was made on the WorldView-2 image as it was provided already corrected by DigitalGlobe™. The image was referenced to the Universal Transverse Mercator (UTM zone 36 South) projection and WGS-84 Geodetic datum. The WorldView-2 image was spectral subsets to separate the four SB and four AB subsets. The separated WorldView-2 subsets together with the 8B were then compared for classifying endangered tree species using SVM and ANN supervised learning classification algorithms.

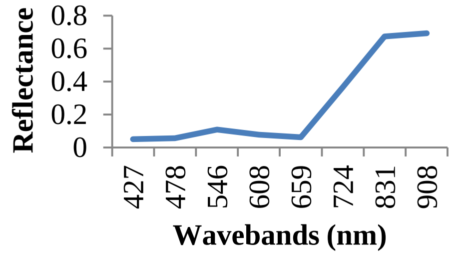
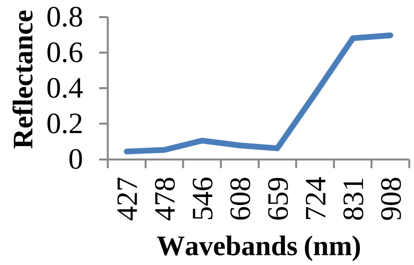
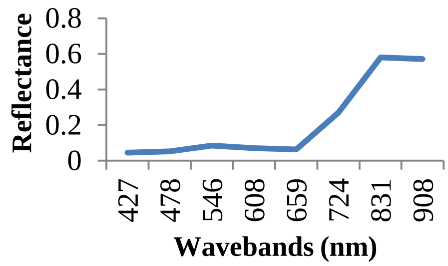
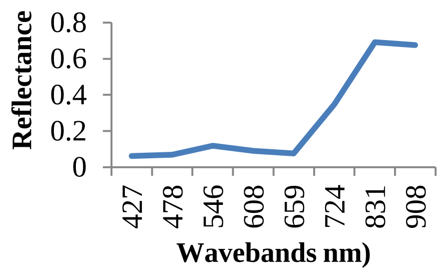
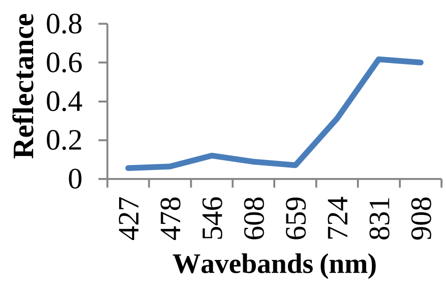
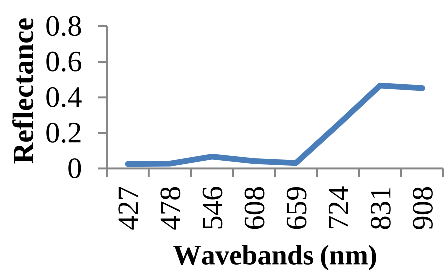
Species	Field Photograph	Spectral curve
Albizia adianthifolia		
Ekebergia capensis		
Harpephyllum caffrum		
Hymenocardia ulmoides		
Sclercarya birrea		
Trichilia dregeana		

Figure 3.1: Field photographs and the corresponding average spectral reflectance curves of the six tree species extracted from WorldView-2 image pixels (*n* = 44 for each spectrum) located at the centre of tree crowns.

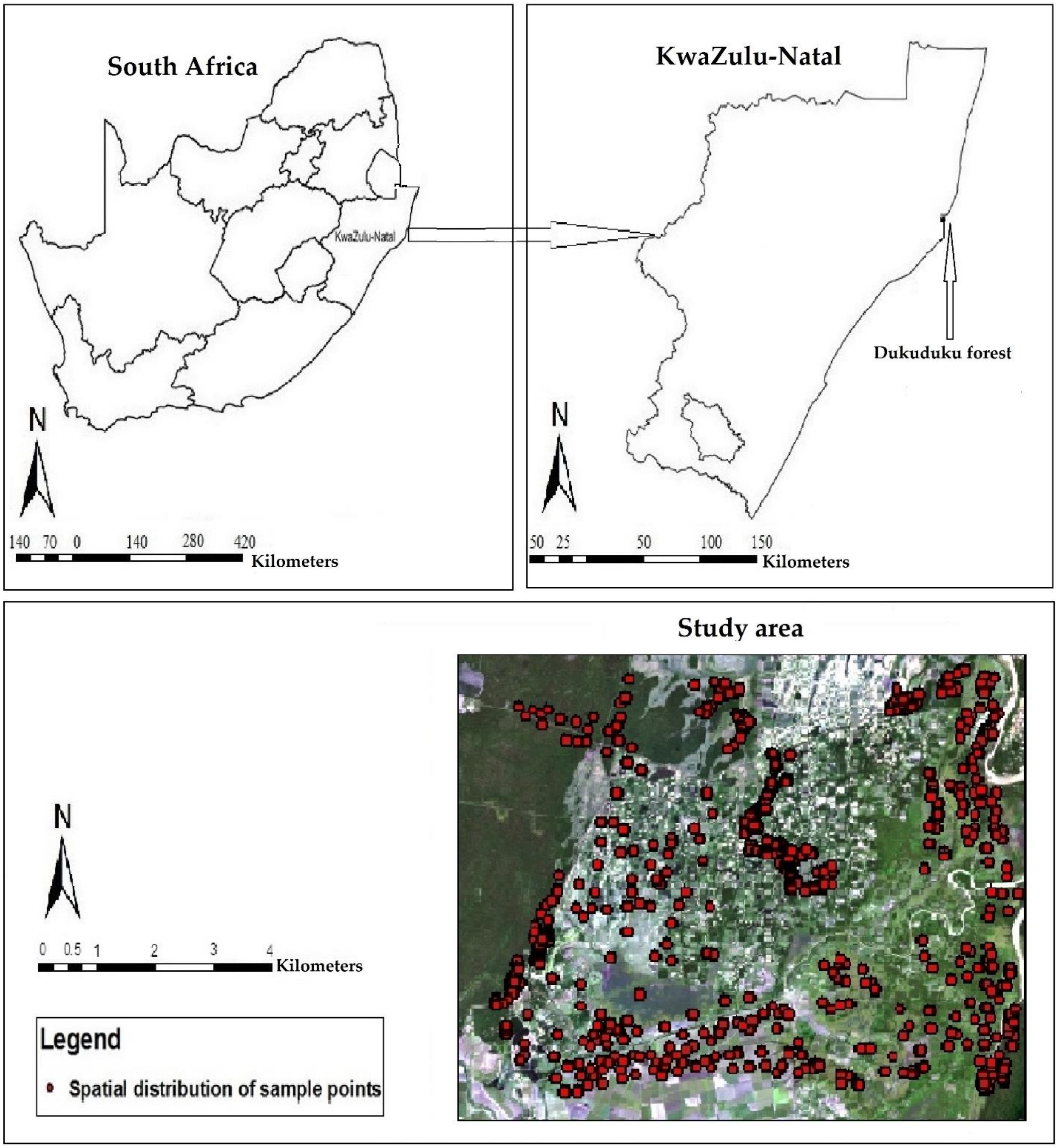


Figure 3.2: The location of the Dukuduku indigenous forest in KwaZulu-Natal province, South Africa and field sample locations overlaid in a true-color WorldView-2 image

3.3 Statistical Analysis

It is noted that a sufficient number of training samples are a pre-requisite for a successful classification (Lu and Weng, 2007). It is also reported that the sample points should be distributed consistently across the spatial extent of the classes to provide a representative description of the overall population (Foody and Mathur, 2004). In the current study, the SVM and ANN classification algorithms were trained on 70% (583) of a randomly selected holdout sample and final accuracy assessments were evaluated using the remaining 30% (244) of the dataset. The ground-truth training samples were overlaid over WorldView-2 spectral subsets, and classification signatures were then created for the six selected endangered tree species and land use/cover classes in the area. After assessing and adjusting signatures, SVM and ANN supervised classification methods were employed.

3.3.1 Support Vector Machines (SVM) Classification Algorithm

SVM classification algorithm is a binary learning technique that analyzes data and recognizes patterns (Cortes and Vapnik, 1995). However, numerous pattern recognition applications require multiple classes. Hence, multi-class SVM problems are solved by generating multiple binary classifiers (Yavuz and Cevikalp, 2008). Amendments are effected to the simple SVM binary classifiers to run as a multi-class classifier using procedures like OAO and OAA. In each class, OAA adopts one binary SVM to alienate a class member from others. Conversely, OAO uses a binary SVM from each pair of classes to separate members of one class from others. The SVM algorithm is then assigned the correct class by using a voting mechanism (Mathur and Foody, 2008; Krahwinkler and Rossman, 2011). The algorithm requires no assumption about the data distribution and uses very efficient principles in order not to overfit the test or new data sample (Cortes and Vapnik, 1995; Burges, 1998; Brown *et al.*, 1999; Everingham *et al.*, 2007). In addition, SVM attempts to maximize the margin, i.e. the distance between the data points of each class, to the optimal separating linear hyperplane axes created from each variable (Petropoulos *et al.*, 2011). In a two class experiment, the algorithm sets two supporting hyperplanes in the boundaries and searches to maximize the margin between them. Sample points lying on the supporting hyperplanes are named support vectors and in the middle of the margin is the optimal hyperplane. The algorithm aims to determine a linear discriminant function with maximum margin to discriminate each class.

However, many classes are not linearly separable, hence SVM projects vectors into a high-dimensional feature space by means of a kernel trick and fits the optimum hyperplane that discriminates classes using an optimization function (Hornik *et al.*, 2006; Yang, 2011).

Polynomial and radial basis function kernels are the most commonly used functions for classifying remotely-sensed data (Huang *et al.*, 2002; Dixon and Candade, 2008). In comparison, a number of researchers have found that a radial basis function outperforms polynomial kernel in classification of remotely-sensed data (Huang *et al.*, 2002; Hornik *et al.*, 2006; Yang, 2011). Furthermore, radial basis function is computationally fast and easy to implement and requires optimizing only two parameters. These are the *C* which is a value for standardizing the error of misclassified data points in the training dataset samples, and λ which is the kernel width parameter of the radial basis function (Hornik *et al.*, 2006).

In the current study, three WorldView-2 spectral subsets and radial basis function were used to find an optimal hyperplane that can differentiate amongst the six endangered tree species in the Dukuduku forest. The *C* and λ parameters of the radial basis function were optimized in order to avoid over-fitting problems (Hornik *et al.*, 2006). The regularization of the two parameters was performed using a 10-fold cross validation method (Hornik *et al.*, 2006; Hsu *et al.*, 2009; Yang, 2011). The training dataset was divided into 10 subsets of equal sizes. SVM models were then trained on nine subset samples, and tested using the removed one and the process was repeated ten times until all subset samples had served as test samples. The pair parameter that minimizes the classification error was then considered as the optimal value for the final classification process. The OAO procedure was used to implement a multiclass SVM model as suggested by Hsu and Lin (2002) who stated that this scheme is more symmetric than OAA with regard to class sizes. The e1071 library version 2.15.2 in R statistical packages (R Development Core, 2012) was employed for optimizing SVM parameters. The optimal SVM parameters were then input into the ENVI software to implement SVM classification algorithm in order to map the endangered tree species and other classes on WorldView-2 image.

3.3.2 Artificial Neural Networks (ANN) Classification Algorithm: Multilayer Perceptron

In machine learning and related fields, an ANN is a nonparametric classification technique that does not depend on an assumption of data normality (Atkinson and Tatnall, 1997; Foody, 2004b; Dixon and Candade, 2008). An ANN is a mathematical model that attempts to simulate the structure and functional aspects of biological neural linkages. It comprises of an interconnected group of artificial neurons and processes information using a connectionist approach for computation (Xiu and Liu, 2003). An ANN is originally designed as pattern recognition and data analysis tools that mimic the neural storage and analytical operations of the brain. Like SVM, an ANN approach has a distinct advantage over other classification methods. The algorithm is advantageous in that it is nonparametric and therefore requires little or no prior knowledge on the distribution of input data (Benediktsson and Sveinsson, 1997). Moreover, an ANN fits arbitrary decision-boundary to separate amongst the data points and therefore produces high classification accuracy (Bishop, 1995; Licciardi *et al.*, 2012).

Various models of ANN have been used in remote sensing studies such as radial basis function, back propagation and multilayer perceptron (Werbos, 1974; Atkinson and Tatnall, 1997; Tang *et al.*, 2003; Dixon and Candade, 2008; Lottering and Mutanga, 2012; Liu *et al.*, 2013). The multilayer perceptron is a commonly used ANN structure that comprises of an input layer and an output layer and one or more hidden layers of nonlinearly-activating nodes (Atkinson and Tatnall, 1997; Foody, 2004b; García Nieto *et al.*, 2012). The nodes in each layer connect with a certain synaptic weight to all the nodes in the next layer (García Nieto *et al.*, 2012). Perceptron learning occurs through changes in the linkages weights after items of data are processed. A multilayer perceptron is a feed forward ANN model that maps input data onto a set of appropriate output. It is an adjustment of the standard linear perceptron that uses three or more layers of neurons (nodes) with nonlinear activation functions (Kavzoglu and Mather, 2003; Foody, 2004b; Xiong *et al.*, 2010). ANN has been widely used in classifying land cover and tree species that are not linearly separable in the original spectral space (Zhang *et al.*, 1997; Xiu and Liu, 2003; Dixon and Candade, 2008; Liu *et al.*, 2013). Moreover, multilayer perceptron model using the standard back propagation algorithm is one of the well-known ANN structures of algorithm. This algorithm used standard back propagation for supervised learning from Tanagra software1.4 (Rakotomalala, 2005).

In the present study, ANN classification algorithm was used as a supervised nonlinear classification algorithm to map endangered tree species and other classes using WorldView-2 data. Numerous variations of internal networks structure, input data, and learning algorithms have been tested to define optimal classifier features. Table 3.1 shows a list of the parameters used to train all ANN models. The input layers consisted of eight for 8B and four for both SB and AB, whilst two hidden layers were found optimum for all models (Table 3.1). The structure of the hidden layers was also tested to assess the necessary number of hidden layers and number of required nodes per layer. This was tested by manually changing the number of nodes. The train-test parameters were excluded to obtain overall test error rate and then viewed to get confusion matrix for classifying endangered tree species and other classes using Tanagra software (Rakotomalala, 2005; Wahbeh *et al.*, 2011). The ANN was trained with a back propagation -training algorithm and one hidden layer (Ingram *et al.*, 2005). The back propagation algorithm is a supervised method that uses the gradient descent technique, which adjusts weights to minimize the classification error and optimizes ANN parameters (Richter *et al.*, 2011; Lottering and Mutanga, 2012). In the present study, the number of hidden layers was set to two and the number of training iterations was set to a default value of 1000 (Rakotomalala, 2005). The optimum number of nodes was established after manually changing the number of nodes (Table 3.1). The optimal parameters were then input into the ENVI software in order to map the endangered tree species and land use/cover on WorldView-2 image (ENVI, 2009).

Table 3.1: Parameters for the best trained and artificial neural networks (ANN) used for mapping endangered tree species

Model	**inputs**	**Hidden**	**Profile**
All WorldView-2 eight bands (8B) subset	8	2	MLP 8:8-2-1:1
Four WorldView-2 standard bands (SB) subset	4	2	MLP 4:4-2-1:1
Four WorldView-2 additional bands (AB) subset	4	2	MLP 4:4-2-1:1

MLP = multilayer perceptron

3.3.3 Accuracy Assessment

The accuracy of SVM and ANN classification algorithms were evaluated using the 30% (n = 244) holdout sample of the dataset. A confusion matrix was constructed to compare the true class with the class assigned by SVM and ANN and to calculate the overall accuracy, PA and UA (Congalton and Green, 2008; Xiong *et al.*, 2010). Overall accuracy is the overall probability that test sample points on the image have been classified properly. PA, which is expressed as a percentage (%), denotes the likelihood of a certain class being correctly classified, while UA refers to the probability that a sample is labeled as a specific class and the classifier accurately assigns it such a class. It has become customary in remote sensing studies to report the kappa index of agreement for accuracy assessment purposes since kappa also compares two maps that show a set of categories. However, recent studies have shown some limitations of kappa since it gives information that is redundant or misleading for practical decision making (Pontius and Millones, 2011). These researchers recommend using a more suitable and simpler approach that focuses on two parameters of disagreement between maps in terms of the quantity (Quantity disagreement; QD) and spatial allocation (Allocation disagreement; AD) of the categories instead of Kappa variants. The two useful parameters were calculated from the cross-tabulation matrices to assess reliability of each classification algorithm. The QD is the amount of difference between the number of test data points and the predicted ones, while the AD describes the number of expected classes that have less than optimal spatial location in comparison to the test data.

According to accuracy metrics achieved for each algorithm in each accuracy assessment method, a statistical analysis can be performed to test if there was any significant difference between the classification results of SVM and ANN classification algorithms. Therefore, the study performed McNemar's test to examine whether there were any significant differences amongst the confusion matrices of the two classification algorithms (SVM and ANN). McNemar's test is a nonparametric test based upon standardized normal test statistic calculated from error matrices of two algorithms given as follows (Foody, 2004a; Leeuw *et al.*, 2006):

$$Z = \frac{f12 - f21}{\sqrt{f12 + f21}} \qquad \text{Equation 3. 1}$$

where $f12$ denotes the number of samples that are misclassified on the first confusion matrix but correctly classified on the second confusion matrix. $f21$ denotes the number of samples that are misclassified on the second confusion matrix but correctly classified on the first confusion matrix. A difference in accuracy between the confusion matrices of two algorithms that used different WorldView-2 spectral subsets is statistically significant ($p \leq 0.05$) if a Z score is more than 1.96 (Foody, 2004a; Leeuw *et al.*, 2006).

3.4. Results

3.4.1 Accuracy Assessment

Table 3.2 shows the summary accuracy assessment results for the WorldView-2 eight bands (8B), WorldView-2 standard bands (SB) and WorldView-2 additional bands (AB) based on the SVM classifier. The overall classification accuracy for the 8B, SB and AB classifications were respectively, 77%, 65% and 70%. In terms of quantity disagreement, 8B and SB classifications (11%) were slightly lower than AB which had 16%. However, the allocation disagreement for SB significantly high at 24%, while it was relatively low for 8B (12%) and AB (14%). The individual class accuracies were generally high for the 8B, while individual class accuracies were low for SB and AB. The individual UA and PA for each endangered tree species achieved by the two classification algorithms are shown in Figures 3.3 and 3.4. When WorldView-2 8B subset was classified using SVM, the UA for some endangered tree species (*Albizia adianthifolia*, *Hymenocardia ulmoides*, *Harpephyllum caffrum* and *Sclercarya birrea*) was relatively higher and ranged between 73.17% and 85.71% while for *Trichilia dregeana* and *Ekebergia capensis* the accuracy was less than 55% (Figure 3.3a).

The accuracy assessment results for the WorldView-2 8B, WorldView-2 SB and WorldView-2 AB based on the ANN classifier were presented in Table 3.3. A higher overall classification accuracy was achieved using WorldView-2 8B (75%) followed by WorldView-2 AB (70%) and WorldView-2 SB (64%). The ANN classifier obtained high quantity disagreement values for AB (11%) followed by 8B and SB that have achieved the same values of quantity disagreement (9%), but these values are less than those obtained using SVM (Table 3.3). Furthermore, the Tables also show relatively high allocation disagreement values for SB followed by AB and 8B that obtained the lowest values when both SVM and ANN algorithms were used (Tables 3.2 and 3.3). Additionally, the use of WorldView-2 8B subset and ANN resulted in relatively higher UA for *Harpephyllum caffrum* (73.68%) and *Sclercarya birrea* (81.81%) whereas all other four endangered tree species (*Albizia adianthifolia*, *Hymenocardia ulmoides*, *Ekebergia capensis* and *Trichilia dregeana*) achieved UA of less than 65% (Figure 3.3b). Regarding the PA, the two classification algorithms (SVM and ANN) mapped *Albizia adianthifolia* with an accuracy ranging between 80% and 100% (Figure 3.4). Overall, all endangered tree species were classified with fairly higher PA, except for *Hymenocardia ulmoides* (Figures 3.3 and 3.4). Additionally, all land use/cover obtained individual accuracies of more than 50% when SVM, ANN and all WorldView-2 subsets were employed (Figures 3.3 and 3.4), except grassland class (45.45%) when SVM and WorldView-2 SB were used (Figure 3.4a).

Table 3.2: Confusion matrix of support vector machines (SVM) classification algorithm using (a) WorldView-2 eight bands, (b) WorldView-2 standard bands and (c) WorldView-2 additional bands for the 30% test datasets. The confusion matrix includes overall accuracy (OA), quantity disagreements (QD) and allocation disagreements (AD) for six endangered tree species and land use/cover classes

Field data	**AA**	**EC**	**HC**	**HU**	**ScB**	**TD**	**GL**	**PF**	**SC**	**OFS**	**Total**
a) WorldView-2 eight bands											
AA	**30**	1	0	8	1	0	0	1	0	0	41
EC	0	**16**	1	0	2	8	1	1	1	0	30
HC	0	0	**16**	1	0	0	2	0	0	0	19
HU	0	0	0	**10**	0	0	0	0	2	1	13
ScB	0	1	1	0	**18**	0	0	1	0	0	21
TD	0	3	0	2	3	**12**	1	0	2	0	23
GL	0	0	0	0	0	0	**16**	3	1	0	20
PF	0	0	0	0	0	0	0	**15**	0	0	15
SC	0	0	0	0	0	0	1	0	**14**	1	16
OFS	0	0	0	0	0	0	1	1	2	**42**	46
Total	30	21	18	21	24	20	22	22	22	44	**244**
OA (%)	77										
QD (%)	11										
AD (%)	12										
b) WorldView-2 standard bands											
AA	**25**	5	5	5	6	4	0	1	2	0	53
EC	1	**11**	2	0	1	2	1	2	1	0	21
HC	2	0	**11**	2	0	0	3	1	0	0	19
HU	1	1	0	**10**	1	2	1	2	0	0	18
ScB	1	2	0	0	**14**	0	0	1	0	0	18
TD	0	2	0	4	2	**12**	1	0	1	1	23
GL	0	0	0	0	0	0	**10**	0	2	0	12
PF	0	0	0	0	0	0	4	**13**	2	1	20
SC	0	0	0	0	0	0	2	0	**13**	2	17
OFS	0	0	0	0	0	0	0	2	1	**40**	43
Total	30	21	18	21	24	20	22	22	22	44	**244**
OA (%)	65										
QD (%)	11										
AD (%)	24										

Table 3.2 (Continued)

c) WorldView-2 additional bands											
AA	**30**	8	5	10	7	8	0	0	0	0	68
EC	0	**11**	4	0	1	0	1	1	1	0	19
HC	0	0	**9**	1	0	0	1	0	0	0	11
HU	0	0	0	**10**	0	0	0	0	2	1	13
ScB	0	0	0	0	**14**	0	0	2	0	1	17
TD	0	2	0	0	2	**12**	1	0	1	0	18
GL	0	0	0	0	0	0	**14**	0	1	0	15
PF	0	0	0	0	0	0	0	**17**	0	3	20
SC	0	0	0	0	0	0	2	0	**16**	1	19
OFS	0	0	0	0	0	0	1	2	1	**38**	42
Total	30	21	18	21	24	20	20	22	22	44	**244**
OA (%)	70										
QD (%)	16										
AD (%)	14										

AA = *Albizia adianthifolia,* EC = *Ekebergia capensis,* HC = *Harpephyllum caffrum,* HU = *Hymenocardia ulmoides,* ScB = *Sclercarya birrea,* TD = *Trichilia dregeana*, GL = grassland, SC = sugarcane, PF = plantation forests and OFS = other forest species

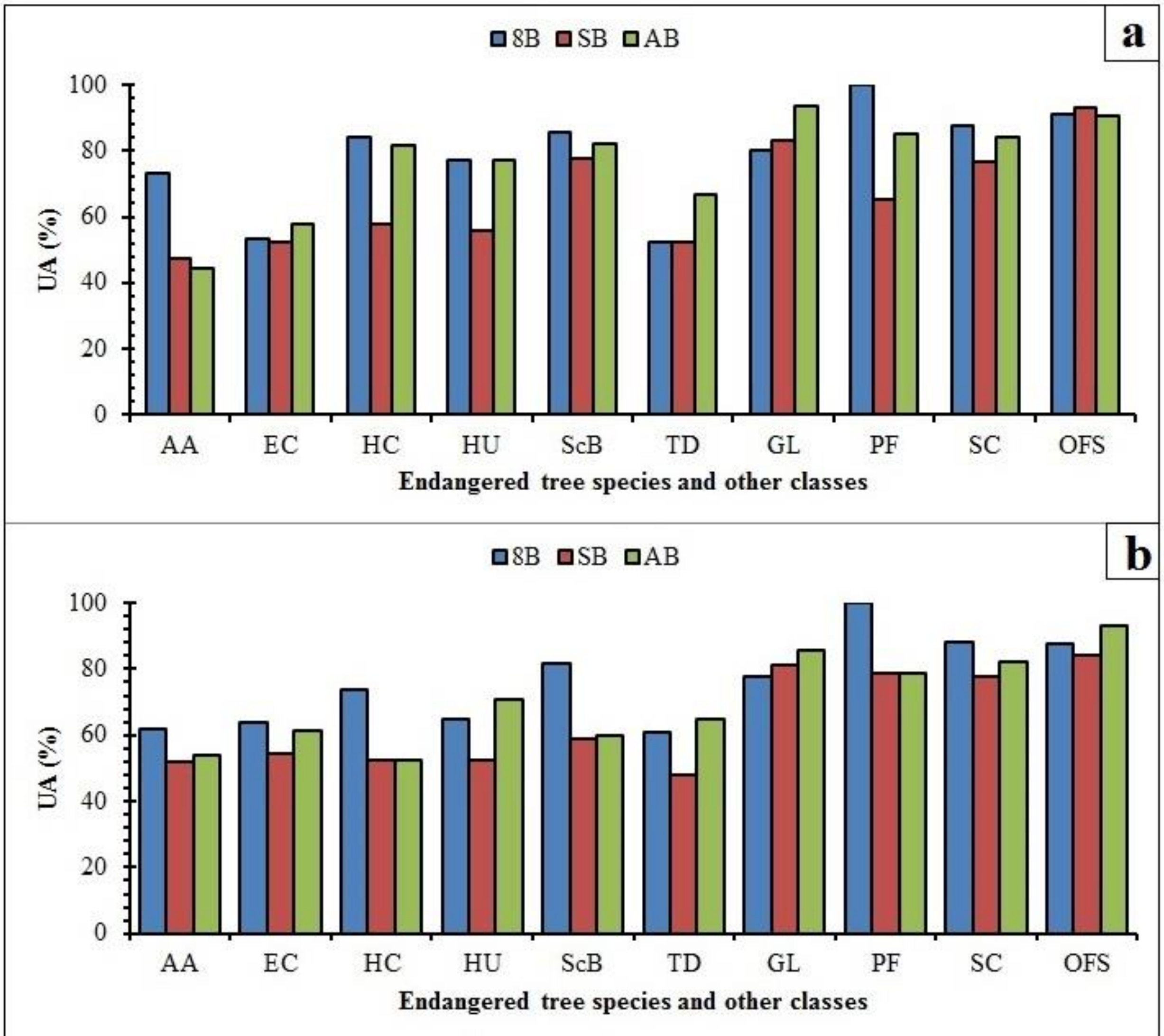
Table 3.3: Confusion matrix of artificial neural networks (ANN) classification algorithm using (a) WorldView-2 eight bands, (b) WorldView-2 standard bands and (c) WorldView-2 additional bands for the 30% test datasets. The confusion matrix includes overall accuracy (OA), quantity disagreements (QD) and allocation disagreements (AD) for six endangered tree species and land use/cover classes

Field data	**AA**	**EC**	**HC**	**HU**	**ScB**	**TD**	**GL**	**PF**	**SC**	**OFS**	**Total**
a) WorldView-2 eight bands											
AA	**29**	1	2	7	2	0	2	1	0	3	47
EC	0	**14**	1	0	1	3	1	1	1	0	22
HC	0	1	**14**	1	1	1	1	0	0	0	19
HU	1	0	0	**11**	1	1	2	0	0	1	17
ScB	0	2	0	0	**18**	1	0	1	0	0	22
TD	0	3	1	2	1	**14**	1	0	1	0	23
GL	0	0	0	0	0	0	**14**	1	1	2	18
PF	0	0	0	0	0	0	0	**18**	0	0	18
SC	0	0	0	0	0	0	0	0	**15**	2	17
OFS	0	0	0	0	0	0	1	0	4	**36**	41
Total	30	21	18	21	24	20	22	22	22	44	**244**
OA (%)	75										

Table 3.3 (Continued)

Field data	**AA**	**EC**	**HC**	**HU**	**ScB**	**TD**	**GL**	**PF**	**SC**	**OFS**	**Total**
a) WorldView-2 eight bands											
QD (%)	09										
AD (%)	16										
b) WorldView-2 standard bands											
AA	**24**	2	1	7	4	1	0	0	2	5	46
EC	1	**12**	2	0	1	2	1	2	1	0	22
HC	1	1	**11**	1	3	1	1	1	0	1	21
HU	2	2	0	**10**	2	2	0	0	0	1	19
ScB	2	1	1	1	**13**	3	0	1	0	0	22
TD	0	3	3	2	1	**11**	1	0	1	1	23
GL	0	0	0	0	0	0	**13**	2	1	0	16
PF	0	0	0	0	0	0	0	**15**	2	2	19
SC	0	0	0	0	0	0	2	0	**14**	2	18
OFS	0	0	0	0	0	0	4	1	1	**32**	38
Total	30	21	18	21	24	20	22	22	22	44	**244**
OA (%)	64										
QD (%)	09										
AD (%)	27										
c) WorldView-2 additional bands											
AA	**28**	6	3	7	6	2	0	0	0	0	52
EC	1	**11**	1	0	1	1	1	1	1	0	18
HC	0	1	**11**	1	2	2	1	2	1	0	21
HU	0	0	0	**12**	1	1	0	0	3	0	17
ScB	1	2	2	0	**12**	1	0	2	0	0	20
TD	0	1	1	1	2	**13**	1	0	1	0	20
GL	0	0	0	0	0	0	**12**	1	1	0	14
PF	0	0	0	0	0	0	3	**15**	0	1	19
SC	0	0	0	0	0	0	2	0	**14**	1	17
OFS	0	0	0	0	0	0	1	1	1	**42**	45
Total	30	21	18	21	24	20	21	22	22	44	**244**
OA (%)	70										
QD (%)	11										
AD (%)	19										

AA = *Albizia adianthifolia*, EC = *Ekebergia capensis*, HC = *Harpephyllum caffrum*, HU = *Hymenocardia ulmoides*, ScB = *Sclercarya birrea*, TD = *Trichilia dregeana*, GL = grassland, SC = sugarcane, PF = plantation forests and OFS = other forest species



Albizia adianthifolia = AA, *Ekebergia capensis* = EC, *Harpephyllum caffrum* = HC, *Hymenocardia ulmoides* = HU, *Sclercarya birrea* = ScB and *Trichilia dregeana* = TD, GL = grassland, SC = sugarcane, PF = plantation forest and OFS = other forest species

Figure 3.3: User's accuracy (%) of the studied six endangered tree species achieved by support vector machines (a) and artificial neural networks (b) classification algorithms when all WorldView-2 eight bands (8B), standard bands (SB) and additional bands (AB) were used

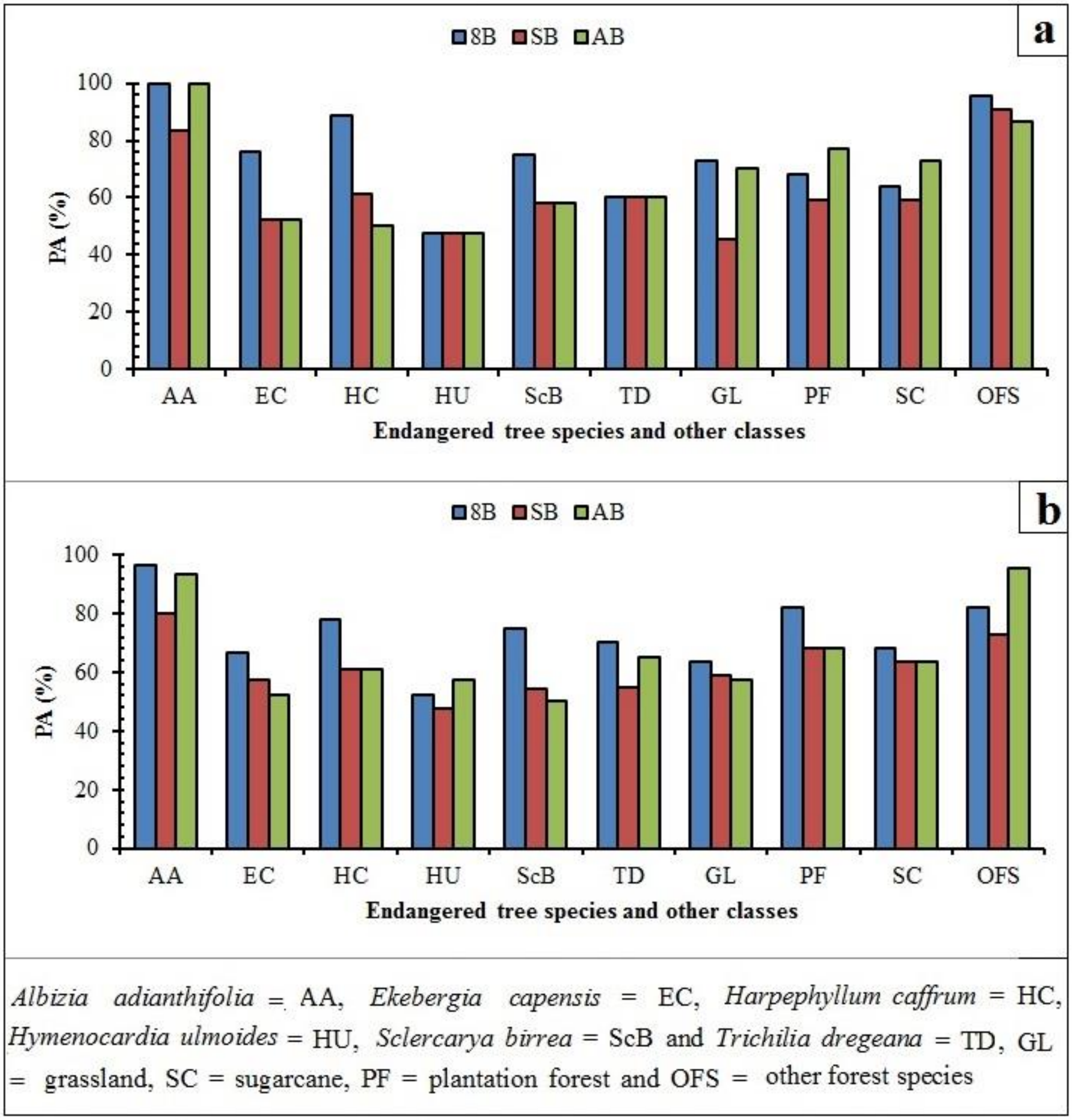


Figure 3.4: Producer's accuracy (%) of the studied six endangered tree species achieved by support vector machines (a) and artificial neural networks (b) classification algorithms when all WorldView-2 eight bands subset (8B), standard bands (SB) and additional (AB) were used

3.4.2 Classification Results

Results of the grid search and 10-fold cross validation method indicated optimal values of λ and *C* for SVM, respectively, of 1 and 10 for WorldView-2 8B, 1 and 1000 for WorldView-2 SB, and 1 and 100 for WorldView-2 AB. When these optimal values were input into SVM algorithm, minimum overall classification errors of 38.40%, 39.40%, and 36.90% for 8B, SB and AB, respectively were obtained. Figures 3.5, 3.6 and 3.7 show the classification maps (land use/cover and endangered tree species) of the study area. The maps show nearly similar spatial distribution of endangered tree species in the study area. Figures 3.5, 3.6 and 3.7 also show the spatial distribution of the six endangered tree species and other classes in the study area when WorldView-2 8B, SB and AB were classified using SVM and ANN. The main visual difference between the maps is that a relatively homogeneous map was produced when the WorldView-2 8B was used as compared with other WorldView-2 spectral subsets (Figures 3.5, 3.6 and 3.7). The maps also show that the study area was mainly surrounded by grassland, sugarcane and plantation forest, while the forest and grassland on the north eastern part of the study area were relatively patchy.

The results in Table 3.4 show areas under each endangered tree species and other classes obtained from different WorldView-2 spectral subsets using SVM and ANN classification algorithms. The Table demonstrates that the plantation class have the largest area (4063.5ha) when WorldView-2 8B subset and SVM classifier were used, while grassland occupied most of the study area (2472.1ha) when SB subset and SVM classifier were deployed. With regard to endangered tree species, results show that the *Albizia adianthifolia* have the largest area when WorldView-2 AB subset and SVM (1316.3ha) as well as ANN (1315.1ha) classifiers were used, while the *Harpephyllum caffrum* achieved the smallest area (478.1ha) when WorldView-2 SB subset and SVM were employed (Table 3.4).

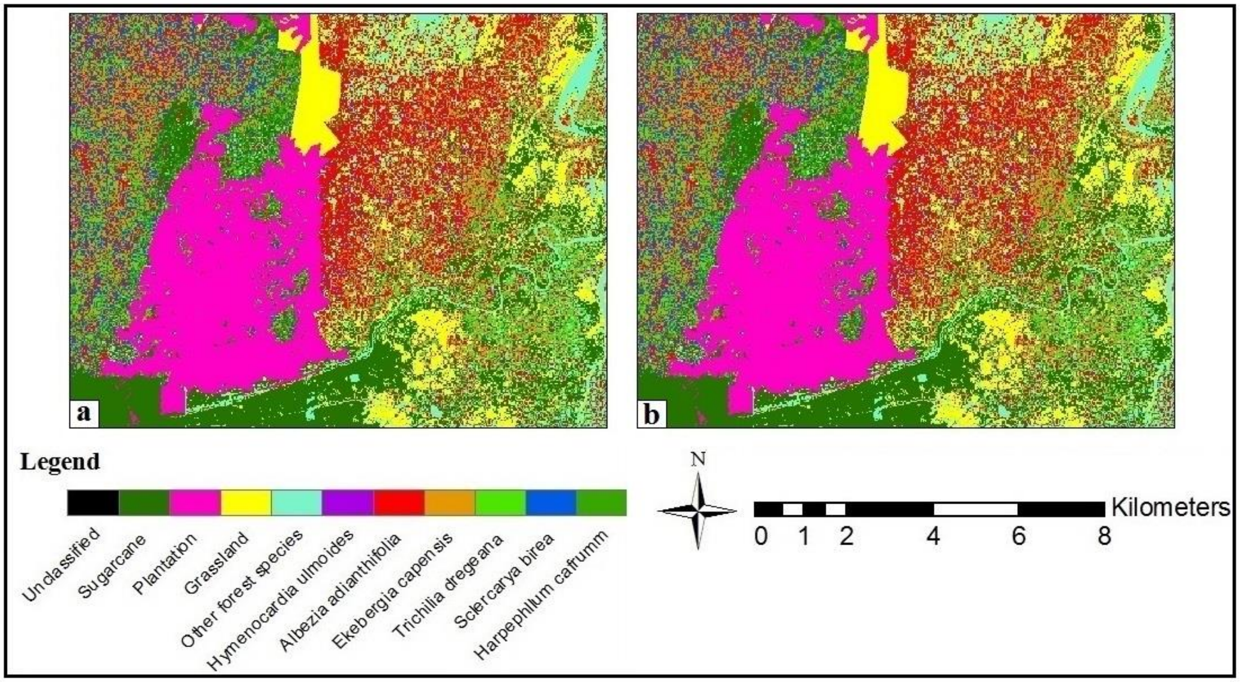


Figure 3.5: Classification maps obtained using all eight WorldView-2 bands (8B): (a) support vector machines algorithm and (b) artificial neural networks algorithm

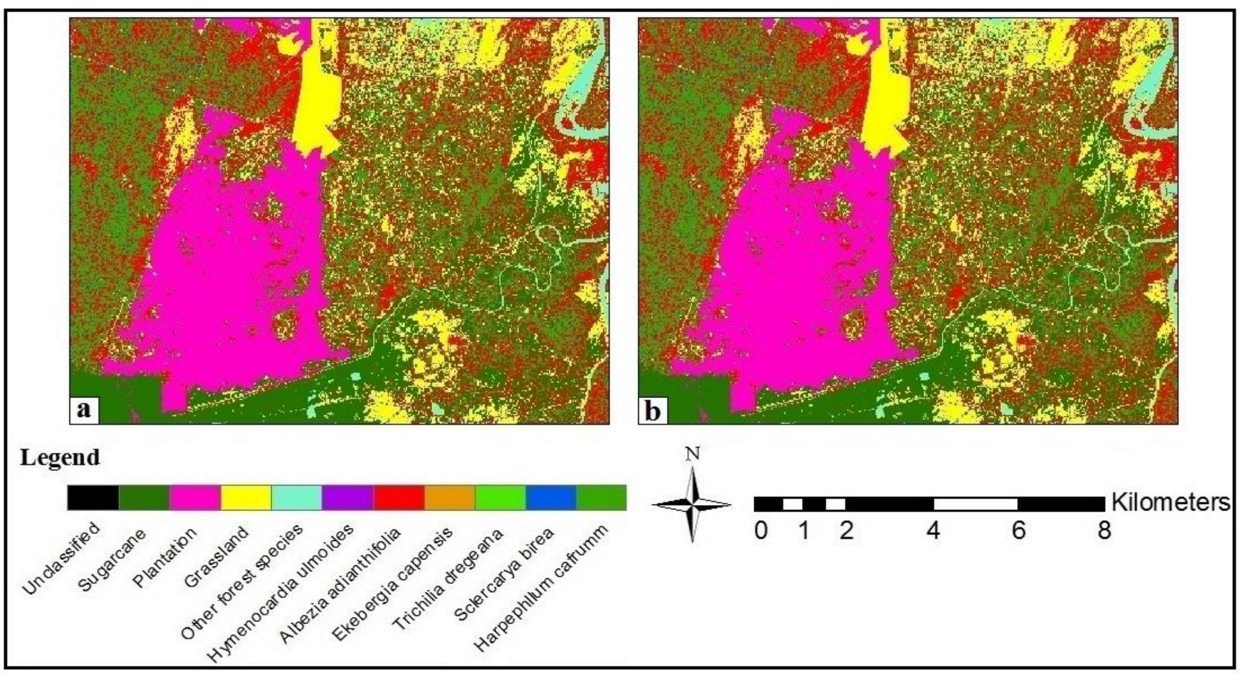


Figure 3.6: Classification maps obtained using four standard WorldView-2 bands (SB): (a) support vector machines algorithm and (b) artificial neural networks algorithm

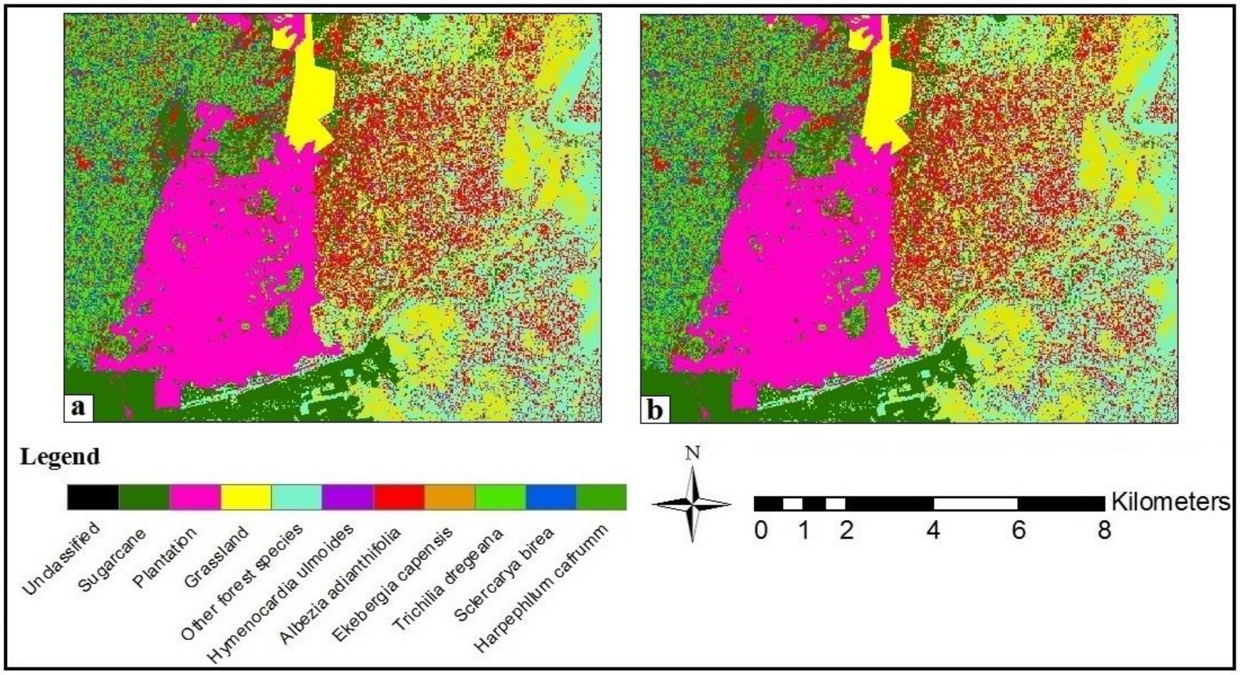


Figure 3.7: Classification maps obtained using four additional WorldView-2 bands (AB): (a) support vector machines algorithm and (b) artificial neural networks algorithm

Table 3.4: Area under each endangered tree species and land use/cover classes in the study area obtained using WorldView-2 data, support vector machines (SVM) and artificial neural networks (ANN) classification algorithms

Class	**Area (ha)**					
	Support vector machines			**Artificial neural networks**		
	8B	**SB**	**AB**	**8B**	**SB**	**AB**
Albizia adianthifolia	1308.8	1311.9	1316.3	1312.4	1311.2	1315.1
Ekebergia capensis	632.0	634.0	632.1	633.2	634.0	628.2
Harpephllum cafrum	486.1	478.1	486.9	488.8	489.4	487.0
Hymenocardia ulmoides	869.5	863.4	864.3	868.2	865.0	864.8
Sclercarya birrea	1109.3	1112.2	1111.1	1110.1	1112.2	1110.6
Trichilia dregeana	929.7	932.5	932.0	930.1	930.5	931.3
Other forest species	1058.0	1059.9	1061.0	1059.0	1058.2	1064.0
Sugarcane	781.2	783.2	783.1	779.2	784.9	783.5
Plantation	4063.5	4051.7	4051.4	4057.0	4056.3	4051.2
Grassland	2461.0	2472.1	2460.7	2461.1	2457.3	2463.1

8B = WorldView-2 eight bands subset, SB = WorldView-2 standard bands subset and AB = WorldView-2 additional bands subset

According to McNemar's test, there were significant differences (Z ≥ 1.96) at 95% confidence level amongst the confusion matrices of SVM and ANN classification algorithms using WorldView-2 8B, SB, and AB spectral subsets (Table 3.5). Regardless of the WorldView-2 subset used, SVM significantly outperformed ANN for mapping the endangered tree species and land use/cover classes in the study area.

Table 3.5: McNemar's test result for comparing classification confusion matrices obtained from support vector machines (SVM) and artificial neural networks (ANN) algorithms using WorldView-2 8B, SB, and AB

All WorldView-2 eight bands (8B) subset				
		ANN		
		Correctly Classified	Misclassified	Total
	Correctly Classified	97	03	100
SVM	Misclassified	02	32	34
	Total	99	35	134
	Z score	5.15		
Four WorldView-2 standard bands (SB) subset				
		ANN		
		Correctly Classified	Misclassified	Total
	Correctly Classified	80	03	83
SVM	Misclassified	00	51	51
	Total	80	54	134
	Z score	7.14		
Four WorldView-2 additional bands subset (AB)				
		ANN		
		Correctly Classified	Misclassified	Total
	Correctly Classified	82	4	86
SVM	Misclassified	5	43	48
	Total	87	47	134
	Z score	5.48		

SVM = support vector machines classification algorithm and ANN = artificial neural networks classification algorithm

3.5 Discussion and Conclusions

This study examined the performance of SVM and ANN algorithms and the potential of new generation multispectral WorldView-2 data to map the spatial extent of endangered tree species in the Dukuduku indigenous forest of South Africa. Other land use/cover types in the study area were also classified in order to map target tree species within different land use/cover patterns. Tree species maps obtained from commonly used medium-spatial-resolution multispectral satellite (30m to 100m) have often been less accurate for operational application (Foody, 2002; Lu and Weng, 2007). Conversely, the use of imagery from very high spatial resolution sensors (<5 m) has its own limitations in terms of cost, availability, processing, and high dimensionality. Limitations that characterize fine and medium scale remotely-sensed data prevent the application and combination of field data with remotely-sensed data (Lu and Weng, 2007). Against this background, this study highlights the utility of multispectral WorldView-2 imagery for mapping endangered tree species in a fragmented Dukuduku forest. The study demonstrates that WorldView-2 data are effective for classifying endangered tree species using SVM and ANN classification algorithms. The spatial location of six selected endangered tree species has been mapped at 2m spatial resolution. Furthermore, the study indicates that the eight bands of WorldView-2 are found to be very suitable for endangered tree species identification and mapping. The results indicate the significant improvement in the overall accuracy of the classification results that can be largely attributed to the eight bands of WorldView-2 (Tables 3.2 and 3.3). This result is in conformity with Cho *et al.* (2013) and Omer *et al.* (2014) who concluded that using WorldView-2 data with advanced classification methods for mapping land use/cover in a fragmented ecosystem could lead to improved classification accuracy. There are two reasons that may have led to the delineation of endangered tree species in Dukuduku area with relatively high classification accuracy. Firstly, with regard to the machine learning classification algorithms SVM is a robust and versatile method that produces accurate classification results when it is employed for mapping vegetation cover and tree species using remotely-sensed data without having to rely on any statistical assumptions (Dixon and Candade, 2008; Xiong *et al.*, 2010; Immitzer *et al.*, 2012; Ghosh and Joshi, 2014). Secondly, ANN is also a robust and reliable algorithm that yields accurate classification results (Table 3.4). The accuracy increased when the hidden layers were increased from 1 to 2 (Huang and Huang, 1991; Duhoux *et al.*, 2001; Nguyen and Chan, 2004; Lu and Weng, 2007). Moreover, tree morphological characteristics like canopy geometry and structural features as well as leaf color of the six endangered tree species are seemingly different (Figure 3.1). That could have resulted in very different spectral features captured by WorldView-2 data and enabled accurate species classification maps.

The comparison between SVM and ANN were employed to investigate their ability for tree species mapping. Both SVM and ANN algorithms achieved comparable overall accuracies. SVM produced higher classification accuracy than ANN by about 2% when WorldView-2 8B was used (Tables 3.2 and 3.3). However, the relatively low overall classification accuracies (ranged between 64% and 65%) obtained using WorldView-2 SB questioned the classification protocol employed and the spectral variability captured by a fewer number of WorldView-2 bands. Combining other forest species in one class could have confused the delineation of six select endangered tree species when only four WorldView-2 bands (SB or AB) were analyzed. Nonetheless, that level of accuracy is quite common in the remote sensing image classification studies that looked at mapping vegetation at tree species level (Omar, 2010; Immitzer *et al.*, 2012; Engler *et al.*, 2013). More ground -truth samples could be required in order to improve the accuracy in such a case (Omar, 2010; Omer *et al.*, 2014).

It is important to note, however, that SVM and ANN were unable to fully deal with the high spectral variation inherent in some tree species (e.g. *Hymenocardia ulmoides*). This is a common problem when mapping heterogeneous landscapes using high spatial resolution images based on per-pixel classification techniques (Lu and Weng, 2007). In this regard, an object-based classification approach would produce higher classification accuracy (e.g. Immitzer *et al.* (2012)). Although the six endangered tree species could be separated accurately using only the four standard bands, the use of the WorldView-2 additional bands led to a considerable improvement in the classification accuracy. That is expected when the advanced imaging systems such as WorldView-2 data with additional bands are used. For instance, the overall accuracy increased from 65% and 64% to 77% and 75% for SVM and ANN, respectively, when new bands were added. The new WorldView-2 bands are useful for differentiating vegetated surfaces and are valuable in vegetation identification (Kimes *et al.*, 1998; Dlamini, 2010; Omar, 2010; Immitzer *et al.*, 2012; Ghosh and Joshi, 2014). Zhou *et al.* (2012) mentioned that the WorldView-2 additional wavebands are expected to provide an increase of up to 30% in classification accuracy. The new bands are also strongly related to vegetation and tree species characteristics. For example, the yellow band is intended for the detection of 'yellowness' in vegetation (Immitzer and Atzberger, 2014) like senescent tree crowns (see *Hymenocardia ulmoides* crown in Figure 3.1). In remote sensing, the red edge is the channel of abrupt change in the leaf reflectance between 680 and 780 nm, due to the combined effects of strong chlorophyll absorption in red bands and high reflectance in the NIR bands (Horler *et al.*, 1983; DigitalGlobe, 2010). The reflectance increases beginning at about 685 nm and an asymptotic reflectance reached at wavelengths beyond 760 nm for each endangered tree species (Figure 3.1). Many spectral features of vegetation are found within the red edge position that is related with changes in chlorophyll content (Filella and Penuelas, 1994; Lichtenthaler *et al.*, 1996; Marchisio *et al.*, 2010). The NIR-2 band that partly overlaps the standard NIR-1 band but is less affected by atmospheric influence is expected to enable a relatively accurate endangered tree species classification.

The low UA for the *Albizia adianthifolia* and PA for *Hymenocardia ulmoides* indicate that there is a probability that pixels classified as *Albizia adianthifolia* and *Hymenocardia ulmoides* may not actually exist on the ground. That could also be due to spectral overlaps between *Albizia adianthifolia* and *Hymenocardia ulmoides* and other forest species. The relatively high AD shown in Tables 3.2 and 3.3 of the confusion matrix was expected since pixels covered by multi-classes could probably be mismatched in terms of spatial pattern between test ground-truth instances and predicated test samples. However, the study employed an independent holdout test sample (30%) for assessing the performance of the classification models (SVM and ANN) based on the recommendation by Adelabu *et al.* (2015) while Atzberger *et al.* (2015) noted that bootstrapping approach produced relatively representative and repeatable accuracy measures when the performance of predictive models is evaluated.

The study showed that SVM outperformed ANN in distinguishing amongst endangered tree species in a fragmented landscape. SVM offers more benefits as compared to other classification models such as ANN. SVM is paired with the kernel trick, exploring and finding tuning kernels that create appropriate feature spaces where the linear classification is able to classify data created by nonlinear phenomena (Mountrakis *et al.*, 2011; Richter *et al.*, 2011; Rullan-Silva *et al.*, 2013). On the other hand, ANN could encounter an over-fitting problem and result in low classification accuracy on the test dataset as shown in Tables 3.2 and 3.3 (Kimes *et al.*, 2000). However, researchers have shown that ANN compares favorably with the established supervised machine learning classification algorithms like SVM for tree species mapping (Combal *et al.*, 2003; Ingram *et al.*, 2005; Dorigo *et al.*, 2007; Xiong *et al.*, 2010; Ghosh and Joshi, 2014). Since the purposively subset WorldView-2 bands was used to test the utility of the SB and AB for mapping endangered tree species, the study did not use any feature selection method like RF to select a fewer number of bands that might classify endangered tree species with a comparable accuracy to the one produced by the 8B. Although, studies have demonstrated the robustness of RF as variable selection and classification approach in tree species mapping (Omar, 2010; Immitzer *et al.*, 2012; Adelabu *et al.*, 2013) a combination of RF as a feature selection method and SVM as a classification approach would have yielded optimum endangered tree species mapping results.

In summary, the study findings are promising for accurate classification of endangered tree species in a fragmented ecosystem using multispectral WorldView-2 bands, SVM and ANN classification algorithms. Moreover, the relatively accurate classification results achieved with SVM and WorldView-2 8B subset in this study provide reliable information on tree species in the Dukuduku area that could be used in the design of management plans and policies as a basis for assessing and monitoring natural resources, ecological fragmentation and ecosystem functions and services.

3.6 Acknowledgments

The author is grateful to the University of KwaZulu-Natal, South Africa and the University of Khartoum, Sudan for funding this study. The author would also like to acknowledge the Inkanyamba Development Trust and Manukelana Arts and Indigenous Nursery in KwaZulu-Natal for facilitating the field data collection. My appreciation extends to the R development core team and Tanagra team for their open source packages for the statistical analysis. Special thanks to Dr. Riyad Ismail for his helpful assistance during data analysis.

CHAPTER FOUR

Empirical Estimation of Leaf Area Index (LAI) of Endangered Tree Species in Intact and Fragmented Indigenous Forest Ecosystems Using WorldView-2 Data and Two Robust Machine Learning Algorithms

This chapter is based on:

G. **Omer**, O. Mutanga, E.M. Abdel-Rahman and E. Adam, "Empirical prediction of leaf area index (LAI) of endangered tree species in intact and fragmented indigenous forests ecosystems using WorldView-2 data and two robust machine learning algorithms" *Remote Sensing*, 8 (4), 1-26, 2016.

ABSTRACT

Leaf area index (LAI) is an important biophysical input variable for forest ecosystems and ecological modeling, as it plays a key role for the forest productivity and structural characteristics. The ground-based methods like the handheld optical instruments used to estimate LAI are expensive and time-consuming. The advent of very high spatial resolution multispectral data and robust machine learning regression algorithms like support vector machines (SVM) and artificial neural networks (ANN) has provided an opportunity to estimate LAI at tree species level. The objective of the present study was, therefore, to test the utility of spectral vegetation indices (SVIs) calculated from the multispectral WorldView-2 data for predicting LAI at tree species level using the SVM and ANN machine learning regression algorithms. The study further tested whether there were significant differences between intact and fragmented (open) indigenous forests LAI at tree species level. The study shows that LAI of six endangered tree species could be accurately estimated using the fragmented forest data compared with the intact forest data. Specifically, the study shows that relatively accurate LAI predictions were achieved for *Hymenocardia ulmoides* using the fragmented stratum data and SVM regression model based on a validation data set (R^2_{Val} = 0.75, $RMSE_{Val}$ = 0.05 (1.37% of the mean)). The study further shows that the SVM regression approach achieved more accurate models for estimating the LAI of the six endangered tree species compared with the ANN regression method. It is concluded that the successful application of the WorldView-2 data, SVM and ANN methods for predicting LAI of six endangered tree species in the Dukuduku indigenous forest could help in making informed decisions and policies regarding management, protection and conservation of these endangered tree species.

Keywords: Leaf area index, tree species, indigenous forest, WorldView-2, support vector machines, artificial neural networks

4.1 Introduction

Indigenous forests in South Africa cover about 0.2% of the country's land surface (Ndlovu, 2013). In KwaZulu-Natal province, coastal lowland indigenous forests occur in small, fragmented and largely scattered patches in relatively dry landscapes (Cho *et al.*, 2012; Ndlovu, 2013). An example of such fragmented forests in KwaZulu Natal, South Africa is the Dukuduku forest. Dukuduku indigenous forest is one of the largest and the best preserved remnants of South African coastal forests (Ntombela, 2003). However, forest in the Dukuduku area is highly threatened by the rapid growth of informal human settlements and agricultural systems (Ndlovu, 2013). As one of the key features of landscape in South Africa, careful indigenous forest monitoring, management and conservation is critical. Therefore, a multi-disciplinary research is needed to optimally manage and conserve the endangered tree species in these indigenous forests as they support livelihoods of millions of people in the poor rural communities. Indigenous forests provide a number of benefits like food resources, fence poles, and traditional medicine to treat human diseases in these poor rural communities (Eldeen, 2005; Brendler *et al.*, 2010).

One way of managing and monitoring indigenous forest ecosystems is to estimate the trees' structural (e.g. height), biophysical (e.g. LAI) and biochemical (e.g. foliar N concentrations) traits. In general, these tree characteristics are measures and proxies for ecosystem resilience, services, conservation, landscape integrity and environmental health. The tree structural and bio-physiological traits could also be used to study the effect of climate change on the indigenous forest ecosystems. For instance, indigenous forests LAI and biomass are indicators of the amount of sequestered carbon which provides a relatively cheap means for offsetting significant shares of the annual greenhouse gas emissions (Jordan, 1969; Laurance *et al.*, 1998; Beets *et al.*, 2011; Itkonen, 2012). Forest LAI is an important biophysical measure for modeling the energy and mass exchange characteristics between the land surface and the atmosphere of terrestrial ecosystems (Asner *et al.*, 2003). LAI is one of the most useful indicators of vegetation development and health for informing the forest management practices with respect to adjustments and requirements and used indirectly as an input variable for primary production, forest growth and yield models (Asner *et al.*, 2003). Therefore, estimating LAI at tree species level is a necessary and valuable information for tree monitoring, management, conservation and ecosystem services in terms of productivity and health status in indigenous forest ecosystems. On the other hand, the different land use/cover types and other ecological threats like land degradation can also significantly affect tree biophysical and structural properties, as well as productivity (i.e. net primary productivity: NPP) in the indigenous forest ecosystems. Specifically, studies have shown that LAI, which is a dimensionless variable that is defined as the total one-sided surface area of all leaves in the canopy per unit ground area (Bréda, 2003), is an important biophysical variable that determines forest photosynthetic capacity, and therefore the NPP, biomass and other ecosystem processes (Gower *et al.*, 1999; Leuning *et al.*, 2008; Boegh *et al.*, 2013).

Commonly, LAI is estimated through destructive and non-destructive ground-based methods (Chason *et al.*, 1991). However, these ground-based methods of estimating LAI, like the handheld optical instruments, are time-consuming, laborious, subjective and expensive, particularly when carried out in large fragmented landscapes. Hence, studies have sought complementary approaches that use rapid, up-to-date, cost-effective and synoptic data for estimating and modeling LAI (Itkonen, 2012; Cho *et al.*, 2014; Tarantino *et al.*, 2015). Development in remote sensing technologies, data, and processing as well as analytical approaches have made it possible to explicitly and accurately estimate LAI of forests and croplands (Darvishzadeh *et al.*, 2009; Pope and Treitz, 2013; Cho *et al.*, 2014; Atzberger *et al.*, 2015). These studies estimated LAI using different empirical (e.g. linear and polynomial regressions) and physical (e.g. inversion of radiative transfer models) approaches. For example, the physical approaches have been used to estimate LAI of forests in different landscapes (Abuelgasim *et al.*, 2006; Darvishzadeh *et al.*, 2009; Pope and Treitz, 2013; Atzberger *et al.*, 2015). Physically-based retrieval methods, which refers to inversion of radiative transfer models against remote sensing observations (Gobron *et al.*, 2000; Houborg and Boegh, 2008), while the empirical approaches that include multiple linear regressions based on more than two bands also have been applied in predicting LAI (Kokaly and Clark, 1999; Kovacs *et al.*, 2004; Pope and Treitz, 2013). However, identifying suitable variables for developing a multiple regression approach is often critical because some variables are either weakly correlated with LAI or are highly correlated to each other (Kovacs *et al.*, 2004). Moreover, the major challenge is that the multiple linear regressions have often produced limited results, mainly because of their requirements to satisfy some statistical conditions or to assume normal distribution of the input dataset as well as it suffers from multi-collinearity (Cohen *et al.*, 2003; Schlerf *et al.*, 2005). Given this problem, a powerful method for identifying the most useful vegetation indices to improve the prediction of LAI using machine learning regression algorithms is really required. In addition, the aforementioned studies utilized remotely-sensed data of varying spectral (multispectral and hyperspectral) and spatial (fine and medium) resolutions and relatively accurate LAI estimation models were obtained. However, from previous studies and available literature no study has modeled LAI at tree species level in indigenous forest ecosystem in South Africa. Estimating LAI at species level could help resource managers to understand the impact of various socio-ecological mechanisms on indigenous tree species and the vulnerability of these trees to external and internal perturbations.

Specifically, previous studies have mostly focused on the use of SVIs that combined the advent of two or three wavebands as opposed to the use of spectral features at a single waveband on modeling forests LAI (Kovacs *et al.*, 2005; Yang *et al.*, 2006; Leuning *et al.*, 2008; Tillack *et al.*, 2014). SVIs are a mathematical combination of different spectral bands, commonly located in the visible and NIR regions of the electromagnetic spectrum (Viña *et al.*, 2011). SVIs have become one of the most important sources of information for monitoring vegetation, tree species, and other forest biophysical traits. The advantage of the SVIs is to improve the information contained in the spectral reflectance by detecting the spectral variability that might be due to different vegetation, plant, canopy and leaf physiological, and morphological characteristics (Moulin, 1999; Viña *et al.*, 2011). Moreover, SVIs are efficient remotely-sensed variables in reducing the noise in the spectral data due to, for example, the ambient atmospheric conditions, sun view angles, canopy geometry, shading, and soil background (Gilabert *et al.*, 2002). Therefore, a number of SVIs have been developed and tested to estimate indigenous forest LAI. It was found that SVIs were suitable for detecting the within forest LAI spatial variability (Moulin, 1999; Cho *et al.*, 2008b). The most commonly used SVI, the normalized different vegetation index (NDVI), simple ratio index (SRI) and soil adjusted vegetation index (SAVI) are calculated from multispectral data of low and medium spatial resolutions like Landsat and MODIS (Moderate Resolution Imaging SpectroRadiometer) and utilized for estimating forest LAI (Chen, 1996; Gilabert *et al.*, 2002; Yang *et al.*, 2006; Leuning *et al.*, 2008). However, because of low and medium spatial resolutions, the previous studies could not estimate the LAI at tree species level. Estimating LAI at tree species level is a necessary and provides valuable information for tree characterization, monitoring, management, conservation and ecosystem services in terms of productivity and health status in indigenous forest ecosystem.

On the other hand, the newly launched multispectral sensors like Sentinel-2, WorldView-2, WorldView-3, and Pleiades provide data of fine to medium pixel size that can also capture the vegetation spectral properties at some unique portions of the electromagnetic spectrum (e.g. red edge). The fine spatial resolution of these newly launched sensors and the inclusion of the additional bands in calculating the SVI offer a great opportunity for estimating LAI at tree species level. The additional bands (e.g. yellow and red edge), which were previously contained in the hyperspectral data, could overcome the limitations of the conventional bands (e.g. red) while reducing the unnecessary redundancy in the hyperspectral remotely-sensed data (Davi *et al.*, 2006; Soudani *et al.*, 2006; Yang, 2011; Adelabu *et al.*, 2013). Studies have revealed that the key bands like the red edge of multispectral data are useful to characterize spatial variability changes in vegetation biophysical parameters like LAI (Tarantino *et al.*, 2015).

Where the empirical regression methods are concerned, most of the above-mentioned studies either employed multiple linear regressions or machine learning approaches like RF to estimate LAI at forest level. However, the major challenge with conventional empirical methods that include multiple linear regressions, is that they assume a normal distribution on the response variables and suffer from multi-collinearity (Cohen *et al.*, 2003; Schlerf *et al.*, 2005). The use of machine learning methods has therefore been regarded as efficient and robust protocols for estimating forest biophysical traits in the field of remote sensing (Thissen *et al.*, 2004; Cho and Skidmore, 2009; Liu *et al.*, 2013). Particularly, these methods which make no assumption of the input predictor variables distribution have increasingly offered a better capability to analyze remote sensing data (Chan and Paelinckx, 2008; Dalponte *et al.*, 2009). In particular, there is a lack of knowledge on whether the high resolution multispectral data (e.g. WorldView-2) could be employed to estimate LAI of individual tree species in two different indigenous forest ecosystems (e.g. fragmented and intact forest ecosystems). Moreover, ecologists might need to test whether there are significant differences between LAI of tree species grown in intact and fragmented forest ecosystems in order to assess the ecosystems' services and resilience in the value chain. The empirical estimation of tree LAI in such complex and dynamic forest ecosystems using remotely-sensed data may require efficient and robust machine learning regression algorithms like SVM and ANN. SVM is a well-known machine learning algorithm which has frequently been used to change the nonlinear regression problem into a linear projection by a variety of kernel methods (Cristianini and Shawe-Taylor, 2000; Cherkassky and Ma, 2004). The advantage of SVM compared to other conventional linear and nonlinear regression methods is that it offers excellent generalization abilities. The SVM algorithm also offers scarce solutions where only the most relevant sample of the training data are weighted causing low computational cost (Cortes and Vapnik, 1995; Gilabert *et al.*, 2002). Previously, this algorithm has been applied to relate SVIs to various vegetation biophysical traits like LAI with various success (Verrelst *et al.*, 2012a; Liang *et al.*, 2013).

On the other hand, ANN regression comprises of an interconnected group of artificial neurons and processes information using a connectionist method for calculation (Atkinson and Tatnall, 1997; Xiu and Liu, 2003; Singh and Chauhan, 2009). The approach therefore offers a very efficient regression method to simulate the relationship between SVIs and LAI (Smith, 1993; Kimes *et al.*, 1998; Fang and Liang, 2003). From the available literature and to the researcher's knowledge, no study utilized SVIs calculated from WorldView-2 data, SVM and ANN regression to predict LAI in tropical indigenous forests at species level. In the present study, the utility of SVIs calculated from the multispectral WorldView-2 data was tested for predicting LAI at tree species level using the SVM and ANN machine learning regression algorithms. The LAI of six endangered tree species grown in the Dukuduku intact and fragmented (open) indigenous forest ecosystems was estimated. The study also tested whether there are significant differences between LAI of the six endangered tree species grown in intact and fragmented indigenous forest ecosystems.

4.2 Methodology

4.2.1 Sampling Procedure and Field Data Collection

A field campaign was carried out between 1st and 7th December 2013 following stratified purposive sampling method to collect LAI measurements from the six selected endangered tree species (*Albizia adianthifolia*, *Ekebergia capensis*, *Harpephyllum caffrum*, *Hymenocardia ulmoides*, *Sclercarya birrea* and *Trichilia dregeana*). A handheld Leica Geosystem GS20 GPS of sub-meter (0–0.25 m) accuracy (Geosystems, 2004) was used to geo-locate the sample trees. A handheld LAI-2200 plant canopy analyzer (PCA) was used to estimate LAI of each sample tree under overcast sky conditions at low solar elevation, i.e. around early morning (8:00-10:00, Greenwich Mean Time: GMT +2) and late afternoon (15:00-18:00, GMT +2) with 180^0 view restrictor on the sensor (Soudani *et al.*, 2006; Lottering and Mutanga, 2012). To avoid direct sunlight on the sensor, it was necessary for the operator to take samples of below and above canopy radiation in the opposite direction to the sun. For each sample plot, one above canopy measurement was taken by walking to an adjacent open field. Next, five below canopy measurements were performed on the individual trees at regular space points around each tree diameter from which the average sample plot LAI was calculated. In total, 563 samples were collected from both the fragmented (n = 300) and intact (n = 263) indigenous forest strata (Figure 4.1). For each endangered tree species, the sample points were 58 and 67 (*Albizia adianthifolia*), 47 and 37 (*Ekebergia capensis*), 41 and 44 (*Harpephyllum caffrum*), 39 and 56 (*Hymenocardia ulmoides*), 40 and 59 (*Sclercarya birrea*), and 38 and 37 (*Trichilia dregeana*) in the intact and fragmented strata, respectively.

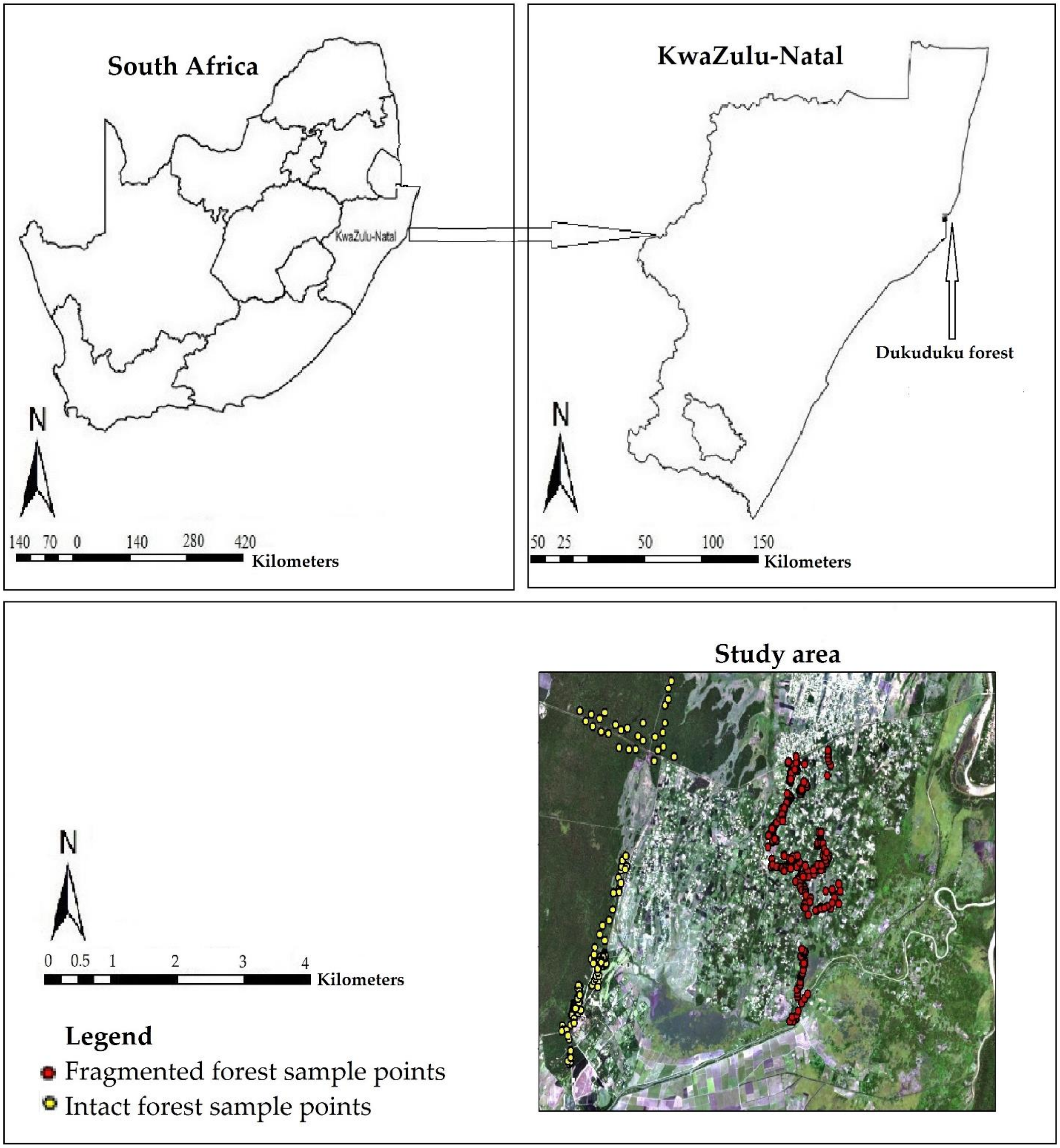


Figure 4.1: The location of the Dukuduku indigenous forest in KwaZulu-Natal province, South Africa and field sample locations overlaid in a true-color WorldView-2 image

4.2.2 Satellite Image Acquisition and Pre-Processing

WorldView-2 image was acquired on 1st December 2013 under clear-sky conditions. The WorldView-2 is the first multispectral commercial satellite with eight wavelength and senses in the 400 –1400 nm spectral range (50-180 nm). The spatial resolution of the multispectral bands is 2.0 m along with a panchromatic band of 0.5 m pixel size with a swath of 16.4 km at nadir (DigitalGlobe, 2010). The WorldView-2 image consists of four conventional bands and four additional bands. Therefore, the sensor has the spectral and spatial resolutions that meet many applications like predicting and monitoring forest structural and biophysical variables at species level (Ozdemir and Karnieli, 2011; Mutanga *et al.*, 2015). The image was atmospherically corrected and transformed at canopy reflectance using the QUAC procedure in ENVI 4.7 software (ENVI, 2009). QUAC performs in-scene based atmospheric correction at the visible and near- to-shortwave infrared (VNIR-SWIR) region of the electromagnetic spectrum for multi- and hyperspectral imagery. QUAC determines atmospheric composition parameters directly from the information contained within the image (Pixel spectra), thus allowing for the retrieval of accurate reflectance spectra (Shen *et al.*, 2005; Zengeya *et al.*, 2013). The acquired image was geometrically corrected (UTM zone 36 South using WGS-84 geodetic datum) by digitalGlobe™.

4.2.3 Spectral Vegetation Indices (SVIs)

SVIs are usually derived from a combination of two or more image bands and are expected to provide more stable information about the vegetation compared to spectral information at a single band (Jackson and Huete, 1991; Jawak and Luis, 2013). SVIs based on absorption and reflectance in the visible and NIR regions (e.g. NDVI) have been widely used for estimating biophysical parameters (e.g. LAI) of agricultural and natural ecosystem (Chen, 1996; Pope and Treitz, 2013; Tillack *et al.*, 2014; Pu and Cheng, 2015). In the current study, after the WorldView-2 image was processed, 24 SVIs were computed (Table 4.1) and utilized to predict the LAI of the six endangered tree species. These indices were selected based on previous studies that predicted forest biophysical traits like LAI and biomass (Lu, 2006; Tillack *et al.*, 2014; Pu and Cheng, 2015; Ojoyi *et al.*, 2016). The study investigated the use of all SVIs combined together (n = 24) for predicting the LAI of the endangered tree species in each forest stratum (i.e. intact and fragmented forests).

Table 4.1: Summary of the WorldView-2-derived spectral vegetation indices (SVIs) used in present study

No.	Vegetation index	Abbreviation	Equation	Reference
1	Simple Ratio Index	SRI	NIR1/RED	(Jordan, 1969)
2	Normalized Difference Vegetation Index	NDVI	NIR1 – RED/NIR1 + RED	Rouse *et al.* (1973)
3	Ratio Vegetation Index	RVI	RED/NIR1	(Richardson and Weigand, 1977)
4	Transformed Vegetation Index	TVI	√NIR1 – RED/NIR1+RED +0.5	(Deering and Rouse, 1975)
5	Non-Linear Index	NLI	(NIR2-RED)/(NIR2 + RED)	(Goel and Qin, 1994)
6	Atmospherically Resistant Vegetation Index	ARVI	(NIR2-(2*RED-BLUE)/(NIR2 +(2*RED-BLUE)	(Kaufman and Tanre, 1992)
7	Structure-Insensitive Pigment Index	SIPI	(NIR1–BLUE)/(NIR1–RED EDGE)	(Peluelas *et al.*, 1995)
8	Renormalized Difference Index	RDI	(NIR1-RED) /(NIR1+RED)½	(Roujean and Breon, 1995)
9	Green Normalized Difference Vegetation Index	GNDVI	(NIR1-GREEN)/(NIR1+ GREEN)	(Gitelson and Merzlyak, 1996)
10	Modified Simple Ratio*	MSR*	(NIR1/RED-1)/(NIR1/RED)½+1	(Chen, 1996)
11	Pigment Specific Simple Ratio (Chlorophyll a)	PSSRa	NIR1/RED EDGE	(Blackburn, 1998)
12	Pigment Specific Simple Ratio(Chlorophyll b)	PSSRb	NIR1/RED	(Blackburn, 1998)
13	Plant Senescence Reflectance Index	PSRI	(RED EDGE–BLUE)/NIR1	(Merzlyak *et al.*, 1999)
14	Enhanced Vegetation Index	EVI	2.5 *((NIR1 − RED)/ (NIR1+ 6* RED −7.5* BLUE + 1))	(Huete *et al.*, 1999)
15	Modified Chlorophyll Absorption in Reflectance Index	MCARI	[(RED EDGE–RED)–0.2(RED EDGE–GREEN)](RED EDGE/RED)	(Daughtry *et al.*, 2000)
16	Modified Simple Ratio	MSR	(NIR1 − BLUE)/(RED − BLUE)	(Sims and Gamon, 2002)
17	Normalized Difference Index	NDI	(NIR1 – RED)/ (NIR1 + RED)	(Sims and Gamon, 2002)

Table 4.1 (Continued)

No.	Vegetation index	Abbreviation	Equation	Reference
18	Transformed Chlorophyll Absorption in Reflectance Index	TCARI	3[(RED EDGE–RED)–0.2(RED EDGE–GREEN)(RED-EDGE/RED)]	(Haboudane *et al.*, 2002)
19	Visible Atmospherically Resistant Index	VARI	(GREEN-RED)/(GREEN+RED-BLUE)	(Gitelson *et al.*, 2002)
20	Visible Green Index	VGI	(GREEN-RED)/(GREEN+RED)	(Gitelson *et al.*, 2002)
21	Modified Normalized Difference	MND	(NIR1–BLUE)/(NIR1+RED EDGE–2BLUE)	(Sims and Gamon, 2002)
22	Carotenoid Reflectance Index	CRI	(1/BLUE)-(1/RED EDGE)	(Gitelson *et al.*, 2002)
23	Green Index	GI	(NIR1/RED) - 1	(Gitelson *et al.*, 2005)
24	Red Index	RI	(NIR1/RED) -1	(Gitelson *et al.*, 2005)

Blue, green, red, red edge, near infrared 1 and near infrared 2 are WorldView-2 bands 2, 3, 5, 6,7 and 8, respectively

4.3 Statistical Analysis

4.3.1 Descriptive Statistics and an Independent *t*-test

The field LAI data were described using the mean and standard deviation (SD) statistics. The data were then tested for normality using the Shapiro-Wilk test (Royston, 1982). An independent *t*-test was then performed with 95% confidence levels ($p \leq 0.05$) to test if there were significant differences in the endangered tree species LAI between the intact and fragmented indigenous forest strata.

4.3.2 Support Vector Machines (SVM) Regression Algorithm

SVM algorithm, which was invented by Cortes and Vapnik (1995), is based on statistical learning theory and can be regarded as the same type of networks, corresponding precisely to the same type of solution but trained in a different way and therefore with different values of the weight after the training (Evgeniou *et al.*, 1999; Gilabert *et al.*, 2002; Durbha *et al.*, 2007). SVM is very specific learning algorithms characterized by the usage of kernels, absence of local minima, sparseness of the solution and capacity control obtained by acting on the margin, or on a number of support vectors (Karamouz *et al.*, 2009). Originally, SVM was developed to solve the classification problems but it was later extended to handle regression problems (Vapnik, 1995; Marabel and Alvarez-Taboada, 2013). The support vector regression algorithm converts the nonlinear regression problem into a linear relationship by using the kernel functions to map the original input space into a new feature space with higher dimensions (Cortes and Vapnik, 1995; Chen *et al.*, 2012). In particular, the SVM regression aims to estimate an unknown continuous-valued function based on a finite number of noisy samples (Cherkassky and Ma, 2002; Marabel and Alvarez-Taboada, 2013). Basically, it makes use of structural minimization principle which is known to have good generalization performance for different dataset sizes as contrasted to empirical risk minimization employed by other method like ANN (Camps-Valls *et al.*, 2006; Akande *et al.*, 2014). There are different variations of SVM that use different optimization methods due to two parameters that are commonly used. These two parameters are ε-insensitive zone (ε-SVM) and regularization parameter (C-SVM) (Cherkassky and Ma, 2002). The accuracy of the SVM is highly dependent on a correct setting of the meta-parameters (ε-SVM and C-SVM). The parameter ε-SVM controls the width of the epsilon-insensitive zone for the training dataset (Kohestani and Hassanlourad, 2015). Hence, the value of ε-SVM can affect the number of support vectors used to construct the regression function. In other words, the bigger the epsilon, the fewer support vectors are selected (Cherkassky and Ma, 2004; Kohestani and Hassanlourad, 2015). Conversely, bigger ε-SVM values result in more 'flat' estimates (Marabel and Alvarez-Taboada, 2013), while the parameter C-SVM determines the balance between the model complexity and the degree to which the larger deviations (than epsilon) are tolerated in the optimization (Cherkassky and Ma, 2002; Marabel and Alvarez-Taboada, 2013). Therefore, the larger values of C-SVM aim at minimizing the empirical risk regardless of the complexity of the model (Marabel and Alvarez-Taboada, 2013). In this regard, both C-SVM and ε-SVM values affect model complexity. A more detailed description of SVM method can be found in Cortes and Vapnik (1995), Cherkassky and Ma (2002) and Ben-Hur and Weston (2010).

In the present study, the SVM regression method was used to estimate the LAI of the six endangered tree species, using the Vapnik's ε-insensitive loss function to minimize the training errors. In order to project the data into a new space, a radial basis function was used, followed by optimization procedure to find the number of support vectors for the best performance. Moreover, the optimal values of the two parameters *C*-SVM and ε-SVM of the radial basis function were obtained using a 10-fold cross validation method and grid search on the training dataset (Hsu *et al.*, 2009; Yang, 2011). The training dataset was divided into 10 subsets of equal size, SVM regression models were then trained on the nine subset samples, and tested on the removed one and the process was repeated ten times until all subset samples had served as test samples (Omer *et al.*, 2015a). The pair parameter that minimizes the prediction error was then considered as the best values for the final prediction performance. The analysis was carried out using the e1071 library version 2.15.2 in R statistical packages (R Development Core, 2015).

4.3.3 Artificial Neural Networks (ANN) Regression Algorithm

ANN is one of the first developed nonparametric machine learning regression techniques. It is a powerful approach that can not only analyze complex relationships but also does not depend on an assumption of data normality (Atkinson and Tatnall, 1997; Foody, 2004b). The ANN is a mathematical model that simulates the structure and functional aspects of the biological neural connections. It consists of an interconnected group of artificial neurons and processes information using a connectionist method for calculation (Xiu and Liu, 2003; Singh and Chauhan, 2009). Several models of ANN such as radial basis function, back propagation and multilayer perceptron have been applied for analyzing remotely-sensed data for a variety of applications like forestry and ecological modeling (Foody *et al.*, 2003; Corne *et al.*, 2004; Ingram *et al.*, 2005). Radial basis function neural networks have proved to be a good function for analyzing a wide variety of remotely-sensed data since it reduces the computational time required for the training process (Foody, 2004b; Boegh *et al.*, 2013). The approach requires one input variable which is the 'distance' between the weight and input nodes. The back propagation is the multilayer feed forward neural networks method which comprises of a series of simple connected neurons, or nodes, between input and output layers (Atkinson and Tatnall, 1997) while the multilayer perceptron is a commonly used ANN structure that consists of an input layer, an output layer and one or more hidden layers of nonlinearly-activating nodes (Atkinson and Tatnall, 1997; García Nieto *et al.*, 2012). The nodes are connected by a certain synaptic weight to all nodes in the next layer and the perceptron learning occurs through the changes in the connections weights after the input protector (SVIs) are processed (García Nieto *et al.*, 2012). The multilayer perceptron is a feed forward ANN model that projects input data onto a set of suitable output by using three or more layers of nodes with nonlinear activation functions (Atkinson and Tatnall, 1997; Xiong *et al.*, 2010). ANN has been widely used in modeling vegetation and tree species traits that are not linearly predictable in the original remotely-sensed variables (Menzies *et al.*, 2007; Liu *et al.*, 2013).

In the present study, ANN regression algorithm was employed using the multilayer perceptron modeling approach to estimate the LAI of the six endangered tree species using the 24 SVIs presented in Table 4.1. Many trials of internal networks structure, input data, and learning algorithms have been tested to define the optimal regression features. The structure of the hidden layers was tested to assess the necessary number of hidden layers and the number of required nodes per layer. This was tested by manually changing the number of nodes in the hidden layer.

4.3.4 Validation

To validate the performance of the SVM and ANN regression algorithms, the reference data were randomly split into 70% (210 and 184 for the intact and fragmented strata, respectively) for calibration and 30% (90 and 79 for the intact and fragmented strata, respectively) for validation based on the recommendation made by Adelabu *et al.* (2015). Moreover, the calibration dataset was used for optimizing the SVM and ANN regression algorithms, while the validation dataset was used to examine the performance and reliability of the prediction models. One-to-one relationships between the measured and predicted LAI values were fitted and the coefficient of determination (R^2), root mean square errors (RMSE) and bias were then calculated (Equations 4.1, 4.2, and 4.3). The RMSE provides direct estimates of the modeling errors expressed in the original measurement units, the lower value of RMSE indicates a good predictive model performance (Akande *et al.*, 2014).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{1=n}^{n} (y_i - \hat{y}_i)^2} \qquad \text{Equation 4. 1}$$

$$\mathrm{RMSE}(\%) = \frac{\sqrt{\frac{1}{n}\sum_{1=n}^{n}\left(y_i - \hat{y}_i\right)^2}}{\bar{y}} \times 100 \qquad \text{Equation 4. 2}$$

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \qquad \text{Equation 4. 3}$$

where y_i is the measured LAI, $\hat{y}_i$ is the predicted LAI, $\bar{y}$ is the mean of the measured LAI, n is the number of sample data in the validation dataset.

4.4 Results

4.4.1 Descriptive Statistics and an Independent *t*-test

A Shapiro–Wilk normality test showed that the LAI data for the six endangered tree species grown in the fragmented and intact indigenous forest strata were normally distributed (p = 0.03 for the intact forest stratum and p = 0.04 for the fragmented forest stratum). Figure 4.2 shows the descriptive statistics of the LAI for the six endangered tree species in the fragmented and intact indigenous forest strata. The result of the independent *t*-test showed that fragmented forest obtained significantly higher ($p \leq 0.05$) mean LAI compared to the intact indigenous forest strata (Figure 4.2). With regard to the individual tree species, there is a great variability in LAI among the tree species and the highest mean LAI were achieved for *Albizia adianthifolia* (4.19) and *Trichilia dregeana* (3.94) in intact forest stratum, while the lowest mean values were obtained for *Albizia adianthifolia* (2.03) in fragmented forest stratum (Figure 4.2). Also, the descriptive statistics of the combined (aggregated) LAI across the six endangered tree species in the intact and fragmented forest ecosystems are shown in Figure 4.3. The result of the independent *t*-test showed that the two forest ecosystems (fragmented and intact) revealed a significant ($p \leq 0.05$) combined LAI.

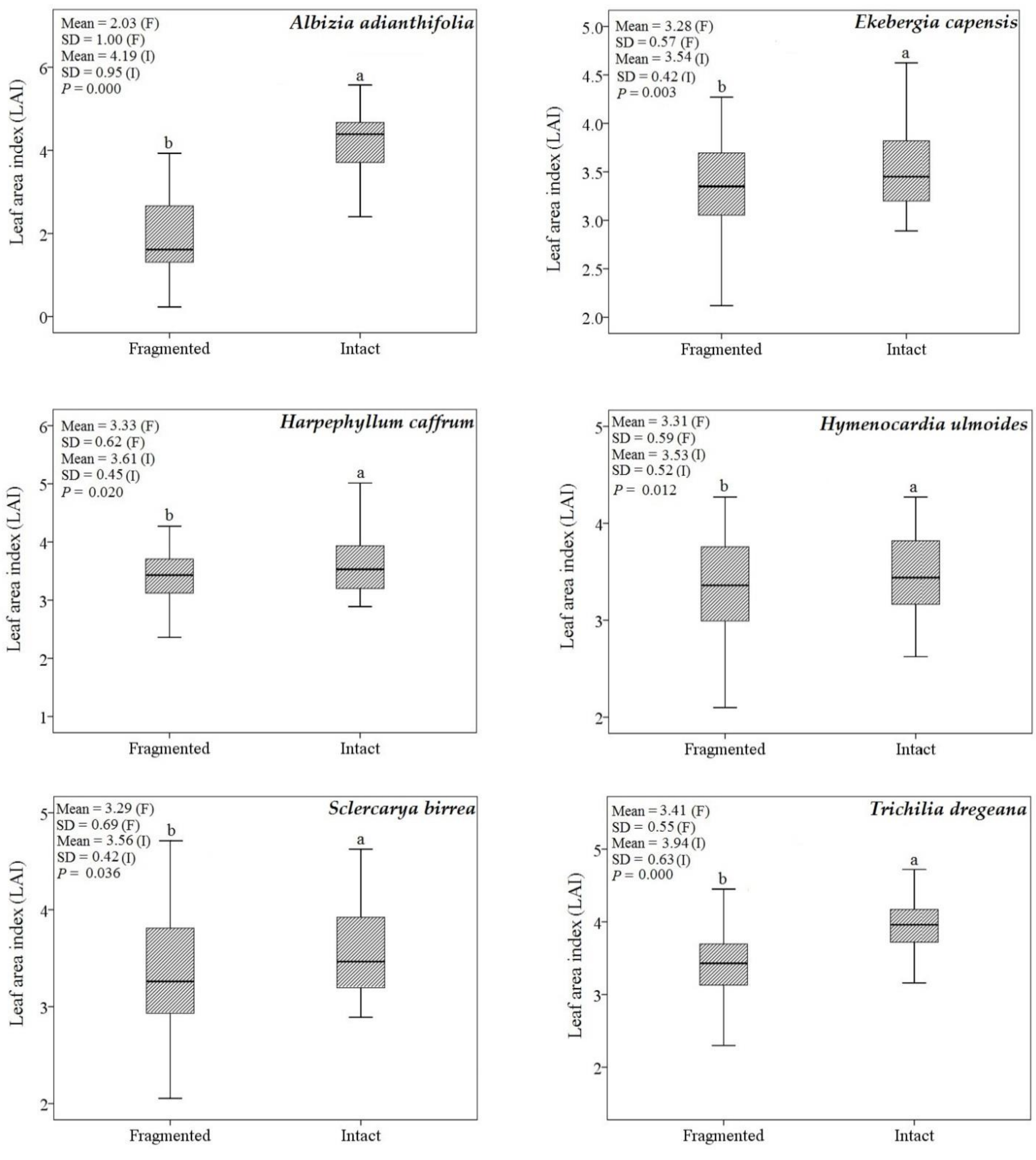


Figure 4.2: Descriptive statistics of the measured LAI of the six endangered tree species in both the intact (I) and fragmented (F) forest ecosystems. LAI data for each tree species in the both forest ecosystems (I and F) with a different letter are significantly different ($p \leq 0.05$) from each other

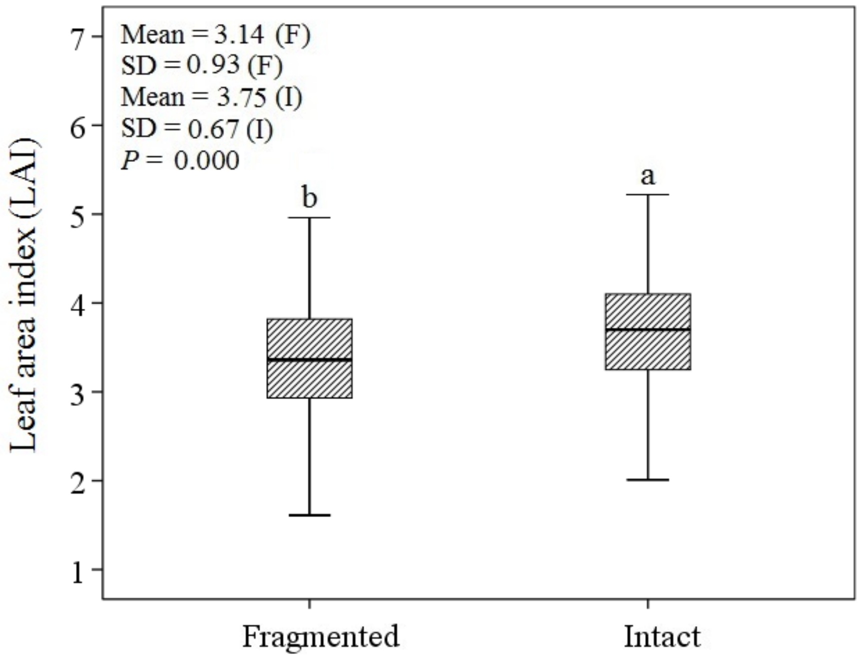


Figure 4.3: Descriptive statistics of the measured LAI of the combined (aggregated) six endangered tree species datasets in both the intact (I) and fragmented (F) forest ecosystems. Combined LAI data with a different letter are significantly different ($p \leq 0.05$) from each other

4.4.2 Support Vector Machines (SVM) and Artificial Neural Networks (ANN) Regression Models

Table 4.2 shows the optimum parameters for both the SVM and ANN regression methods. The 10-fold cross validation method and grid search approaches resulted in optimal ε-SVM and C-SVM values of 1 and 100, respectively, for all endangered tree species in the two forest strata, except for the *Albizia adianthifolia* and *Sclercarya birrea* in fragmented forest stratum (1 and 1000), *Hymenocardia ulmoides*, *Sclercarya birrea* and *Trichilia dregeana* in intact forest stratum (1 and 10). The table also shows that the input layers for ANN regression method ranged between 1 and 6 for the six endangered tree species in the fragmented forest stratum and between 4 and 6 for the species in the intact forest stratum, while the hidden layers were varied between 2 to 5 and 4 to 9 for the fragmented and intact forest strata, respectively. The optimal ε-SVM and C-SVM values of 1 and 100 respectively, for combined fragmented and intact data also presented in Table 4.2.

Table 4.2: The optimal parameters for the best trained SVM and ANN regression models used for estimating the LAI of the six endangered tree species in the fragmented and intact indigenous forest strata

	Support vector machines (SVM)					
Endangered tree species	**Fragmented forest stratum**			**Intact forest stratum**		
	***ε*-SVM**	***C*-SVM**		***ε*-SVM**	***C*-SVM**	
AA	1.0	1000		1.0	100	
EC	1.0	100		1.0	100	
HC	1.0	100		1.0	100	
HU	1.0	100		1.0	10	
ScB	1.0	1000		1.0	10	
TD	1.0	100		1.0	10	
Combined data	1.0	100		1.0	100	
	Artificial neural networks (ANN)					
	inputs	**Hidden**	**Profile**	**inputs**	**Hidden**	**Profile**
AA	3.0	05	MLP 3:3-5-1:1	4.0	06	MLP 4:4-6-1:1
EC	5.0	04	MLP 5:5-4-1:1	6.0	08	MLP 6:6-8-1:1
HC	2.0	03	MLP 2:2-3-1:1	4.0	09	MLP 4:4-9-1:1
HU	1.0	02	MLP 1:1-2-1:1	5.0	06	MLP 5:5-6-1:1
ScB	1.0	02	MLP 1:1-2-1:1	4.0	04	MLP 4:4-4-1:1
TD	6.0	05	MLP 6:6-5-1:1	6.0	04	MLP 6:6-4-1:1
Combined data	3.0	04	MLP 3:3-4-1:1	4.0	05	MLP 4:4-5-1:1

AA = *Albizia adianthifolia*, EC = *Ekebergia capensis*, HC = *Harpephyllum caffrum*, HU = *Hymenocardia ulmoides*, ScB = *Sclercarya birrea*, TD = *Trichilia dregeana*, *ε*-SVM = ε-insensitive zone, *C*-SVM = regularization parameter and MLP = multilayer perceptron

The results of training (calibrating) both the SVM and ANN regression approaches are presented in Table 4.3. All the SVM and ANN models explained more than 70% of the variance ($R^2_{\mathrm{Cal}} \leq$ 0.70) in the tree LAI in the fragmented forest stratum, except for the *Sclercarya birrea* when the SVM and ANN models were trained and for the *Trichilia dregeana* using the ANN regression method. For the intact forest stratum, the results showed R^2_{Cal} values of more than 0.70 for all endangered tree species when SVM regression model was fitted, while the ANN regression models resulted in R^2_{Cal} values of less than 0.70 for all tree species except for the *Harpephyllum caffrum* (Table 4.3). In general, the SVM regression models yielded relatively better results compared to the ANN models. On the other hand, models developed using the fragmented forest data fitted LAI data more accurately compared with the models developed using the intact forest data (Table 4.3).

Table 4.3: Coefficient of determination (R^2_{Cal}) and root mean square errors ($RMSE_{Cal}$) for the SVM and ANN regression models when calibrated using the data collected from the fragmented and intact forest strata

	Support vector machines (SVM)					
Endangered tree species	**Fragmented forest stratum**			**Intact forest stratum**		
	R^2_{Cal}	$RMSE_{Cal}$	$RMSE_{Cal}$%	R^2_{Cal}	$RMSE_{Cal}$	$RMSE_{Cal}$ %
AA	0.79	0.09	2.06	0.83	0.13	1.93
EC	0.75	0.05	0.87	0.77	0.05	1.27
HC	0.82	0.04	1.09	0.78	0.04	1.35
HU	0.86	0.03	0.73	0.73	0.06	0.98
ScB	0.64	0.08	1.12	0.72	0.09	3.50
TD	0.72	0.05	1.73	0.69	0.06	1.72
Combined data	0.81	0.06	1.89	0.77	0.08	2.58
Artificial neural networks (ANN)						
AA	0.74	0.07	4.69	0.62	0.09	2.18
EC	0.70	0.03	0.91	0.63	0.05	1.33
HC	0.78	0.04	1.14	0.71	0.05	1.39
HU	0.84	0.03	0.78	0.59	0.07	2.00
ScB	0.66	0.04	1.15	0.67	0.05	1.31
TD	0.67	0.06	1.81	0.62	0.08	1.99
Combined data	0.74	0.07	2.41	0.62	0.10	4.02

$_{Cal}$ = Calibration dataset, AA = *Albizia adianthifolia*, EC = *Ekebergia capensis*, HC = *Harpephyllum caffrum*, HU = *Hymenocardia ulmoides*, ScB = *Sclercarya birrea*, and TD = *Trichilia dregeana*

4.4.3 Model Validation

Figures 4.4, 4.5, 4.6 and 4.7 show the one-to-one relationships between the measured and predicted LAI for all models developed in the present study. When the performance of the SVM prediction models was assessed, the results showed that the LAI could be better estimated for the *Hymenocardia ulmoides* trees grown in the fragmented forest ecosystem as indicated by the relatively higher R^2_{Val}, and lower error metrics (Figure 4.4) while for the intact forest ecosystem (Figure 4.5) the best model was achieved for predicting the LAI of the *Albizia adianthifolia* trees (R^2_{Val} = 0.80 and $RMSE_{Val}$ = 2.10% of the mean). The slope in all other predictive models was deviated from the expected one-to-one relationship and the models either overestimated or underestimated the LAI measurements. On the other hand, the best ANN regression model was achieved for estimating the LAI of the *Hymenocardia ulmoides* trees (Figure 4.6) in the fragmented forest stratum (R^2_{Val} = 0.71 and $RMSE_{Val}$ = 1.52% of the mean) and for *Harpephyllum caffrum* trees (Figure 4.7) in the intact forest stratum (R^2_{Val} = 0.71 and $RMSE_{Val}$ = 1.57% of the mean). It is interesting to note that most of the SVM models developed using the fragmented data overestimated the LAI in all tree species, except for *Sclercarya birrea*. Furthermore, the results also showed that the LAI could be better estimated for the combined fragmented forest dataset (Figure 4.8) compared with combined intact forest dataset (Figure 4.9). Since the SVM regression method achieved relatively more accurate prediction models, predicted LAI maps were produced using the regression algorithm for the combined intact and fragmented data (Figure 4.10). The maps can be compared to the WorldView-2 image which depicts the intact and fragmented forest ecosystems. The maps show the spatial variation in the forest LAI in the fragmented and intact forest ecosystems.

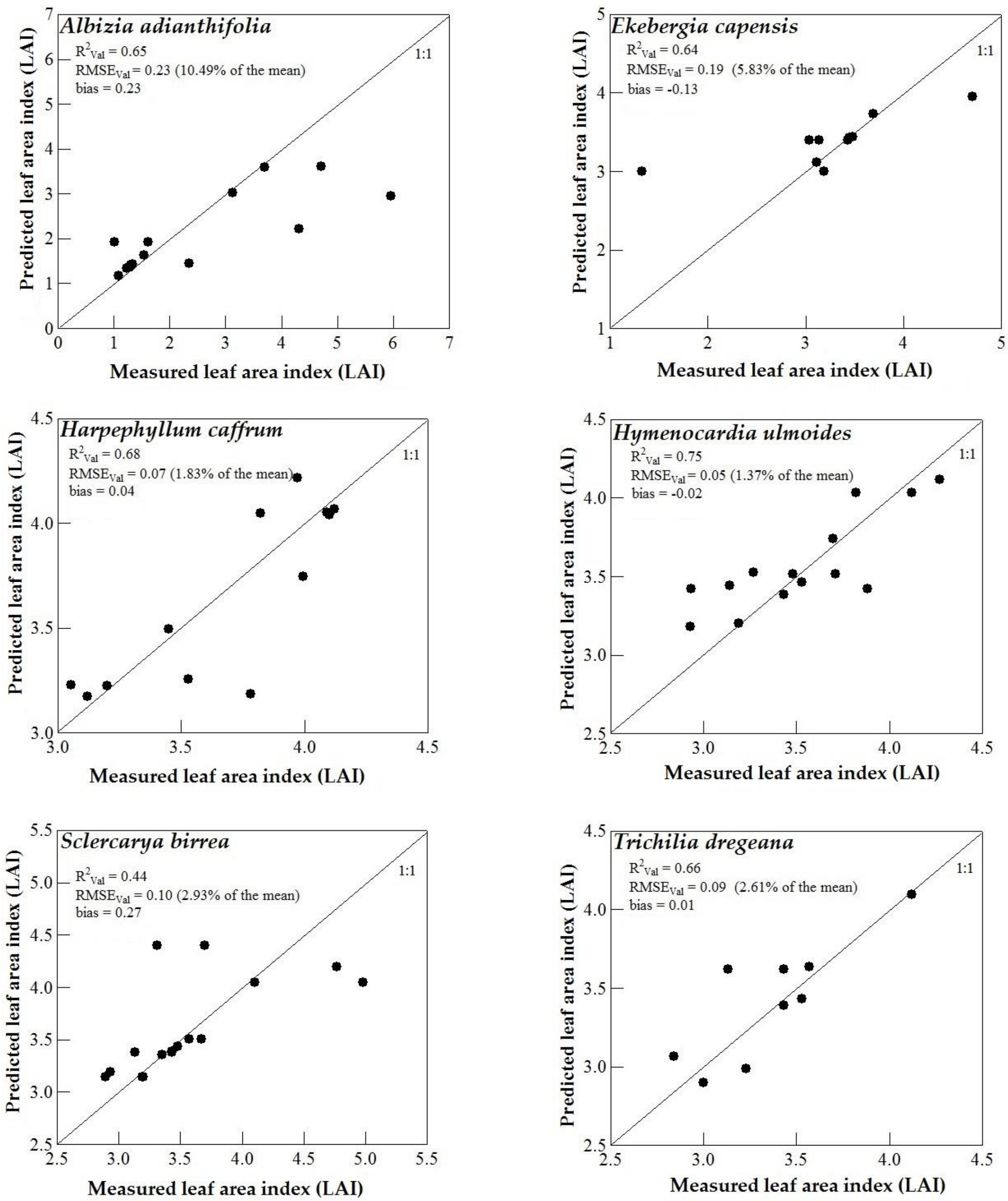


Figure 4.4: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using support vector machines (SVM) regression algorithm and fragmented indigenous forest data

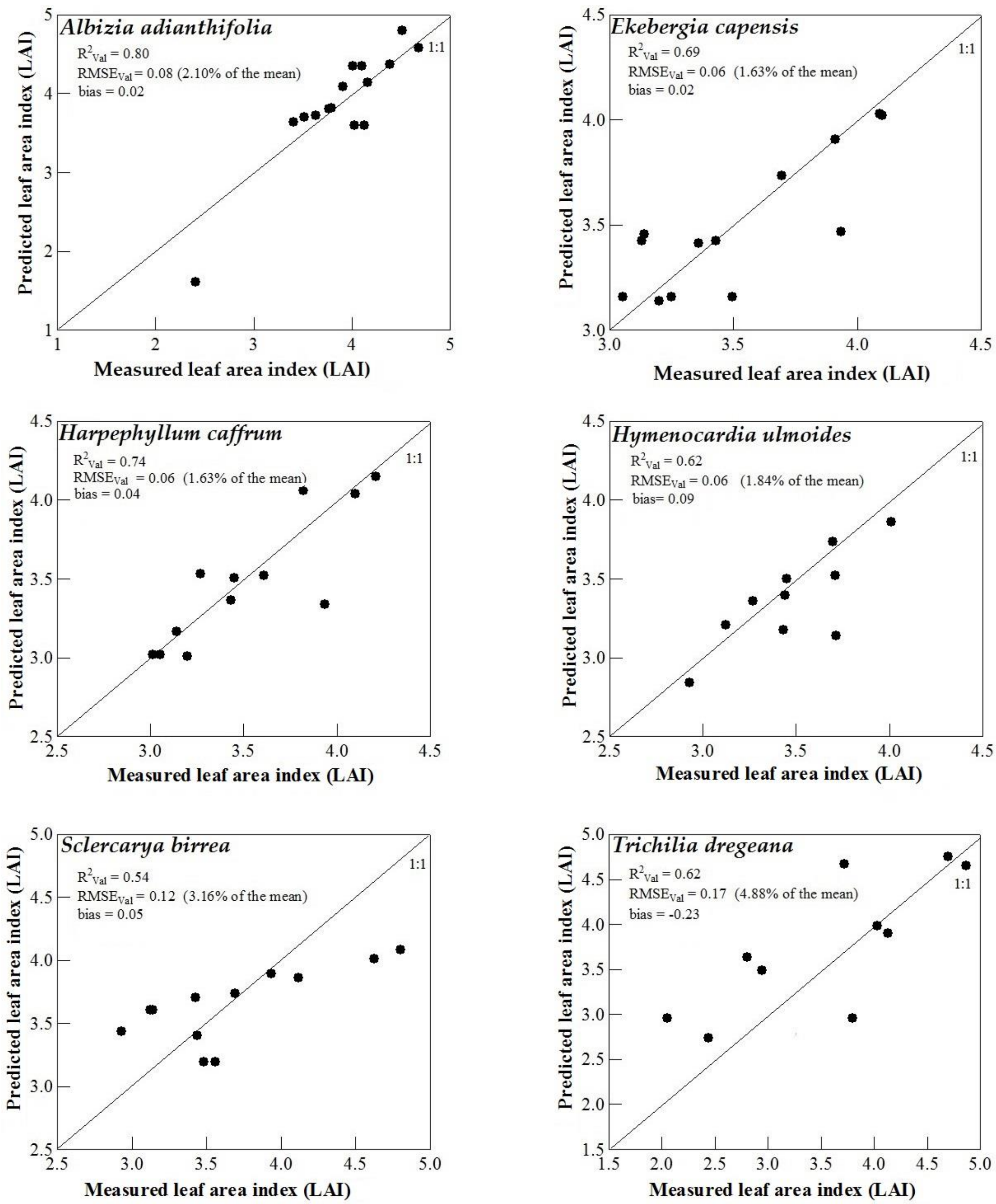


Figure 4.5: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using support vector machines (SVM) regression algorithm and intact indigenous forest data

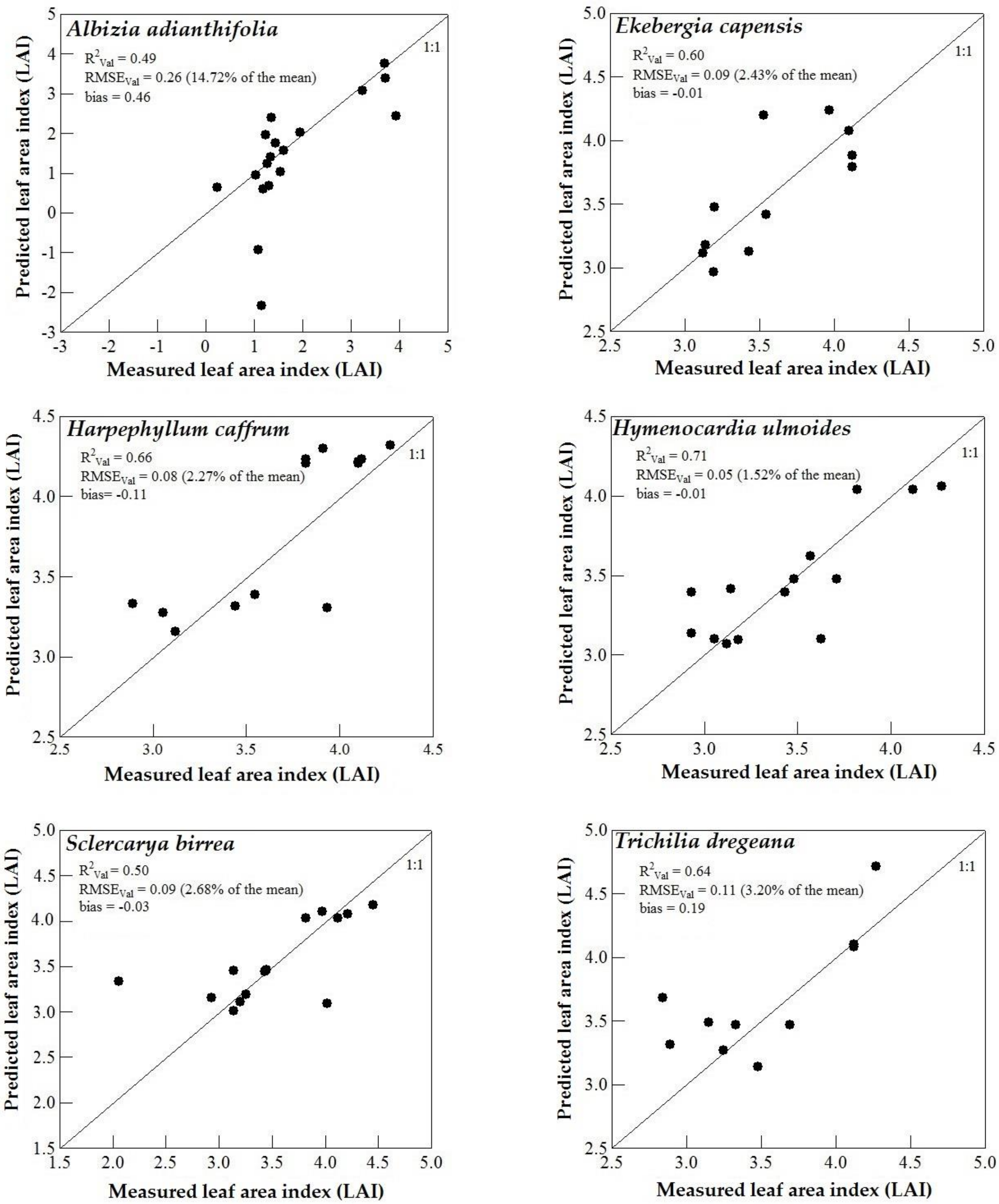


Figure 4.6: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using artificial neural networks (ANN) regression algorithm and fragmented indigenous forest data

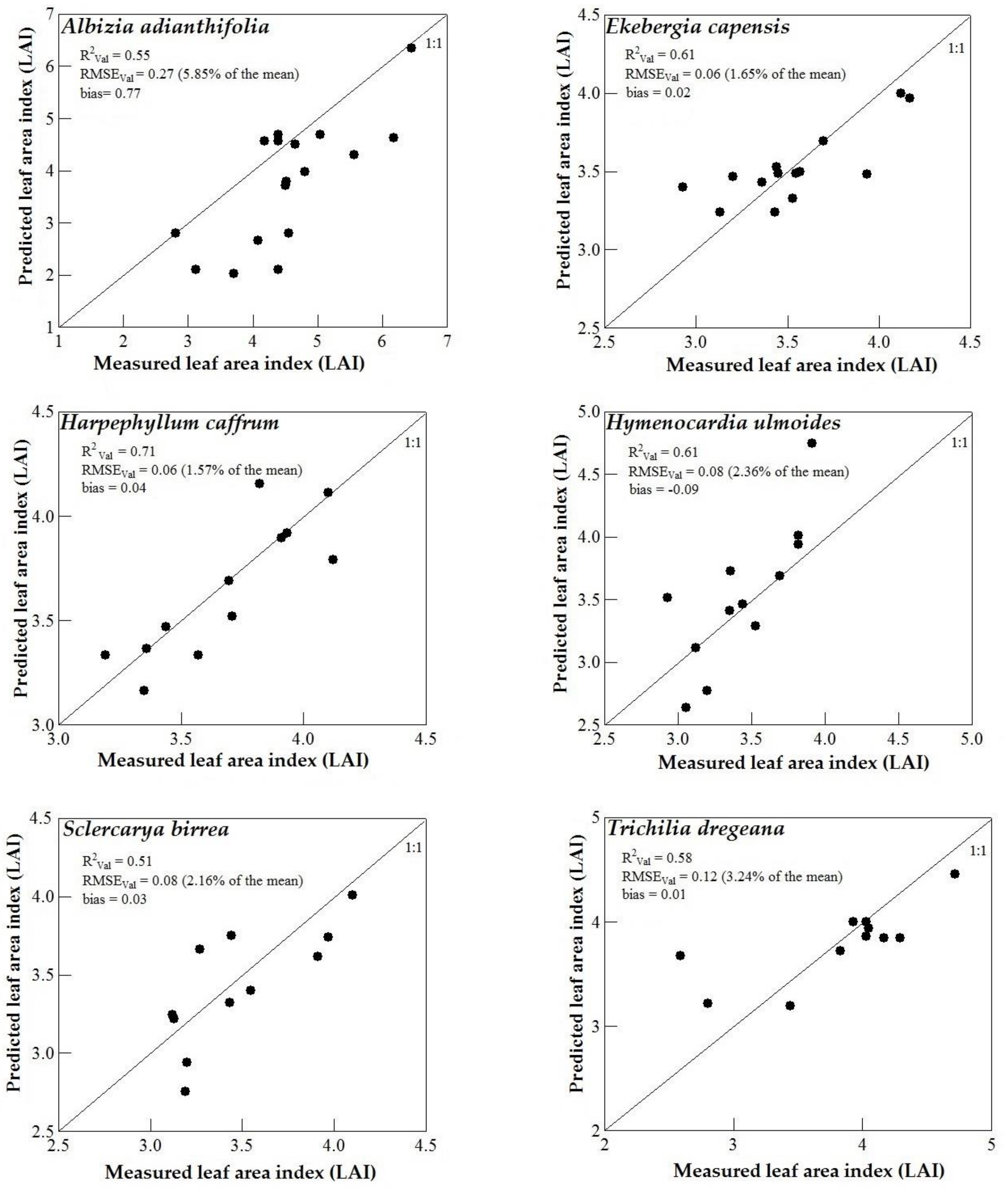


Figure 4.7: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using artificial neural networks (ANN) regression algorithm and intact indigenous forest data

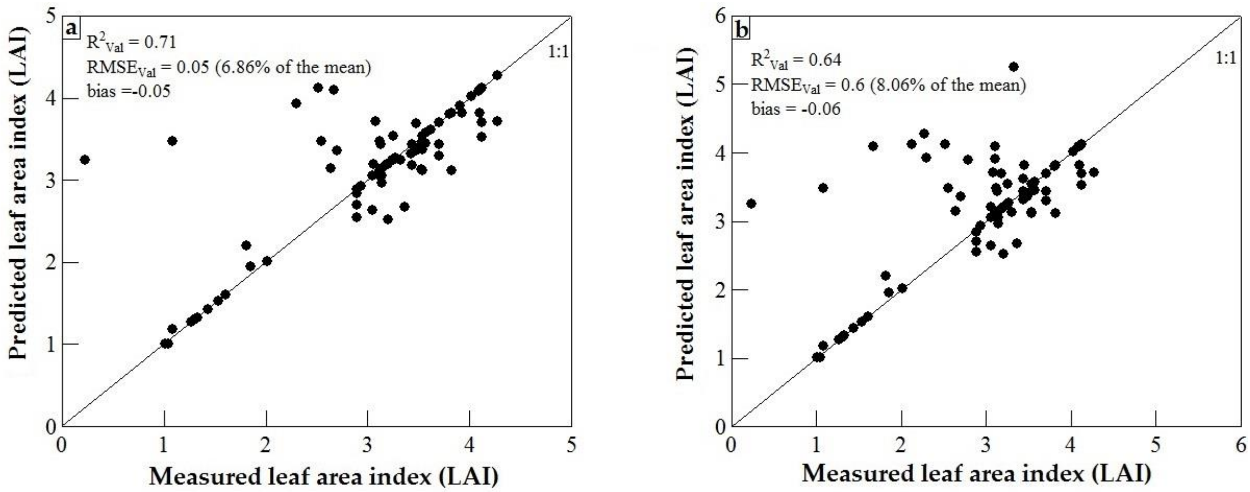


Figure 4.8: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using combined fragmented indigenous forest data and (a) support vector machines and (b) artificial neural networks

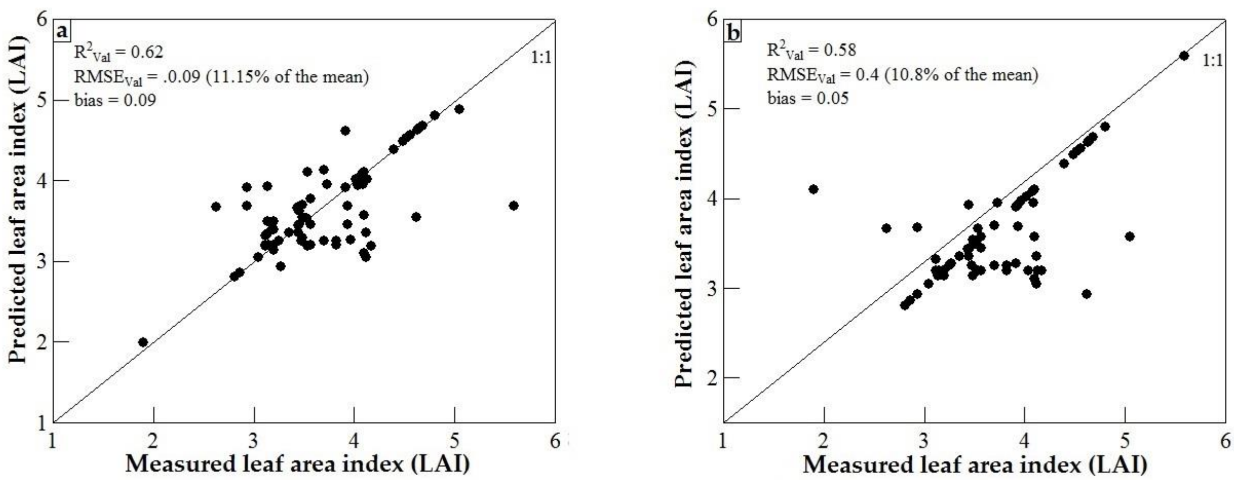


Figure 4.9: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using combined intact indigenous forest data and (a) support vector machines and (b) artificial neural networks

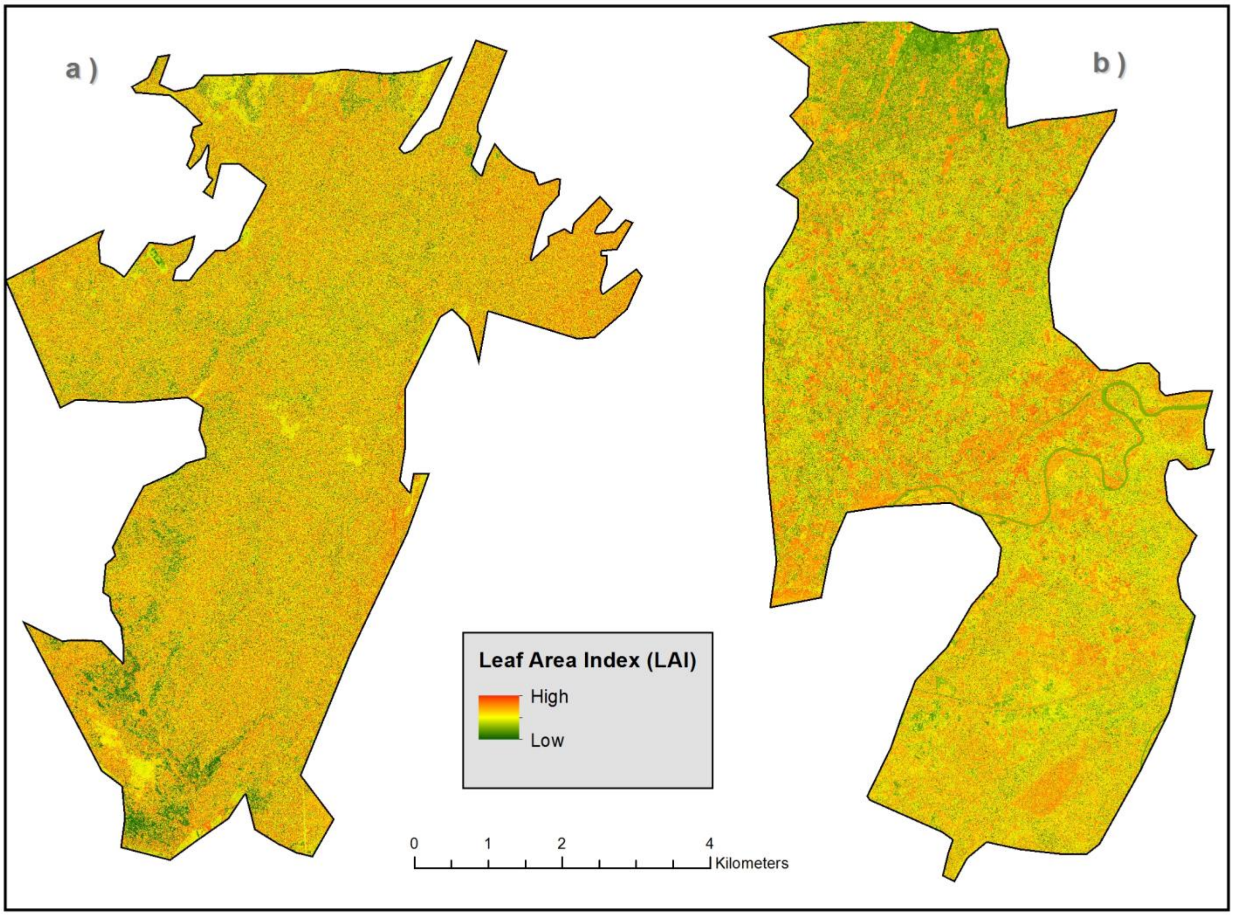


Figure 4.10: Leaf area index predicted map of the indigenous Dukuduku forest area. The map was produced using the support vector regression algorithm and a) the intact forest data and b) the fragmented forest data

4.5 Discussion

The present study aimed at testing the relationships between LAI of six endangered tree species and SVIs derived from WorldView-2 imagery captured from fragmented and intact indigenous forest ecosystems. The machine learning SVM and ANN regression algorithms were tested for deriving predictive models that can accurately estimate the LAI of the six endangered tree species. The study also tested the null hypothesis that LAI of the six endangered tree species grown in the intact and fragmented indigenous forest ecosystems are not significantly different. The use of the optical sensors with different types of spectral and spatial resolution (i.e. fine, medium and coarse) has attained different degrees of success for LAI estimation (Abuelgasim *et al.*, 2006; Yang *et al.*, 2006; Leuning *et al.*, 2008). However, estimating LAI using SVIs such as

NDVI and SRI computed from medium-spatial-resolution multispectral satellites (10 m to 100 m) is constrained by an asymptotic relationship with LAI in densely vegetated areas like indigenous forest ecosystems (Mutanga and Skidmore, 2004b; Davi *et al.*, 2006). Therefore, the study explored the use of the fine spatial resolution WorldView-2 sensor and the inclusion of the additional red edge band in calculating the SVIs for estimating LAI at tree species level.

The study shows that LAI at individual tree species level can accurately be estimated in fragmented and intact forest ecosystems. These results are consistent with other studies that demonstrated the utility of WorldView-2 in predicting LAI at different spatial scales of a landscape (Pope and Treitz, 2013; Pu and Cheng, 2015; Tarantino *et al.*, 2015). These findings therefore support the assertion that the potential utility of WorldView-2 spectral variables offer improved predictions accuracy of vegetation biophysical characteristics such as LAI in indigenous ecosystems (Mutanga *et al.*, 2012; Pope and Treitz, 2013; Tillack *et al.*, 2014; Tarantino *et al.*, 2015). The successful use of WorldView-2 data in predicting LAI at tree species level could be due to the fine pixel size (2m) that is required to capture the spectral properties of each individual tree species. This is in agreement with the findings of other studies like Pu and Cheng (2015) who found that LAI predictive models generated using WorldView-2 data in a mixed forest ecosystem performed better than those derived using the relatively low spatial resolution, Landsat 5 TM data.

Although, all the 24 SVIs (Table 4.1) combined together were utilized to predict LAI of the endangered tree species, it hypothesized that the red edge band which was included in some of the SVI could have enhanced the performance of the LAI predictive models. The red edge band which is the inflection point in the slope that connects the reflectance in the red band in the NIR spectral range (Pu *et al.*, 2003; Mutanga and Skidmore, 2007; Herrmann *et al.*, 2010) is more sensitive to the vegetation biophysical properties like chlorophyll content as compared to other regions of the electromagnetic spectrum (Pope and Treitz, 2013; Tillack *et al.*, 2014). These studies found that the vegetation indices calculated from the red edge and NIR have relatively stronger correlation with LAI in different landscapes. Chlorophyll content can be one of the vegetation biochemicals that has a direct relationship with LAI (Mutanga and Skidmore, 2004b; Mutanga and Skidmore, 2007; Herrmann *et al.*, 2010). In general, this result is in conformity with Mutanga *et al.* (2012) who concluded that the vegetation indices derived from WorldView-

2 data involving the additional red edge band can improve the predicting accuracies of vegetation biophysical properties (e.g. biomass) compared with the indices that only include conventional bands.

The tree LAI in the fragmented forest (open) was significantly higher than that in the intact forest, and the predictive models for estimating LAI in fragmented forest (Figures 4.4 and 4.6) outperformed those for predicting LAI in the intact forest (Figures 4.5 and 4.7). The higher LAI in the fragmented forest is expected since the trees are grown with relatively less competition from their surrounding plants compared with the trees in the intact forest and they could have optimally utilized the climatic and soil elements that are required for their growth. In the intact forest, the target endangered tree species could have been mixed with other tree species within the sample plots and could also have resulted in mixed spectral features (SVI). Hence, the mixed spectral features in the intact forest ecosystem might have hindered the performance of the LAI predictive models. Moreover, in a few cases in the intact forest ecosystem and due to the difficulty of taking measurements close to the stem area of the individual trees, the LAI measurements were collected from sites where two or more tree species were overlapped. The spectral features from the overlapped sites could also have resulted in mixed spectral features due to different trees structural and biophysical traits (Omer *et al.*, 2015a). Similarly, it is interesting to note that in the intact forest stratum, the study sampled trees along the roads and open paths, hence it is expected that one side of trees could have received more sunlight than the other side of the trees. Since the LAI is a light-dependent biophysical trait, the variation in light alongside tree crowns could have confounded the prediction of LAI in the intact forest and the performance of the models developed when intact forest data were utilized.

The study utilized two optimized learning nonlinear regression methods (SVM and ANN) to estimate LAI in the Dukuduku indigenous forest at tree species level. Tree biophysical parameters in such a complex and dynamic natural ecosystem might possibly not be modeled using a linear relationship. The nonlinear SVM and ANN regression approaches explained the high variability in the trees' LAI in the complex Dukuduku landscape and resulted in predictive models of relatively high accuracy. The study also parametrized the two regression approaches to get the best meta parameters for estimating LAI (Cherkassky and Ma, 2004; Verger *et al.*, 2008). The results showed that different optimal parameters were required to estimate LAI in the fragmented and intact forest ecosystems. This was expected since the study employed empirical statistical approaches for deriving the predictive models under two different forest ecosystems. This result is in conformity with other studies that reported different optimal settings for SVM and ANN under different levels and complexities of landscapes (Camps-Valls *et al.*, 2006; Verger *et al.*, 2008; Lottering and Mutanga, 2012; Verrelst *et al.*, 2012b; Liang *et al.*, 2013).

Furthermore, the study shows that the LAI predictive models derived using SVM regression performed relatively better than those derived using ANN regression. Other studies also noted the superiority of SVM models for predicting forests and crops LAI (Camps-Valls *et al.*, 2006; Pope and Treitz, 2013; Tillack *et al.*, 2014). The superiority of SVM models for predicting endangered tree species LAI when compared with the ANN models could also be due to the fact that SVM regression usually makes use of structural minimization principle which is known to have the ability to produce accurate predictive models (Cristianini and Shawe-Taylor, 2000; Camps-Valls *et al.*, 2006; Akande *et al.*, 2014). Meanwhile, ANN regression approach employs model functions like radial basis function that are relatively biased when used with input remotely-sensed variables and can deviate from what has been presented during the training stage (Kimes *et al.*, 1998; Atzberger, 2004; Baret and Buis, 2008). Furthermore, ANN regression is often referred to as a black-box technique that could encounter an overfitting problem on the test dataset (Kimes *et al.*, 2000; Qiu and Jensen, 2004). ANN also requires a relatively long processing time during the training phase due to manual adjustments of the hidden layers nodes. However, SVM was optimized using a 10-fold cross validation method, while ANN optimal parameters were obtained using a trial and error approach. Further studies should employ the same method to calibrate and optimize SVM and ANN regression methods when they are compared for their performance in predicting forest biophysical traits.

Overall, the results are promising for accurate prediction of LAI at tree species level in the Dukuduku forest ecosystem. However, the results should be interpreted with some caution as snapshot data was used at specific environmental conditions and forest ecosystems. Further studies should explore the transferability of the present models to other points in space or time. The LAI estimates should also be utilized to study and model other forest biophysical (e.g. biomass, NPP) and metro-physiological (e.g. evapotranspiration) traits using process-based physical models.

4.6 Conclusions

The present study shows a successful application of high spatial resolution WorldView-2 data and the machine learning SVM and ANN regression methods for estimating LAI of six endangered tree species in fragmented and intact Dukuduku indigenous forest ecosystems in South Africa. The results showed that 60% ($R^2_{Val} > 0.60$) of the variation in LAI of the endangered tree species could be explained by the predictive models when data in the fragmented forest ecosystem were utilized. On the other hand, the results revealed that a maximum R^2_{Val} of 0.80 could be obtained for estimating the LAI of the endangered tree species in intact forest ecosystems. In general, LAI predictive models developed using the fragmented forest data performed more accurately ($RMSE_{val}$ ranged between 1.37% and 14.72% of the mean) compared with the models developed using the intact forest data ($RMSE_{val}$ ranged between 1.57% and 5.85% of the mean) and SVM regression approach achieved relatively more accurate LAI estimation models compared with ANN regression.

Overall, the successful application of the WorldView-2 data, SVM and ANN for predicting LAI of six endangered tree species in the Dukuduku indigenous forest could help in making informed decisions and policies regarding management, protection and conservation of these endangered tree species. The strength of the multispectral WorldView-2 data, however, needs to be further tested for other biophysical (e.g. biomass) and biochemical parameters (e.g. leaf N, C_N) in indigenous and tropical forest ecosystems within a heterogeneous landscape where the biophysical and biochemical traits are highly variable. The findings of this study, do however, provide the necessary insight and motivation to the remote sensing community, ecologists and forest managers to shift toward identifying the most suitable and readily available remotely-sensed necessary for reliable and accurate indigenous forest management and monitoring protocols particularly in a fragmented ecosystem.

4.7 Acknowledgments

This research was supported by the University of KwaZulu-Natal, South Africa and the University of Khartoum, Sudan. The author’s appreciation is extended to the R development core team for their open source packages for the statistical analysis. The author would also like to thank Mr. Romano Lottering for his great assistance during the data analysis.

CHAPTER FIVE

Mapping Leaf Nitrogen and Carbon Concentrations of Intact and Fragmented Indigenous Forest Ecosystems Using Empirical Modeling Techniques and WorldView-2 Data

This chapter is based on:

G. Omer, O. Mutanga, E.M. Abdel-Rahman, K. Peerbhay and E. Adam, "Mapping leaf nitrogen and carbon concentrations of intact and fragmented indigenous forest ecosystems using empirical modeling techniques and WorldView-2 data". Under preparation.

ABSTRACT

Forest nitrogen (N) and carbon (C_N) are among the most important biochemical components of tree organic matter, and the estimation of their concentrations can help to monitor the nutrient uptake processes and health of forest trees. Traditionally, these tree biochemical components are estimated using costly, labour intensive, and time-consuming. The use of very high spatial resolution multispectral data and advanced machine learning regression algorithms such as support vector machines (SVM) and artificial neural networks (ANN) provide an opportunity to accurately estimate foliar N and C_N concentrations over intact and fragmented forest ecosystems. In the present study, the utility of spectral vegetation indices calculated from WorldView-2 imagery for mapping leaf N and C_N concentrations of fragmented and intact indigenous forest ecosystems was explored using SVM and ANN regression algorithms. The study further tested whether there were significant differences in the leaf N and C_N concentrations between the intact and fragmented indigenous forest ecosystems. The study showed that the intact forest obtained significantly higher (p = 0.03) mean values for N as compared to the fragmented indigenous forest. There was no significant difference (p = 0.55) in the C_N mean concentration between the intact and fragmented indigenous forest strata. The results further showed that the foliar N and C_N concentrations could be more accurately estimated using the fragmented stratum data compared with the intact stratum data. Specifically, the results showed that the most accurate N predictions were achieved when the fragmented data and support vector machines were utilized (R^2_{Val} = 0.77, $RMSE_{Val}$ = 1.07% of the mean). In addition, the most accurate foliar C_N predictions were achieved for the fragmented data using the SVM regression method (R^2_{Va}l = 0.67, $RMSE_{Val}$ = 1.64% of the mean). Overall, SVM regressions achieved more accurate models for estimating forest foliar N and C_N concentrations in the fragmented and intact indigenous forests compared to the ANN regression method. It is concluded that the successful application of the WorldView-2 data integrated with SVM can provide an accurate framework for mapping the concentrations of biochemical elements in two indigenous forest ecosystems.

Keywords: Intact indigenous forest, fragmented indigenous forest, nitrogen, carbon, WorldView-2, support vector machine, artificial neural networks

5.1 Introduction

Indigenous forests are a source of valued resources that highly contribute to rural communities in southern Africa (Shackleton and Shackleton, 2004; Eldeen and van Staden, 2007; van Wyk, 2008). Indigenous forests play a vital role in the nutrient and carbon cycling of ecosystems. Leguminous forest species, for example, provide a substantial amount of nitrogen (N) to the other flora habitats, and contribute to preventing soil erosion (Vitousek and Sanford, 1986; Eldeen, 2005; Brendler *et al.*, 2010). The degradation of indigenous forest ecosystems is a significant contributing factor to climate change since expanding carbon (C_N) storage in indigenous forests has been identified as a potential measure to mitigate global warming (DeFries *et al.*, 2000; de Chazal and Rounsevell, 2009). Different acreage of forest cover store different amount of C_N and the changes in forest cover can be used to monitor the changes in sequestrated C_N (Mushtaq and Malik, 2014). Estimates of the leaf chemistry of key tree species in indigenous forests allow for a better understanding of ecosystem functioning, when many biochemical processes, like photosynthesis, respiration and litter composition, are related to the chemical composition of tree species. The N and C_N concentrations are considered to be among the most important chemical components of green foliage. Foliar N is closely related to the rate of maximum photosynthetic and can help measure ecosystem productivity (Huber *et al.*, 2008). Similarly, C_N is closely related to biomass and may assist in managing critical indigenous forest resources sustainably. One unique fragmented forest in KwaZulu-Natal is the Dukuduku forest. This forest provides basic resources, including medicinal products to the surrounding communities and valuably contributes to the ecosystem services concept (Shackleton and Shackleton, 2004; Eldeen and van Staden, 2007; van Wyk, 2008). Dukuduku indigenous forest provides different products and usable materials for human needs such as fence poles and construction, raw material, and livestock browsing (Eldeen, 2005; Cho *et al.*, 2012; Mlambo, 2013). The forest is also one of the best preserved remnants of coastal forests in the country (van Wyk *et al.*, 2006). Despite the aforementioned importance, the coastal forest species in the Dukuduku area face growing threats and pressure by the rapid growth of informal human settlements and agricultural activities. Moreover, the spatial coverage in the Dukuduku indigenous forest is being constantly reduced by the extreme variability of climate and other intensive land use interventions (van Wyk *et al.*, 2006; Ndlovu, 2013).

As one of the critically fragmented forest landscapes in South Africa, Dukuduku requires careful monitoring, management and conservation. Therefore, a multi-disciplinary study is required to optimally manage and conserve the endangered tree species in these indigenous forests that support the livelihoods of millions of people (Eldeen, 2005; Cho *et al.*, 2013). One of the best ways to improve the management and monitoring of indigenous forest ecosystems is to estimate certain tree structural (e.g. tree diameter), biophysical (e.g. LAI) and biochemical (e.g. foliar N and C_N concentrations) attributes. Generally, these tree characteristics are proxies for ecosystem resilience, services, conservation and forest health. For example, N is taken up by plants in the form of nitrates which are used in the synthesis of components that include chlorophyll, C_N fixing enzyme ribulose biphosphate carboxylase and inert structural components in cell tissue (Cho *et al.*, 2013). On the other hand, a large amount of C_N is allocated to cellulose (65%) and lignin (20%) (Elvidge, 1990), which are intensely impacted on forest growth, and the terrestrial ecosystem C_N cycle (Malhi *et al.*, 2011; McMurtrie and Dewar, 2013). In this context, indigenous forests constitute one of the main sinks of atmospheric carbon dioxide (CO_2), and take up between 15% to 25% of annual global green gas emissions (Jensen *et al.*, 1999). The impact of forest disturbance on N and C_N stocks at the different landscape levels has been widely studied. For instance, a recent study by Cho *et al.* (2013) asserted that the forest foliar N stock in fragmenting ecosystems is affected by the conversion of intact forests into grassland. This conversion reduces the leaf N stock in the landscape. In addition, the ecological consequences of forest fragmentation include C_N sequestration and canopy nutrient stocks (Vasconcelos and Luizão, 2004; Broadbent *et al.*, 2008). However, from the available literature and to the researcher's knowledge, no study has estimated and mapped forest leaf N and C_N concentrations at forest stratum level, especially within fragmented and intact forest ecosystems. Estimating and mapping leaf N and C_N concentrations at forest stratum level could be utilized by forest and resource managers to understand and manage the functioning, health and the changes experienced by the indigenous forest ecosystems.

Traditional measurements of indigenous forest biochemical traits involve costly, laborious and time-consuming analytical approaches which are spatially constrained over large areas. Hence, studies sought complementary approaches that use rapid, up-to-date, and cost-effective for estimating and modeling forest foliar N and C_N concentrations (Adjorlolo *et al.*, 2013; Cho *et al.*, 2013; Mutanga *et al.*, 2015). The use of remotely-sensed data provide rapid and synoptic approaches for estimating forest foliar biochemical contents (Huber *et al.*, 2008; Cho *et al.*, 2013). Previous studies achieved acceptable accuracies for estimating forest leaf N and C_N concentrations using empirical approaches and SVIs calculated from multispectral (Hansen and Schjoerring, 2003; Zhang *et al.*, 2006) and hyperspectral (Foody *et al.*, 1996; Ferwerda *et al.*, 2005; Siegmann *et al.*, 2013) datasets. Most of the aforementioned studies have employed hyperspectral data for mapping foliar N and C_N concentrations in plantation forestry. However, very little attention has been given to estimating foliar N and C_N concentrations in different indigenous forest ecosystems (e.g. fragmented versus intact forest strata). In general, the operational use of hyperspectral data for estimating vegetation biochemical components is constrained by the availability, quality, and cost of airborne and spaceborne hyperspectral images (Mutanga *et al.*, 2012). The high dimensionality and co-linearity associated with hyperspectral datasets also limit their use for estimating vegetation biochemical traits. Estimating biochemical parameters at varying forest ecosystem levels could help resource managers to understand the impact of various socio-ecological mechanisms on indigenous forest species and the vulnerability of these ecosystems to external and internal perturbations.

As an alternative data source, the recent improvement of multispectral satellites like Sentinel-2, WorldView-3 and WorldView-2, are designed with additional bands that offer a great opportunity for estimating forest foliar N and C_N concentrations. In particular, these new satellites provide high spatial resolution imagery that could be suitable for estimating forest foliar N and C_N concentrations at ecosystem level. The use of WorldView-2 data, for instance, has provided accurate vegetation biochemical estimates (Gobron *et al.*, 2000; Mutanga *et al.*, 2015). As mentioned above, WorldView-2 imagery has a reasonable number of spectral wavebands that are configured within distinguishable portions of the electromagnetic spectrum to overcome the limitations of the conventional bands of other multispectral sensors like QuickBird, SPOT, and Landsat, while reducing the unnecessary redundancy as contained in the hyperspectral data (Adjorlolo *et al.*, 2013; Ramoelo *et al.*, 2014; Mutanga *et al.*, 2015). The additional WorldView-2 bands like the red edge are known to have a positive relationship with chlorophyll and N concentrations (Cho and Skidmore, 2006). Furthermore, the researchers have commonly focused on use of SVIs which combine the advent of two or more bands as opposed to the use of spectral features at a single band on modeling forests biochemical parameters (Peluelas *et al.*, 1995; Ferwerda *et al.*, 2005; Cho and Skidmore, 2009; Adjorlolo *et al.*, 2013).

In general, previous studies have either employed conventional (e.g. multiple linear regressions) or advanced machine learning (e.g. RF) regression approaches for estimating forest biochemical components (Peterson *et al.*, 1988; Serrano *et al.*, 2002; Karimi *et al.*, 2008; Ramoelo *et al.*, 2014; Mutanga *et al.*, 2015). However, the major challenge with the conventional empirical methods is that they assume a normal distribution on the response variables and suffer from multi-collinearity (Cohen *et al.*, 2003). The use of advanced machine learning regression methods has therefore been regarded as efficient and robust protocols for estimating forest biochemical components in the field of remote sensing (Huang *et al.*, 2004; Thissen *et al.*, 2004; Wang *et al.*, 2009; Liu *et al.*, 2013). Particularly, these methods do not require any statistical assumptions and efficiently handled multi-collinearity in the input predictor variables. The present study hypothesized that there is a lack of knowledge on whether or not high resolution WorldView-2 multispectral data with additional bands could be employed for estimating forest foliar N and C_N concentrations of different indigenous forest ecosystems (e.g. fragmented and intact forest ecosystems). Moreover, forest ecologists might need to test whether there are significant differences between N and C_N concentrations of tree species in intact and fragmented forest ecosystems.

The study also tested the use of the efficient and robust SVM and ANN regression algorithms for estimating forest N and C_N concentrations. SVM is a universal learning method, introduced by Vapnik (1995), which uses kernel functions to project the input data space to a high-dimensional feature space (Shao and He, 2011). The method is dependent on the structural risk minimization as an alternative of the empirical risk minimization, that can cause the solution to be captured in a local minimum and the networks overfitted. The structural risk minimization reduces the empirical error and model complexity simultaneously, which can improve the generalization ability of the SVM for regression problems in many applications. The SVM algorithm avoids overfitting and multi-dimensional problems when dealing with spectral data, and produces accurate forest biochemical estimates (Vapnik, 1995; Burges, 1998; Zhang *et al.*, 2008). The algorithm was applied to relate SVIs to various vegetation biochemical traits such as N and C_N (Ferwerda *et al.*, 2005; Tian *et al.*, 2011). On the other hand, the ANN algorithm comprises an interconnected group of artificial neurons and processes information using a connectionist method for calculation (Singh and Chauhan, 2009). The algorithm offers a very efficient tool to simulate the relationship between SVIs and biochemical permeates and has been utilized for estimating forest biochemical contents like N and C_N using remotely-sensed data (Jensen *et al.*, 1999; Wang *et al.*, 2009; Liu *et al.*, 2010). However, to the best of the researcher's knowledge, few studies have been done to compare the performance of the SVM and ANN algorithms to estimate forest foliar N and C_N concentrations in tropical indigenous forest ecosystems. Therefore, this study aimed to test the utility of SVIs computed from WorldView-2 imagery for predicting foliar N and C_N concentrations of fragmented and intact indigenous forest ecosystems using the SVM and ANN machine learning regression algorithms. The study further tested whether there were significant differences between the tree foliar N and C_N concentrations between the intact and fragmented indigenous forest ecosystems.

5.2 Methodology

5.2.1 Field Data Collection

The field data collection was carried out between 1st and 7th December 2013. The stratified purposive sampling approach was adopted to collect leaf samples from six tree species grown in fragmented and intact Dukuduku indigenous forest ecosystems. The six tree species were purposively selected for analysis in the present study. These trees are regarded as rare and endangered species because of their rapid harvesting and removal (van Wyk *et al.*, 2006). A measuring tape and compass, with a handheld Leica Geosystem GS20 GPS of sub-meter (0–0.25 m) accuracy (Geosystems, 2004), were used to geo-locate the sample trees. The leaf samples of the six tree species were collected across the study area and then clipped with a pair of hand scissors and sealed in brown paper bags. The leaf samples were then pooled across the tree species in each forest ecosystems; the fragmented and intact forest strata and then stored in a cool box for transportation and chemical analysis. In total, 170 samples were collected from both the intact (n = 85) and fragmented (n = 85) indigenous forest strata (Figure 5.1).

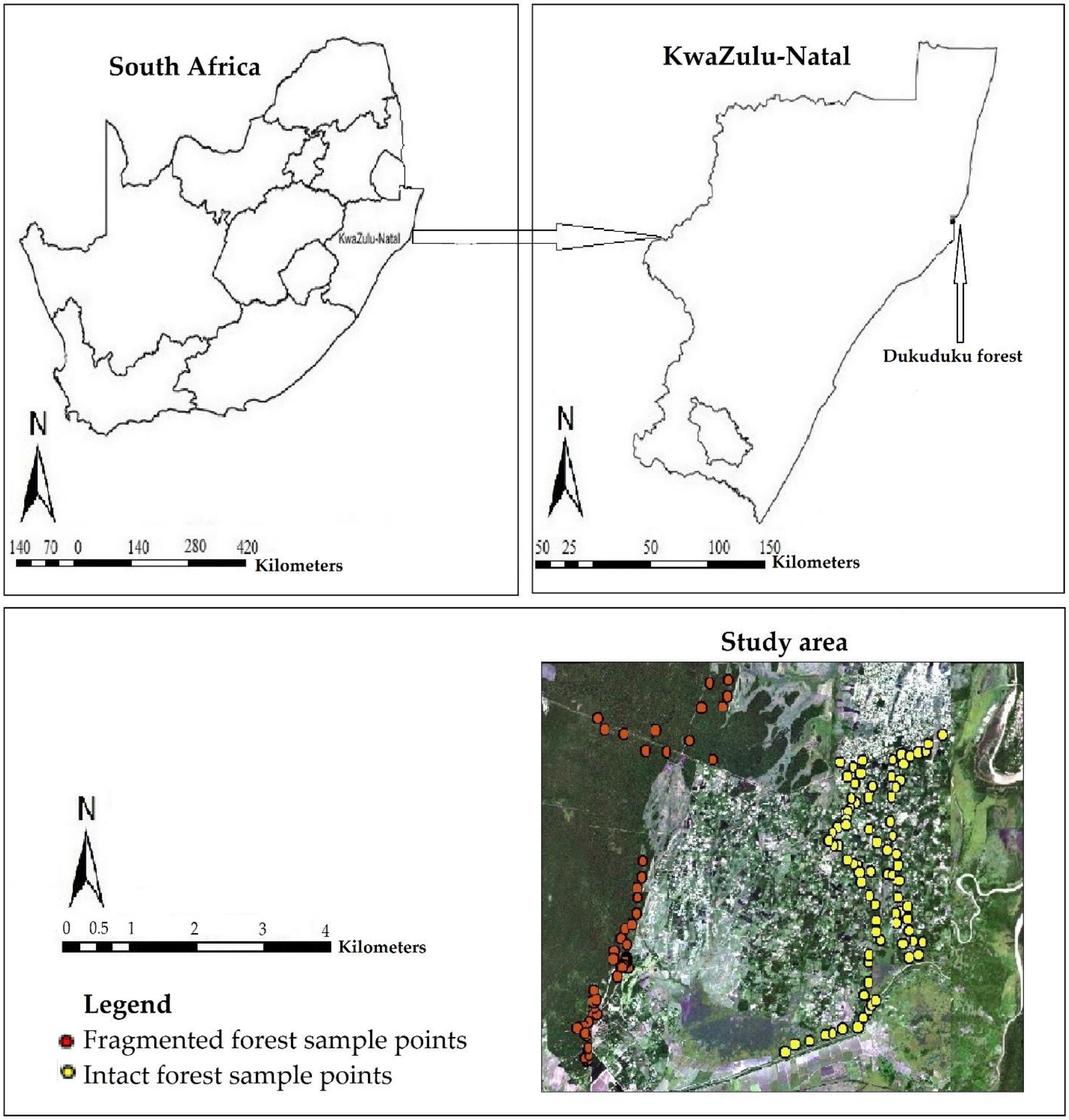


Figure 5.1: The location of the Dukuduku indigenous forest in KwaZulu-Natal province, South Africa and field sample locations overlaid in a true-color WorldView-2 image

5.2.2 Chemical Analysis

The harvested leaf samples of each forest stratum were subjected to chemical analysis following the Kjeldahl procedure (Kjeldahl, 1883). The samples were first oven-dried at 70°C for 48 hours and then mill-crushed to about 1 mm. The dried and crushed samples were then processed for complete feed analysis to estimate the N and C_N concentrations in percentage. For a detailed description of Kjeldahl procedure, the readers are referred to, for example, Kjeldahl (1883), Peluelas *et al.* (1995) and Sáez-Plaza *et al.* (2013).

5.2.3 Satellite Image Acquisition and Pre-Processing

A cloud-free multispectral WorldView-2 imagery covering the study area was acquired on 1st December 2013. The WorldView-2 image consists of eight wavebands, spanning the wavelength range of 400-1040 nm with a spatial resolution of 2 m and swath width of 16.4 km at nadir (DigitalGlobe, 2010). The spectral ranges of the WorldView-2 eight bands include four standard bands (blue: 450 - 510 nm, green: 510 – 580 nm, red: 630–690 nm, and near infrared-1: 770–895 nm) and four additional bands (coastal blue: 400 – 450 nm, yellow: 585-625 nm, red edge: 705–745 nm, and a new near infrared-2: 860–1040 nm) (DigitalGlobe, 2010). The sensor has the spectral and spatial resolutions that meet many forest applications such as predicting and monitoring forest biochemical variables at forest ecosystem level (Mutanga *et al.*, 2015). The image was atmospherically corrected and transformed at canopy reflectance using the QUAC procedure in ENVI 4.7 software (ENVI, 2009). QUAC performs in-scene based atmospheric correction at the visible and near- to-shortwave infrared (VNIR-SWIR) region of the electromagnetic spectrum for multi-and hyperspectral imagery (Shen *et al.*, 2005). The acquired image was already geometrically corrected by DigitalGlobe™.

5.2.4 Spectral Vegetation Indices (SVIs) Derived from WorldView-2 Data

After the WorldView-2 image was processed, 24 SVIs were calculated (Table 4.1 in Chapter 4) and utilized to predict the leaf N and C_N concentrations of the pooled tree species data across the fragmented and intact indigenous forest ecosystems. These indices are known features for vegetation chlorophyll and other biochemical properties (Haboudane *et al.*, 2002; Wu *et al.*, 2008; Adjorlolo *et al.*, 2013; Mutanga *et al.*, 2015).

5.3 Statistical Analysis

5.3.1 Descriptive Statistics and Independent *t*-test

The Shapiro-Wilk test (Royston, 1982) was used to test the normality in the response variables (e.g. foliar N and C_N concentrations). An independent *t*-test was then used with 95% confidence level ($p \leq 0.05$) to test if there were significant differences in the leaf N and C_N concentrations between the intact and fragmented indigenous forest strata.

5.3.2 Support Vector Machines (SVM) Algorithm

SVM is a relatively new learning system based on statistical learning theory (Vapnik, 1995; Durbha *et al.*, 2007). SVM can be observed as the same type of networks corresponding precisely to the same type of solution but trained in a different way and then with different values of the weight after the training (Vapnik, 1995; Gilabert *et al.*, 2002; Durbha *et al.*, 2007). The method is characterized by the usage of kernels, absence of local minima, and sparseness of the solution and capacity control obtained by acting on the margin, or on support vectors number (Karamouz *et al.*, 2009). Initially, SVM was developed to solve classification problems and later extended to handle regression problems (Cortes and Vapnik, 1995; Marabel and Alvarez-Taboada, 2013). The SVM regression approach mostly converts the nonlinear regression problem into a linear relationship using the kernel functions (Cortes and Vapnik, 1995; Chen *et al.*, 2012). In particular, the goal of SVM is to estimate an unknown continuous-valued function based on a finite number of noisy samples (Marabel and Alvarez-Taboada, 2013). Basically, it makes use of structural minimization principle which is known to have accurate generalization performance for different datasets size as contrasted to empirical risk minimization employed by other approaches like ANN (Camps-Valls *et al.*, 2006; Akande *et al.*, 2014). The SVM algorithm uses two meta-parameters (ε-SVM and C-SVM) to optimize the method. The performance of the SVM therefore depends on the correct setting of these two meta-parameters (Cherkassky and Ma, 2004). Specifically, the parameter ε-SVM controls the width of the epsilon-insensitive zone to fit the training dataset (Kohestani and Hassanlourad, 2015). Hence, the value of ε-SVM can affect the number of support vectors to make the regression function. In other words, the bigger the epsilon, the fewer support vectors are selected (Cherkassky and Ma, 2004; Kohestani and Hassanlourad, 2015). On the other hand, a bigger ε-SVM values result in more ‘flat’ estimates, while the C-SVM determines the balance between the model complexity and the degree to which the larger deviations (than epsilon) are tolerated in the optimization (Marabel and Alvarez-Taboada, 2013). Therefore, the larger values of *C*-SVM aim at minimizing the empirical risk regardless of the complexity of the model. In this regard, both ε-SVM and *C*-SVM values affect the model complexity.

In the present study, the SVM algorithm was used to estimate the foliar N and C_N concentrations of the fragmented and intact ecosystems, using the Vapnik's ε-insensitive loss function to minimize the training errors (Marabel and Alvarez-Taboada, 2013). In order to project the data into a new space, a radial basis function was used, followed by an optimization procedure to find the number of support vectors for the best performance (Hsu *et al.*, 2009; Richter *et al.*, 2011). Moreover, the optimal values of the *C*-SVM and ε-SVM parameters of the radial basis function was performed using a 10-fold cross validation method and grid search on the training dataset (Hsu *et al.*, 2009; Shao and He, 2011). The training dataset was divided into 10 subsets of equal size, whereby SVM models were then trained on the 9 subsets samples, and tested on the removed one and the process was repeated 10 times until all subset samples had served as test samples. The pair parameter that minimizes the prediction error was then considered as the best values for the final prediction performance. The analysis was carried out using the e1071 library version 2.15.2 in R statistical packages (R Development Core, 2015).

5.3.3 Artificial Neural Networks (ANN) Algorithm

ANN regression algorithm which was introduced by Atkinson and Tatnall (1997) is a nonparametric machine learning procedure that is suitable for analyzing complex relationships. ANN is a computation model that tries to simulate the structure and functional features of biological neural connections. It comprises of an interconnected group of artificial neurons and processes information using a connectionist method for calculation. Various models of ANN algorithm (e.g. multilayer perceptron, radial basis function, and back propagation) have been used for analyzing remotely-sensed data for a variety of applications like forestry modeling (Atkinson and Tatnall, 1997; Foody, 2004b; Wang *et al.*, 2009; Liu *et al.*, 2013; Omer *et al.*, 2015a). Radial basis function neural network has been demonstrated to be an accurate function for analyzing a large variety of remotely-sensed data since it reduces the computational time required for the training process (Foody, 2004b; Boegh *et al.*, 2013). The approach needs one input variable which is the 'distance' between the weight and input nodes. The back propagation is a multilayer feed forward neural networks method which comprises of a series of simple connected nodes, or neurons between input and output layers (Atkinson and Tatnall, 1997), while the multilayer perceptron is a commonly used ANN structure that consists of an input layer, an output layer and one or more hidden layers of nonlinearly-activating nodes (Atkinson and Tatnall, 1997; García Nieto *et al.*, 2012). The nodes are connected by a certain synaptic weight to all nodes in the next layer and the perceptron learning occurs through changes in the networks weights after the input protector (SVIs in this study) are processed (García Nieto *et al.*, 2012). The multilayer perceptron is a feed forward ANN model that projects input data onto a set of suitable output by using three or more layers of nodes with nonlinear activation functions (Atkinson and Tatnall, 1997; Xiong *et al.*, 2010). ANN algorithm has been extensively used in modeling vegetation traits such as biochemical parameters (N and C_N) that are not linearly predictable in the original remotely-sensed variables (Papale and Valentini, 2003; Melesse and Hanley, 2005; Wang *et al.*, 2009; Liu *et al.*, 2013).

In the current study, ANN regression algorithm using the multilayer perceptron modeling approach was performed to estimate leaf N and C_N concentration of intact and fragmented forest ecosystems using the 24 SVIs derived from WorldView-2 imagery. Various trials of internal networks structure, input data, and learning algorithms have been tested to define the optimal regression features. The structure of the hidden layers was tested to assess the necessary number of hidden layers and the number of required nodes per layer. This was used by manually changing the number of nodes in the hidden layer. The ANN was then trained with a back propagation learning algorithm and one hidden layer to model the respective biochemical parameters.

5.3.4 Validation

To test the performance of the SVM and ANN models, the reference data were randomly split into 70% (59) for calibration and 30% (26) for validation according to the recommendation made by Adelabu *et al.* (2015). One-to-one relationships between the measured and predicted foliar N and C_N concentrations were fitted and R^2_{Val}, $RMSE_{Val}$ and bias were then calculated (see Equations 4.1, 4.2, and 4.3 in Chapter 4).

5.4 Results

5.4.1 Descriptive Statistics and Independent *t*-test

The results obtained from Shapiro–Wilk normality test indicated that the leaf N and C_N concentrations in the fragmented and intact indigenous forest strata were normally distributed (p = 0.05 for the intact forest stratum and p = 0.04 for the fragmented forest stratum). Table 5.1 shows the descriptive statistics of the fragmented and intact indigenous forest leaf N and C_N concentrations. The independent *t*-test showed that the intact forest stratum obtained significantly higher (p = 0.03) mean foliar N value compared to the fragmented indigenous forest stratum. There was no significant difference (p = 0.55) in the mean foliar C_N concentration between the intact and fragmented indigenous forest strata. The higher mean values for foliar N (1.84%) and C_N (45.16%) concentrations were obtained for the intact forest ecosystem, while the least mean values for N (1.78%) and C_N (44. 95%) concentrations were obtained for the fragmented forest stratum (Table 5.1). Furthermore, the descriptive statistics of the combined (aggregated) N and C_N across the two forest strata is also shown in Table 5.1.

Table 5.1: Descriptive statistics of the measured leaf nitrogen (N) and carbon (C_N) concentration (%) obtained from intact and fragmented indigenous forest strata as well as the combined stratum data. Means with the same letter are not significantly different ($p \geq 0.05$) from each other according to the independent *t*-test

			Nitrogen (N)			
Forest stratum (ecosystem)	**No. of sample**	**Mean**	**Min**	**Max**	**SD**	***P* value**
Intact	85	1.84a	0.27	4.42	0.89	0.03
Fragmented	85	1.78b	0.23	3.99	0.87	
Combined	170	1.81	0.23	4.42	0.88	
			Carbon (C_N)			
Intact	85	45.16a	38.60	49.73	1.99	0.55
Fragmented	85	44.95a	38.23	50.77	2.58	
Combined	170	44.88	38.60	50.77	2.07	

SD = Standard deviation

5.4.2 Support Vector Machines (SVM) and Artificial Neural Networks (ANN) Regression Models

The optimum parameters obtained for both SVM and ANN regressions are shown in Table 5.2. The 10-fold cross validation and grid search approach results in optimal *ε*-SVM and *C*-SVM values of 1 and 100, respectively for N and C_N in both the intact and fragmented strata, except for C_N (1 and 1000) in the intact forest stratum, and N and C_N when the combined data were used (1 and 10). Table 5.2 also shows that the input layers for ANN algorithm ranged between 8 and 10 for the intact, fragmented and combined data when N concentration was estimated, 9 for intact and 11 for both fragmented and combined data when C_N was estimated, while the number of hidden layers varied between 3 to 5 for both N and C_N in the two forest strata (ecosystems).

Table 5.2: The optimal parameters for the best trained SVM and ANN models used for estimating the nitrogen (N) and carbon (C_N) concentrations of the fragmented and intact indigenous forest strata as well as combined data

Regression model						
Support vector machines (SVM)						
Forest stratum (ecosystem)	**Nitrogen (N)**			**Carbon (C_N)**		
	***ε*-SVM**	***C*-SVM**		***ε*-SVM**	***C*-SVM**	
Intact	1.0	100		1.0	1000	
Fragmented	1.0	100		1.0	100	
Combined	1.0	10		1.0	10	
Artificial neural networks (ANN)						
	Nitrogen (N)			**Carbon (C_N)**		
	inputs	**Hidden**	**Profile**	**inputs**	**Hidden**	**Profile**
Intact	08	3	MLP 8:8-3-1:1	09	5	MLP 9:9-5-1:1
Fragmented	10	4	MLP 10:10-4-1:1	11	4	MLP 11:11-4-1:1
Combined	09	4	MLP 9:9-4-1:1	11	5	MLP 11:11-5-1:1

ε-SVM = ε-insensitive zone, *C*-SVM = regularization parameter, MLP = multilayer perceptron

The results of calibrating both the SVM and ANN regression approaches in intact, fragmented and combined data are presented in Table 5.3. Both the SVM and ANN models explained more than 80% of the variance ($R^2_{Cal} \leq 0.80$) in the forest foliar N concentration, except when the fragmented forest data and ANN regression method were utilized. For the forest foliar C_N concentration, the results showed R^2_{Cal} values ranged between 0.64 and 0.71 for SVM models, while the ANN models resulted in R^2_{Cal} range of 0.55 and 0.59 (Table 5.3). In general, the SVM models yielded relatively more accurate results for estimating forest foliar N and C_N concentrations compared to the ANN. On the other hand, models developed using the intact forest data fitted the training data more accurately compared with the models developed using the fragmented forest data (Table 5.3).

Table 5.3: Coefficient of determination (R^2_{Cal}) and root mean square errors ($RMSE_{Cal}$) for the SVM and ANN regression models when calibrated using the data collected from the fragmented and intact forest strata

	Nitrogen (N)					
Forest stratum (ecosystem)	**Support vector machines**			**Artificial neural networks (ANN)**		
	R^2_{Cal}	$RMSE_{Cal}$	$RMSE_{Cal}$%	R^2_{Cal}	$RMSE_{Cal}$	$RMSE_{Cal}$%
Intact data	0.93	0.03	1.23	0.89	0.05	1.94
Fragmented data	0.83	0.07	2.68	0.73	0.11	3.20
Combined data	0.93	0.04	1.47	0.91	0.04	1.32
	Carbon (C_N)					
Intact data	0.69	0.12	0.27	0.57	0.16	0.35
Fragmented data	0.64	0.17	0.45	0.55	0.29	0.66
Combined data	0.71	0.09	0.23	0.59	0.14	0.32

$_{Cal}$ = Calibration dataset

5.4.3 Model Validation

One-to-one relationships between measured and predicted forest foliar N and C_N concentrations for all the predictive models are shown in Figures 5.2, 5.3, and 5.4. N models derived using the fragmented data and SVM regression approach performed more accurately ($RMSE_{Val}$ = 1.07% of the mean) compared with models developed using the intact data and ANN regression methods ($RMSE_{Val}$ = 5.56% of the mean). In contrast, the ANN regression method achieved a more accurate model compared with SVM using the combined fragmented and intact data (Figure 5.4).

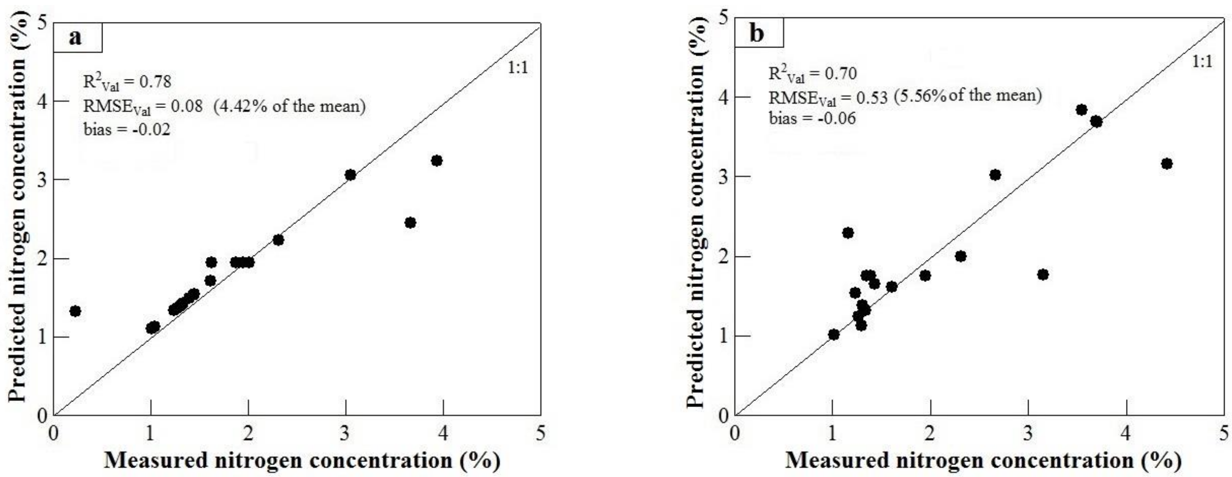


Figure 5.2: One-to-one relationships between measured and predicted forest leaf nitrogen (N) concentration (%) based on an independent validation dataset (30%) using the intact indigenous forest data and (a) support vector machines and (b) artificial neural networks

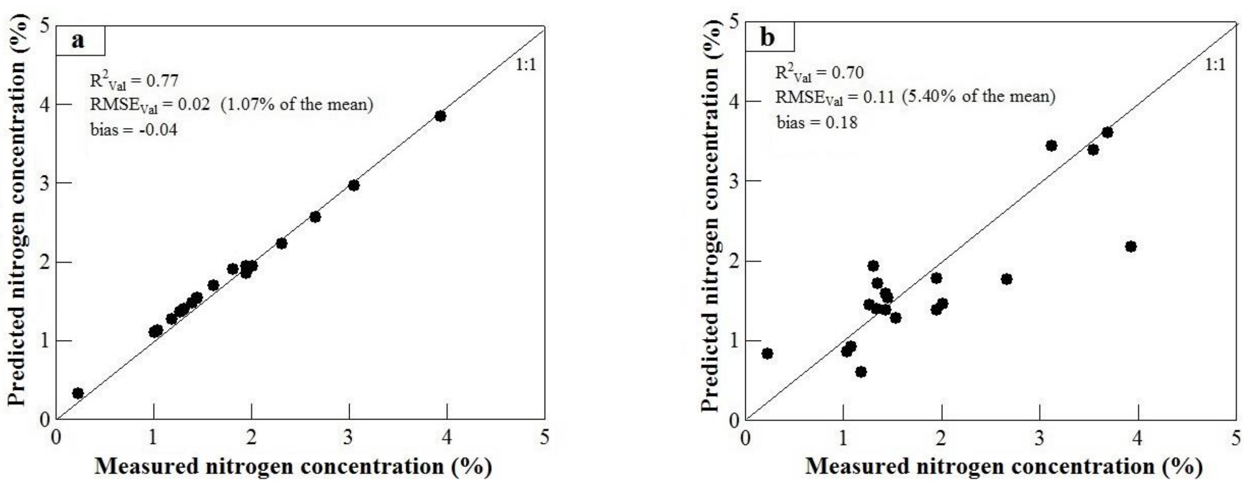


Figure 5.3: One-to-one relationships between measured and predicted forest leaf nitrogen (N) concentration (%) based on an independent validation dataset (30%) using the fragmented indigenous forest data and (a) support vector machines and (b) artificial neural networks

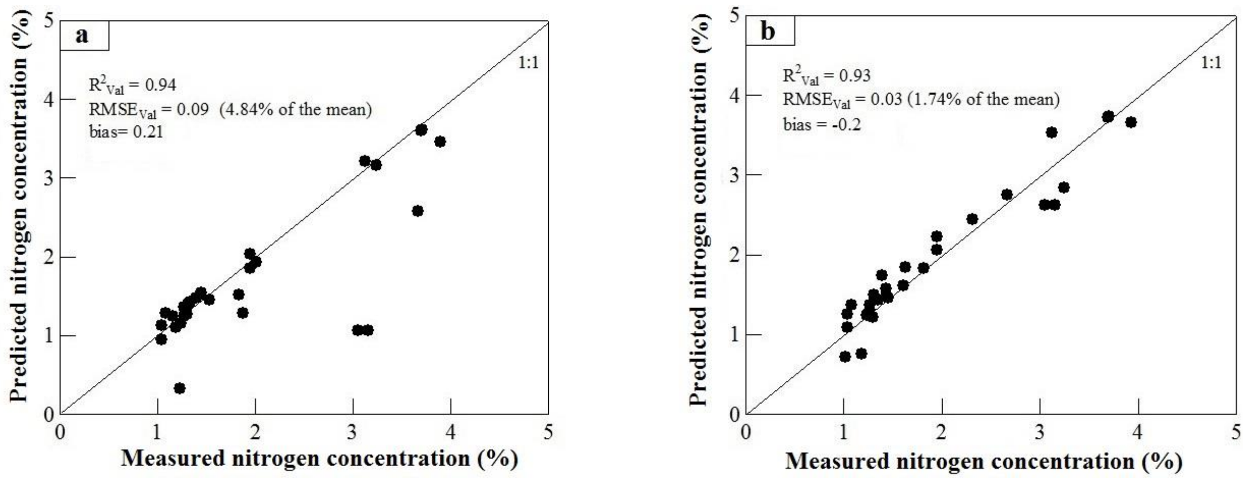


Figure 5.4: One-to-one relationships between measured and predicted forest leaf nitrogen (N) concentration (%) based on an independent validation dataset (30%) using the combined intact and fragmented data and (a) support vector machines and (b) artificial neural networks

Figures 5.5, 5.6 and 5.7 show the results for validating the forest foliar C_N estimation models. When the performance of the prediction models was assessed, the results showed that the forest foliar N and C_N concentrations could be better estimated in the fragmented forest stratum as indicated by the relatively higher R^2_{Val} (0.67), and lower error metrics ($RMSE_{Val}$ = 1.64% of the mean) (Figure 5.6a). The gradient in most of the other predictive models deviated from the expected one-to-one relationship and the models either overestimated or underestimated the forest foliar C_N estimates. Likewise, the SVM models for estimating forest foliar C_N concentrations outperformed the ANN ones.

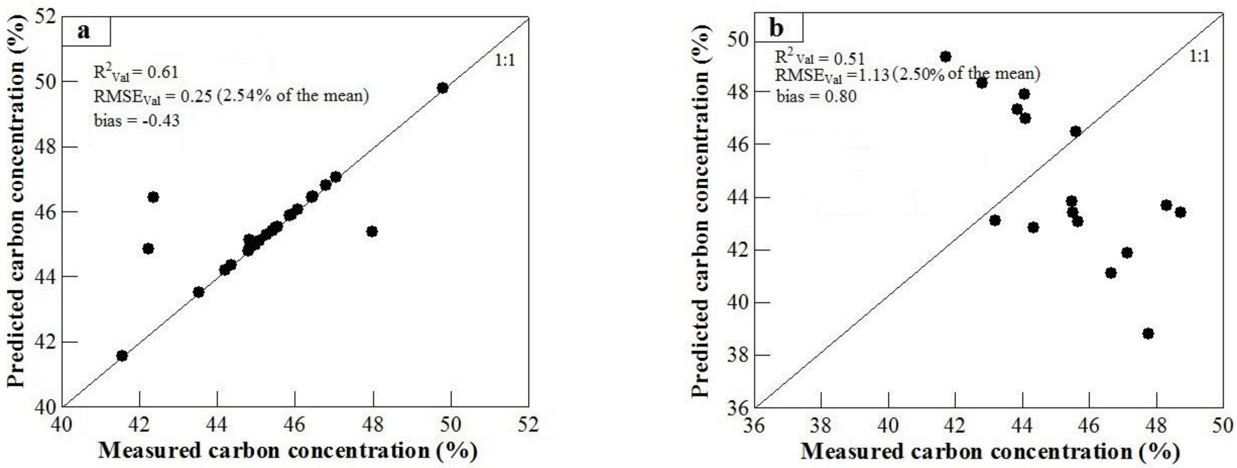


Figure 5.5: One-to-one relationships between measured and predicted forest leaf carbon (C_N) concentration (%) based on an independent validation dataset (30%) using the intact indigenous forest data and (a) support vector machines and (b) artificial neural networks

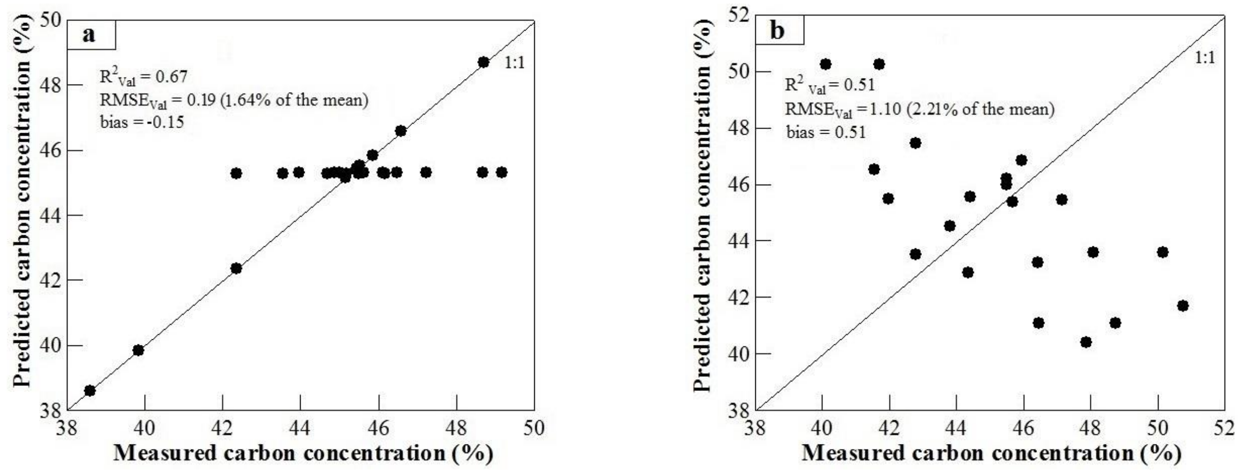


Figure 5.6: One-to-one relationships between measured and predicted forest leaf carbon (C_N) concentration (%) based on an independent validation dataset (30%) using the fragmented indigenous forest data and (a) support vector machines and (b) artificial neural networks

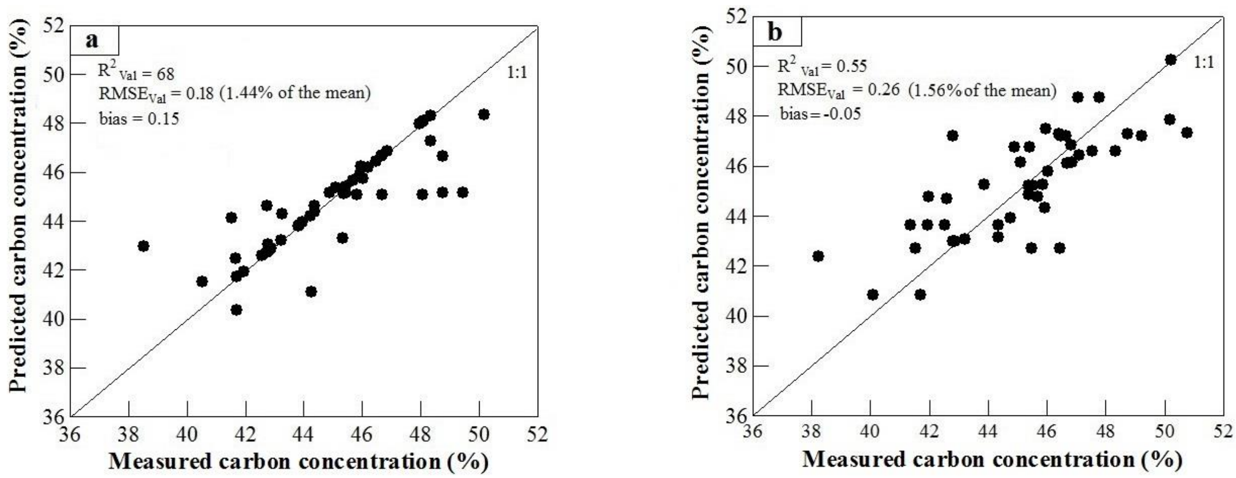


Figure 5.7: One-to-one relationships between measured and predicted forest leaf carbon (C_N) concentration (%) based on an independent validation dataset (30%) using the combined intact and fragmented data (a) support vector machines and (b) artificial neural networks algorithms

Since the SVM regression method produced the best results, predicted forest foliar N and C_N concentrations maps were produced using the WorldView-2 imagery and SVM regression algorithm (Figures 5.8 and 5.9). The maps can be compared to the WorldView-2 image which depicts the intact and fragmented forest areas. The maps show the spatial variation in the forest foliar N and C_N concentrations in both intact and fragmented forest ecosystems. In general, the forest foliar N and C_N spatial patterns are more compact in the intact forest compared with the fragmented forest and higher foliar N concentrations were observed towards the northern part of both forest landscapes (Figure 5.8). However, there is not a distinct trend in forest foliar C_N concentrations across the study area (Figure 5.9).

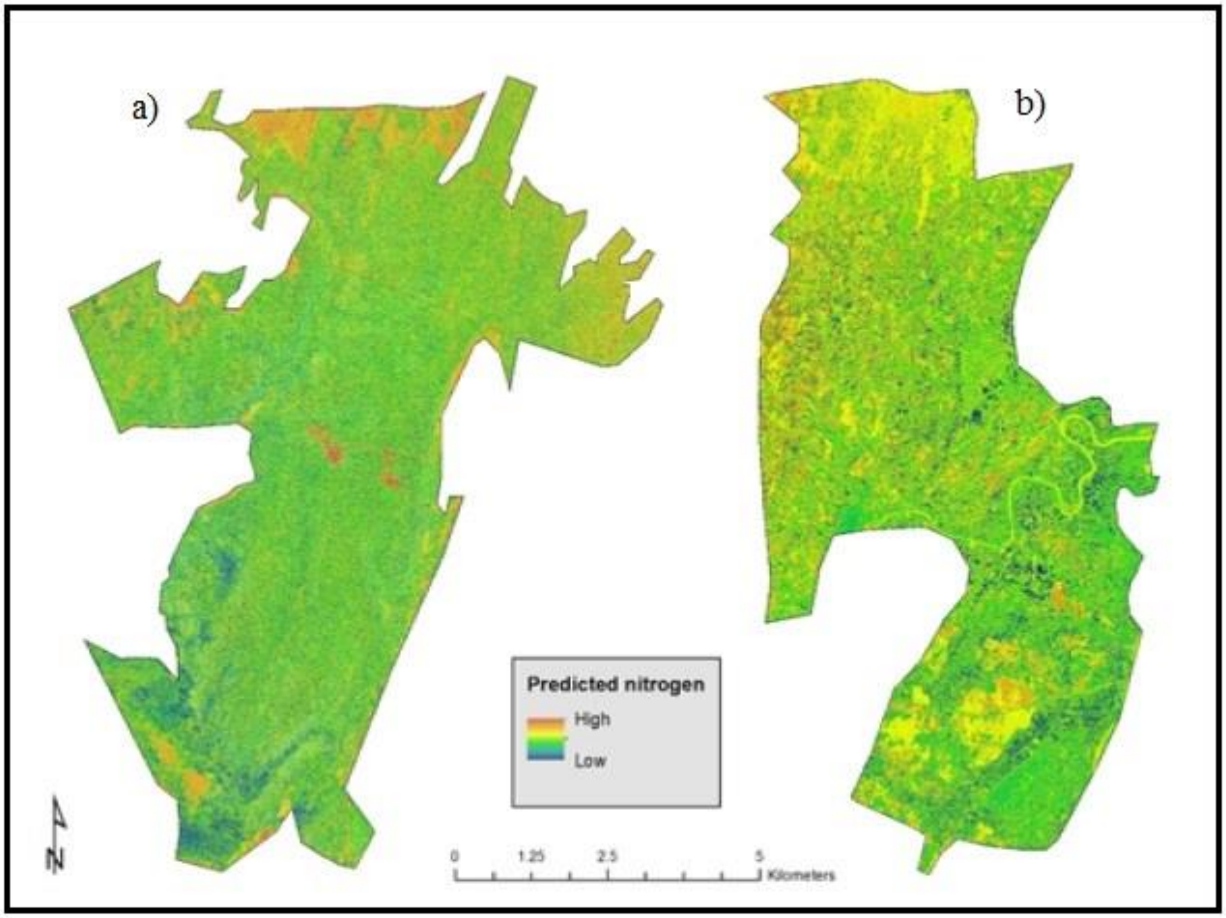


Figure 5.8: Forest foliar nitrogen (N) concentration (%) map of the indigenous Dukuduku forest area. The map was produced using the support vector machines regression algorithm and a) the intact forest data and b) the fragmented forest data

Figure 5.9: Forest foliar carbon (C_N) concentration (%) map of the indigenous Dukuduku forest area. The map was produced using the support vector machines regression algorithm and a) the intact forest data and b) the fragmented forest data

5.5 Discussion

The relationships between forest foliar N and C_N concentrations and SVIs calculated from the high spatial resolution WorldView-2 imagery acquired from fragmented and intact indigenous forest ecosystems was tested in the present study. The machine learning SVM and ANN regressions were employed for deriving predictive models that can accurately estimate the forest foliar N and C_N concentrations in the intact and fragmented forest ecosystems. The study also investigated the null hypothesis that forest foliar N and C_N concentrations in the intact and fragmented indigenous forest ecosystems were not significantly different.

The use of the optical sensors with different spectral and spatial resolutions for estimating plantation forests' and crops' foliar N and C_N concentrations has achieved different degrees of success (Martin and Aber, 1997; Abdel-Rahman *et al.*, 2010; Mutanga *et al.*, 2012). However, estimating foliar N and C_N concentrations using SVIs such as NDVI and SRI calculated from medium-spatial-resolution multispectral satellites (10 m to 100 m) is constrained by the lack of spectral information that can mimic the spectral properties of the N and C_N, particularly in densely vegetated and fragmented indigenous forest ecosystems (Kumar *et al.*, 2001; Miphokasap *et al.*, 2012). Also, the medium and coarse pixel size of such optical multispectral data might not capture the spectral responses of the target trees/crops or forest ecosystems as they could be mixed with other confound spectral responses within the pixels. This study, therefore, explored the use of the fine spatial resolution (2 m) multispectral WorldView-2 data which is characterized by the inclusion of some additional bands like yellow and red edge for estimating forest foliar N and C_N concentrations in fragmented and intact indigenous forest ecosystems. The study shows that the forest foliar N and C_N concentrations at each stratum can be accurately estimated. This result is in agreement with other studies that revealed the utility of WorldView-2 data in predicting biochemical concentrations like N and C_N in different types of savannah landscapes (Adjorlolo *et al.*, 2013; Zengeya *et al.*, 2013; Mutanga *et al.*, 2015). The successful application of WorldView-2 data for predicting N and C_N concentration at a forest strata (intact and fragmented) level could be due to the fine pixel size (2 m) of WorldView-2 that is needed to capture the spectral properties of each forest stratum. Moreover, the inclusion of the red edge band in calculating some of the SVIs could also have contributed to the derivation of such accurate forest foliar N and C_N estimates (Mutanga *et al.*, 2015). The red edge band which is the inflection point in the slope connecting the reflectance in the red and in the NIR spectral range is more sensitive to the vegetation biochemical properties like chlorophyll and N content as compared to other regions of the electromagnetic spectrum (Pu *et al.*, 2003; Mutanga and Skidmore, 2007; Herrmann *et al.*, 2010). In this regard, this finding is in conformity with Mutanga *et al.* (2015) who concluded that the vegetation indices calculated from the red edge band of WorldView-2 data can improve the prediction accuracy of vegetation biochemical properties compared with the indices calculated from the other conventional bands.

The study also shows that the leaf N and C_N concentrations in the fragmented forest ecosystem were significantly lower than those in the intact forest ecosystem. This finding is consistent with the findings by Cho *et al.* (2013) who concluded that indigenous forest fragmentation leads to significant losses in foliar N as most of the land use/cover classes, like agro-urban and grassland systems, resulting from forest fragmentation obtained least foliar N concentration when compared to the intact indigenous forest. In addition, the degraded forest patches and intensive cropping systems like maize and sugarcane in the fragmented stratum (Cho *et al.*, 2013) could have led to poor soil fertility and hence less available N to the forest ecosystem. Less available N can lead to a poor photosynthetic system and therefore less foliar C_N content in the fragmented ecosystem as opposed to the healthy intact forest ecosystem that could have had relatively more available soil N and an efficient photosynthetic system (i.e. higher foliar C_N content). Future studies should investigate the role of these indigenous forest ecosystems in the nutrition of the surrounding agro-ecosystems (e.g. farm and grass lands). However, the predictive models for estimating fragmented forest foliar N and C_N concentrations outperformed those for predicting intact forest foliar N and C_N concentrations. The species composition and tree density in the intact forest ecosystem could have been higher compared to the fragmented forest ecosystem. This could have confounded the performance of the intact forest foliar N and C_N predictive models, since the models were derived at forest ecosystem (stratum) level using tree level field samples. In particular, the field data were collected from only six target species and the predictive models were derived using the pooled leaf samples across the six tree species in each forest stratum. The spectral features of the target tree species and other tree species were found to be different from each other (Omer *et al.*, 2015a). These discrepancies in the performance of the two forest strata (fragmented and intact) predictive models are in accordance with the finding of Omer *et al.* (2016) who noted that models for estimating tree LAI in a fragmented indigenous forest ecosystem performed better than those developed using the intact indigenous forest data.

The study utilized two optimized learning nonlinear regression methods (SVM and ANN) to estimate forest foliar N and C_N concentrations in the Dukuduku indigenous forest ecosystem. Forest biochemical parameters in such a complex and dynamic natural ecosystem might possibly not be modeled in a linear relationship. The nonlinear SVM and ANN methods could have explained a great variability in the forest foliar N and C_N concentrations and resulted in predictive models of relatively better performance. The study also parametrized the two regression approaches to get the best generalization estimates of forest foliar N and C_N concentrations during the learning process (Cherkassky and Ma, 2004). The results showed that different optimal parameters were required to estimate the forest foliar N and C_N concentrations in the fragmented and intact forest ecosystems. These optimal parameters correspond to the best performance for the trained SVM and ANN regression models. That was expected since the study employed empirical statistical approaches for deriving the predictive models under two different forest ecosystems.

In general, it was found that the SVM models for estimating forest foliar N and C_N concentrations outperformed the ANN models in both the fragmented and intact forest ecosystems. This finding is consistent with some studies that reported the superior performance of SVM regression models for estimating vegetation biochemical contents compared with other machine learning regression approaches. For example, Karimi *et al.* (2008) noted that SVM regression is an efficient approach for projecting the input data to a higher dimensional space and commonly is not affected by the collinearity. Zhai *et al.* (2013) also reinforce the accurate performance of the SVM algorithm when compared with the partial least squares regression for predicting vegetation biochemical components. Moreover, SVM is considered as a solution for those cases in which a significant amount of nonlinear information is present. Furthermore, SVM regression commonly makes use of the structural minimization principle which is known to have the ability to produce accurate predictive models (Cristianini and Shawe-Taylor, 2000), while ANN regression method employs model functions like radial basis function that are relatively biased when performed with input remotely-sensed variables (Kimes *et al.*, 1998; Baret and Buis, 2008).

Forest foliar N and C_N maps (Figures 5.8 and 5.9) were produced using the SVM regression algorithm as it achieved the most accurate predictive models. The maps show distinct variations in the forest N and C_N concentrations between the intact and fragmented forest ecosystems. Notwithstanding, the accuracy of the maps could have been underlain by the heterogeneity in the Dukuduku indigenous forest. The final map thus offers the potential for linking forest foliar N and C_N variability with their health conditions, even in the landscapes with disparate intact and fragmented area. Mapping forest leaf N and C_N concentrations at the scale of the WorldView-2 image (2 m) in the natural forest ecosystems of Africa would allow for a stratum-to-stratum assessment of leaf N and C_N stocks because of the relatively small crown of some tree species. Many forest ecosystems (e.g. intact and fragmented) consist of a mixture of different tree species and understory vegetation that might create a high variability of leaf N and C_N content. It can therefore be inferred that the successful mapping of the foliar N and C_N concentrations in the fragmented and intact forests might provide a better understanding of the required land management practices or what is expected from such forest ecosystems and their role in the value chain. This is even more relevant in the indigenous forest region of Africa, where a large number of the population depends on the indigenous forest ecosystems for their livelihoods.

Overall, the study provides promising results for accurate forest foliar N and C_N mapping in the Dukuduku intact and fragmented forest ecosystems. However, the results should be interpreted with some care as data at specific environmental conditions and at a fixed location were used. Further studies should explore the transferability of the present models to other points in space or time. The forest foliar N and C_N estimates should also be integrated with process-based physical models for a further understanding and assessment of intact and fragmented forests. However, one of the shortfalls of using the WorldView-2 data for modeling purposes is the cost implications associated with acquiring imagery over a large area or at a landscape level. Therefore, mapping forest foliar N and C_N concentrations over broader extents would require further investigation involving the testing of cost effective medium resolution satellite sensors like Sentinel-2. However, there would have to be some modifications to the respective methodology to account for the spatial configuration of the intact and fragmented indigenous forest ecosystems.

5.6 Conclusions

In the present study, the utility of the high spatial resolution multispectral WorldView-2 imagery with SVM and ANN regression algorithms was investigated for mapping forest foliar N and C_N concentrations in the intact and fragmented Dukuduku indigenous forest ecosystems. The results showed that the N and C_N predictive models developed using the fragmented forest data ($RMSE_{Val}$ ranged between 1.07% and 5.40% of the mean) outperformed models developed using the intact forest data ($RMSE_{Val}$ ranged between 2.50% and 5.56% of the mean). It is also concluded that the SVM regression approach achieved relatively more accurate forest foliar N and C_N estimation models compared with the ANN. Overall, the study provides valuable information that could be utilized by forest managers and ecologists to understand the functioning, health and key changes of the Dukuduku indigenous forest ecosystem. The study also presents an opportunity for comparing the competing machine learning regression methods for analyzing the newly launched multispectral imagery for mapping forest foliar N and C_N concentrations across different forest ecosystems. The strength of the WorldView-2 multispectral sensor, however, needs to be further investigated for mapping other biochemicals (e.g. P and K) in indigenous forest ecosystems within a fragmented landscape.

5.7 Acknowledgments

The author would like to thank the University of KwaZulu-Natal, South Africa and the University of Khartoum, Sudan for funding this study. My appreciation is extended to the R development core team for their open source packages for the statistical analysis. The author is very grateful to the many individuals who have helped with the data collection and processing and for their valuable comments on the manuscript.

CHAPTER SIX

SYNTHESIS AND RECOMMENDATIONS

Remote sensing of Endangered Tree Species in a Fragmented Indigenous Forest Ecosystem: A Synthesis

6.1 Introduction

What is the benefit of remote sensing for land use/cover and endangered tree species in fragmented indigenous forest ecosystems? Research in fragmented indigenous forest ecosystems revealed that land use/cover and tree species (e.g. endangered tree species) classification are the most important research fields in ecological applications of remote sensing. Moreover, fragmented landscapes in different sites of Africa are described by the removal of the tree species for other land uses such as pasture and agricultural activities, particularly in Dukuduku indigenous forest, South Africa (van Wyk *et al.*, 2006). In the study area, the existence of endangered tree species is threatened by human encroachment activities that have resulted in an over utilization of resources which has subsequently affected the forest ecosystem (Cho *et al.*, 2013; Omer *et al.*, 2015a). Endangered tree species require sound conservation and management protocols that need intensive fieldwork to geo-locate as well as monitor endangered tree species characteristics and estimate their coverage and distribution (Rushton *et al.*, 2004; Pouteau *et al.*, 2012). Therefore, more accurate information from forest surveys is required for monitoring and distinguishing the endangered tree species from other land use/cover classes in order to improve sustainable indigenous forest management practices. In efforts to minimize the potential loss of tree species in indigenous forest ecosystems of southern Africa, therefore, an integrated management strategy is needed to combine mapping and monitoring methods.

Monitoring and mapping the general spatial distribution of vegetation and tree species' biophysical and biochemical traits (e.g. LAI, N, and C_N) over large fragmented areas in indigenous forest ecosystem using conventional approaches is challenging as it is complex, very expensive, and labour intensive. Hence, there is a need for methods that consider the financial logistics, real time detection and advanced techniques for monitoring endangered tree species in indigenous forest ecosystems. In this regard, remote sensing has an advantage of being able to meet data requirements, and has proven to be a cost-effective, commendable and reliable opportunity for monitoring and mapping forest species characteristics. Moreover, earth observation data are able to produce timely and accurate information for use when mapping the spatial distribution of vegetation species parameters especially in indigenous forest ecosystems where data collection may be difficult.

The aim of the study was to investigate the utility of multispectral WorldView-2 data for mapping endangered tree species and other land use/cover types in the fragmented Dukuduku indigenous forest ecosystem in South Africa. The study further explored the possibility to estimate biophysical and biochemical traits of the six endangered tree species.
The specific objectives of the current study were to:

(1) Investigate the utility of high spatial resolution multispectral WorldView-2 data and advanced machine learning classification algorithms for mapping the land use/cover classes in a fragmented Dukuduku indigenous forest ecosystem;

(2) Examine the utility of the advanced multispectral WorldView-2 data for mapping endangered tree species in the fragmented Dukuduku indigenous forest ecosystem using machine learning classification algorithms;

(3) Test the utility of spectral vegetation indices (SVIs) calculated from the multispectral WorldView-2 data for predicting endangered tree species LAI in the fragmented and intact indigenous forest ecosystems using machine learning regression algorithms; and

(4) Map fragmented and intact indigenous forest leaf N and C_N concentrations using multispectral WorldView-2 spectral variables and machine learning regression algorithms.

All of the above objectives have been achieved in this study.

6.2 Summary of the Findings

6.2.1 Exploring the Capability of WorldView-2 High Spatial Resolution Data for Classifying Land Use/Cover Classes

Multispectral data such as SPOT and Landsat TM cover large areas of the earth’s surface at repeated time intervals, making remote sensing more effective than conventional approaches for mapping of land use/cover. Recently, the developments of high spatial resolution multispectral data (IKONOS) have brought great opportunities for monitoring and mapping land use/cover (Pu and Landry, 2012). Multispectral remotely-sensed data have high spatial resolution and offer relatively more bands but with lower spatial resolution. The low spatial resolution multispectral sensors cause the problem of spectral overlap and mixed pixels between the different classes and might not accurately map land use/cover in a fragmented indigenous forest ecosystem (Foody, 2002; Cho *et al.*, 2012). On the other hand, the development in multispectral sensors, such as WorldView-2, containing relatively fewer additional bands like yellow and red edge, make mapping land cover and vegetation at species level possible (Dlamini, 2010; Omer *et al.*, 2014). In the present study, WorldView-2 spectral subsets were tested for mapping different land use/cover classes in a fragmented Dukuduku indigenous forest of South Africa using two leaning classification algorithms; SVM and ANN (Chapter 2). The relatively accurate land use/cover maps obtained using the SVM classification algorithms and three WV2 subsets are presented in Figure 6.1. The main visual difference between the maps is that the relatively homogenous map was produced when the WorldView-2 8B (Figure 6.1a) was used as compared with WorldView-2 SB (Figure 6.1b) and WorldView-2 AB (Figure 6.1c). The maps also show that the fragmented Dukuduku forest was mainly surrounding by forests plantation and grassland, while the grassland on the north eastern of the study area was fragmented. The SVM classifier was able to classifying the eight land use/cover classes with an overall accuracy of 78%, 51% and 64% for WorldView-2 8B, WorldView-2 SB and WorldView-2 AB, respectively. Overall, the use of WorldView-2 8B significantly outperformed the WorldView-2 SB and WorldView-2 AB subsets for classifying land use/cover classes.

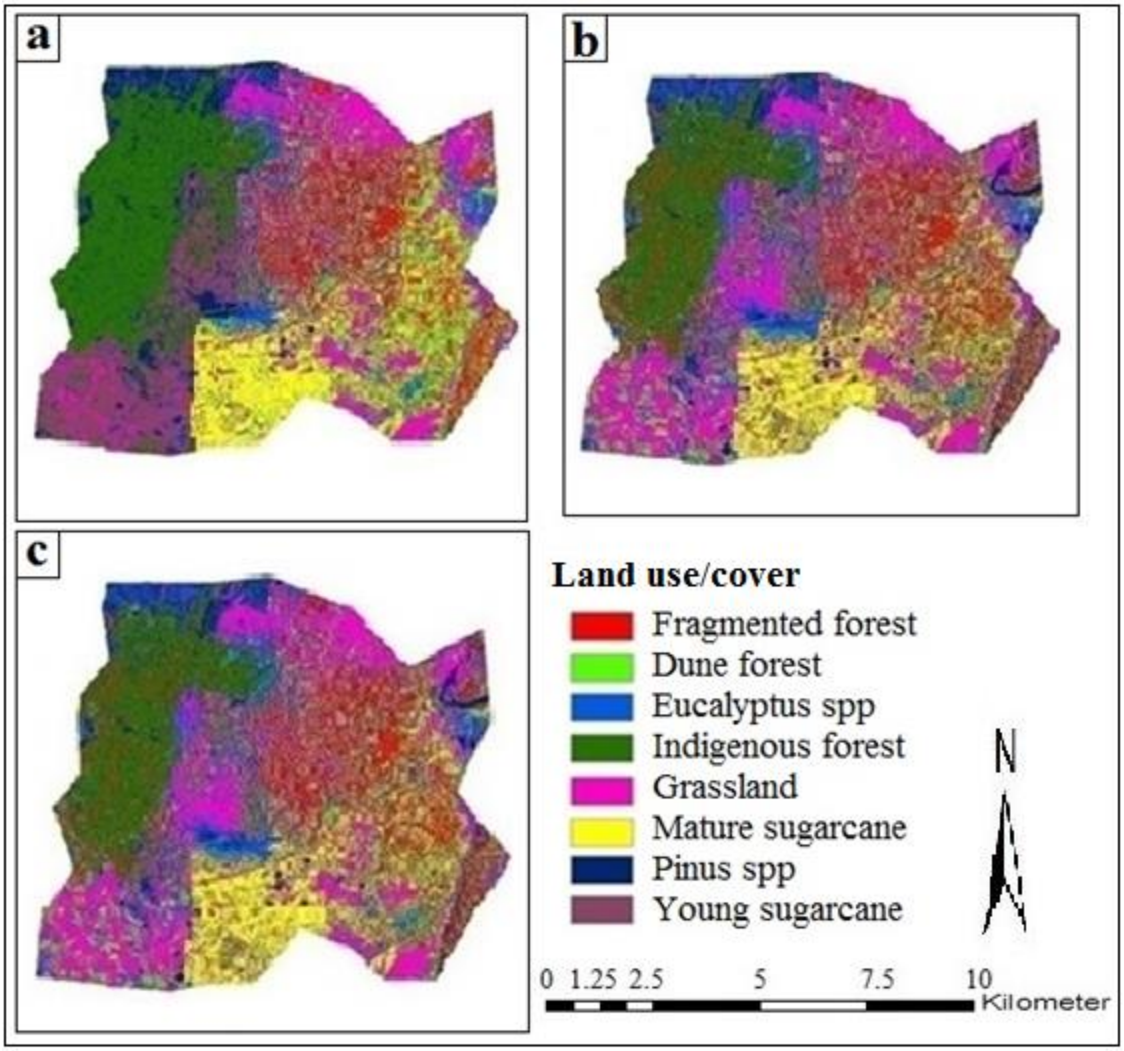


Figure 6.1: Land use/cover classification maps obtained using support vector machines classifier: (a) all eight WorldView-2 bands, (b) four standard WorldView-2 bands and (c) four additional WorldView-2 bands

6.2.2 Evaluating the Utility of Multispectral WorldView-2 Data for Mapping the Endangered Tree Species

Remotely-sensed data were regarded as a useful source of information for monitoring and mapping vegetation and forest species communities (Clark *et al.*, 2005; Pignatti *et al.*, 2009). Mapping of tree species, however, still faces some challenges in relation to ambiguous classes used. Multiple objects within a pixel can lead to spectral confusion and poor discrimination amongst different cover classes (Aplin, 2003; Cingolani *et al.*, 2004). Specifically, these challenges hinder the tree species classification when multispectral data are captured in fragmented ecosystems (Cho *et al.*, 2012). Moreover, multispectral sensors like SPOT and Landsat TM have high spatial resolution and offer relatively more bands but with low spatial resolution. Use of low spatial resolution multispectral sensors is challenging because of the problem of spectral overlap and mixed pixels between tree species due to broader and fewer spectral measurements. Conversely, the advent of WorldView-2 data with additional bands like red edge makes vegetation mapping at species level possible (Omar, 2010; Omer *et al.*, 2015b). This study evaluated the ability of WorldView-2 for classifying the endangered tree species and other land use/cover classes using advanced machine learning ANN and SVM classification algorithms (Chapter 3). The endangered tree species and land use/cover maps produced using WorldView-2 8B with the SVM and ANN methods are presented in Figure 6.2. The maps show nearly similar spatial distribution of endangered tree species in the Dukuduku area. The main visual difference between the maps is that a relatively homogenous map was obtained when the WorldView-2 8B was used. The machine learning classification algorithms were able to map the six target species and land use/cover with an overall accuracy of 77% for SVM and 75% for ANN and WorldView-2 8B. Overall, this study demonstrated that the SVM and ANN methods with WorldView-2 8B have the potential to map endangered tree species in the Dukuduku indigenous forest.

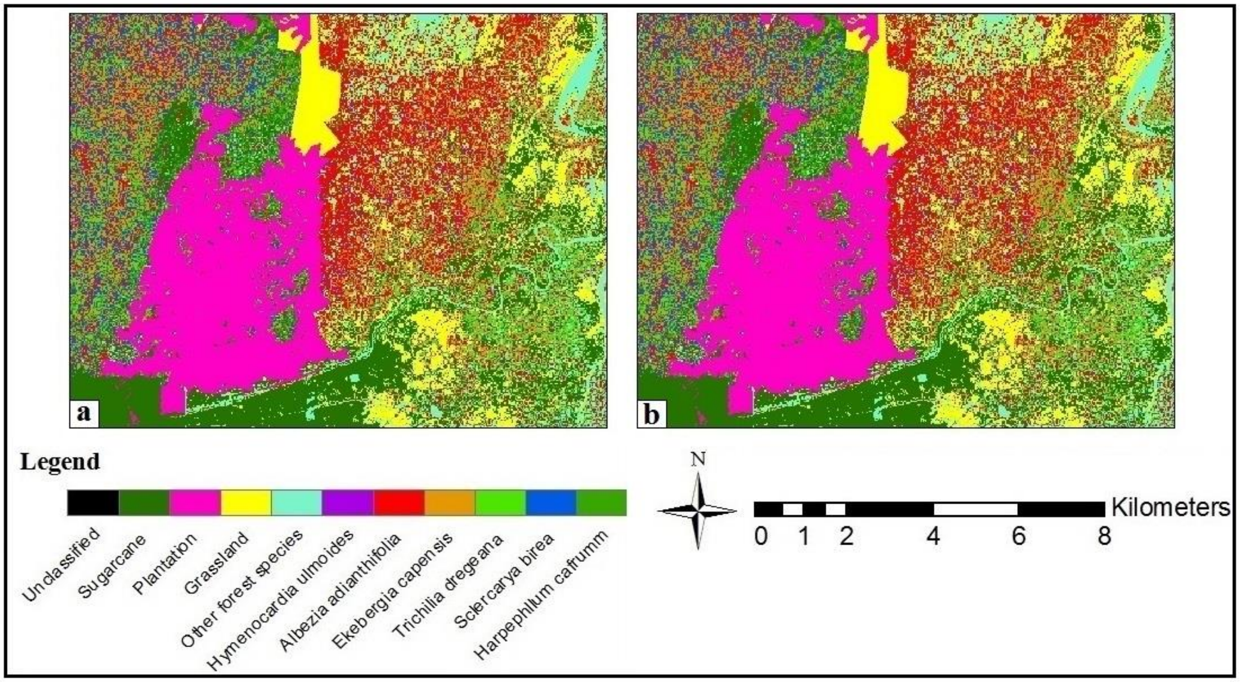


Figure 6.2: Classification maps obtained using all eight WorldView-2 bands (8B): (a) support vector machines algorithm and (b) artificial neural networks algorithm

6.2.3 Quantifying Leaf Area Index of Endangered Tree Species Using WorldView-2 Data and Two Machine Learning Regression Algorithms

Development in remote sensing technologies and analytical approaches makes it possible to explicitly and accurately estimate forests LAI (Abuelgasim *et al.*, 2006; Darvishzadeh *et al.*, 2008; Atzberger *et al.*, 2015). Studies have used remotely-sensed data of varying spectral (multispectral and hyperspectral) and spatial (fine and medium) resolutions and relatively accurate LAI estimation models were obtained. However, because of the low and medium spatial resolutions, the previous studies could not model the LAI at tree species level. On the other hand, most of the previous studies have used multiple linear regressions to estimate vegetation biophysical parameters (Kovacs *et al.*, 2004; Adjorlolo *et al.*, 2013; Pope and Treitz, 2013). However, the major challenge is the unreliable and poorly performing predictive models when multiple linear regressions is employed, mainly because of its requirements to assume a normally distributed response (output) variable as well as encounters multi-collinearity problem (Cohen *et al.*, 2003; Schlerf *et al.*, 2005; Adjorlolo *et al.*, 2013). From the available literature, no previous study has modeled LAI at tree species level using the newly launched multispectral sensors like

WorldView-2 and advanced machine learning regression algorithms like SVM and ANN in a fragmented indigenous forest. The machine learning methods are regarded as being effective for estimating the biophysical contents of vegetation and tree species. In this study (Chapter 4), LAI of six endangered tree species was estimated in fragmented and intact indigenous forest ecosystems using SVIs calculated from WorldView-2 image and the SVM as well as ANN regressions. The study further tested whether there were significant differences between LAI of the six endangered tree species in the intact and fragmented indigenous forest ecosystems. The study shows that the fragmented forest obtained significantly higher ($p \leq 0.05$) mean LAI as compared to the intact indigenous forest. The results showed that LAI could be accurately estimated using the fragmented stratum data compared with the intact stratum data. With regard to the individual tree species, the results showed that the LAI was better estimated for the *Hymenocardia ulmoides* tree species grown in the fragmented forest ecosystem as indicated by the relatively higher R^2_{Val} (0.75) and lower error metrics like RMSE $_{Val}$ (1.37% of the mean) when SVM was employed (Figure 6.3). In general, SVM regression methods achieved relatively more accurate LAI estimates compared with the ANN regression. It is concluded that tree LAI can be estimated using the machine learning regression algorithms and WorldView-2 multispectral data.

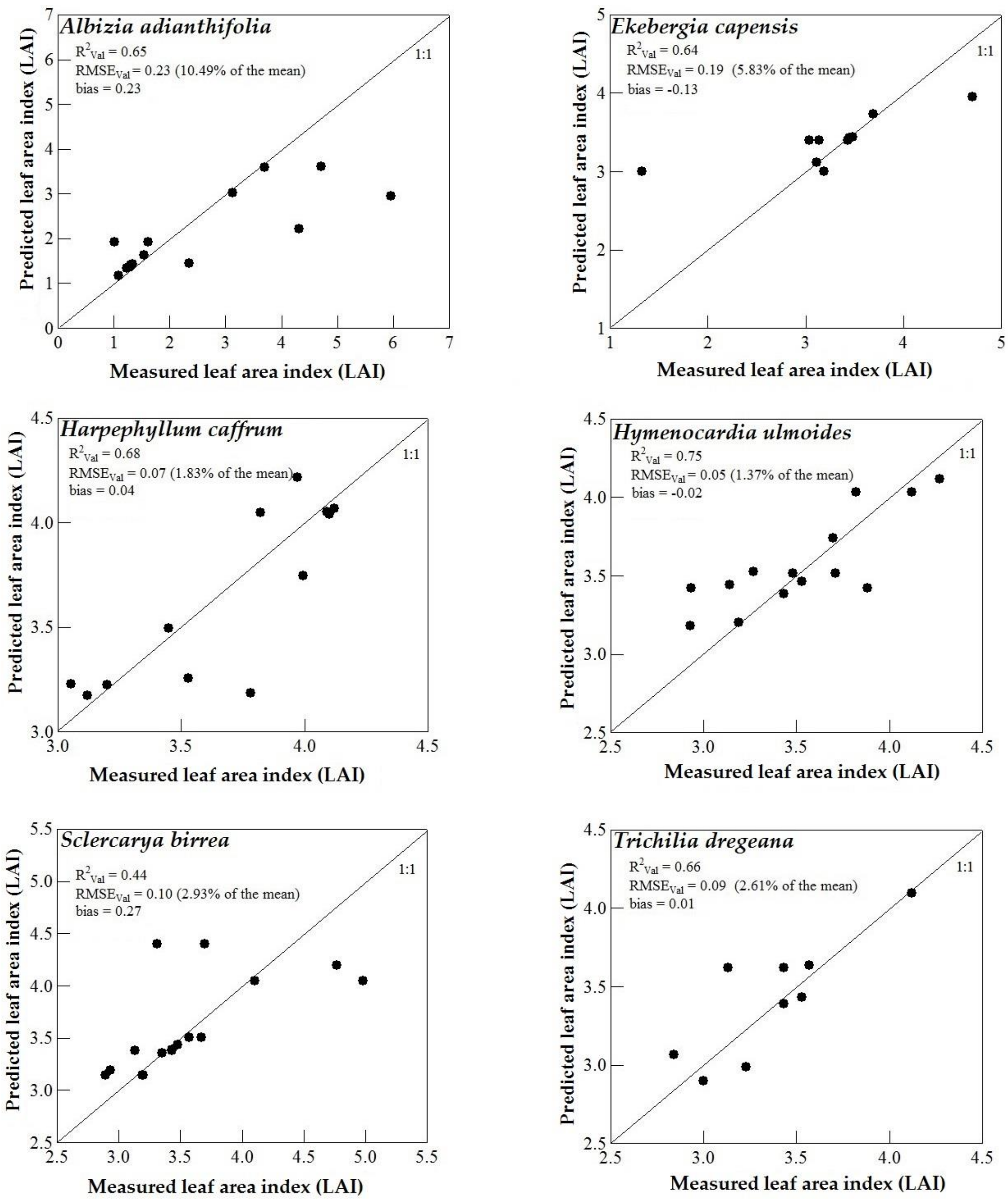


Figure 6.3: One-to-one relationships between measured and predicted LAI based on an independent validation dataset (30%) using support vector machines (SVM) regression algorithm and fragmented indigenous forest data

6.2.4 Evaluating the Reliability and Robustness of Worldview-2 Data and Machine learning Regression Algorithms for Mapping Fragmented and Intact Forest Leaf Nitrogen and Carbon Concentrations

Forest leaf N and C_N are among the most important biochemical components of tree organic matter, and the estimation of their concentrations can help to monitor the nutrient uptake processes and forest health. Monitoring indigenous forest biochemical traits (e.g. N and C_N) using conventional approaches are costly and time-consuming. Hence, there is a need for complementary approaches that use rapid, up-to-date, and cost-effective for estimating and mapping forest foliar N and C_N concentrations (Adjorlolo *et al.*, 2013; Mutanga *et al.*, 2015). In this context, remotely-sensed data provide rapid and synoptic approaches for estimating forest foliar N and C_N concentrations (Martin and Aber, 1997; Huber *et al.*, 2008; Cho *et al.*, 2013). Studies have noted that plantation forest leaf N and C_N concentrations can be estimated using empirical approaches and multispectral (Hansen and Schjoerring, 2003; Zhang *et al.*, 2006) as well as hyperspectral (Foody *et al.*, 1996; Ferwerda *et al.*, 2005; Tian *et al.*, 2011; Siegmann *et al.*, 2013) datasets. However, the indigenous forests also play an important role for ecosystem services and the value chain in general. Therefore, their foliar biochemical components need to be monitored and mapped. In particular, estimating forest foliar N and C_N concentrations of different forest ecosystems like the fragmented and intact Dukuduku forests could help resource managers to understand the impact of various socio-ecological mechanisms on indigenous forest species and the vulnerability of these ecosystems to external and internal perturbations. In addition, machine learning algorithms (e.g. SVM and ANN) are regarded as efficient and robust regression approaches for deriving models in such complex and dynamic forest ecosystems (Omer *et al.*, 2015a). To the best of the researcher's knowledge, there is a lack of literature on mapping fragmented and intact indigenous forest foliar N and C_N concentrations using the advanced SVM and ANN learning regression methods. The present study explored the use of WorldView-2 spectral variables and SVM as well as ANN regression approaches for mapping fragmented and intact indigenous forest leaf N and C_N the concentrations (Chapter 5). The chapter further investigated whether there are significant differences in the leaf N and C_N concentrations between the intact and fragmented indigenous forest ecosystems.

The results showed that the intact forest obtained significantly higher (p = 0.03) mean foliar N concentration as compared to the fragmented indigenous forest. However, there was no significant difference (p = 0.55) in the mean C_N concentration between the intact and fragmented indigenous forest. In addition, the results showed that forest foliar N and C_N concentrations could be more accurately estimated using the fragmented stratum data compared with the intact stratum data. The results also revealed that relatively accurate foliar N predictions were achieved for the fragmented forest data using the SVM (R^2_{Val} = 0.77, $RMSE_{Val}$ = 1.07% of the mean) and ANN (R^2_{Val} = 0.70, $RMSE_{Val}$ = 5.40% of the mean) regression methods (Figure 6.4). Furthermore, more accurate foliar C_N predictions were achieved for the fragmented data using SVM (R^2_{Val} = 0.67, $RMSE_{Val}$ = 1.64% of the mean) and ANN (R^2_{Val} = 0.51, $RMSE_{Val}$ = 2.21% of the mean) methods (Figure 6.5) compared with the intact data. The study further demonstrated that the SVM regression approach achieved relatively more accurate models for estimating the forest leaf N and C_N concentrations in the fragmented and intact indigenous forests as compared to the ANN regression method. Hence, predicted forest foliar N and C_N concentrations maps were produced using the SVM regression algorithm and the WorldView-2 spectral vegetation indices. The maps of forest foliar N (Figure 6.6) and C_N (Figure 6.7) show the spatial variation in the forest foliar N and C_N concentrations in the intact and fragmented forest ecosystems. The forest foliar N and C_N spatial patterns were more compact in the intact forest compared with the fragmented forest and higher N concentrations were observed towards the northern part of both forest landscapes (Figure 6.6). However, there was no distinct trend in C_N concentrations across the study area (Figure 6.7). Overall, the results obtained from this study demonstrated the robustness and usefulness of the SVM regression for mapping forest foliar N and C_N concentrations in the fragmented and intact Dukuduku indigenous forest ecosystems.

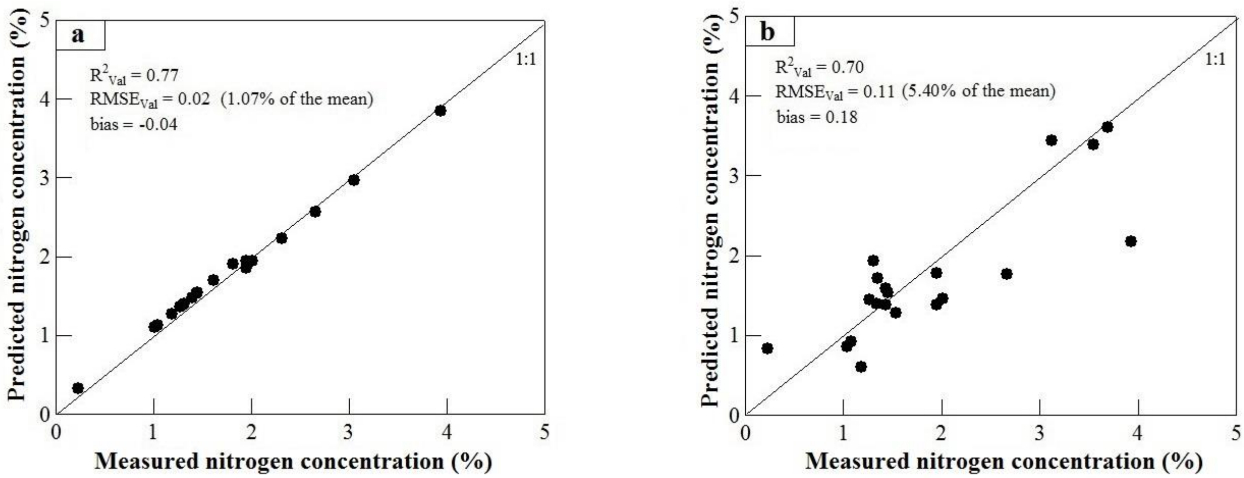


Figure 6.4: One-to-one relationships between measured and predicted forest leaf nitrogen concentration (%) based on an independent validation dataset (30%) using fragmented indigenous forest data and (a) support vector machines and (b) artificial neural networks

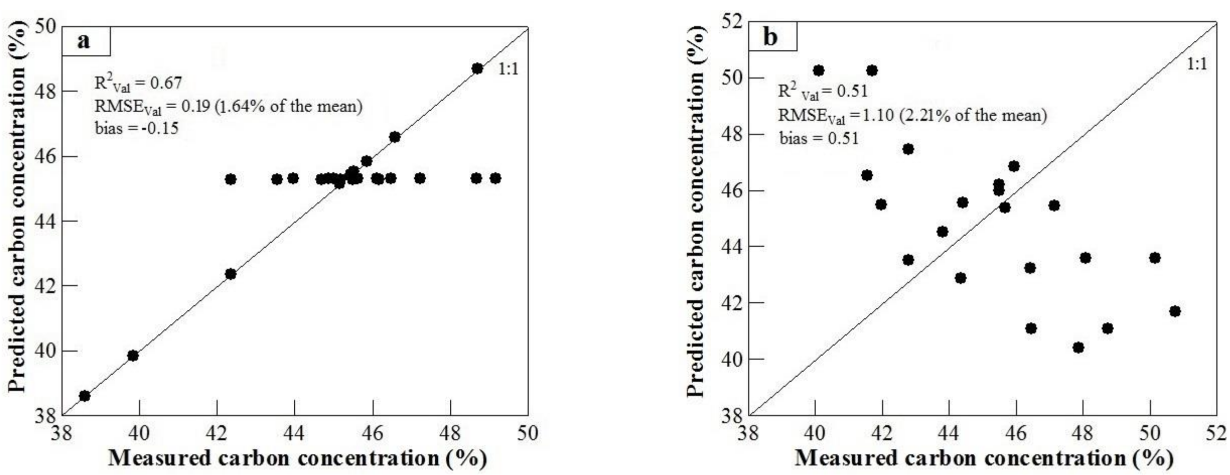


Figure 6.5: One-to-one relationships between measured and predicted forest leaf carbon concentration (%) based on an independent validation dataset (30%) using fragmented indigenous forest data and (a) support vector machines and (b) artificial neural networks.

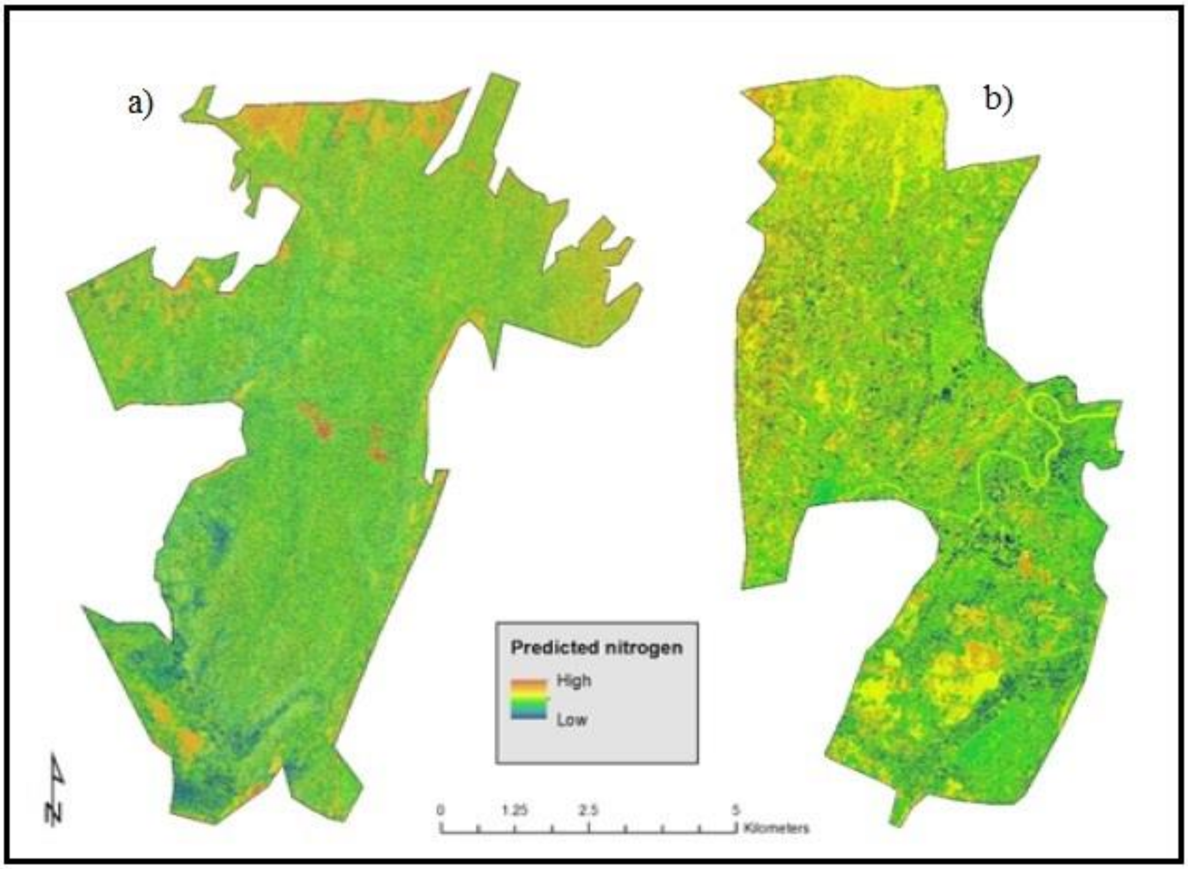


Figure 6.6: Forest foliar nitrogen (N) concentration (%) predicted map of the indigenous Dukuduku forest area. The map was produced using the support vector machine regression algorithm and a) the intact forest data and b) the fragmented forest data

Figure 6.7: Forest foliar carbon (C_N) concentration (%) predicted map of the indigenous Dukuduku forest area. The map was produced using the support vector machine egression algorithm and a) the intact forest data and b) the fragmented forest data

6.3 Conclusions

The aim of the study was to investigate the utility of the multispectral WorldView-2 data for mapping endangered tree species and other land use/cover types in the fragmented Dukuduku indigenous forest ecosystem in South Africa. The study further explored the possibility of estimating biophysical and biochemical traits of six endangered tree species. The research carried out from Chapters 2 to 6 in this study showed that there is a potential for using multispectral WorldView-2 imagery to map eight land use/cover and six selected endangered tree species. The possible use of multispectral WorldView-2 data for predicting biophysical and biochemical parameters has also been demonstrated (Chapters 4 and 5). The findings reported in this study are that the information contained in the multispectral WorldView-2 data and the use of advanced classification and regression algorithms can accomplish these tasks. The main conclusions were based on the following findings from the different objectives addressed in this study:

1. The application of high spatial resolution multispectral data with additional bands such as WorldView-2 and the machine learning SVM classification algorithm has a high potential for classifying different land use/cover classes in a fragmented indigenous forest ecosystem (Chapter 2);

2. The application of multispectral WorldView-2 data has a high potential for mapping six endangered tree species, using two machine learning classification algorithms (SVM and ANN). The result implies that the challenges facing researchers in mapping tree species in indigenous forest ecosystem could be minimized through the use of multispectral data with additional bands (Chapter 3);

3. There is a potential for using WorldView-2 spectral subsets for mapping the general land use/cover classes (Chapter 2) and endangered tree species (Chapter 3) using SVM and ANN methods. However, the use of the WorldView-2 8B produced more accurate maps than those of the SB and AB subsets;

4. The inclusion of the four WorldView-2 AB can improve the classification results of the land use/cover classes and endangered tree species in a fragmented indigenous forest ecosystem (Chapters 2 and 3);

5. LAI of endangered tree species in intact and fragmented indigenous forest ecosystems could be accurately predicted using SVIs derived from multispectral WorldView-2 data. Moreover, LAI predictive models developed using the fragmented forest data performed more accurately compared with the models developed using the intact forest data (Chapter 4); and

6. Forest foliar N and C_N concentrations could be accurately predicted using the multispectral WorldView-2 spectral variables under fragmented and intact indigenous forest ecosystems. Moreover, the predictive models derived using the fragmented forest data performed more accurately compared with those developed using the intact forest

data (Chapter 5). However, the intact forest foliar N and C_N concentrations were significantly higher than the fragmented forest.

Overall, the successful application of the WorldView-2 data, SVM and ANN classification and regression approaches in endangered tree species and their biophysical and biochemical variables in the Dukuduku indigenous forest could help in making informed decisions and policies regarding management, protection and conservation of these rare and endangered tree species. The strength of the WorldView-2 multispectral data, however, needs to be further tested for other biophysical (e.g. biomass, NPP), metro-physiological (e.g. evapotranspiration) and biochemical (e.g. Chlorophyll) traits using process-based physical models in indigenous and tropical forest ecosystems within a heterogeneous and fragmented landscape where the biophysical and biochemical traits are highly variable. The findings of this study, however, provide the necessary insights and motivation to the remote sensing community, forest managers and ecologists to move toward identifying the most suitable and readily available remote sensing satellites necessary for accurate and reliable indigenous forest monitoring specifically in a fragmented and heterogeneous landscape.

6.4 Recommendations

Remote sensing is an integral and essential tool for the collection of data necessary to support decisions and action programs to improve productivity and health status in indigenous forest ecosystems. While not all seek to use remote sensing for monitoring forest health and conservation have proven successful, several have been shown to meet data requirements, and have been demonstrated to be accurate, precise, and provide up-to-date spatial information on the current status of indigenous forest vegetation and cost-effective alternatives to ground data acquisition. In this regard, the study expects that the results of this thesis could be used to support precision indigenous forest analysis and develop effective and sustainable forest management systems. The results from this study contribute to the existing research in general, and further support scientific knowledge of ecological and forest management practices in Africa and other sites around the world which are more sensitive to climate changes. Furthermore, the results of this study could lay the foundation for possible management practices that will enable efficient and sustainable use of the resources emanating from endangered tree species. In this context, the following recommendations could be considered for future research work:

1. In this study eight course land use/cover classes were mapped using WorldView-2 data with a fine pixel size. The intra-classes variability could have exceeded the fine pixel size of the WorldView-2 image. Hence, remotely-sensed data of medium spatial resolution (e.g. 10 or 20 m) could yield relatively better classification results. Furthermore, this study did not consider the scale of the fragmentation which should match the image pixel size to accurately map the land use/cover classes, since the study area is a fragmented landscape. Future studies should look at the scale of the fragmentation using fragstats metrics extracted from multi-temporal satellite data. Scale of fragmentation should then be matched with an appropriate image resolution to derive more accurate land use/cover classification maps.

2. Newly launched multispectral data contain additional bands and do not need complex processing techniques. In this regard, the utility of multispectral satellites other than WorldView-2 (e.g. WorldView-3, Sentinel-2, RapidEye and Pleiades) for mapping endangered tree species in intact and fragmented forest ecosystems should be tested. In addition, it is also of interest for future research to assess the ability of Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) for detecting and classifying the endangered tree species.

3. The present study focused on determining the possibility of the spectral discrimination of endangered tree species (n = 6) and their biophysical characteristic like LAI in a fragmented and intact forest ecosystems. Biochemical parameters like N and C_N across Dukuduku indigenous forest strata have also been predicted. Further research should investigate and estimate the biophysical (e.g. biomass, NPP), metro-physiological (e.g. evapotranspiration) and biochemical (e.g. chlorophyll) traits for each of the endangered tree species in fragmented and intact forest ecosystems using process-based physical models.

4. The availability of multispectral sensors will allow mapping and estimating of some vegetation and tree species characteristics in indigenous forest ecosystems. This includes the biochemical variables such as chlorophyll that are important in monitoring the health

of tree species. This will help to establish a fundamental understanding of the spatial distribution of endangered tree species functions and quality which could lead to the development of early warning systems to detect any subtle changes in the indigenous forest systems, such as signs of stress, and could lead to the development of techniques to classify forest conditions like healthy or disturbed ecosystem based on their species quality and quantity.

5. In order for remote sensing approaches to become operational for mapping endangered tree species, further studies are needed to investigate the optimal spatial resolution (pixel size) that could better distinguish among the endangered tree species in highly diverse ecosystems like tropical and indigenous forests.

6. The reliability of the internal accuracy assessments of machine learning SVM and ANN algorithms for classifying land use/cover and endangered tree species was tested in this study. More research is still needed on the strength of SVM and ANN as compared to other algorithms like RF, which has proved to be successful in remote sensing classification and regression approaches. Since the purposively subset WorldView-2 band was used to test the utility of the SB and AB for mapping endangered tree species, the study did not use any feature selection method like RF to select a fewer number of bands that might classify endangered tree species with a comparable accuracy to the one yielded by the 8B. Further research should test the use of other machine learning methods like RF as variable selection and classification approach for tree species mapping.

Finally, one of the shortfalls of WorldView-2 is the cost implication associated with acquiring imagery over a regional or landscape level. Therefore, estimation and mapping biophysical (e.g. LAI) and biochemical parameters (e.g. C_N or N) over broader extents would require further investigation involving the testing of cost effective medium resolution satellite sensors such as Landsat 8. However, there would have to be some modifications to the respective methodology to account for the spatial configuration of the intact and fragmented indigenous forests.

References

Abdel-Rahman, E.M., Ahmed, F.B. and van den Berg, M., 2010. Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy. *International Journal of Applied Earth Observation and Geoinformation* 12, S52-S57.

Abdel-Rahman, E.M., Way, M., Ahmed, F., Ismail, R. and Adam, E., 2013. Estimation of thrips (Fulmekiola serrata Kobus) density in sugarcane using leaf-level hyperspectral data. *South African Journal of Plant and Soil* 30, 91-96.

Abuelgasim, A.A., Fernandes, R.A. and Leblanc, S.G., 2006. Evaluation of national and global LAI products derived from optical remote sensing instruments over Canada. *IEEE Transactions on Geoscience and Remote Sensing* 44, 1872-1884.

Adam, E., Mutanga, O., Odindi, J. and 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.

Adam, E., Mutanga, O. and Rugege, D., 2010. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecology and Management* 18, 281-296.

Adam, E., Mutanga, O., Rugege, D. and Ismail, R., 2012. Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing* 33, 552-569.

Adelabu, S., Mutanga, O. and Adam, E., 2015. Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto International* 30, 810-821.

Adelabu, S., Mutanga, O., Adam, E. and 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, 073480-1-073480-13.

Adjorlolo, C., Mutanga, O. and Cho, M.A., 2013. Estimating canopy nitrogen concentration across C3 and C4 grasslands using WorldView-2 multispectral data and the random forest algorithm, 2nd International Conference on IEEE Agro-Geoinformatics (Agro-Geoinformatics), pp. 286-291.

Agrawal, G. and Sarup, J., 2011. Comparision of QUAC and FLAASH atmospheric correction modules on EO-1 Hyperion data of Sanchi. *International Journal of Advanced Engineering Sciences and Technologies* 4, 178-186.

Akande, K.O., Owolabi, T.O., Twaha, S. and Olatunji, S.O., 2014. Performance comparison of SVM and ANN in predicting compressive strength of concrete. *IOSR Journal of Computer Engineering (IOSR-JCE)* 16, 88-94.

Alsubaie, N.M., 2012. The Potential of Using Worldview-2 Imagery for Shallow Water Depth Mapping, M.Sc Thesis, University of Calgary, Alberta, Canada, 85 pp.

Aplin, P., 2003. Comparison of simulated IKONOS and SPOT HRV imagery for classifying urban areas. *Remotely sensed cities, V.Mesev, Ed. London: Taylor & Francis*, 23-45.

Asner, G.P., Scurlock, J.M. and A Hicke, J., 2003. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecology and Biogeography* 12, 191-205.

Atkinson, P.M. and Tatnall, A., 1997. Introduction neural networks in remote sensing. *International Journal of remote sensing* 18, 699-709.

Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote sensing of environment* 93, 53-67.

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A. and le Maire, G., 2015. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation* 43, 19-31.

Baillie, J.E., Hilton-Taylor, C. and Stuart, S.N., 2004. IUCN red list of threatened species: a global species assessment. World Conservation Union.

Baret, F. and Buis, S., 2008. Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems, In: Liang, S. (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application, Springer, pp. 171-200.

Beets, P.N., Reutebuch, S., Kimberley, M.O., Oliver, G.R., Pearce, S.H. and McGaughey, R.J., 2011. Leaf area index, biomass carbon and growth rate of radiata pine genetic types and relationships with LiDAR. *Forests* 2, 637-659.

Ben-Hur, A. and Weston, J., 2010. A user's guide to support vector machines, Data mining techniques for the life sciences. Springer, pp. 223-239.

Benediktsson, J. and Sveinsson, J., 1997. Feature extraction for multisource data classification with artificial neural networks. *International Journal of Remote Sensing* 18, 727-740.

Benitez-Malvido, J., 1998. Impact of forest fragmentation on seedling abundance in a tropical rain forest. *Conservation Biology* 12, 380-389.

Bennett, K.P. and Campbell, C., 2000. Support vector machines: hype or hallelujah? *ACM SIGKDD Explorations Newsletter* 2, 1-13.

Bishop, C.M., 1995. Neural networks for pattern recognition, New York, NY, USA: Oxford University Press.

Blackburn, G.A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *International Journal of Remote Sensing* 19, 657-675.

Boegh, E., Houborg, R., Bienkowski, J., Braban, C.F., Dalgaard, T., Dijk, N.v., Dragosits, U., Holmes, E., Magliulo, V. and Schelde, K., 2013. Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop-and grasslands in five European landscapes. *Biogeosciences* 10, 6279-6307.

Bréda, N.J., 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany* 54, 2403-2417.

Breiman, L., 2001. Random forests. *Machine learning* 45, 5-32.

Brendler, T., Eloff, J., Gurib-Fakim, A. and Phillips, L., 2010. Green gold: success stories using southern African plant species. Association of African Medicinal Plants Standards (AAMPS) Publishing, Mauritius.

Broadbent, E.N., Asner, G.P., Keller, M., Knapp, D.E., Oliveira, P.J.C. and Silva, J.N., 2008. Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. *Biological Conservation* 141, 1745-1757.

Brown, M., Gunn, S.R. and Lewis, H.G., 1999. Support vector machines for optimal classification and spectral unmixing. *Ecological Modelling* 120, 167-179.

Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2, 121-167.

Camps-Valls, G., Bruzzone, L., Rojo-Álvarez, J.L. and Melgani, F., 2006. Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters* 3, 339-343.

Chan, J.C.-W. and 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.

Chason, J.W., Baldocchi, D.D. and Huston, M.A., 1991. A comparison of direct and indirect methods for estimating forest canopy leaf area. *Agricultural and Forest Meteorology* 57, 107-128.

Chen, G., Hay, G.J. and St-Onge, B., 2012. A GEOBIA framework to estimate forest parameters from lidar transects, Quickbird imagery and machine learning: A case study in Quebec, Canada. *International Journal of Applied Earth Observation and Geoinformation* 15, 28-37.

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, P., Liew, S.C., Lim, R. and Kwoh, L.K., 2011. Mapping coastal ecosystems of an offshore landfill island using WorldView-2 high resolution satellite imagery, Proc. 34th. International Symposium on Remote Sensing of Environment, 10-15 April 2011, Sydney, Australia.

Chen, Q., 2011. Comparison of Worldview-2 and IKONOS-2 Imagery for Identifying Tree Species in the Habitat of an Endangered Bird Species in Hawaii; 8-Band Research Challenge; DigitalGlobe: Longmont, CO, USA.

Cherkassky, V. and Ma, Y., 2002. Selection of meta-parameters for support vector regression, Artificial Neural Networks—ICANN 2002. Springer, pp. 687-693.

Cherkassky, V. and Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 17, 113-126.

Cho, M., Naidoo, L., Mathieu, R. and Asner, G., 2011. Mapping savanna tree species using Carnegie Airborne Observatory hyperspectral data resampled to WorldView-2 multispectral configuration, Proceedings of the 34th International Symposium on Remote Sensing of Environment, pp. 10-15, Sydney, Australia.

Cho, M., Ramoelo, A. and Holloway, J., 2014. Estimating leaf area index (LAI) by inversion of Prosail radiative transfer model using SPOT 6 imagery, The 10th conference of the African Association of Remote Sensing of the Environment (AARSE), Johannesburg, South Africa, pp. 50-55.

Cho, M. and Skidmore, A., 2009. Hyperspectral predictors for monitoring biomass production in Mediterranean mountain grasslands: Majella National Park, Italy. *International Journal of Remote Sensing* 30, 499-515.

Cho, M., Skidmore, A. and Atzberger, C., 2008a. Towards red-edge positions less sensitive to canopy biophysical parameters for leaf chlorophyll estimation using properties optique spectrales des feuilles (PROSPECT) and scattering by arbitrarily inclined leaves (SAILH) simulated data. *International journal of remote sensing* 29, 2241-2255.

Cho, M., Sobhan, I., Skidmore, A. and de Leeuw, J., 2008b. Discriminating species using hyperspectral indices at leaf and canopy scales. *The International Archives of the Spatial Information Sciences* XXXVII. Part B7, 369-376.

Cho, M.A., Debba, P., Mutanga, O., Dudeni-Tlhone, N., Magadla, T. and Khuluse, S.A., 2012. Potential utility of the spectral red-edge region of SumbandilaSat imagery for assessing indigenous forest structure and health. *International Journal of Applied Earth Observation and Geoinformation* 16, 85-93.

Cho, M.A., Ramoelo, A., Debba, P., Mutanga, O., Mathieu, R., van Deventer, H. and Ndlovu, N., 2013. Assessing the effects of subtropical forest fragmentation on leaf nitrogen distribution using remote sensing data. *Landscape ecology* 28, 1479-1491.

Cho, M.A. and Skidmore, A.K., 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment* 101, 181-193.

Cihlar, J., 2000. Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing* 21, 1093-1114.

Cingolani, A.M., Renison, D., Zak, M.R. and Cabido, M.R., 2004. Mapping vegetation in a heterogeneous mountain rangeland using Landsat data: an alternative method to define and classify land-cover units. *Remote Sensing of Environment* 92, 84-97.

Civco, D.L., 1993. Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Science* 7, 173-186.

Clark, M.L., Roberts, D.A. and Clark, D.B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote sensing of environment* 96, 375-398.

Cohen, W.B., Maiersperger, T.K., Gower, S.T. and Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment* 84, 561-571.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragnere, A., Myneni, R., Knyazikhin, Y. and Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote sensing of environment* 84, 1-15.

Congalton, R. and Green, K., 2008. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, Boca Raton, Florida, USA.

Corne, S.A., Carver, S.J., Kunin, W.E., Lennon, J.J. and van Hees, W.W., 2004. Predicting forest attributes in southeast Alaska using artificial neural networks. *Forest Science* 50, 259-276.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning* 20, 273-297.

Cristianini, N. and Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press: Cambridge, UK.

Dalponte, M., Bruzzone, L., Vescovo, L. and Gianelle, D., 2009. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment* 113, 2345-2355.

Darvishzadeh, R., Atzberger, C., Skidmore, A. and Abkar, A., 2009. Leaf Area Index derivation from hyperspectral vegetation indicesand the red edge position. *International Journal of Remote Sensing* 30, 6199-6218.

Darvishzadeh, R., Skidmore, A., Schlerf, M. and Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment* 112, 2592-2604.

Daughtry, C., Walthall, C., Kim, M., De Colstoun, E.B. and McMurtrey, J., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment* 74, 229-239.

Davi, H., Soudani, K., Deckx, T., Dufrêne, E., Le Dantec, V. and François, C., 2006. Estimation of forest leaf area index from SPOT imagery using NDVI distribution over forest stands. *International Journal of Remote Sensing* 27, 885-902.

de Chazal, J. and Rounsevell, M.D.A., 2009. Land-use and climate change within assessments of biodiversity change: A review. *Global Environmental Change* 19, 306-315.

Deering, D. and Rouse, J., 1975. Measuring'forage production' of grazing units from Landsat MSS data, the10th International Symposium on Remote Sensing of Environment, Ann Arbor, Mich, pp. 1169-1178.

DeFries, R., Hansen, M., Townshend, J., Janetos, A. and Loveland, T., 2000. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology* 6, 247-254.

DigitalGlobe, 2010. Whitepaper: The Benefits of the 8 Spectral Bands of WorldView-2, USA, Retrieved. vol 8, 1-12.

Dixon, B. and Candade, N., 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? *International Journal of Remote Sensing* 29, 1185-1206.

Dlamini, W.M., 2010. Multispectral detection of invasive alien plants from very high resolution 8-band satellite imagery using probabilistic graphical models, Digital Globe ® 8Bands Research Challenge. 1-17.

Dorigo, W., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R. and Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation* 9, 165-193.

Duhoux, M., Suykens, J., De Moor, B. and Vandewalle, J., 2001. Improved long-term temperature prediction by chaining of neural networks. *International Journal of Neural Systems* 11, 1-10.

Durbha, S.S., King, R.L. and Younan, N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment* 107, 348-361.

Eeley, H.A.C., Lawes, M.J. and Piper, S.E., 1999. The influence of climate change on the distribution of indigenous forest in KwaZulu-Natal, South Africa. *Journal of Biogeography* 26, 595-617.

Eeley, H.A.C., Lawes, M.J. and Reyers, B., 2001. Priority areas for the conservation of subtropical indigenous forest in southern Africa: a case study from KwaZulu-Natal. *Biodiversity and Conservation* 10, 1221-1246.

Eldeen, I. and van Staden, J., 2007. Antimycobacterial activity of some trees used in South African traditional medicine. *South African Journal of Botany* 73, 248-251.

Eldeen, I.M.S., 2005. Pharmacological Investigation of some trees used in South African. Traditional Medicine, PhD Thesis, KwaZulu-Natal University, Pietermaritzburg, South Africa.

Elsharkawy, A., Elhabiby, M. and El-Sheimy, N., 2012. Improvement in the Detection of Land Cover Classes Using the WorldView-2 Imagery, International Scientific ASPRS Annual Conference, Sacramento, California.

Elvidge, C.D., 1990. Visible and near infrared reflectance characteristics of dry plant materials. *Remote Sensing* 11, 1775-1795.

Engler, R., Waser, L.T., Zimmermann, N.E., Schaub, M., Berdos, S., Ginzler, C. and Psomas, A., 2013. Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. *Forest Ecology and Management* 310, 64-73.

ENVI, 2009. Environment for Visualising Images. ITT Industries, Inc, Boulder, Colorado, USA.

Everingham, Y., Lowe, K., Donald, D., Coomans, D. and Markley, J., 2007. Advanced satellite imagery to classify sugarcane crop characteristics. *Agronomy for sustainable development* 27, 111-117.

Evgeniou, T., Pontil, M. and Poggio, T., 1999. A unified framework for regularization networks and support vector machines.

Fang, H. and Liang, S., 2003. Retrieving leaf area index with a neural network method: Simulation and validation. *IEEE Transactions on Geoscience and Remote Sensing* 41, 2052-2062.

Fennell, C., Light, M., Sparg, S., Stafford, G. and Van Staden, J., 2004. Assessing African medicinal plants for efficacy and safety: agricultural and storage practices. *Journal of ethnopharmacology* 95, 113-121.

Ferwerda, J.G., Skidmore, A.K. and Mutanga, O., 2005. Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *International Journal of remote sensing* 26, 4083-4095.

Filella, I. and Penuelas, J., 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing* 15, 1459-1470.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C. and Gibbs, H.K., 2005. Global consequences of land use. *Science* 309, 570-574.

Foody, G., 2004a. Thematic Map Comparison. *Photogrammetric Engineering & Remote Sensing* 70, 627-633.

Foody, G., 2004b. Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. *International Journal of Remote Sensing* 25, 3091-3104.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment* 80, 185-201.

Foody, G.M., Boyd, D.S. and 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. and Mathur, A., 2004. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment* 93, 107-117.

Foody, G.M., Palubinskas, G., Lucas, R.M., Curran, P.J. and Honzak, M., 1996. Identifying terrestrial carbon sinks: classification of successional stages in regenerating tropical forest from Landsat TM data. *Remote Sensing of Environment* 55, 205-216.

Fuller, D.O., 2001. Forest fragmentation in Loudoun County, Virginia, USA evaluated with multitemporal Landsat imagery. *Landscape Ecology* 16, 627-642.

García Nieto, P., Martínez Torres, J., Araújo Fernández, M. and Ordóñez Galán, C., 2012. Support vector machines and neural networks used to evaluate paper manufactured using Eucalyptus globulus. *Applied Mathematical Modelling* 36, 6137-6145.

Geosystems, L., 2004. Leica Geosystems GS20 Field Guide-1.1.0en. Leica Geosystems AG CH-9435 Heerbrugg, Switzerland.

Ghosh, A. and Joshi, P., 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.

Gilabert, M., González-Piqueras, J., Garcıa-Haro, F. and Meliá, J., 2002. A generalized soil-adjusted vegetation index. *Remote Sensing of Environment* 82, 303-310.

Gitelson, A.A., Kaufman, Y.J., Stark, R. and Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment* 80, 76-87.

Gitelson, A.A. and Merzlyak, M.N., 1996. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology* 148, 494-500.

Gitelson, A.A., Vina, A., Ciganda, V., Rundquist, D.C. and Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters* 32.

Gobron, N., Pinty, B., Verstraete, M.M. and Widlowski, J. L., 2000. Advanced vegetation indices optimized for up-coming sensors: Design, performance, and applications. *IEEE Transactions on Geoscience and Remote Sensing* 38, 2489-2505.

Goel, N.S. and Qin, W., 1994. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. *Remote Sensing Reviews* 10, 309-347.

Goldblati, P., 1978. An analysis of the flora of southern Africa: its characteristics, relationships, and orgins. *Annals of the Missouri Botanical Garden*, 369-436.

Gower, S.T., Kucharik, C.J. and Norman, J.M., 1999. Direct and indirect estimation of leaf area index, f APAR, and net primary production of terrestrial ecosystems. *Remote Sensing of Environment* 70, 29-51.

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J. and Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment* 81, 416-426.

Hamilton, A.C., 2004. Medicinal plants, conservation and livelihoods. *Biodiversity and Conservation* 13, 1477-1517.

Hansen, P. and Schjoerring, J., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment* 86, 542-553.

Herold, M., Mayaux, P., Woodcock, C., Baccini, A. and Schmullius, C., 2008. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment* 112, 2538-2556.

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V. and Bonfil, D., 2010. Assessment of leaf area index by the red-edge inflection point derived from VENµS bands, Proceedings of the ESA Hyperspectral Workshop, Frascati (Italy), pp. 1-7.

Heumann, B.W., 2011. An object-based classification of mangroves using a hybrid decision tree—Support vector machine approach. *Remote Sensing* 3, 2440-2460.

Horler, D., Dockray, M. and Barber, J., 1983. The red edge of plant leaf reflectance. *International Journal of Remote Sensing* 4, 273-288.

Hornik, K., Meyer, D. and Karatzoglou, A., 2006. Support vector machines in R. *Journal of statistical software* 15, 1-28.

Houborg, R. and Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment* 112, 186-202.

Hsu, C.W., Chang, C.C. and Lin, C.J., 2009. A practical guide to support vector classification. "Technical Note", Department of Computer Science and Information Engineering, National Taiwan Univ, Taiwan.

Hsu, C.W. and Lin, C.J., 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* 13, 415-425.

Huang, C., Davis, L. and Townshend, J., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23, 725-749.

Huang, S.C. and Huang, Y.F., 1991. Bounds on the number of hidden neurons in multilayer perceptrons. *IEEE Transactions on Neural Networks* 2, 47-55.

Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R. and Foley, W.J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment* 93, 18-29.

Huber, S., Kneubühler, M., Psomas, A., Itten, K. and Zimmermann, N.E., 2008. Estimating foliar biochemistry from hyperspectral data in mixed forest canopy. *Forest Ecology and Management* 256, 491-501.

Huete, A., Justice, C. and Van Leeuwen, W., 1999. MODIS vegetation index (MOD13). *Algorithm theoretical basis document version 3; University of Virginia: Charlottesville, VA, USA* 3, 213.

Hutchings, A., 1996. Zulu medicinal plants: an inventory. University of Natal Press, South Africa.

Immitzer, M. and Atzberger, C., 2014. Early Detection of Bark Beetle Infestation in Norway Spruce (Picea abies, L.) using WorldView-2 Data. *Photogrammetrie-Fernerkundung-Geoinformation* 2014, 351-367.

Immitzer, M., Atzberger, C. and 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.

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

Itkonen, P., 2012. Estimating leaf area index and aboveground biomass by empirical modeling using SPOT HRVIR satellite imagery in the Taita Hills, SE Kenya, University of Helsinki, Helsinki, Finland, 142 pp.

Jackson, R.D. and Huete, A.R., 1991. Interpreting vegetation indices. *Preventive Veterinary Medicine* 11, 185-200.

Jäger, A.K., Hutchings, A. and van Staden, J., 1996. Screening of Zulu medicinal plants for prostaglandin-synthesis inhibitors. *Journal of Ethnopharmacology* 52, 95-100.

Jawak, S.D. and Luis, A.J., 2013. Very high-resolution satellite data for improved land cover extraction of Larsemann Hills, Eastern Antarctica. *Journal of Applied Remote Sensing* 7, 073460-073460.

Jensen, J., Qiu, F. and Ji, M., 1999. Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *International Journal of Remote Sensing* 20, 2805-2822.

Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 663-666.

Karamouz, M., Ahmadi, A. and Moridi, A., 2009. Probabilistic reservoir operation using Bayesian stochastic model and support vector machine. *Advances in water resources* 32, 1588-1600.

Karimi, Y., Prasher, S., Madani, A. and Kim, S., 2008. Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. *Canadian Biosystems Engineering* 50, 13-20.

Kätsch, C., 2006. Monitoring Woodlands from space–The Possible Role of Modern Remote Sensing & Geo-Informatics in Monitoring of Southern African Woodlands and Indigenous Forest, Proceedings of the IV. Natural Forests & Savannah Woodlands Symposium, pp. 6.5-9.5.

Kaufman, Y.J. and Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing* 30, 261-270.

Kavzoglu, T. and Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation* 11, 352-359.

Kavzoglu, T. and Mather, P., 2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing* 24, 4907-4938.

Kimes, D., Knyazikhin, Y., Privette, J., Abuelgasim, A. and Gao, F., 2000. Inversion methods for physically-based models. *Remote Sensing Reviews* 18, 381-439.

Kimes, D., Nelson, R., Manry, M. and Fung, A., 1998. Review article: Attributes of neural networks for extracting continuous vegetation variables from optical and RADAR measurements. *International Journal of Remote Sensing* 19, 2639-2663.

Kjeldahl, J., 1883. Neue methode zur bestimmung des stickstoffs in organischen körpern. *Fresenius' Journal of Analytical Chemistry* 22, 366-382.

Kohestani, V. and Hassanlourad, M., 2015. Modeling the mechanical behavior of carbonate sands using artificial neural networks and support vector machines. *International Journal of Geomechanics* 16, 04015038.

Kokaly, R.F. and Clark, R.N., 1999. Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. *Remote Sensing of Environment* 67, 267-287.

Kovacs, J., Flores-Verdugo, F., Wang, J. and Aspden, L., 2004. Estimating leaf area index of a degraded mangrove forest using high spatial resolution satellite data. *Aquatic Botany* 80, 13-22.

Kovacs, J., Wang, J. and Flores-Verdugo, F., 2005. Mapping mangrove leaf area index at the species level using IKONOS and LAI-2000 sensors for the Agua Brava Lagoon, Mexican Pacific. *Estuarine, Coastal and Shelf Science* 62, 377-384.

Krahwinkler, P. and Rossman, J., 2011. Using decision tree based multiclass support vector machines for forest mapping, IEEE International Geoscience and Remote Sensing Symposium., Vancouver, Canada, pp. 307-318.

Kumar, L., Schmidt, K.S., Dury, S. and Skidmore, A.K., 2001. Imaging spectrometry and vegetation science. In: F. van der Meer, de Jong, S.M. (Editor), Imaging Spectrometry. Kluwer Academic, Dordrecht, The Netherlands, pp. 111–155.

Larsen, M., 2007. Single tree species classification with a hypothetical multi-spectral satellite. *Remote Sensing of Environment* 110, 523-532.

Laurance, W.F., Laurance, S.G. and Delamonica, P., 1998. Tropical forest fragmentation and greenhouse gas emissions. *Forest Ecology and Management* 110, 173-180.

Lawrence, R.L., Wood, S.D. and Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment* 100, 356-362.

Leeuw, J.D., Jia, H., Yang, L., Liu, X., Schmidt, K. and Skidmore, A., 2006. Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing* 27, 223-232.

Leuning, R., Zhang, Y., Rajaud, A., Cleugh, H. and Tu, K., 2008. A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resources Research* 44.

Liang, D., Guan, Q., Huang, W., Huang, L. and Yang, G., 2013. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat. *Transactions of the Chinese Society of Agricultural Engineering* 29, 117-123.

Licciardi, G.A., Villa, A., Dalla Mura, M., Bruzzone, L., Chanussot, J. and Benediktsson, J.A., 2012. Retrieval of the height of buildings from WorldView-2 multi-angular imagery using attribute filters and geometric invariant moments. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5, 71-79.

Lichtenthaler, H., Gitelson, A. and Lang, M., 1996. Non-destructive determination of chlorophyll content of leaves of a green and an aurea mutant of tobacco by reflectance measurements. *Journal of Plant Physiology* 148, 483-493.

Liu, M., Liu, X., Li, M., Fang, M. and Chi, W., 2010. Neural-network model for estimating leaf chlorophyll concentration in rice under stress from heavy metals using four spectral indices. *biosystems engineering* 106, 223-233.

Liu, M., Wang, M., Wang, J. and Li, D., 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical* 177, 970-980.

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

Louw, C., Regnier, T. and Korsten, L., 2002. Medicinal bulbous plants of South Africa and their traditional relevance in the control of infectious diseases. *Journal of Ethnopharmacology* 82, 147-154.

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. *International journal of remote sensing* 27, 1297-1328.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28, 823-870.

Lyons, K., Brigham, C., Traut, B. and Schwartz, M., 2005. Rare species and ecosystem functioning. *Conservation Biology* 19, 1019-1024.

Malhi, Y., Doughty, C. and Galbraith, D., 2011. The allocation of ecosystem net primary productivity in tropical forests. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, 3225-3245.

Mansour, K. and Mutanga, O., 2012. Classifying increaser species as an indicator of different levels of rangeland degradation using WorldView-2 imagery. *Journal of Applied Remote Sensing* 6, 063558-1-063558-17.

Marabel, M. and Alvarez-Taboada, F., 2013. Spectroscopic determination of aboveground biomass in grasslands using spectral transformations, support vector machine and partial least squares regression. *Sensors* 13, 10027-10051.

Marchisio, G., Pacifici, F. and Padwick, C., 2010. On the relative predictive value of the new spectral bands in the WorldWiew-2 sensor, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 2723-2726.

Martin, M.E. and Aber, J.D., 1997. High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological applications* 7, 431-443.

Mathers, P., 1999. Computer Processing of Remotely-Sensed Images. John Wiley & Sons.

Mathur, A. and Foody, G., 2008. Multiclass and binary SVM classification: Implications for training and classification users. *IEEE Geoscience and Remote Sensing Letters* 5, 241-245.

McCarthy, M.J. and Halls, J.N., 2014. Habitat mapping and change assessment of coastal environments: An examination of WorldView-2, QuickBird, and IKONOS satellite imagery and airborne LiDAR for mapping barrier island habitats. *ISPRS International Journal of Geo-Information* 3, 297-325.

McMurtrie, R.E. and Dewar, R.C., 2013. New insights into carbon allocation by trees from the hypothesis that annual wood production is maximized. *New Phytologist* 199, 981-990.

Melesse, A.M. and Hanley, R.S., 2005. Artificial neural network application for multi-ecosystem carbon flux simulation. *Ecological Modelling* 189, 305-314.

Menzies, J., Jensen, R., Brondizio, E., Moran, E. and Mausel, P., 2007. Accuracy of Neural Network and Regression Leaf Area Estimators for the Amazon Basin. *GIScience & Remote Sensing* 44, 82-92.

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B. and Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia plantarum* 106, 135-141.

Miphokasap, P., Honda, K., Vaiphasa, C., Souris, M. and Nagai, M., 2012. Estimating canopy nitrogen concentration in sugarcane using field imaging spectroscopy. *Remote Sensing* 4, 1651-1670.

Mlambo, E., 2013. The project of the conservation and restoration of the Dukuduku forest between Mtubatuba and St Lucia. Inkanyamba Development Trust and Manukelana Arts and Indigenous Nursery St. Lucia, South Africa, Dukuduku project report, 1-28.

Moulin, S., 1999. Impacts of model parameter uncertainties on crop reflectance estimates: a regional case study on wheat. *International Journal of Remote Sensing* 20, 213-218.

Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 247-259.

Muñoz-Villers, L. and López-Blanco, J., 2008. Land use/cover changes using Landsat TM/ETM images in a tropical and biodiverse mountainous area of central-eastern Mexico. *International Journal of Remote Sensing* 29, 71-93.

Mushtaq, H. and Malik, T., 2014. Accounting carbon dioxide emission and stratification of carbon stock in Western Ghats, India. a geospatial approach. *International Journal of Remote Sensing & Geoscience (IJRSG)* 3, 1-5.

Mutanga, O., 2005. Discriminating tropical grass canopies grown under different nitrogen treatments using spectra resampled to HYMAP. *International Journal of Geoinformatics* 1, 21-32.

Mutanga, O., Adam, E., Adjorlolo, C. and Abdel-Rahman, E.M., 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. *International Journal of Applied Earth Observation and Geoinformation* 34, 178-187.

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

Mutanga, O. and Skidmore, A., 2004a. Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment* 90, 104-115.

Mutanga, O. and Skidmore, A.K., 2004b. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing* 25, 3999-4014.

Mutanga, O. and Skidmore, A.K., 2007. Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing* 62, 34-42.

Naidoo, L., Cho, M., Mathieu, R. and Asner, G., 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing* 69, 167-179.

Navulur, K., 2009. Enhance your Analysis with WorldView-2's Eight Spectral Bands. *GisDevelopment: http://www. gisdevelopment. net/technology/rs/ma09_210. htm*.

Ndlovu, N.B., 2013. Quantifying indigenous forest change in Dukuduku from 1960 to 2008 using GIS and remote sensing techniques to support sustainable forest management planning, Stellenbosch University, Stellenbosch, South Africa.

Nguyen, H.H. and Chan, C.W., 2004. Multiple neural networks for a long term time series forecast. *Neural Computing & Applications* 13, 90-98.

Nitze, I., Schulthess, U. and Asche, H., 2012. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proceedings of the 4th GEOBIA*, 7-9.

Ntombela, T.E., 2003. The impact of subsistence farming and informal settlement on Dukuduku Forest as a tourist resource, M.Sc Thesis, University of Zululand, Richards Bay, South Africa113 pp.

Ojoyi, M., Mutanga, O., Odindi, J. and Abdel-Rahman, E.M., 2016. Application of topo-edaphic factors and remotely sensed vegetation indices to enhance biomass estimation in a heterogeneous landscape in the eastern Arc Mountains of Tanzania. *Geocarto International* 31, 1-21.

Omar, H., 2010. Commercial Timber Tree Species Identification Using Multispectral Worldview2 Data, Digital Globe® 8Bands Research Challenge, Longmont, CO, USA. 2–3, 2-13.

Omer, G., Mutanga, O., Abdel-Rahman, E.M. and Adam , E., 2014. Potential utility of the WorldView-2 multispectral data and support vector machines algorithm to classifying land use/cover in Dukuduku landscape, Kwazulu-Natal, South Africa, The 10th conference of the African Association of Remote Sensing of the Environment (AARSE), Johannesburg, South Africa, pp. 309-318.

Omer, G., Mutanga, O., Abdel-Rahman, E.M. and Adam, E., 2015a. 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* PP, 1-16.

Omer, G., Mutanga, O., Abdel-Rahman, E.M. and Adam , E., 2015b. Exploring the utility of the additional WorldView-2 bands and support vector machines in mapping land use/land cover in a fragmented ecosystem, South Africa. *South African Journal of geomatics* 4, 414-133.

Omer, G., Mutanga, O., Abdel-Rahman, E.M. and Adam, E., 2016. Empirical prediction of leaf area index (LAI) of endangered tree species in intact and fragmented indigenous forests ecosystems using WorldView-2 data and two robust machine learning algorithms. *Remote Sensing* 8, 1-26.

Otukei, J. and Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation* 12, S27-S31.

Ozdemir, I. and Karnieli, A., 2011. Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *International Journal of Applied Earth Observation and Geoinformation* 13, 701-710.

Pal, M., 2003. Random forests for land cover classification, IEEE International Geoscience and Remote Sensing Symposium, 2003 (IGARSS '03). Proceedings. 2003, pp. 3510-3512 vol.3516.

Pal, M., 2006. Support vector machine-based feature selection for land cover classification: a case study with DAIS hyperspectral data. *International Journal of Remote Sensing* 27, 2877-2894.

Pal, M., 2009. Kernel methods in remote sensing: a review. *ISH Journal of Hydraulic Engineering* 15, 194-215.

Pal, M. and Mather, P., 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing* 26, 1007-1011.

Paola, J. and Schowengerdt, R., 1995. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of remote sensing* 16, 3033-3058.

Papale, D. and Valentini, R., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology* 9, 525-535.

Peerbhay, K.Y., Mutanga, O. and Ismail, R., 2014. Investigating the Capability of Few Strategically Placed Worldview-2 Multispectral Bands to Discriminate Forest Species in KwaZulu-Natal, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote sensing* 7, 307-316.

Peluelas, J., Baret, F. and Filella, I., 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica* 31, 221-230.

Penner, J.E., 1994. Atmospheric chemistry and air quality. *Changes in land use and land cover: a global perspective*, 175-209.

Peterson, D.L., Aber, J.D., Matson, P.A., Card, D.H., Swanberg, N., Wessman, C. and Spanner, M., 1988. Remote sensing of forest canopy and leaf biochemical contents. *Remote Sensing of Environment* 24, 85-108.

Petropoulos, G.P., Kalaitzidis, C. and 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.

Petropoulos, G.P., Kontoes, C. and Keramitsoglou, I., 2011. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using Support Vector Machines. *International Journal of Applied Earth Observation and Geoinformation* 13, 70-80.

Pignatti, S., Cavalli, R.M., Cuomo, V., Fusilli, L., Pascucci, S., Poscolieri, M. and Santini, F., 2009. Evaluating Hyperion capability for land cover mapping in a fragmented ecosystem: Pollino National Park, Italy. *Remote Sensing of Environment* 113, 622-634.

Pilger, N., 2008. Coupling lidar and high-resolution digital imagery for biomass estimation in mixed-wood forest environments. ASPRS Annual conference Portland, Oregon, USA, pp. 8.

Pontius, R.G. and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 4407-4429.

Pope, G. and Treitz, P., 2013. Leaf area index (LAI) estimation in boreal mixedwood forest of Ontario, Canada using light detection and ranging (LiDAR) and Worldview-2 imagery. *Remote Sensing* 5, 5040-5063.

Pouteau, R., Meyer, J. Y., Taputuarai, R. and Stoll, B., 2012. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological informatics* 9, 37-46.

Pu, R. and Cheng, J., 2015. Mapping forest leaf area index using reflectance and textural information derived from WorldView-2 imagery in a mixed natural forest area in Florida, US. *International Journal of Applied Earth Observation and Geoinformation* 42, 11-23.

Pu, R., Gong, P., Biging, G.S. and Larrieu, M.R., 2003. Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. *IEEE Transactions on Geoscience and Remote Sensing* 41, 916-921.

Pu, R. and Landry, S., 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment* 124, 516-533.

Qiu, F. and Jensen, J., 2004. Opening the black box of neural networks for remote sensing image classification. *International Journal of Remote Sensing* 25, 1749-1768.

Quackenbush, L.J., Ke, Y. and Kroll, C.N., 2006. Investigating new advances in forest species classification: establishing a baseline, Proceedings of 2006 ASPRS Annual Conference, pp. 1-5.

R Development Core, T., 2012. R: A language and environment for statistical computing. Vienna, Austria.

R Development Core, T., 2015. R: A language and environment for statistical computing. Vienna, Austria.

Rakotomalala, R., 2005. TANAGRA: a free software for research and academic purposes, Proceedings of EGC, pp. 697-702.

Ramoelo, A., Cho, M., Mathieu, R., Madonsela, S., van de Kerchove, R., Kaszta, Z. and Wolff, E., 2014. 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*.

Ray, T.W., 1995. Remote monitoring of land degradation in arid/semiarid regions, PhD Thesis, California Institute of Technology, California, US.

Ray, T.W. and Murray, B.C., 1996. Nonlinear spectral mixing in desert vegetation. *Remote Sensing of Environment* 55, 59-64.

Richardson, A.J. and Weigand, C., 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing* 43.

Richter, K., Atzberger, C., Vuolo, F. and D'Urso, G., 2011. Evaluation of Sentinel-2 spectral sampling for radiative transfer model based LAI estimation of wheat, sugar beet, and maize. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4, 458-464.

Roujean, J. L. and Breon, F.-M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* 51, 375-384.

Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS, 3rd ERTS Symposium, Washington, DC (NASA), pp. 309 – 317

Royston, J., 1982. Algorithm AS 181: the W test for normality. *Applied Statistics*, 176-180.

Rullan-Silva, C., Olthoff, A., de la Mata, J.D. and Pajares-Alonso, J., 2013. Remote Monitoring of Forest Insect Defoliation-A Review. *Forest Systems* 22, 377-391.

Rushton, S., Ormerod, S. and Kerby, G., 2004. New paradigms for modelling species distributions? *Journal of applied ecology* 41, 193-200.

Sáez-Plaza, P., Michałowski, T., Navas, M.J., Asuero, A.G. and Wybraniec, S., 2013. An overview of the Kjeldahl method of nitrogen determination. Part I. Early history, chemistry of the procedure, and titrimetric finish. *Critical Reviews in Analytical Chemistry* 43, 178-223.

Schlerf, M., Atzberger, C. and Hill, J., 2005. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment* 95, 177-194.

Serrano, L., Penuelas, J. and Ustin, S.L., 2002. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sensing of Environment* 81, 355-364.

Sewram, V., Raynor, M.W., Mulholland, D.A. and Raidoo, D.M., 2000. The uterotonic activity of compounds isolated from the supercritical fluid extract of Ekebergia capensis. *Journal of pharmaceutical and biomedical analysis* 24, 133-145.

Shackleton, C. and Shackleton, S., 2004. The importance of non-timber forest products in rural livelihood security and as safety nets: a review of evidence from South Africa. *South African Journal of Science* 100, 658-664.

Shafri, M., Zulhaidi, H. and Ramle, F., 2009. A comparison of support vector machine and decision tree classifications using satellite data of Langkawi Island. *Information Technology Journal* 8, 64-70.

Shao, Y. and He, Y., 2011. Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. *Soil Research* 49, 166-172.

Shen, S.S., Bernstein, L.S., Adler-Golden, S.M., Sundberg, R.L., Levine, R.Y., Perkins, T.C., Berk, A., Ratkowski, A.J., Felde, G. and Hoke, M.L., 2005. Validation of the quick atmospheric correction (QUAC) algorithm for VNIR-SWIR multi-and hyperspectral imagery,"Proceedings of the SPIE, Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery XI, vol. 5806, L. S. Bernstein and P. E. Lewis, Eds. Orlando, FL, USA. 668–678.

Siegmann, B., Jarmer, T., Lilienthal, H., Richter, N., Selige, T. and Höfled, B., 2013. Comparison of narrow band vegetation indices and empirical models from hyperspectral remote sensing data for the assessment of wheat nitrogen concentration, Proceedings of the EARSeL 8th SIG-Imaging Spectroscopy Workshop.

Sims, D.A. and Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment* 81, 337-354.

Singh, Y. and Chauhan, A.S., 2009. Neural networks in data mining. *Journal of Theoretical and Applied Information Technology* 5, 36-42.

Smith, J., 1993. LAI inversion using a back-propagation neural network trained with a multiple scattering model. *IEEE Transactions on Geoscience and Remote Sensing* 31, 1102-1106.

Soudani, K., François, C., Le Maire, G., Le Dantec, V. and Dufrêne, E., 2006. Comparative analysis of IKONOS, SPOT, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. *Remote sensing of environment* 102, 161-175.

Sundnes, F., 2013. The Past in the Present: Struggles Over Land and Community in Relation to the Dukuduku Claim for Land Restitution, South Africa, Forum for Development Studies. Taylor & Francis, pp. 69-86.

Tang, L., Tian, L.F. and Steward, B.L., 2003. Classification of broadleaf and grass weeds using Gabor wavelets and an artificial neural network. *Transactions of the ASAE* 46, 1247.

Tarantino, E., Novelli, A., Laterza, M. and Gioia, A., 2015. Testing high spatial resolution WorldView-2 imagery for retrieving the leaf area index, 3rd International Conference on Remote Sensing and Geoinformation of the Environment, pp. 95351N-95351N-8.

Thissen, U., Pepers, M., Üstün, B., Melssen, W. and Buydens, L., 2004. Comparing support vector machines to PLS for spectral regression applications. *Chemometrics and Intelligent Laboratory Systems* 73, 169-179.

Tian, Y., Yao, X., Yang, J., Cao, W., Hannaway, D. and Zhu, Y., 2011. Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground-and space-based hyperspectral reflectance. *Field crops research* 120, 299-310.

Tillack, A., Clasen, A., Kleinschmit, B. and Förster, M., 2014. Estimation of the seasonal leaf area index in an alluvial forest using high-resolution satellite-based vegetation indices. *Remote Sensing of Environment* 141, 52-63.

Vaiphasa, C., Skidmore, A., de Boer, W. and Vaiphasa, T., 2007. A hyperspectral band selector for plant species discrimination. *ISPRS Journal of Photogrammetry and Remote Sensing* 62, 225-235.

Van Aardt, J. and Wynne, R., 2007. Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing* 28, 431-436.

van Wyk, B.E., 2008. A broad review of commercially important southern African medicinal plants. *Journal of Ethnopharmacology* 119, 342-355.

van Wyk, G., Everard, D., Midgley, J. and Gordon, I., 2006. Classification and dynamics of a southern African subtropical coastal lowland forest. *South African Journal of Science* 62, 133-142.

Vapnik, V., 1995. The Nature of Statistical Learning Theory; Springer: New York, NY, USA, 1995.

Vapnik, V., 1998. Statistical learning theory. Support Vector Machines for Pattern Recognition. John Wiley & Sons, New York.

Vasconcelos, H.L. and Luizão, F.J., 2004. Litter production and litter nutrient concentrations in a fragmented Amazonian landscape. *Ecological applications* 14, 884-892.

Verger, A., Baret, F. and Weiss, M., 2008. Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment* 112, 2789-2803.

Verikas, A., Gelzinis, A. and Bacauskiene, M., 2011. Mining data with random forests: A survey and results of new tests. *Pattern Recognition* 44, 330-349.

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J. and Moreno, J., 2012a. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Transactions on Geoscience and Remote Sensing* 50, 1832-1843.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G. and Moreno, J., 2012b. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3. *Remote Sensing of Environment* 118, 127-139.

Viña, A., Gitelson, A.A., Nguy-Robertson, A.L. and Peng, Y., 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment* 115, 3468-3478.

Vitousek, P.M. and Sanford, R., 1986. Nutrient cycling in moist tropical forest. *Annual review of ecology and systematics* 17, 137-167.

Vogelmann, J.E., 1995. Assessment of forest fragmentation in southern New England using remote sensing and geographic information systems technology. *Conservation Biology* 9, 439-449.

Von Maltitz, G., Mucina, L., Geldenhuys, C., Lawes, M., Eeley, H., Adie, H., Vink, D., Fleming, G. and Bailey, C., 2003. Classification system for South African indigenous forests: an objective classification for the Department of Water Affairs and Forestry. *Environmentek report ENV-PC* 17, 1-28.

Wahbeh, A.H., Al-Radaideh, Q.A., Al-Kabi, M.N. and Al-Shawakfa, E.M., 2011. A comparison study between data mining tools over some classification methods. *International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence*, 18-26.

Wang, Y., Wang, F., Huang, J., Wang, X. and Liu, Z., 2009. Validation of artificial neural network techniques in the estimation of nitrogen concentration in rape using canopy hyperspectral reflectance data. *International Journal of Remote Sensing* 30, 4493-4505.

Watt, J. and Breyer-Brandwijk, M., 1962. The Medicinal and Poisonous Plants of Southern and Eastern Africa (2nd Edn.) Livingstone. *Edinburgh, London*, 205-206.

Werbos, P., 1974. Beyond regression: New tools for prediction and analysis in the behavioral sciences, Harvard University, Cambridge, MA.

Wiersum, K., Dold, A., Husselman, M. and Cocks, M., 2006. Cultivation of medicinal plants as a tool for biodiversity conservation and poverty alleviation in the Amatola region, South Africa. *Frontis* 17, 43-57.

Wu, C., Niu, Z., Tang, Q. and Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology* 148, 1230-1241.

Xiong, Y., Zhang, Z. and Chen, F., 2010. Comparison of artificial neural network and support vector machine methods for urban land use/cover classifications from remote sensing images A Case Study of Guangzhou, South China, IEEE International Conference on Computer Application and System Modeling (ICCASM), pp. V13-52-V13-56.

Xiu, L. and Liu, X., 2003. Current status and future direction of the study on artificial neural network classification processing in remote sensing. *Remote Sensing Technology and Application* 18, 339-345.

Yang, W., Huang, D., Tan, B., Stroeve, J.C., Shabanov, N.V., Knyazikhin, Y., Nemani, R.R. and Myneni, R.B., 2006. Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the terra MODIS sensor: 2000-2005. *IEEE Transactions on Geoscience and Remote Sensing* 44, 1829-1842.

Yang, X., 2011. Parameterizing support vector machines for land cover classification. *Photogrammetric engineering and remote sensing* 77, 27-37.

Yavuz, H.S. and Cevikalp, H., 2008. A new distance measure for hierarchical clustering, 16th IEEE Signal Processing, Communication and Applications Conference, SIU, pp. 1-4.

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O. and Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology* 396, 128-138.

Zengeya, F.M., Mutanga, O. and Murwira, A., 2013. Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savanna landscape. *International Journal of Applied Earth Observation and Geoinformation* 21, 513-524.

Zhai, Y., Cui, L., Zhou, X., Gao, Y., Fei, T. and Gao, W., 2013. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods. *International Journal of Remote Sensing* 34, 2502-2518.

Zhang, J. H., Ke, W., Bailey, J. and Ren-Chao, W., 2006. Predicting nitrogen status of rice using multispectral data at canopy scale. *Pedosphere* 16, 108-117.

Zhang, X., Li, C. and Yuan, Y., 1997. Application of Neural Networks to dentify vegetation types from satellite images. AI Applications, 11, 99-106.

Zhang, Y., Cong, Q., Xie, Y. and Zhao, B., 2008. Quantitative analysis of routine chemical constituents in tobacco by near-infrared spectroscopy and support vector machine. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 71, 1408-1413.

Zhang, Y., Gao, J. and Wang, J., 2007. Detailed mapping of a salt farm from Landsat TM imagery using neural network and maximum likelihood classifiers: a comparison. *International Journal of Remote Sensing* 28, 2077-2089.

Zhou, X., Jancso, T., Chen, C. and Verone, M.W., 2012. Urban Land Cover Mapping Based on Object Oriented Classification Using WorldView 2 Satellite Remote Sensing Images, International Scientific Conference on Sustainable Development and Ecological Footprint, Sporon, Hungary, pp. 1-10.