Mapping *Prosopis glandulosa* (mesquite) invasion in the arid environment of South Africa using remote sensing techniques

NYASHA FLORENCE MURERIWA

0604748V

Supervisor: Dr Elhadi Adam

A dissertation submitted to the School of Geography, Archaeology and Environmental Studies, Faculty of Science, University of the Witwatersrand in fulfilment of the academic requirements for the degree of Master of Science in Environmental Sciences

March 2016

Johannesburg

South Africa

Abstract

Decades after the first introduction of the *Prosopis* spp. (mesquite) to South Africa in the late 1800s for its benefits, the invasive nature of the species became apparent as its spread in regions of South Africa resulting in devastating effects to biodiversity, ecosystems and the socio-economic wellbeing of affected regions. Various control and management practices that include biological, physical, chemical and integrated methods have been tested with minimal success as compared to the rapid spread of the species. From previous studies, it has been noted that one of the reasons for the low success rates in mesquite control and management is a lack of sufficient information on the species invasion dynamic in relation to its very similar co-existing species. In order to bridge this gap in knowledge, vegetation species mapping techniques that use remote sensing methods need to be tested for the monitoring, detection and mapping of the species spread. Unlike traditional field survey methods, remote sensing techniques are better at monitoring vegetation as they can cover very large areas and are time-effective and cost-effective. Thus, the aim of this research was to examine the possibility of mapping and spectrally discriminating *Prosopis glandulosa* from its native co-existing species in semi-arid parts of South Africa using remote sensing methods.

The specific objectives of the study were to investigate the spectral separability between *Prosopis glandulosa* and its co-existing species using field spectral data as well as to upscale the results to different satellites resolutions. Two machine learning algorithms (Random Forest (RF) and Support Vector Machines (SVM)) were also tested in the mapping processes. The first chapter of the study evaluated the spectral discrimination of *Prosopis glandulosa* from three other species (*Acacia karoo*, *Acacia mellifera* and *Ziziphus mucronata*) in the study area using in-situ spectroscopy in conjunction with the newly developed guided regularized random forest (GRRF) algorithm in identifying key wavelengths for multiclass classification. The GRRF algorithm was used as a method of reducing the problem of high dimensionality associated with hyperspectral data. Results showed that there was an increase in the accuracy of discrimination between the four species when the full set of 1825 wavelengths was used in classification (79.19%) as compared to the classification used by the 11 key wavelengths identified by GRRF (88.59%). Results obtained from the second chapter showed that it is possible to spatially discriminate mesquite from its co-existing acacia species and other general land-cover types at a

2 m resolution with overall accuracies of 86.59% for RF classification and 85.98% for SVM classification. The last part of the study tested the use of the more cost effective SPOT-6 imagery and the RF and SVM algorithms in mapping *Prosopis glandulosa* invasion and its co-existing indigenous species. The 6 m resolution analysis obtained accuracies of 78.46% for RF and 77.62% for SVM.

Overall it was concluded that spatial and spectral discrimination of *Prosopis glandulosa* from its native co-existing species in semi-arid South Africa was possible with high accuracies through the use of (i) two high resolution, new generation sensors namely, WorldView-2 and SPOT-6; (ii) two robust classification algorithms specifically, RF and SVM and (iii) the newly developed GRRF algorithm for variable selection and reducing the high dimensionality problem associated with hyperspectral data.

Some recommendations for future studies include the replication of this study on a larger scale in different invaded areas across the country as well as testing the robustness of the RF and SVM classifiers by making use of other machine learning algorithms and classification methods in species discrimination.

Preface

The research work described in this dissertation was carried out in the School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, from May 2014 to March 2016 under the supervision of Doctor Elhadi Adam (School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, South Africa).

I would like to declare that the research work reported in his dissertation has never been submitted in any form for any degree or diploma in any tertiary institution. It, therefore, represents my original work. Where use has been made of the work from other authors or organisations it is duly acknowledged within the text or references chapter.

Nyasha Florence Mureriwa March 2016

As the candidate supervisor, I certify the above statement and have approved this dissertation for submission.

Doctor Elhadi Adam Signed: Date: 24/03/16

Declaration 1 – Plagiarism

I, Nyasha Florence Mureriwa, declare that:

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons.
4. This dissertation does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted then:
   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 dissertation does not contain text, graphics, or tables copied and pasted from the internet, unless specifically acknowledged and the source being detailed in the dissertation and I the References section.

Nyasha Florence Mureriwa

Declaration 2 – Publications and Manuscripts

1. **Mureriwa**, N., Adam, E., Sahu, A., & Tesfamichael, S. (2016). Examining the Spectral Separability of *Prosopis glandulosa* from Co-Existing Species Using Field Spectral Measurement and Guided Regularized Random Forest. Remote Sensing, *8*(2), 144; doi: 10.3390/rs8020144.
2. **Mureriwa**, N. and Adam, E. (in review). Mapping *Prosopis glandulosa* (mesquite) invasion in the semi-arid environment of South Africa using very high resolution WorldView-2 imagery and machine learning classifiers. Journal of Arid Environment.
3. **Mureriwa**, N. and Adam E. (in preparation). Cost effective approach for mapping P*rosopis* invasion in arid South Africa using Spot-6 Imagery and machine learning classifiers.
4. **Mureriwa** N. and Adam E. (2015). Mapping *Prosopis glandulosa* (mesquite) invasion and its co-existing species in the semi-arid environment of South Africa using Worldview-2 imagery and machine learning classifiers. *Conference proceedings. The 36th Asian Conference on Remote Sensing 2015. Crowne Plaza, Quezon City, Metro Manila, Philippines.*
5. **Mureriwa**, N., Adam, E., Sahu, A., & Tesfamichael, S. (2015). Spectral discrimination of *Prosopis glandulosa* (mesquite) in arid environment of South Africa: testing the utility of in situ hyperspectral data and guided regularized random forest algorithm. (*ACRS 2015*) *Conference proceedings. The 36th Asian Conference on Remote Sensing 2015. Crowne Plaza, Quezon City, Metro Manila, Philippines.*

Signed:

Dedication

To my beloved mother Eunice Muzenda, for her unending love, support, encouragement and belief in my success long before I could even imagine it. Thank you. May your soul rest in everlasting peace.

To my dearest parents Joachim and Agnes Murerïwa, for your undying support, patience, love and motivation. Thank you. I could not have achieved success without you.

To my uncle Wellington Runoza and sister Audrey Zvikaramba for always being there for me. You were taken from the world before I could finish my studies. May your souls rest in peace.

Acknowledgements

Firstly, I would like to thank the Lord Almighty for guiding and blessing me throughout my studies to be able to achieve my goals.

I would like to thank my family and friends for standing by me patiently throughout my studies and understood when I missed some important events due to study-commitments. I am highly appreciative to my parents for believing in me and always supporting my dreams. Their unconditional love, without which my success would not have been possible. To my sisters Priscilla, Marcia and Cynthia, thank you for your prayers, constant encouragement, motivation, help with my responsibilities when I could not be there and laughter when I took things too seriously. I thank my aunts, uncles, cousins and grandparents for their love and support for my success throughout the years. I thank my friends for being kind and understanding for being absent a lot throughout these past two years and giving me a helping hand or just listening to me when I needed them to.

I am extremely grateful to my supervisor, Dr. Elhadi Adam, who took me under his wing and gave me encouragement, went on field work with me to the very arid and far regions of Limpopo and the Northern Cape Provinces in not so pleasant weather and conditions. He additionally facilitated the use of the Spectroradiometer needed for part of my study without which a bulk part of my research would not have been acquired. He taught me how to be an independent researcher and gave me a first look into data acquisition and processing in areas I had no experience in. I greatly appreciated how he taught me how to sharpen my critical thinking and scientific writing which led to our papers being accepted and published in international journals and presented at a major international conference, all of which would not have been possible without his guidance and commitment to my success.

I thank my co-authors Dr. Solomon Tesfamichael (University of Johannesburg, South Africa) for his assistance in the collection and analysis of hyperspectral data and Mr. Anshuman Sahu (Hitachi America Big Data Laboratory, Santa Clara, California, USA) for his assistance with the hyperspectral data statistical analysis.

I extend my gratitude to Mr. Nico Byleveldt of the Department of Environmental Affairs, Northern Cape, South Africa and Mr. Shorty Thebe of the Department of Water Affairs, Kimberley, South Africa for their assistance in locating the appropriate study area as well as organising access to the study area.

I appreciate the assistance provided by the C. E. Moss Herbarium Department at the School of Animal, Plant and Environmental Sciences, University of the Witwatersrand to confirm the species identification.

I thank the staff of the Department of Geography, Archaeology and Environmental Studies at the University of the Witwatersrand for their support and assistance in acquiring funding for conferences where I gained substantial knowledge and a pool of new friends and mentors.

I also thank the South African National Space Agency (SANSA) for providing the SPOT-6 imagery and the Southern Mapping Company, South Africa for providing the WorldView-2 imagery of the study area.

For my data processing my gratitude goes to the R Development Team and for their open source package that was used for data analysis; the ESRI Incorporated team for the ArcGIS software where image pre-processing was conducted; the Exelis Visual Information Solutions team for providing the ENVI software used for image pre-processing and classification; and the EnMAP-Box development team for the open source software used to carry out the classification of the imagery.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

General Introduction

1.1. Background of *Prosopis* spp. Invasion

Invasive alien species (IAS) is an introduced, alien, exotic, non-indigenous, or non-native species, (which is a species living outside its native distributional range) and has arrived there by human activity, either deliberate or accidental (Groom et al. 2006; Mampholo 2006). Accidental introductions occur when species are dispersed by human transport such as airplanes and ships into new geographical regions (Itholeng 2007). They establish and spread impacting negatively on biodiversity, agriculture, water resources, and human health – these impacts therefore have a direct and indirect impact on economic growth and livelihoods (Witt 2010). One such introduction is that of the woody plant called *Prosopis* (mesquite).

Most introductions of *Prosopis* were intentional, although there have been accidental cross-border introductions between neighbouring counties. *Prosopis* was introduced for many reasons: to provide fodder and shade in the arid areas of South Africa and Australia; for dune stabilization, afforestation and fuel wood supply in Sudan; for live fencing in Malawi; initially to rehabilitate old quarries and later for afforestation and the provision of fuelwood and fodder in Kenya; for fuelwood production and rehabilitating degraded soil in India; for local greening, ornamental cultivation and soil stabilization in many Middle Eastern countries; and for vegetation trials in Spain (Chikuni et al. 2005; Choge and Chikamai 2004; Elfadl and Luukkanen 2006; Ghazanfar 1996; Laxén 2007; Pasiecznik et al. 2001; van Klinken et al. 2007; Zimmermann 1991). *Prosopis* was possibly first introduced unintentionally into Botswana, Nigeria and Yemen through livestock trading with neighbouring countries (Geesingis et al. 2004; Pasiecznik et al. 2001).

Despite the positive characteristics of the mesquite listed above, over time the tree showed to have negative impacts such as: the trees form extensive impenetrable thickets over large areas; it overruns grazing land; negatively affects biodiversity (plants within these dense infestations no longer provide useful services) and it excessively consumes surface and ground water (Pasiecznik 1999). Hence, in 2004 *Prosopis* was rated in the world's top 100 least wanted species by the Invasive Species Specialist Group of the IUCN. Millions of hectares of rangeland have already been invaded, and the process is still occurring in South Africa, Australia and coastal Asia (Pasiecznik 1999). Invasion has already occurred in northern Sudan where the Gash Delta of the Atbara River has been almost completely taken over by *Prosopis glandulosa*

(Mwangi and Swallow 2008). Thus, Sudan has passed a law to eradicate it (Update 1997). In the Awash basin of Ethiopia, it is aggressively invading pastoral areas in the Middle and Upper Awash Valley, and Eastern Harerge. It is one of the three top priority invasive species in Ethiopia and has been declared a noxious weed.

Because of its negative effects on ecosystems and land use, environmental management policies have been put in place over the years to control mesquite invasion in many countries. These include mechanical removal of the plant, felling and herbicide treatment of cut stumps, foliar spraying of saplings and burning (Harding 1987; van Klinken et al. 2009). For over fifty years, ranchers in south-western USA and Argentina tried a range of techniques to eradicate or control *Prosopis glandulosa* (Pasiecznik 1999). There are high costs associated with its eradication and a cost effective program is yet to be found. This effective management requires up to date temporal and spatial information about the spatial distribution of mesquite invasion and its negative impacts on the ecosystem services (Nie et al. 2012).

Traditionally, mapping the spatial distribution of vegetation species generally needs intensive fieldwork, including visual estimation and identification of species quality and quantity all of which are costly and time-consuming and sometimes impossible to accomplish due to poor accessibility or large coverage (Hoshino et al. 2012). On the other hand, remote sensing techniques offer an economic and effective technique, producing timely and accurate information for mapping vegetation species (McGlynn and Okin 2006).

1.2. Statement of the problem

The International Union for the Conservation of Nature (IUCN) declared *Prosopis glandulosa* to be one of the world's worst invasive species (Bromilow 2010; Henderson 2001; Mwangi and Swallow 2005). In the 1800s, six *Prosopis* species from Central America were introduced to the arid parts of South Africa for fodder, fuel and shade (Harding, 1987; Wise et. al., 2012). Here they have hybridised and spread rapidly. In the Northern Cape Province, for example, *Prosopis* invasions have increased from 127 821 ha in 1974 to 1 473 953 ha (~ 4% of the Province) in 2007, roughly doubling between 2004 and 2007 (Van den Berg et al. 2014; Wise et al. 2012b). Invasive alien plants are estimated to occupy at least 10 million hectares of land in South Africa with an average annual rate of spread (mainly by animal movement that feed on the seed of *Prosopis*) of at least 5% (van Wilgen et al. 2012). It is the deep rooted desert adapted shrub *Prosopis glandulosa* which is a major cause for concern in the arid and semi-arid parts of the country, especially the North West, Northern Cape, the Free State, Western Cape and parts of the Eastern Cape Province. In these areas of South Africa, *Prosopis* invasions are spreading rapidly at average annual rates in excess of 15% in upland areas and up to 30% in riparian areas (Van den Berg et al. 2014) thereby threatening water supply to groundwater dependent communities (Dzikiti et al. 2013a).

Prosopis invasions also have a variety of negative social, ecological and economic impacts. They alter ecosystem services such as water supply, hydrological functioning, grazing potential and soil quality (Dzikiti et al. 2013b; Le Maitre et al. 2011; Nie et al. 2012; van Klinken et al. 2007). Native biodiversity in many parts of the world has also been negatively impacted by invasive *Prosopis* species (Dean et al. 2002; El-Keblawy and Al-Rawai 2006; Kaur et al. 2012; Steenkamp and Chown 1996).

1.3. Research objectives

The main aim of this study is to examine the possibility of discriminating and mapping *Prosopis glandulosa* and its native co-existing species in semi-arid South Africa.

The specific objectives of this study are as follows:

1. To investigate the usefulness of in situ spectroscopic data in discriminating *Prosopis glandulosa* from three other co-existing species.
2. To test the utility of the newly developed guided regularized random forest (GRRF) to accurately discriminate amongst *Prosopis glandulosa* and its co-existing species (multiclass classification).
3. To examine if WorldView-2 imagery and two machine learning algorithms (Random Forest (RF) and Support Vector Machines (SVM)) can map *Prosopis glandulosa* invasion and its co-existing species.
4. To explore the cost-effectiveness of using SPOT-6 imagery to map *Prosopis glandulosa* invasion and its co-existing indigenous species using machine learning algorithms.

1.4. Dissertation outline

In order to achieve the objectives of this research, this dissertation is presented in six chapters organised as a collection of research papers submitted to international peer reviewed journals. The study consists of the introduction (Chapter 1), literature review (Chapter 2), three core chapters (Chapter 3, 4 and 5) that form publishable papers of which one (Chapter 3) has been published, one is under peer review (Chapter 4) and one is in preparation (Chapter 5). Each paper has been written as a stand-alone journal article that can be read individually from the rest of the dissertation but drawing to the overall objectives of the study. As a consequence, there is some repetition of content between chapters especially in the "Introduction" and "Method" sections as well as the introduction (Chapter 1) and literature review (Chapter 2). The final chapter of the dissertation is the overall conclusion of the research (Chapter 6). The content of the six chapters is:

Chapter 2 is the detailed literature review that provides information on the ecology of *Prosopis* spp. as well as its introduction history all over the world and specifically in South Africa is explored. Additionally, the species' negative invasion impacts, management and control are examined with the gaps in research in the spatial and spectral discrimination of *Prosopis* from its co-existing species being highlighted.

Chapter 3 contains the spectral discrimination of *Prosopis glandulosa* and its co-existing species using field spectroscopy and machine learning algorithms. This chapter evaluates the spectral discrimination of *Prosopis glandulosa* from its co-existing species by using *in-situ* hyperspectral data, traditional random forest (RF) and the newly developed guided regularized random forest algorithm (GRRF). The problem of high dimensionality associated with hyperspectral data is reduced by applying the GRRF algorithm to the total wavelengths selected by the traditional RF (n =1825) to reduce them to minimum key wavelengths (n = 11). The change in overall accuracy is investigated when the two algorithms are applied to the hyperspectral data.

Chapter 4 investigates the mapping of mesquite and its co-existing species in semi-arid South Africa by making use of new generation, high resolution, WorldView-2 imagery together with the random forest (RF) and support vector machines (SVM) classifiers. Spatial discrimination at species level is investigated to determine which of the classifiers is more robust for monitoring invasive spread of *Prosopis glandulosa*.

Chapter 5 assesses the cost-effectiveness of using new generation, high resolution, SPOT-6 imagery to map *Prosopis glandulosa* and its co-existing species. Following on Chapter 4 that uses expensive WorldView-2 imagery for mapping, this chapter expands the species-level spatial discrimination of mesquite by using free SPOT-6 data. The random forest ensemble as well as support vector machines are used as classifiers and their overall accuracies are evaluated.

Finally, Chapter 6 combines the results of each of the individual chapters and provides an overall conclusion. Recommendations for future research using remote sensing techniques to better monitor mesquite invasion via species-level discrimination are explored. In addition, ways to aid in the current management and control methods that have not been as successful as initially anticipated in controlling the once beneficial species are investigated.

Lastly, a single reference list is provided at the end of the dissertation.

CHAPTER TWO

Literature Review

2.1 Introduction

Invasive alien species (also known as non-indigenous or non-native species) are currently a major focus for biological conservationists, governments, farmers, ecologists and environmental managers world-wide (Joshi et al. 2004). These biological invasions pose a great threat to biodiversity as well as human activities and have been identified as a major non-climatic driver of global change (Huang and Asner 2009; Shackleton et al. 2014b). Various initiatives to control and better manage invasion have been practiced in communities internationally to guarantee lasting effects by ensuring that people make informed choices (Lowe et al. 2000). For the development of better control strategies, knowledge of the species areal extent, location and spread dynamics is imperative (Mack et al. 2000; Robinson et al. 2016; Schlesinger et al. 1990).

Remote sensing offers ways of detecting, monitoring and mapping biological invaders by extracting useful information without any physical contact with the invaders (Huang and Asner 2009; Joshi et al. 2004). These techniques offer more cost effective and timely ways of mapping species more accurately as opposed to traditional methods of vegetative mapping that were time consuming and difficult to achieve due to large coverage and poor accessibility (Akasheh et al. 2008; Hoshino et al. 2012; Li and Fox 2012). Table 2.1 below summarises some studies conducted on invasive species and the remote sensing methods used.

Table 2.1: Comparison of applications of remote sensing on invasive species

SPECIES	COUNTRY	TYPE OF DATA	CLASSIFICATION METHOD/ ALGORITHM APPLIED	ACCURACY	AUTHOR
Water hyacinth (*Eichhornia crassipes*)	Uganda, Kenya and Tanzania	Landsat, IKONOS, JERS and Radarsat	Change detection, ISODATA	93%	(Albright et al. 2004)
Mesquite *(Prosopis)*	South Africa	Landsat, NOAA, MODIS, SPOT-5	NDVI classification, EVI classification	72%	Van den Berg et al. (2014)
Mesquite *(Prosopis)*	Kenya	Landsat	Sub-pixel classification in ERDAS imagine; change detection and surveys	84% (surveys)	Zeila (2011)
Mesquite *(Prosopis)*	Australia	WorldView-2	Variation between 3 different band sets	78.7%; 90.5% and 88.1%	Robinson et al. (2016)
Leafy spurge (*Euphorbia esula* L) and Spotted knapweed (*Centaurea maculosa* Lam.)	North America	Hyperspectral imagery	Breiman Cutler Classification (BCC) – random forest	Leafy spurge: 86% Spotted knapweed: 84%	(Lawrence et al. 2006)
Spotted knapweed (*Centaurea maculosa* Lam.) and Babysbreath *(Gypsophila paniculata)*	Western United States	Hyperspectral imagery - AISA	Spectral Angle Mapper algorithm (SAM) at 1, 2, 3, 4, 5 and 10 degree angles	Spotted knapweed: 57% Babysbreath: 97%	(Lass et al. 2009)
Leafy spurge (*Euphorbia esula* L)	North America	AVIRIS imagery; ASD spectrometer	Spectral Angle Mapper algorithm (SAM)	74%	(Hunt 2009)

Prosopis glandulosa Torr. var. *torreyana* (Honey mesquite) is one of the 44 species of the genus *Prosopis* (Fabaceae family). It is a multistemed acacia-like shrub (Figure 2.1) that can grow to about 14 m in height (Henderson 2001) . The tree is comprised of compound dark green leaves (with high tannin content); reddish-brown branchlets with axial thorns; and an extensive lateral and tap root system (Figure 2.1) that can go as deep as 50 m below the surface to reach deep water tables as well as surface water with a lateral root extension of up to 30m (Dzikiti et al. 2013b; Hoshino et al. 2012; Matthews and Brand 2004; Pasiecznik et al. 2004; Pasiecznik et al. 2001). Additionally, honey mesquite is fast-growing, evergreen and flowers from June to November with yellow flowers and produces fruit in the form of hanging pods that are green when immature and yellow-purple when mature (Henderson 2001; Masilamani and Vadivelu 1997). Moreover, the species is highly tolerant of harsh conditions such as low rainfall, high temperatures as well as saline, alkaline and infertile soils making it easily adaptable to semi-arid and arid environments (Pasiecznik et al. 2001; Shiferaw et al. 2004; Zimmermann et al. 2006). Most importantly, it has been noted that there is a close resemblance between the flower, pod and leaf morphology of *Prosopis glandulosa*, *Prosopis velutina* (velvet mesquite) and *Prosopis chilensis* (algarroba) due to hybridization as well as a close relation between species which makes distinction difficult (Henderson 2001; Pasiecznik et al. 2001).

(a) (b) (c) (d)

Figure 2.1: *Prosopis glandulosa* (a) Tree acacia-like structure (b) Leaves and branchlets (c) Pods and inflorescence (d) Root system

2.2 History of *Prosopis* and its benefits

According to Zimmermann (1991), the taxa of *Prosopis* is native to South and Central America as well as the Caribbean. It occurs as either a native or an introduced species in the world's arid and semi-arid regions (Rejmánek and Richardson 2013; Shackleton et al. 2014b). From the 1800s, it was introduced to new environments all over the world such as coastal Asia, Australia, Hawaii, India, Sudan and Malawi due to problems of deforestation, desertification and fuelwood shortages in those areas (Pasiecznik 1999; Shackleton et al. 2014b). The species can currently be found as either a native or introduced species in 129 mainland (19 countries in the America's, 40 in Africa, 26 in Asia and 4 in Europe) and island countries and territories (24 island island/atoll countries in the Pacific, Atlantic and Indian Oceans and Australia as well as 18 Caribbean islands)(Shackleton et al. 2014b). The characteristic of *Prosopis* adaptability to harsh environments allowed it to fit well into agroforestry systems and serve the purposes of controlling soil erosion, afforestation of arid lands, sand dune stabilization as well as improving soil fertility (Pasiecznik 1999; Pasiecznik et al. 2001). In South Africa, the plant species was introduced from the late 1800s until 1960 to provide benefits such as nectar for honey production; utilization of its timber in construction and furniture production; provision of shade for livestock due to its wide canopy and use of pods for fodder (Chikuni et al. 2014; Zimmermann et al. 2006; Zimmermann and Pasiecznik 2005). Over the years, through a combination of flooding events, animal movement and deliberate planting, the mesquite seed has spread far and wide across the country and now covers millions of hectares in the Northern Cape, Western Cape, North-West and Free State provinces of South Africa (Van den Berg 2010; Wise et al. 2012b; Zachariades et al. 2011).

On a socio-economic scale, many communities now also look to mesquite as a source of income. In Malawi for example, Chikuni et al. (2005) noted that about 44% of the people in one village relied on products from *Prosopis* as an income source. Similarly, in India about 70% of the villages in the dry regions have their fuel supply dependent on *Prosopis* (Pasiecznik et al. 2001). In Kenya, the local economy in some communities has been boosted by US$1.5 million per year through the sale of fodder and charcoal (Choge et al. 2012). A company in South Africa is utilizing the pods to produce an organic medicine (manna') to stabilize blood sugar levels in humans and is making a profit of about US$100 000 per annum locally. If marketed internationally, the profits have a 10-fold potential (Wise et al. 2012a).

2.3. Negative Impacts

Despite all the benefits of Prosopis, studies have shown that it is an invasive species worldwide with negative social, economic and environmental impacts (Maundu et al. 2009; Mwangi and Swallow 2005; Pasiecznik et al. 2001). Many dynamics of the *Prosopis* species make them successful invaders and these include interspecific hybridization (Zimmermann 1991), they produce large quantities of seeds that remain viable for decades if not damaged, the trees have a rapid growth rate and an ability to coppice after damage, can withstand harsh climatic conditions (both high temperatures and low rainfall) and can grow in all soil types including alkaline, saline and infertile soils (Felker et al. 1981; Pasiecznik et al. 2001; Shiferaw et al. 2004).

2.3.1 Environmental impacts

Mesquite has root systems that allow the efficient utilization of both surface and ground water (up to 50 m in depth). Classified as phreatophytes the species has the ability to develop deep root systems and so can access the saturated zone in the subsurface water in both the dry and rainy season which also devastatingly affects other plants growing in the region (Dzikiti et al. 2013b; Elfadl and Luukkanen 2006).

Prosopis also impacts on biodiversity. In sub-Saharan Africa natural vegetation such as acacia and other riparian thicket species have been degraded (Hoffman et al. 1999). *Acacia erioloba* stands for instance, in the semi-arid and arid regions of South Africa have reportedly died as a direct result of the lowering of the water table by mesquite invasions (Wise et al. 2012b; Woodborne 2004). There has also been a reduction in the species diversity and richness of certain bird species, frugivores, insectivores, raptors and numbers of dung beetles in areas where mesquite invasions have occurred (Dean et al. 2002; Dean and Milton 1999; Steenkamp and Chown 1996).

2.3.2 Socio-economic impacts

Dense and impenetrable thickets are a characteristic of mesquite and this unfavourably impacts on human socio-economic activities. Local communities in Kenya, South Africa, Sudan, Eritrea, Malawi, Ethiopia and Pakistan for example have noted how these thickets provide refuge to thieves thus making communities vulnerable; there is encroachment onto paths, villages, homes, water sources, crop- and pastureland; and injuries due to thorns cause tyre punctures and animal flesh wounds and sometimes death (Chikuni et al. 2014; Choge and Chikamai 2004;

Laxén 2007; Maundu et al. 2009; Mwangi and Swallow 2008). Beneficial native species have also been negatively affected in such a way that grazing land for instance is becoming scarce and some livestock owners in Kenya, around Lake Baringo, claim that they now have to move 40-50km away, from where they reside, in search of grazing (Mwangi and Swallow 2008). Similarly, in Australia, South Africa and coastal Asia millions of rangeland have been invaded and continue to be invaded to this day (Pasiecznik 1999); pastoral areas in the Middle and Upper Awash Valley and Eastern Harege areas of the Awash basin are extremely invaded and the Gash Delta of the Atbara river of Sudan has been completely invaded by *Prosopis* (Mwangi and Swallow 2008).

Moreover, a lot of conflict between communities has arisen due to the introduction of *Prosopis.* Loss of land due to invasions forced crop farmers from Chemonke village in Kenya have had to seek alternative settlement elsewhere, often resulting in conflict with established communities (Mwangi and Swallow 2005); local livestock herders in Mali face the potential of losing their land rights and violent conflict over limited natural resources between neighbouring communities in Kenya and Ethiopia (Djoudi and Brockhaus 2011; Shackleton et al. 2014b) are some of the examples of this conflict impact.

Mwangi and Swallow (2008), noted through surveys conducted in the local communities around Lake Baringo, Kenya that 85-90% of respondents to a questionnaire favoured complete eradication of invasive *Prosopis* species. In another Kenyan study by Maundu et al. (2009) in the areas of Garissa, Loiyangalani, and Baringo it was found that 64%, 79%, and 67% of respondents, respectively, said that life would be better without *Prosopis*. In Sudan, a law has been passed to eradicate the species and over 90% of livestock owners in eastern Sudan regard invasive *Prosopis* as a liability (Update 1997; Zeila 2011). Pastoralists in Ethiopia refer to mesquite as the "Devil Tree" and it is one of the country's three top priority invasive species (Awale and Sugule 2006; Mwangi and Swallow 2005). In Australia the taxa of Prosopis is rated as one of the 20 worst invasive species and is a declared weed in all of its mainland states (Committee 2012). Furthermore, in 2004 the International Union of the Conservation of Nature (IUCN) rated mesquite as one of the world's top 100 most invasive species (Baillie et al. 2004; Mwangi and Swallow 2005). In light of these devastating effects of mesquite, it became clear that measures of control and management of the species were paramount.

2.4. Control and Management of *Prosopis* spp.

2.4.1 Control

A wide range of techniques have been tried and tested for over fifty years world over in countries such as south-western USA, Argentina, Australia and South Arica to control or eliminate *Prosopis* (Pasiecznik et al. 2001).

2.4.1.1. Mechanical control

This consists of the use of physical methods to remove and damage the invasive plant. Methods such as burning, cutting, felling and uprooting the trees have been tried (Geesingis et al. 2004; Harding 1987). Follow-up actions are necessary however, with this method because felled trees often coppice well and seed germination is frequently stimulated by soil disturbance (Van den Berg 2010).

2.4.1.2. Chemical Control

This involves the application of herbicides on the cut tree stumps so that it can move downwards on the stump to prevent regrowth of the tree species. Treatment of cut tree stumps with picloram (Tordon^{TM}) in diesel was the standard method used for many years. Its use was discontinued and replaced by others, however, due to the environmental risks associated with this herbicide as well as the high cost. In this method it is important to have the correct dosage, application method, application time and follow-up for successful results (Van den Berg 2010; van Klinken et al. 2009; Zeila 2011).

2.4.1.3. Biological control

This method consists of the use of natural agents to control other natural species in a specific way so that the species can still be exploited for its beneficial uses. Agents used operate as a supportive measure of other combative measures in place. A few countries are testing this method but Australia and South Africa are the two main countries that have investigated this approach in great depth. Seed-eating beetles have been used to target the seeds and pods of *Prosopis* and thus decrease seed production and dense thicket formation. In South Africa three species of beetles have been introduced and these are *Algarobius prosopis* (LeConte) in 1987, *Algarobius bottimeri* (Kingsolver) in 1990 and *Neltumius arizonensis* (Schaeffer) in 1992 (Coetzer and Hoffmann 1997; Klein et al. 2011; Zimmermann 1991). *Algarobius bottimeri* and *Neltumius arizonensis* failed to establish (Zimmermann et al. 2006) and in general biological control methods have not been very successful in lessening the problem of mesquite (van

Klinken et al. 2009). Additionally, over the past ten years, in Argentina, nine beetle species, four moths, one gall midge (within the species section Algorobia, Chilensis, Sericanthae and Pallidae and Ruscifoliae) and a flower-bud galler (*Asphondylia prosopidis* – Cockerel) in the USA have been considered (Zachariades et al. 2011).

2.4.1.4. Indirect and integrated control

These are methods that help to control the species indirectly so that the species is eventually killed and spread reduced. Examples of these strategies are the use of fire to burn the species all the way to the crown as well as the seeds lying on the soil surface. This has been successful in controlling some mesquite species (Mampholo 2006). When the species are still seedlings it is easier too outcompete them in an area by over-sowing and ploughing the area with beneficial plant species in the area such as grass (Van den Berg 2010).

Most importantly, it has been noted that main spread of *Prosopis* is due to animal migration and droppings containing viable seeds. The digestion process helps germination when expelled seeds are deposited in moist nutrient –rich dung. Thus, techniques such as grazing management which includes restriction of grazing during and immediately after other control methods have been applied and reduction of grazing during seed-drop season (Mampholo 2006; Van den Berg 2010).

2.4.1.5. Utilization

This method involves the use of the mesquite species benefits such as cutting the trees for firewood, furniture production, charcoal production (Choge and Chikamai 2004; Pasiecznik et al. 2006; Shackleton et al. 2014b).

2.4.2 Management

It should be noted that the method of control chosen for use by a country is dependant on a number of characteristics because each country has different requirements, capabilities and needs. Wealth is one of these driving factors. It has been found that poorer countries such as Ethiopia and Kenya for example tend to use more of the utilization and mechanical methods as far as possible whilst wealthier countries like the Middle-eastern countries that also have isolated invasions tend to use chemical and mechanical methods only (Shackleton et al. 2014b).

No country uses biological control only and although Australia and South Africa are the main utilizers of biological control, areas where ‘biological control agents’ are present have been found but are not deliberately used. Egypt (has the seed-feeding beetles—*Coleoptera* and *Burchidae*), Sudan and Yemen (have the *Algarobis prosopis*). These have not been introduced because of concerns that either the beetles might affect the less invasive *Prosopis pallida* populations which are needed by the local communities (in Yemen for example) (Pasiecznik et al. 2006) or the effect of such insects could lead to larger trees and greater pod production (Shackleton et al. 2015; Zachariades et al. 2011). In South Africa, the current management technique is the integrated approach that combines chemical and mechanical control methods along with biological methods through the government-run Working for Water Program established in 1995 (van Wilgen et al. 2012; Zachariades et al. 2011). The program has had success in reducing impacts and density on a small scale in some areas but the magnitude of impacts is still increasing very rapidly with a 35% increase between 1996 and 2008 despite a US$42.7 million (R435.5 million) spent on management (van Wilgen et al. 2012).

Another major factor that determines the management technique used is people's perceptions about the invasive species and these perceptions are shaped by their day-to-day interactions with the species and its effect on their local economies and livelihoods (Binggeli 2001; Pasiecznik et al. 2001). For example due to an earlier advantage of fuel shortage decrease as well as use as a field boundary marker, people in the Rajasthan province of India welcomed the introduction of mesquite but changed their perception when their agricultural lands were colonised (Mwangi and Swallow 2005). Economic benefit is another reason that shapes people's perceptions. If a species is economically beneficial and its management cost does not exceed the benefits, the utilisation of the species will be favoured over its eradication. Other factors that influence people's perceptions are how damaging species is to property, the media's portrayal of the species, the opinion of powerful people in society and whether the species is physically appealing (Shackleton et al. 2015). These varying opinions thus make the management of *Prosopis* a contentious issue.

Generally, these control methods have not been very successful as they are neither cost effective nor technically fruitful. In South Africa, with *Prosopis* invasions estimated to cover about 1.8 million hectares and increasing at 8% per annum there is a potential to invade between 5 to 32 million hectares (Van den Berg 2010; Versfeld et al. 1998). Le Maitre et al. (2011) estimated that US$109.1 million (US$1 = c. R7 in March 2011) would be needed to clear the invaded uplands and US$76.6 million to clear invaded floodplains. Clearing costs per hectare were estimated to vary from US$13–534 depending on the densities of the infestations.

In order to successfully come up with management plans that can effectively eradicate and or control *Prosopis*, the dynamics of the processes underlying plant invasions to reduce negative impacts whilst maximising benefits and opportunities for management invention are necessary (Robinson et al. 2008). This includes a look into the spatial distribution of the invasive species.

2.5. Remote sensing and vegetation species mapping

Traditionally, vegetation species mapping needs intensive field work that is both time consuming and costly and at times unachievable due to poor accessibility (Hoshino et al. 2012; Kent and Coker 1994; Lee and Lunetta 1995). Remote sensing, on the other hand, is a technique that gathers data regularly about the earth's features without actually being in direct contact with those features (Adam and Mutanga 2009). The two main advantages that make remote sensing preferable to field-based methods in landcover classification, are that it has repeat coverage which allows continuous monitoring, and its digital data can be easily integrated into a geographic information system for more analysis which is less costly and less time-consuming (Ozesmi and Bauer 2002; Schmidt and Skidmore 2003; Shaikh et al. 2001). Multispectral data such as LandsatTM and SPOT imagery have been used to identify general vegetation classes or to discriminate broad vegetation communities (Harvey and Hill 2001; Li et al. 2005; May et al. 1997). Multispectral data have been used but the limitation has been lack of the spectral and spatial resolution and mixed pixels and hence low accuracy has been achieved. On the other hand, hyperspectral data often consist of over 100 contiguous bands of 10 nm or less bandwidth. These contiguous bands and narrow ranges lead to the possibility of discriminating and mapping vegetation species more accurately and precisely than the standard multispectral bands (Borges et al. 2007; Schmidt and Skidmore 2003; Ustin et al. 2004). The use hyperspectral data was useful but the limitation is that hyperspectral data is very expensive and difficult to process. Then new generation advanced multispectral data such as WorldView and RapidEye has been found to overcome both these limitations.

Remote sensing is one method that has not been fully utilised as a tool in control. In Kenya, the invasion of *Prosopis* impacts the livelihood of dryland communities and the ecological integrity of the fragile arid and semi-arid lands (Zeila 2011). Remote sensing and Geographical Information System techniques have been employed to investigate the extent of the species' in Garissa County. Since mesquite is evergreen, studies were carried out in Kenya's driest season (September) to make it easier to distinguish mesquite from the surrounding non-deciduous flora. Studies were carried out in 2000 and 2006 and datasets from the two years was then compared. Landsat was used for mapping and a socio-economic survey was done to find out from the community the implications and perceptions of the mesquite. From this study, it was realised that pixel resolution used (30 m x 30 m for multispectral bands) has mixed spectra and thus raises the mixed pixel problem which dims efficient classification using standard classifiers. The study is limited in terms of spectral resolution by using Landsat TM imageries for acquisition of scenes mainly due to the implications in acquiring imageries with much higher spectral resolutions such as QuickBird. Moreover the study recognizes that there is advancement in the techniques for mapping the spread of invasive alien species than the technique used in this study (Zeila 2011).

On the other hand, studies done in Sudan made use of the Normalised Difference Infrared Index (NDII). The reason for this was that with near infrared bands, spectral reflectance shows mesquite to be healthier growing than other plants and thus easier to map. The study compared the spectral reflectance in the infrared of stressed and unstressed canopies (Hoshino et. al., 2012). Landsat5 Thematic Mapper was used and made use of bands 4 (near infrared) and band 7 (short-wave infrared). For comparison, a handheld soil moisture measurement system (Hydrosense) was used to accommodate the problem of backscatter. This study showed the high water usage of mesquite as the results showed a high foliar water content (Hoshino et al. 2012). This study had limitations. There were penetration difficulties into the mesquite trees when using the PALSAR L-band for detection. Thus to date additional studies on the estimation of the mesquite biomass using the PALSAR L-band microwave data are needed. Moreover, due to the density of the trees in the study area, soil moisture retrieval is a challenging problem because of the complicated scattering mechanisms of the mesquite canopy (Hoshino et al. 2012).

Studies carried out in South Africa for the detection, quantification and monitoring of *Prosopis* invasion in the whole of the Northern Cape made use of Landsat for multi-temporal data and MODIS EVI Imagery (Van den Berg 2010). In this study they were able to predict the future possible areas of invasion and highlighted the relationship between habitat and future *Prosopis* invasion although they pointed out that more accurate mapping of the species is still required. In addition, they did note that there is a long standing problem of managers and ecologist is the differentiation between alien and indigenous vegetation in mixed stands, a property that was not catered for in their study. Moreover, with multi-spectral data becoming less expensive, the use of these images should be promoted for appropriate sites for better results. It was highlighted that due to the availability of imagery from the SPOT series (which has been available for use in South Africa since 2008) could be a possible way to continue the change analysis detection and monitoring of the spatial dynamics and clearing programs of *Prosopis* invasion (Van den Berg 2010).

More recently, Robinson et al. (2016) conducted *Prosopis* studies in the Pilbara Region of Australia using WorldView-2 imagery and object-based data analysis to map *Prosopis* from its co-existing eucalypt species and background soil types. With a high spatial resolution of 2m for each of the eight WorldView-2 bands, different band-set combinations were used to evaluate the best band combination that can differentiate *Prosopis* (van Klinken et al. 2007). The highest accuracy was achieved by not making use of all eight bands but of band-subsets with the dual near-infrared bands to be the most informative, followed by the red edge band combinations. Though successful in differentiating average-sized mesquite plants, the study had the limitation of only discriminating between only two plant species and background soil (Peerbhay et al. 2013). Unlike previous studies, this WV-2 imagery has the great potential to develop mesquite management responses in stands of 16 square meters and greater in a grass-free matrix over a heterogeneously soil type distribution (Robinson et al. 2016).

No attempt, to my knowledge, has been made to map *Prosopis* at species level using remote sensing in South Africa. Mapping at this level would increase the understanding of the invasion dynamic of mesquite as its monitoring in relevance to the indigenous co-existing species can be monitored. The high resolution multispectral WorldView-2 data (2 m multispectral resolution; 0.46m pan-sharpened resolution) is able to discriminate between the co-existing species to provide useful information to assist in management and control strategies already in place. Additionally, with the availability of free SPOT-6 data (6 m multispectral resolution and 1.5 m resolution pan-sharpened), cost-effective mapping of mesquite at species level can also be applied to obtain useful information on invasion. This gap in knowledge has provided the basis for this study.

CHAPTER THREE

Spectral discrimination of *Prosopis glandulosa* and its co-existing species using field spectroscopy and guided regularized random forest.

This chapter is based on:

1. **Mureriwa**, N., Adam, E., Sahu, A., & Tesfamichael, S. (2016). Examining the Spectral Separability of *Prosopis glandulosa* from Co-Existing Species Using Field Spectral Measurement and Guided Regularized Random Forest. *Remote Sensing*, *8*(2), 144.
2. **Mureriwa**, N., Adam, E., Sahu, A., & Tesfamichael, S. (2015). Spectral discrimination of *Prosopis glandulosa* (mesquite) in arid environment of South Africa: testing the utility of in situ hyperspectral data and guided regularized random forest algorithm. (*ACRS 2015*) *Conference proceedings. The 36th Asian Conference on Remote Sensing 2015. Crowne Plaza, Quezon City, Metro Manila, Philippines.*

3.1 Abstract

The invasive taxa of *Prosopis* is rated in the world's top-100 unwanted species and lack of spatial data about the invasion dynamics has made the current control and monitoring methods unsuccessful. This study thus tests the use of *in-situ* spectroscopy data with a newly developed algorithm, guided regularized random forest (GRRF) to spectrally discriminate *Prosopis* from coexisting acacia species (*Acacia karoo, Acacia mellifera* and *Ziziphus mucronata*) in arid environment of South Africa. Results show that GRRF was able to reduce the high dimensionality of the spectroscopy data and select key wavelengths (n = 11) for discriminating among the species. These wavelengths are located at 356.3 nm, 468.5 nm, 531.1 nm, 665.2 nm, 1262.3 nm, 1354.1 nm, 1361.7 nm, 1376.9 nm, 1407.1 nm, 1410.9 nm and 1414.6 nm. The use of these selected wavelengths increases the overall classification accuracy from 79.19% and a Kappa value of 0.7201 when using all wavelengths to 88.59% and a Kappa of 0.8524 when the selected wavelengths were used. Based on our relatively high accuracies and ease of use, it is worth considering the GRRF method for reducing the high dimensionality of spectroscopy data. However, this assertion should receive considerable additional testing and comparison before it is accepted as a substitute for reliable high dimensionality reduction.

Keywords: *Prosopis glandulosa*; Spectroscopy; Guided Regularized Random Forest; field spectroscopy; variable selection

3.2. Introduction

Taxa of *Prosopis* (mesquite) cover large areas of the world's hot arid and semi-arid environments as an introduced or native species (Shackleton et al. 2014b). *Prosopis* is a fast-growing, drought and salt-resistant plant with remarkable coppicing power (Zeila 2011). It is a thorny shrub that can grow to about 5 m in height and is evergreen. It fixes nitrogen and is tolerant of arid conditions and saline soils (van Klinken et al. 2007). The spread of the plant is caused mostly by the movement and migratory patterns of livestock through droppings (Awale and Sugule 2006). Mesquite species and their hybrids became invasive in the arid northern parts of South Africa as well as other similar environs of the world because of their adaptability to the harsh climatic conditions, vigorous growth, high seed production leading to large seed banks, the absence of natural seed feeding insects and efficiency of the seed dispersal mechanism (Lloyd et al. 2002). The majority of introductions of mesquite were intentional, but accidental cross-border inductions between neighbouring countries have occurred (Shackleton et al. 2014a). It is, for example, believed that the plant was introduced inadvertently into Botswana, Nigeria and Yemen through livestock trading (Geesingis et al. 2004; Pasiecznik et al. 2001). It was intentionally introduced for a number of reasons such as to provide shade and fodder in the arid areas of Australia and South Africa (Zimmermann 1991); for sand-dune stabilization, afforestation as well as fuelwood supply in Sudan (Ghazanfar 1996); for live fencing in Malawi (Chikuni et al. 2005); for local greening, ornamental cultivation and soil stabilization in many Middle Eastern countries (Ghazanfar 1996); initially to rehabilitate old quarries and later for afforestation and the provision of fuelwood and fodder in Kenya (Choge et al. 2012); for fuelwood production and rehabilitating degraded soil in India (Pasiecznik et al. 2001; van Klinken et al. 2007); and for vegetation trials in Spain (Elfadl and Luukkanen 2006; Laxén 2007).

The plants have negative impacts on ecosystems such as formation of extensive impenetrable thickets over large areas; loss of biodiversity; encroachment onto grazing land; and excessive consumption of surface and ground water (Pasiecznik 1999). Globally, large areas of rangeland have already been lost due to invasion of mesquite, and the problem is still occurring (Pasiecznik 1999). In South Africa, approximately 1.8 million hectares of land has been invaded by the plant and the invasion is increasing at 8% per annum (Le Maitre et al. 2004; Van den Berg 2010; Versfeld et al. 1998) while over a million hectares has been invaded by the plant in Australia with the potential to spread over 70% of Australia's land area (Osmond 2003). Similar problems have been reported in Kenya (Maundu et al. 2009; Witt 2010), Sudan and Ethiopia (Mwangi and Swallow 2005). As a result of such environmental impacts, *Prosopis* was rated in the world's top 100 least wanted species in 2004 by the Invasive Species Specialist Group of the International Union for Conservation of Nature (Baillie et al. 2004). Various methods are used to control mesquite invasion in different countries such as South Africa, Sudan, USA, Argentina and Ethiopia. These include mechanical removal of the plant, which often involves cutting and/or burning of the target plant (Harding 1987; van Klinken et al. 2009); biological control by making use of beetles that feed on the plant (Coetzer and Hoffmann 1997; Zimmermann 1991); chemical control by treating cut tree stumps with herbicides such as picrolam (Zachariades et al. 2011) and finally indirect control which involves a combination of methods such as grazing and over-sowing of an area with beneficial plant species (Mampholo 2006). Generally, mesquite invasion control methods are normally associated with high costs that need to be minimized through efficient management. This efficient management requires up-to-date information about spatial and temporal distribution of mesquite invasion and its negative impacts on the ecosystem services (Nie et al. 2012).

Traditionally methods of mapping the spatio-temporal distribution of vegetation species generally need intensive fieldwork that involves visual observation and identification of species quality and quantity. Such methods are relatively expensive, time-consuming and sometimes impossible to accomplish due to poor accessibility or large coverage (Hoshino et al. 2012). On the other hand, remote sensing methods offer a more efficient and less costly alternative, producing timely and accurate information for mapping vegetation species (Zeila 2011). Few studies have been applied in this area for investigations such as the mapping of *Prosopis* density in South Africa using Landsat and MODIS EVI images (Van den Berg 2010), discriminating between stressed and healthy mesquite canopies using PALSAR L-band data and Normalised Difference Infrared Index (NDII) in Sudan (Hoshino et al. 2012) and mapping the extent of *Prosopis* invasion using Landsat imagery in Kenya (Zeila 2011). However, these studies did not assess the plant at the species level due to lack of spectral and spatial resolutions of remotely-sensed data used. For example, Landsat and PALSAR L-band images have a rather low spatial resolution that prevents them from resolving individual plants. In addition, multispectral data such as Landsat images suffer from the mixed pixel problem where a pixel value represents a combination of objects present within the pixel area. Pixel impurity can be overcome by using hyperspectral data that provides the capability to define surface features with higher spectral and spatial resolutions (Barry et al. 2002; Liew et al. 2002). The use of hyperspectral remote sensing in mapping vegetation species in different landscapes has been well established (Adam et al. 2012b; Artigas and Yang 2005; Cho et al. 2008; Fung et al. 1999; Kumar and Skidmore 1998). Unfortunately, one of the notable problems in hyperspectral data processing is that in most cases, the number of training samples (n) is limited as compared to the large number of hyperspectral spectral bands (p) (Hsu 2007a). This 'small *n* large *p* problem' has been termed the 'curse of dimensionality', which leads to the 'peaking phenomenon' or 'Hughes phenomenon' which introduces multi-collinearity in the input data matrix (Hsu 2007b; Melgani and Bruzzone 2004). The estimation of statistic class parameters are thus rendered inaccurate and unreliable. Furthermore, the computation of such large, collinear data sets becomes time consuming and prohibitive in analysis (Bajcsy and Groves 2004; Hsu 2007a; Kavzoglu and Mather 2002).

In light of this, techniques that reduce the problem of high dimensionality without sacrificing significant information are vital. Feature selection is often considered to be a practical as well as an important method in processing and analysing hyperspectral data (Borges et al. 2007; Pal 2005; Shaw and Manolakis 2002). Over the last few years a random forest algorithm has been commonly used in hyperspectral remote sensing applications as both a classification and feature selection method. A random forest algorithm developed by Breiman (2001) is based on unpruned trees and bootstrap samples of the original data to improve the classification and regression trees (CART) method by combining a large set of decision trees for the final result. Hyperspectral dimensionality reduction has shown to be major successes of random forest algorithms in remote sensing applications. However, studies have shown that random forest provides an internal measure of variable importance but it does not automatically choose the optimal number of variables that yield the best classification accuracy (Adam et al. 2012b). Moreover, the random forest method for variable importance measurement shows a bias towards correlated predictor (Adjorlolo et al. 2013a; Strobl et al. 2008). Deng and Runger (2012) thus proposed a regularization framework that can be applied to random forest (regularized random forest) and boosted trees (regularized boosted trees). The regularization framework avoids selecting a new feature for splitting the data in a tree node when that feature produces similar information to a feature already selected (Deng 2013; Deng and Runger 2012). An added advantage is that the framework builds one model that may considerably reduce the training time (Deng and Runger 2012). A new method that improves on regularized random forest is called Guided Regularized Random Forest (GRRF) that uses the importance scores from an ordinary random forest to guide the feature selection process (Deng and Runger 2013).

The aim of this study was therefore to investigate the possibility of spectral discrimination of *Prosopis* from other co-existing native tree species using *in situ* spectroscopy. The specific objectives of the study were to (a) discriminate the mesquite plant (*Prosopis glandulosa*) from three other species (*Acacia karoo*, *Acacia mellifera* and *Ziziphus mucronata*) in the study area; (b) test the utility of the new developed guided regularized random forest in identifying key wavelengths that accurately discriminate among the tree species (multiclass classification).

3.3. Materials and Methods

3.3.1 Study area

The Northern Cape Province is a vast area covering 363 203 km^2 which is nearly a third of the country's land area. The province is classified as a dry arid region with fluctuating temperatures and varying topographies consisting of six biomes (Mucina and Rutherford 2006). Savanna and Desert biomes dominate the northern part while the west is dominated by the Succulent Karoo biome. The central part of the province is dominated by the Nama Karoo biome. The study area is situated in the north-western part of the province and is about 5 km from the small town of Griekwastad and 170 km from the city of Kimberley. The study area (Figure 3.1) covers plains with a variety of acacia, such as Buffalo-thorn Jujube (*Ziziphus mucronata*), Camel Thorn (*Acacia erioloba*), Sweet-thorn Acacia (*Acacia karoo*) and Black-thorn Acacias (*Acacia mellifera*) and a mixture of grasses such as Kalahari Coach (*Stipagrostis amabilis*), Giant Stick Grass (*Aristida meridionalis*) and Lehmann's Lovegrass (*Eragrostis lehmanniana*) dominating the grassy plains (Van den Berg et al. 2014). The main activity in the study area is animal farming mainly grazing from cattle and goats. Horses and donkeys are also prevalent in the area and are used as a cheaper mode of transportation. These animals ingest the nutritious seed pods of mesquite and excrete viable seeds in their droppings, thus helping to spread mesquite over shorter distances enabling extremely dense invasions of mesquite. As long as the seeds are not damaged by chewing, the process of digestion actually helps germination, especially since the seeds are deposited in moist, nutrient-rich dung (Zeila 2011).

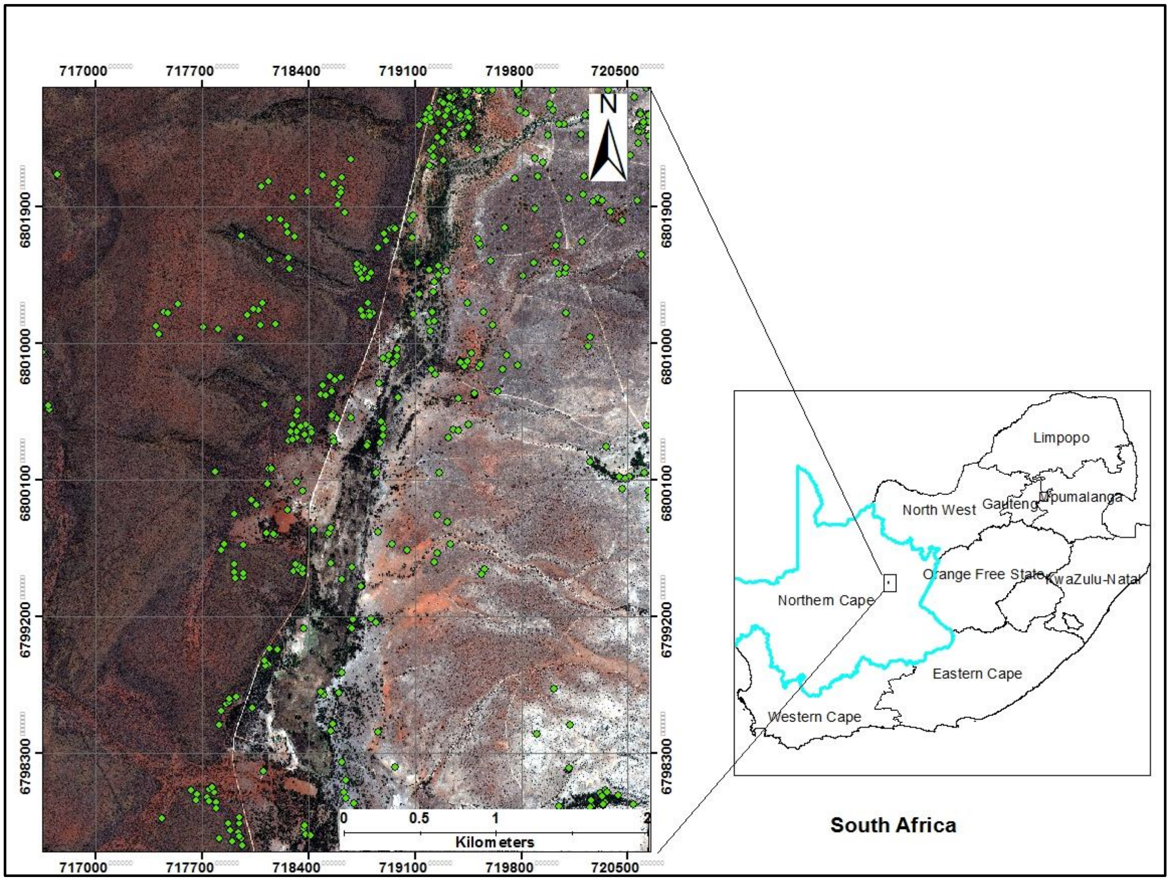


Figure 3.1: A true-colour composite WorldView2 image showing the location of the study area and some of the field samples presented as green dots.

3.3.2 Identification of mesquite and other co-existing tree species

The most common tree species associated with mesquite in the area were identified in the field in summer of 2015 through field surveys. In total, three main co-existing species associated with *Prosopis glandulosa* have been identified as most common tree species and these are *Acacia karoo*, *Acacia mellifera* and *Ziziphus mucronata.* Colour digital photographs of the species were taken, as well as the collection of samples from each of the species, including mesquite and were sent to the C. E. Moss Herbarium Department at the School of Animal, Plant and Environmental Sciences, University of the Witwatersrand to confirm the species identification. *Acacia karoo, Acacia mellifera* and *Ziziphus mucronata* are all indigenous plants in South Africa. They are spread throughout the country but are most dominant in the North-West, Limpopo and Northern Cape Provinces of South Africa. *Acacia mellifera* which is known as Black Thorn in southern Africa usually occurs as a multi-stemmed shrub up to 3 m high and sometimes it can grow as a tree to a height of 7 m (Hagos and Smit 2005). The species is well adapted to dry and arid environmental conditions and it may grow in a variety of soil types ranging from Kalahari sands to heavy and clayey soil (Smit et al. 1999).

Acacia karoo is known as Sweet thorn and it is widely distributed across different habitats of South African region including dry thornveld, river valley scrub, bushveld, woodland, grassland, river banks and coastal dunes of South Africa, Namibia, Angola, Botswana, Zambia and Zimbabwe (Taylor and Barker 2012). *Acacia karoo* may grow as a shrub or small to medium-sized tree to height of 12 m. It is a pioneer species and has the ability to encroach rapidly into grassland grazing areas, and it considered to be the most important woody invader of grasslands in South Africa (Taylor and Barker 2012).

Ziziphus mucronata also known as the Buffalo thorn, is a tropical fruit tree species which is native to the Indo-Malaysian region of South-East Asia, southern Africa, China, Australasia and the Pacific Islands. It is a spiny, evergreen and fast-growing tree with a spreading crown, stipular spines and many drooping branches (Priyanka et al. 2015). The tree may grow to heights between 3 m and 12 m. The leaves are readily eaten by camels, cattle and goats (Priyanka et al. 2015).

3.3.3 Field spectroscopy measurements

Following the identification of the common tree species associated with mesquite, field spectral reflectance measurements were collected at canopy level over four days from 27 to 30 March 2015 between 10:00 am and 02:00 pm under sunny and cloudless conditions. The spectral reflectances were collected from mesquite and the common tree species using the Spectral Evolution® RS-3500 Remote Sensing Portable Spectroradiometer Bundle. The Spectroradiometer has a wavelength range of 350 to 2500 nm with a spectral resolution 1 nm that is resampled from inherent spectral resolutions of 3 nm at 700 nm, 8 nm at 1600 nm and 6 nm at 2100 nm (Evolution 2012). Each vegetation plot (6 m x 6 m) of *Prosopis* and its co-existing species was sampled by cutting three to six branches from the top canopy. Piles of the branches from each sample were placed randomly on top of a black thick cardboard and the leaf reflectance was immediately measured at a nadir-looking angle at approximately 25cm above the branches (Bian et al. 2013). In order to derive representative reflectance spectra for each canopy (Figure 3.2) about 15 to 20 measurements were collected from each pile of branches by moving randomly over each canopy. Due to interferences such as change in atmospheric conditions as well as irradiance of the sun, a white reference spectral measurement was used every 10 to 20 measurements on the calibration panel to counterbalance any changes. The spectral measurements (15 to 20) from each plot were then averaged to represent the spectral reflectance of each vegetation plot (Figure 3.2). In total, 498 vegetation plots were sampled; 133 for *Prosopis glandulosa*, 108 for *Acacia karoo*, 133 for *Acacia mellifera* and 124 for *Ziziphus mucronata* (Table 3.1). In addition to the field spectral measurements, metadata giving information of general weather conditions, land cover class and coordinates were recorded for each point measured by the Spectroradiometer.

Table 3.1: Sample plots of *Prosopis glandulosa* and its co-existing species.

Species	**Training samples (70%)**	**Test samples (30%)**	**Total samples**
Prosopis glandulosa (PR)	93	40	133
Acacia karoo (AK)	76	32	108
Acacia mellifera (AM)	93	40	133
Ziziphus mucronata (ZM)	87	37	124

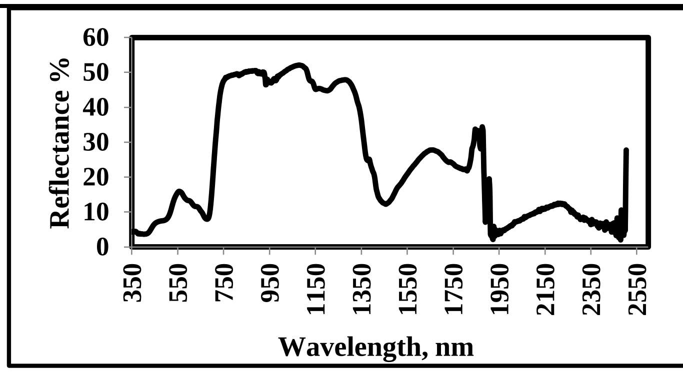
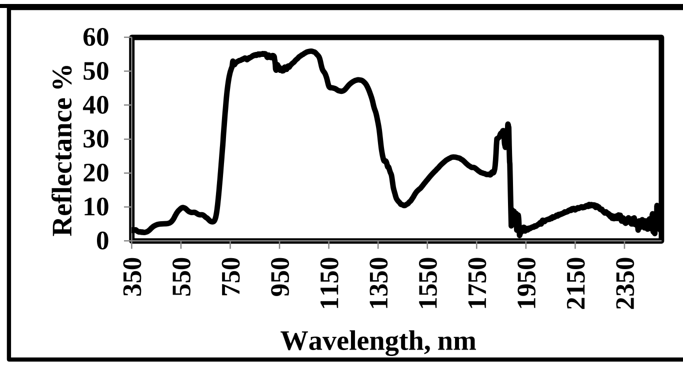
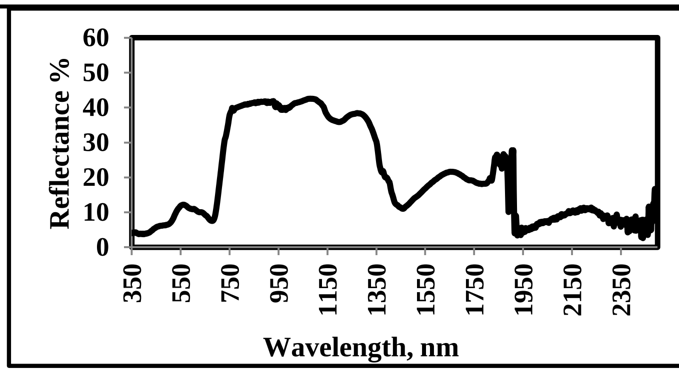
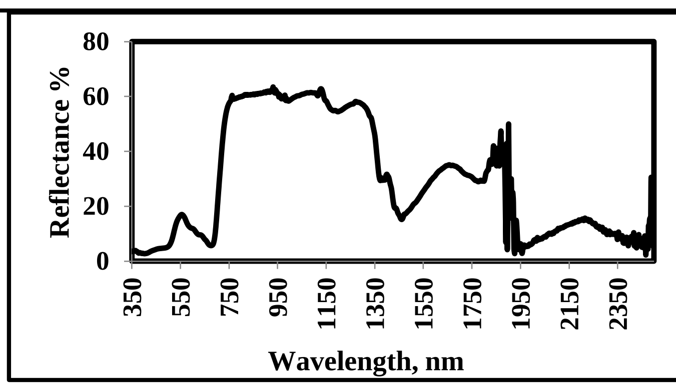
Species	**Species Spectra**
***Prosopis glandulosa* (PR)** 	
***Acacia karoo* (AK)** 	
***Acacia mellifera* (AM)** 	
***Ziziphus mucronata* (ZM)** 	

Figure 3.2: Images and spectra of *Prosopis glandulosa* and its co-existing species.

3.3.4 Field spectroscopy data analysis

Due to noise in the reflectance spectra mainly caused by atmospheric water absorption (Zhao et al. 2007), reflectance values of 325 wavelengths from three spectral regions: between 904.5 – 994.5 nm (100 bands); between 1807.2 – 2027.7 nm (90 bands) and between 2182.4 – 2503.4 nm (135 bands), were removed from the species spectra. Thus, only 1825 wavelengths were used for the spectral analysis. To reduce the 'curse of dimensionality' of hyperspectral data, traditional random forest (RF) (Breiman 2001) and the new guided regularized random forest (GRRF) developed by Deng and Runger (2013) were adapted for variable importance measurements and feature selection respectively (Figure 3.3).

3.3.5 Random forest classifier and variable importance measurement

Over the last decade, the random forest algorithm (RF) has been increasingly used to provide a new means in classifying multispectral and hyperspectral remote sensing data for different applications. RF is an ensemble decision trees developed by Breiman (2001) in the field of machine learning to improve classification and regression trees (CART). The algorithm combines bootstrap sampling to construct a large set of decision trees based on model aggregation ideas. Each tree contributes with a single vote for the assignment of the most frequent class to the input data. The two sources of randomness include; random inputs and random features. The algorithm benefits from the two powerful techniques; bagging and random subspace selection (Lin et al. 2011). Firstly, random forest builds many binary decision trees (*ntree*) to enhance the diversity of the classification trees using several bootstrap samples with replacement that are drawn from the original observations. Each single decision tree contributes with a single vote for the assignment of the most frequent class to the input data. The true classification is determined in accordance with the maximum number of votes from the collection of trees. The samples that are not in the bootstrap sample are called out-of-bag (OOB) sample. The OOB sample (about 30% of the total data) can be used to estimate the misclassification error and variable importance. Secondly, at each node, a given number of input variables (*mtry*) are randomly chosen from a random subset of the features. To ensure a lower similarity (i.e. diversity) between the individual trees and thus a low-bias, each single tree is grown without pruning on the original bootstrap sample (Breiman 2001; Genuer et al. 2010; Lin et al. 2011). To improve the classification accuracy, RF parameters (i.e. *mtry* and *ntree*) have to be optimized (Breiman 2001). The default number of trees (*ntree*) is 500, while the default value for the number of variables (*mtry*) is √P, where P equals the number of predictor variables within a dataset (Breiman 2001).

For this study, a grid-search approach based on the OOB estimate of error was used to find the optimal combination for these two parameters (Tian et al. 2009). The grid search value for *mtry* was varied from 1 to 10 for with a single value interval, while the range of the grid search value for the ntree parameter was varied from 500 (default value) to 10000 with an interval of 500 (20 steps). Additionally, random forest provides an internal measure of variable importance using three different methods namely, the number of times each variable is selected, the Gini importance and the permutation accuracy importance measure (Strobl et al. 2007). In this study the Gini importance measure was adopted. The predictive power of each variable is quantified by a score (called Gini importance or Gini Contrast), depending on the importance it gained over all the trees in the random forest (Breiman 2001). The ensemble does this by using the *Gini* index computed using the following equations 1, 2 and 3. The Gini index at a node φ, denoted by $G\ (\varphi)$, is given as:

$$G\ (\varphi) = \sum_{c=1}^{numclass} \hat{\rho}_c\ (1 - \hat{\rho}_c) \tag{1}$$

Where $\hat{\rho}_c$ is the proportion of observations belonging to class c at node φ. The information gain of feature f_i based on Gini index on node φ is then computed as:

$$IG\ (f_i,\varphi) = G\ (\varphi) - \alpha_L\ G\ (\varphi^L) - \alpha_R\ G\ (\varphi^R) \tag{2}$$

Where φ^L and φ^R denote the left and right child nodes respectively of node φ in a tree, and α_L and α_R are the proportions of observations in the left and right child nodes respectively. As mentioned previously, in an RF model, a random subset of features is chosen at each node and the feature with the highest information gain is used for splitting.

The overall importance score of feature f_i is given by:

$$IS\ (f_i) = \frac{\sum_{\{split(f_i)\}} IG\ (f_i\ ,\varphi)}{ntree} \tag{3}$$

$\{split(f_i)\}$ is the set of all nodes over all trees ($ntree$) where f_i is used for splitting.

Basically, variables associated with the OOB sample are randomly permuted and classification trees are grown on the modified dataset. The permuted feature was used to predict the response and obtain the accuracy. If the wavelength is initially important in the final prediction the accuracy will drop significantly after the permutation. Thus, the difference in prediction accuracy with and without permuting the feature (wavelength) was used in this study to measure the importance of the feature. A key advantage of the random forest variable importance is that it not only deals with the impact of each variable individually, but also looks at multivariate interactions with other variables (Strobl and Zeileis 2008). Several approaches such as Kursa and Rudnicki (2010) and Diaz-Uriarte and Alvarez de Andres (2006) have built on the above measure to identify the relevant set of features. However, they are either computationally expensive or do not find non-redundant set of features (Figure 3.4).

3.3.6 Feature selection using guided regularized random forest

Random forest has been intensively used to reduce the high dimensionality of hyperspectral data while returning relatively good accuracy levels (Abdel-Rahman et al. 2012, 2013; Adam et al. 2012b; Vincenzi et al. 2011). However, these studies have shown that although RF provides insight into the importance of each variable in the classification process, it fails to automatically select the key number of variables that could yield the lowest error rate (Adam et al. 2012b). To address this shortcoming, a regularization framework that can be applied to random forest (regularized random forest) and boosted trees (regularized boosted trees) was developed by Deng and Runger (2013). This regularization framework builds one model that reduces training time significantly by avoiding the selection of a new feature for data splitting in a tree node when that feature produces similar information to features already selected. This method is called the guided regularized random forest (GRRF). GRRF utilizes the raw feature importance scores obtained from an initial RF model. The parameters involved in the GRRF model are $mtry$, $ntree$ and τ which are optimized over a grid search using a 10-fold cross-validation on the training set. The importance score of a feature in RF is obtained by averaging the information gain (based on Gini index) over all nodes across all trees obtained where the feature is used to split on. For the purpose of GRRF, the raw importance scores obtained from RF are normalized for each feature using equations 4, 5 and 6.

$$NORM\_IS(f_i) = IS\ (f_i)/(max_{i=1}^{F}\ IS\ (f_i)) \tag{4}$$

Also the corresponding information gain is computed as:

$$IG_{GRRF}\ (f_i\ ,\varphi) = \begin{cases} IG\ (f_i\ ,\varphi)\ f_i\ \in\ F^* \\ \mu_i IG\ (f_i\ ,\varphi) otherwise \end{cases} \tag{5}$$

Where F^* is the set of indices of features that were used for splitting in previous nodes. For root node, $F^* = \emptyset$. μ_i is an importance co-efficient for feature f_i calculated as:

$$\mu_i = (1 - \tau) + \ \tau IG_{GRRF}\ (f_i\ ,\varphi) \tag{6}$$

τ is the regularization constant. When τ =0, we obtain the same results as from RF.

Similar studies have shown that GRRF is effective in selecting high quality feature subsets while maintaining predictive accuracies (Deng and Runger 2013). Interested readers are referred to, for example Deng and Runger (2012), Deng and Runger (2013) and Deng (2013) for comprehensive description on GRRF theory, principles and mathematical formulation.

3.3.7 Accuracy assessment

The accuracy of the RF classifier was assessed by using the independent test dataset (30%). OOB, which provided an unbiased estimate of the internal RF error, was used to assess the misclassification. A confusion matrix was subsequently constructed to compute the overall accuracy (OA), user’s accuracy (UA), and producer’s accuracy (PA) as criteria for evaluating the generalization ability (accuracy) of the RF classifiers (Mather and Tso 2003). OA is a ratio (%) between the number of correctly classified samples and the number of test samples, while UA represents the likelihood that a sample belongs to specific class and the classifier accurately assigns it such class. PA expresses the probability of a certain class being correctly recognized.

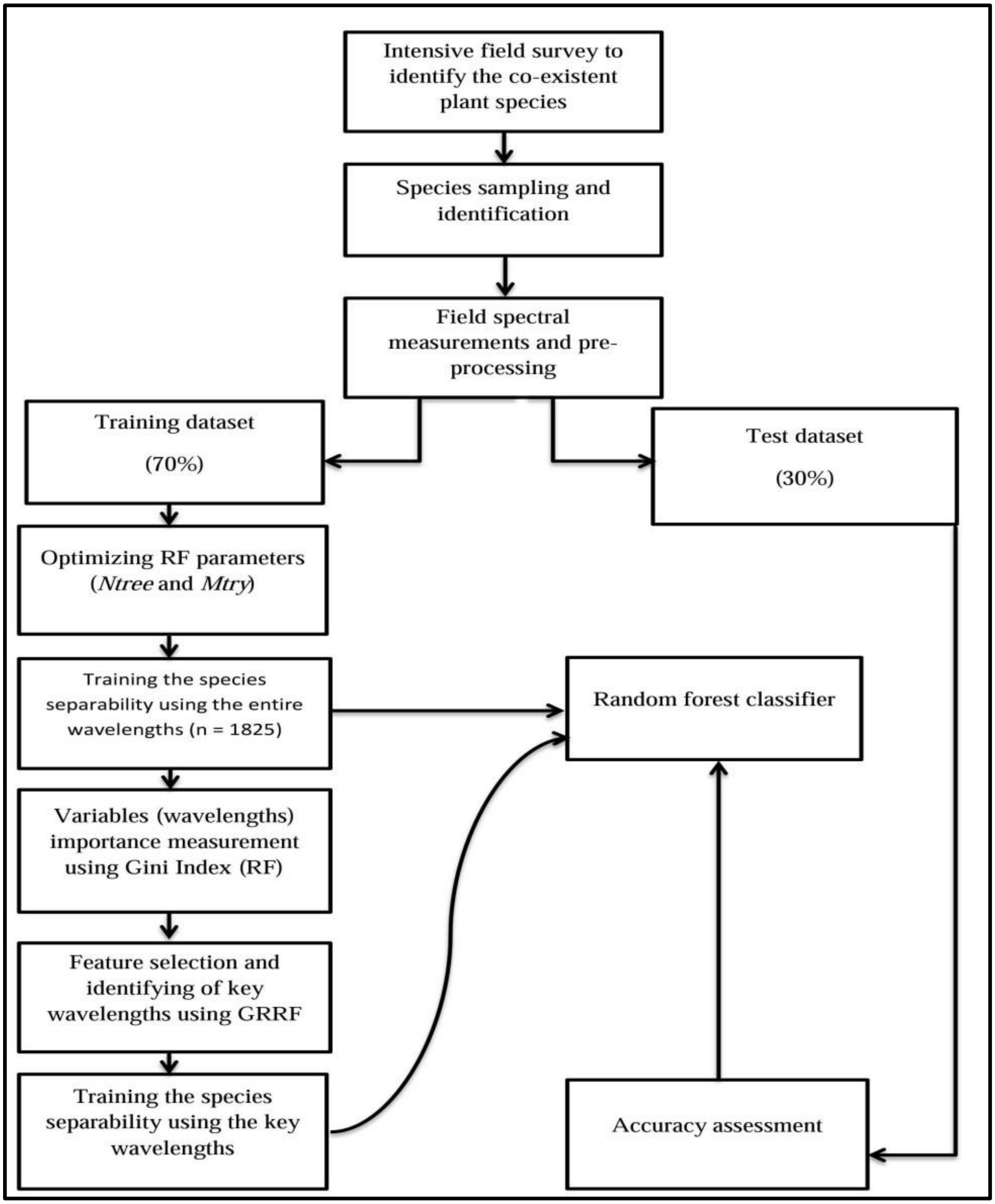


Figure 3.3: Flowchart describing the random forest (RF) and guided regularized random forest (GRRF) models used in this study.

3.4. Results

3.4.1 Variables importance measurement and selection

The ordinary RF classifier was able to determine the importance of each wavelength in discriminating between the four species namely, *Prosopis glandulosa*, *Acacia karoo, Acacia mellifera* and *Ziziphus mucronata* as shown in Figure 3.4. Based on the mean decrease in Gini index, the most important wavelengths are located across the electromagnetic spectrum. For example, the wavelengths 343.7 nm and 719.4 nm are the most important wavelengths in the visible (400 – 700 nm) and red edge (690 - 720 nm) regions respectfully. Many most important wavelengths for discriminating among the species are also found in the near infrared region. These are located between 1399.6 and 1407 nm. Figure 3.4 indicates that the top important wavelength is located at 1410.9 nm.

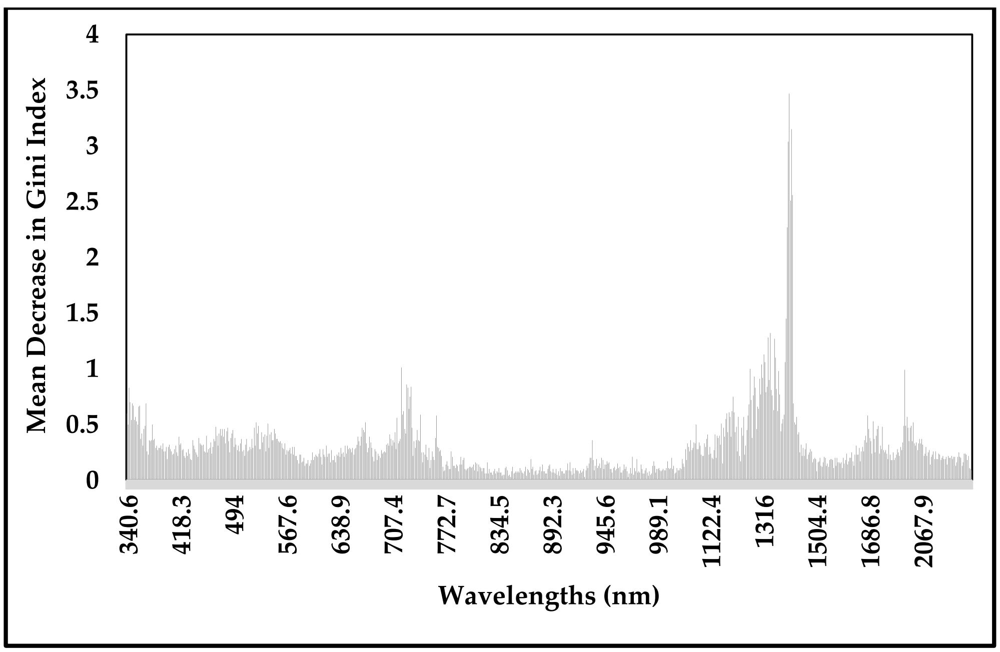


Figure 3.4: The importance of wavelengths as measured by the traditional Random Forest using mean decrease in *Gini index*. The most important variables are those with highest mean index.

These importance scores from the random forest were used to enable GRRF's selection of subset wavelengths that can better discriminate between the four different species. GRRF was able to identify 11 optimal wavelengths that yield lowest OOB error. These optimal wavelengths are located at 356.3 nm, 468.5 nm, 531.1 nm, 665.2 nm, 1262.3 nm, 1354.1 nm, 1361.7 nm, 1376.9 nm, 1407.1 nm, 1410.9 nm and 1414.6 nm (Figure 3.5). These wavelengths were then used as input variables for RF classifier model to discriminate between *Prosopis* and co-existing species.

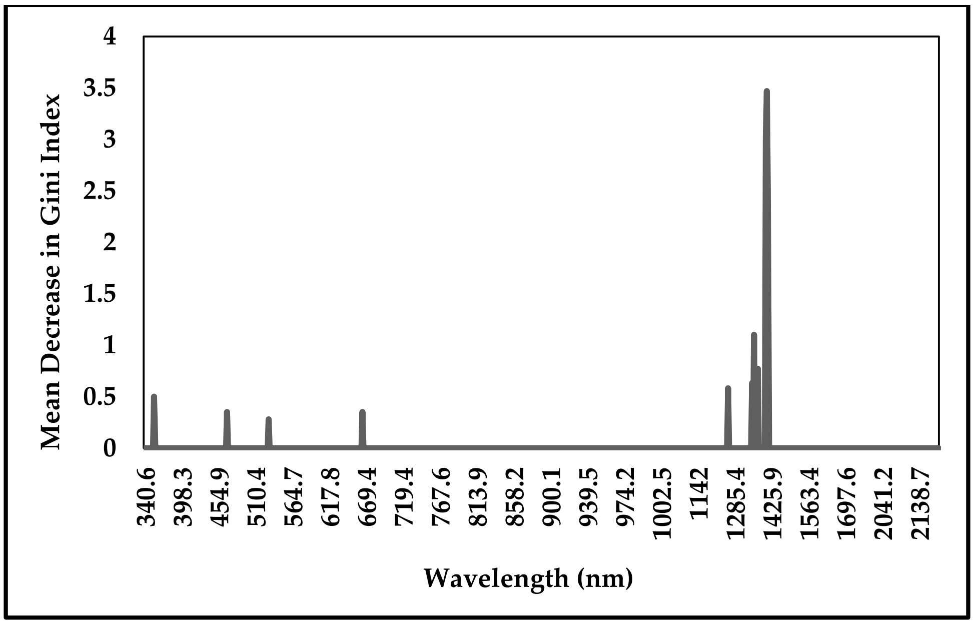


Figure 3.5: Wavelengths selected by Guided regularized random forest based on the importance scores as measured by the traditional Random forest.

3.4.2 Accuracy Assessment

The best wavelengths selected by GRRF (n=11) were input into random forest classifier. The lowest OOB error of 11.41% was obtained using best combination of *ntree* and *mtry*. The classification model yielded an overall accuracy of 88.59% using the selected wavelengths (n=11), compared with an overall accuracy of 79.19% when the total number of wavelengths (n=1825) was used (Table 3.2). A comparison between the producer and user accuracies for the two datasets is shown in Table 3.3 for each vegetation species.

Table 3.2: Confusion matrix showing the overall classification accuracy and kappa statistic for discrimination among the four vegetation species; *Prosopis glandulosa* (PR), *Acacia karoo* (AK), *Acacia mellifera* (AM), and *Ziziphus mucronata* (ZM). The error was calculated using Out-of-Bag method and the test dataset.

Class	**Using 1825 wavelengths**					**Class**	**Using the selected 11 wavelengths**				
	AK	**AM**	**PR**	**ZM**	**Total**		**AK**	**AM**	**PR**	**ZM**	**Total**
AK	25	2	3	2	32	**AK**	27	1	2	2	37
AM	2	34	3	1	40	**AM**	2	36	2	0	40
PR	4	3	30	3	40	**PR**	1	1	36	2	40
ZM	3	1	4	29	37	**ZM**	2	0	2	33	37
Total	34	40	40	35	149	**Total**	32	38	42	37	149
OA = 79.19% Kappa = 0.7201						OA = 88.59% Kappa = 0.8524					

Table 3.3: Producer's accuracy (%) and User's accuracy (%) of the four classes (*Prosopis glandulosa* (PR), *Acacia karoo* (AK), *Acacia mellifera* (AM), and *Ziziphus mucronata* (ZM)) using all the variables (1825 wavelengths) and the most important variables (11 wavelengths).

Class	**Using 1825 wavelengths**		**Class**	**Using 11 wavelengths**	
	Producer's accuracy (%)	User's accuracy (%)		Producer's accuracy (%)	User's accuracy (%)
AK	73.53	78.13	AK	84.38	84.38
AM	85.00	85.00	AM	94.74	90.00
PR	75.00	75.00	PR	85.71	90.00
ZM	82.86	78.38	ZM	89.19	89.19

3.5. Discussion

Many studies have demonstrated the importance of spatial data in managing and controlling invasive plant species (Amaral et al. 2015; Asner et al. 2008; Shouse et al. 2013). Since the 1800s, the invasion of the taxa of *Prosopis* has posed significant threat to species diversity and caused substantial socio-economic damages world-wide (Zimmermann 1991). Many plant invasion control methods namely biological, chemical and mechanical have been tried and tested over the years to reduce the impacts of mesquite with little success as the plant is still spreading at a rate of 8% per annum in South Africa (Le Maitre et al. 2004; Van den Berg 2010; Versfeld et al. 1998). The lack of the timely and accurate spatial data on the dynamics of the spread has been one of the major challenges for control (Wise et al. 2012b). This is due to the complexity of the mesquite ecology such as biology, rapid spread and many uncertainties associated with its niche colonisation (Shiferaw et al. 2004). This study investigated the potential use of hyperspectral data in discriminating mesquite from three co-existing species in an arid environment. Results from this study show that mesquite can be accurately discriminated from other species in an arid environment of South Africa using hyperspectral data and machine learning algorithms.

The study integrated the traditional random forest and the newly developed guided regularized random forest for hyperspectral variable selection in a multiclass classification. The traditional RF was used successfully to provide the variable importance measures to guide the regularised feature selection process. It was expected to find many wavelengths share similar Gini information and score at a node, due to the high autocorrelation between neighbouring wavelengths (1 nm interval) (Kumar et al. 2003). However, the GRRF method reduces the high dimensionality of the hyperspectral data while ensuring that such dimensionality reduction would not cause any loss of important information relevant to the object under study (Adam and Mutanga 2009). Many researchers have used the random forest algorithm as a dimensionality reduction tool in different hyperspectral remote sensing applications (Abdel-Rahman et al. 2014; Adam et al. 2012b; Chan and Paelinckx 2008; Lin et al. 2011; Zhang et al. 2009). However, studies have shown drawbacks on the use of random forest as a tool to measure variable importance as well as variable selection method (Adam et al. 2013; Adjorlolo et al. 2013b). Therefore, in this study we introduced a new developed method which has never been tested before in hyperspectral variables selection. This new developed method (GRRF) was able to eliminate the irrelevant and redundant wavelengths and select key wavelengths (n = 11) out of 1825 wavelength on one iteration with less computational processes. Previous variables selection method was based on using varSelRF to build multiple RF models and iterations to add feature(s) with the highest importance scores(s) (forward variables selection) or to eliminate feature(s) with the least importance scores(s) in a backward variable selection method (Ismail and Mutanga 2011; Mansour et al. 2012). Such methods are computationally expensive and are not applicable in a large number of features (Deng and Runger 2013). The selected wavelengths produced lowest OOB error than the complete feature set (n = 1825 wavelengths). It is also notable that the selected wavelengths are distributed across the entire noise free spectrum. This is because the regularization in GRRF does not select a new feature for splitting the data in a tree node if the new feature is similar in terms of information gain to the one that was already selected (Deng and Runger 2012). Such methods allow the exploration of the rich information content in hyperspectral data across the spectrum region rather than selecting only highly correlated features with redundant information (Adam et al. 2014). The most important wavelengths selected by GRRF were at the visible and red edge (356.3 nm, 468.5 nm, 531.1 nm and 665.2 nm) and the short-wave infra-red (1262.3 nm, 1354.1 nm, 1361.7 nm, 1376.9 nm, 1407.1 nm, 1410.9 nm and 1414.6 nm) regions of the electromagnetic spectrum. The visible region of the spectrum is greatly affected by the selective absorption of the photosynthetic pigments (Ceccato et al. 2001a). The red edge region is the region in which the effect of vegetation biochemical is most relevant (Adjorlolo et al. 2013a). The short-wave infrared (SWIR) are affected by water properties associated with vegetation such as leaf area index, strong leaf or canopy liquid water absorption and macronutrients (Carter 1994; Ceccato et al. 2001b; Ghulam et al. 2007).

The new variable selection method used in this study was first developed and tested by Deng and Runger (2013) in a binary classification. The method has also shown a competitive accuracy performance in multiclass classification in this study. Following the recommendation of Deng and Runger (2013), the selected wavelengths by GRRF (n = 11) were input into RF classifier to discriminate between the *Prosopis* and other species (n = 4). This was due to the fact that the trees in GRRF are not designed independently as feature selection and they may therefore have a higher variance than RF (Deng and Runger 2013). The wavelengths selected by GRRF (n = 11) yielded high classification accuracy in RF classifier compared with the entire wavelengths (n =

1825). This was expected due to the fact that the redundant variables in a model-based analysis decrease the performance of the classifiers because the noise in the redundant data can cause convergence instability of the classification models (Bajcsy and Groves 2004).

This high overall accuracy achieved in this study shows the potential use of hyperspectral remote sensing for mapping *Prosopis* at species level and therefore provides more detail about the spatial dynamics of the *Prosopis* invasion. Such details are useful for effective management of the species (Wise et al. 2012b). Previous attempts of mapping *Prosopis* were carried out using multispectral data such as Landsat and some environmental data to evaluate the susceptibility of certain areas to mesquite invasion (Van den Berg et al. 2014). Such approaches are suitable to characterize *Prosopis* invasion if the plant has large spatial coverage and thus are unable to discriminate the species from other vegetation species at fine scales. In contrast, the use of higher spatial and spectral resolution data such as the one used in this study has a great potential in fighting the invasion of the species, since species-level identification is achieved satisfactorily.

3.6. Conclusions

By considering the results from the study it can be concluded that:

1. One of the major problems in controlling mesquite has been the presence of mixed stands that consist of alien *Prosopis* mixed and indigenous species. *Prosopis glandulosa* can be accurately detected from its co-existing species namely *Acacia karoo* (that is also structurally similar), *Acacia mellifera* and *Ziziphus mucronata* using hyperspectral data. Such potential data could provide environmental managers and ecologists insight into the development of possible appropriate spatio-temporal management practices to better control the invasive spread of mesquite.
2. The problem of high dimensionality associated with spectroscopy data processing can be reduced considerably by making use of the newly developed GRRF method. The new GRRF method created high quality feature variables for the traditional RF classifier and can thus be seen as a more efficient and effective feature selection tool to reduce high dimensionality in spectroscopy data. However, this assertion should receive considerable additional testing and comparison with the commonly used variable selection methods before it is accepted as a substitute for reliable high dimensionality reduction.

3. The wavelengths selected by GRRF showed that the greatest discriminatory power of *Prosopis* from other species across the spectrum regions of mainly visible, red edge and short-wave infrared regions. These wavelengths are located at 356.3 nm, 468.5 nm, 531.1 nm, 665.2 nm, 1262.3 nm, 1354.1 nm, 1361.7 nm, 1376.9 nm, 1407.1 nm, 1410.9 nm and 1414.6 nm.

Overall, the results of this study offer the potential of using remote sensing to guide the physical, biological and chemical control of *Prosopis* invasion. The results of this study still however need to be tested in different landscapes to establish a good understanding of spectral characteristic of *Prosopis* and other co-existing vegetation at species level. In addition, more studies are still needed to upscale these results to airborne or space-borne sensor resolutions to determine the optimal spectral and spatial resolutions to detect *Prosopis* taxa. These studies should consider the canopy structures of the species as well as the understorey and soil background characteristic.

3.7 Acknowledgments

A great thank you to Mr. Nico Byleveldt of the Department of Environmental Affairs, Northern Cape, South Africa; and Mr. Shorty Thebe of the Department of Water Affairs, Kimberley, South Africa for their assistance in locating the appropriate study area as well as organising access to the study area. A special thanks is given to Dr. Solomon Tesfamichael for the assistance in field data collection and analysis. Further gratitude is extended to Dr. Houtao Deng and Mr. Anshuman Sahu of Hitachi America Big Data Laboratory, Santa Clara, California, USA for their assistance in the statistical analysis of the data. We thank the R development core team for the open source packages used for the statistical analysis.

CHAPTER FOUR

Mapping mesquite (*Prosopis glandulosa*) and its co-existing species using a high resolution image

This chapter is based on:

1. **Mureriwa**, N., and Adam E., (2015). Mapping *Prosopis glandulosa* (mesquite) invasion and its coexisting species in the semi-arid environment of South Africa using Worldview-2 imagery and machine learning classifiers. (ACRS 2015) *Conference proceedings*. The 36^{th} Asian Conference on Remote Sensing 2015. Crowne Plaza, Quezon City, Metro Manila, Philippines.
2. **Mureriwa**, N., and Adam, E., (In review). Mapping *Prosopis glandulosa* (mesquite) invasion in the semi-arid environment of South Africa using high resolution WorldView-2 imagery and machine learning classifiers. Journal of Arid Environment.

4.1. Abstract

Prosopis glandulosa is one of the 44 species of *Prosopis* recognized in the world, of which 40 are native to the Americas. Both accidental and deliberate introductions of the *Prosopis* (mesquite) species occurred due to their wide range of economic importance have facilitated their spread outside their place of origin, making them one of the top 100 worst alien weeds according to the International Union for Conservation of Nature (IUCN) ranking. Over the last decade, a suite of new-generation multispectral imagery such as the RapidEye, Sentinel and WorldView series, with high spatial and spectral resolutions, have emerged. This study investigated the ability of the WorldView-2 (WV-2) imagery in mapping the invasion of the infamous mesquite and its discrimination from the co-existing indigenous species in the semi-arid region of the Northern Cape Province of South Africa using the random forest and support vector machines as classifiers. Our results showed that the eight band multispectral WV-2 imagery was able to detect and distinguish *Prosopis glandulosa* effectively from the three co-existing indigenous species of acacia with an overall accuracy of 86 % at 2 m spatial resolution. The findings of this study provide a new insight for an economically feasible approach using the multispectral WV-2 sensor in mapping the encroachment and extent of invasive alien plants with similar accuracy as those of hyperspectral imagery.

Key words: Multispectral satellite; *Prosopis glandulosa*; Invasive species; WorldView 2; Algorithm; Classifiers

4.2. Introduction

Invasive plant species are defined as plants that colonize areas outside their natural ranges with or without human interference (Shiferaw et al. 2004). The introduction of invasive species into new habitats has increasingly been a global phenomenon due to the increase of human mobility and the socio-economic dynamics (Shiferaw et al. 2004). Invasive alien species can become very aggressive by taking over land for example in their new areas of introduction, if the environmental conditions are favourable (Kolar and Lodge 2001). Many studies have shown that alien plant invasions impact the structure and function of ecosystems and cause substantial changes in the socio-economic well-being in many parts of the world (Shackleton et al. 2014b; van Wilgen et al. 2012).

A very good example of such invasive species is the *Prosopis glandulosa* var. *glandulosa* J. Torrey, which is one of the 44 recognized species in the genus (Pasiecznik et al. 2004). *Prosopis* is native to North and South America and has been introduced intentionally to different countries across the world for different purposes (Mwangi and Swallow 2005). Among these are to provide shade and fodder in the arid regions of Australia and South Africa (Zimmermann 1991); for sand-dune stabilization, afforestation as well as fuelwood supply in Sudan (Ghazanfar 1996); for livestock fencing in Malawi (Chikuni et al. 2005); for local greening, ornamental cultivation and soil stabilization in many Middle Eastern countries (Ghazanfar 1996); for rehabilitation of old quarries and later for afforestation and the provision of fuelwood and fodder in Kenya (Choge et al. 2012); for fuelwood production and rehabilitating degraded soil in India (Pasiecznik et al. 2001; van Klinken et al. 2007); and for vegetation trials in Spain (Elfadl and Luukkanen 2006; Laxén 2007). Apart from this, accidental cross-border inductions between neighbouring countries have also occurred (Shackleton et al. 2014b). It is, for example, believed that the plant was introduced inadvertently into Botswana, Nigeria and Yemen through livestock trading (Geesingis et al. 2004; Pasiecznik et al. 2001).

Despite the positive impacts of *Prosopis*, the species tends to form dense impenetrable thickets associated with unfavourable impacts on human economic activities (Pasiecznik et al. 2001). For example, *Prosopis* overruns grazing land; negatively affects biodiversity and excessively consumes surface and ground water (Pasiecznik 1999). Numerous taxa of *Prosopis* are rapidly invading many parts of the world and have sucessfully become dominant by suppressing the native plant species. The rapid invasion of *Prosopis* species is intimately associated with the inherent characteristics of the species, such as large seed production, rapid growth rates and high efficiency in utilizing both surface and ground water. Many *Prosopis* species are able to tolerate extreme temperatures and low rainfall (Shiferaw et al. 2004), and they are not limited by levels of alkaline and saline or infertile soils (Shackleton et al. 2014b). A recent comprehensive global review of *Prosopis* distribution shows that the plant currently occurs in at least 129 countries (Shackleton et al. 2014b). High climatic suitability for potential invasion of *Prosopis* is also found in many countries in Europe and Africa (Maundu et al. 2009). Therefore, the *Prosopis* are listed as one of the top most aggressive invasive species in countries such as Australia, India, Ethiopia, Sudan and South Africa (Shackleton et al. 2014b). Hence, in 2004 *Prosopis* was rated the world's top 100 least wanted species by the Invasive Species Specialist Group of the International Union for Conservation of Nature (IUCN) (Henderson 2001; Mwangi and Swallow 2005). Studies have shown that the negative impacts associated with *Prosopis* were perceived to exceed benefits (Shackleton et al. 2015). Thus, an effective management of intervention to control the existing *Prosopis* invasion and to mitigate its negative impacts is of a paramount necessity.

Different methods have been implemented over the years to control *Prosopis* invasion in a few countries. These include mechanical removal of the plant, chemical control methods such as herbicide treatment of cut stumps, foliar spraying of saplings and burning (Harding 1987; van Klinken et al. 2009) and biological control methods using seed-eating beetles to curb the *Prosopis* further spread (Zachariades et al. 2011). These methods however, have not been very successful due to the high costs and lack of knowledge on key aspects of *Prosopis* species such as spatial dynamics, scale of the invasion, and reasons of their introduction (Shackleton et al. 2014b). Only 13% of the countries with a high invasion rate have detailed spatial data on the distribution or percent cover of *Prosopis*. Therefore, up to date temporal and spatial information on the spatial distribution of *Prosopis glandulosa* (mesquite) invasion and its negative impacts on the ecosystem services is crucial for effective management (Nie et al. 2012).

Field based methods for mapping vegetation species are generally costly and time-consuming and sometimes impossible to accomplish due to poor accessibility or extensive land coverage (Hoshino et al. 2012). On the other hand, remote sensing techniques offer an economic and effective technique that produces timely and accurate information for mapping vegetation species.

Both multispectral and hyperspectral data have been used in mapping vegetation species in different landscapes (Akasheh et al. 2008; Harvey and Hill 2001; Lawrence et al. 2006; Peerbhay et al. 2013; Saatchi et al. 2008). Multispectral data such as Landsat TM and SPOT imagery have been used to identify general vegetation classes and communities (Harvey and Hill 2001; Li et al. 2005). However, the utility of commonly used multispectral data is limited by the lack of spectral and spatial resolutions. On the other hand, the acquisition of narrow and contiguous spectral channels by hyperspectral sensors allow the detection of vegetation at species level, which otherwise would be masked by the broad bands of multispectral sensors (Adam et al. 2010; Goetz 2009). Nonetheless, the use of hyperspectral data comes with its own limitations in terms of cost, time, availability, processing and high dimensionality of data (Goetz 2009).

Over the last decade, a suite of new-generation imagery such as RapidEye, Sentinel series and WorldView (WV) series have emerged. These imageries are characterized by high spatial and spectral resolutions and therefore, provide more details on land cover mapping. Moreover, the development in computer and mathematical sciences have led to more advanced algorithms such as random forest, support vector machines and neural networks, which have greatly improved the digital image processing (Ham et al. 2005; Kumar et al. 2015; Omer et al. 2015b). The aim of this study is therefore to test the use of WorldView-2 imagery and two machine learning algorithms, namely, Random Forest (RF) and Support Vector Machines (SVM) in mapping *Prosopis glandulosa* (mesquite) invasion and the co-existing indigenous species in the semi-arid region of the Northern Cape Province of South Africa.

4.3. Materials and methods

4.3.1. Study area

The study was conducted in the Northern Cape Province of South Africa (Fig. 4.1). The province has an area of about 363 203 km^2 and constitutes a third of the country’s surface area. It is a dry region with extremely fluctuating daily temperatures, varying topographies and is comprised of six biomes namely, the Savanna, Desert, Succulent Karoo, Grassland, Fynbos and Nama Karoo biomes (Mucina and Rutherford 2006). The study area is situated in the north-western part of the province and is about 5 km from the small town of Griekwastad and 170 km from the city of Kimberley. It covers plains with a variety of acacia such as *Acacia erioloba*, *Acacia karoo* and *Acacia mellifera.* The province also consists of a mixture of grasses such as *Stipagrostis amabilis*, *Aristida meridionalis* and *Eragrostis lehmanniana* that dominate the grassy plains of the region (Van den Berg et al. 2014). In addition to these land-cover types, there is also a range of soil types in the area such as the deep-grey calcareous sands, yellow sands and red-yellow apedal soils just to mention a few (Group 1991).

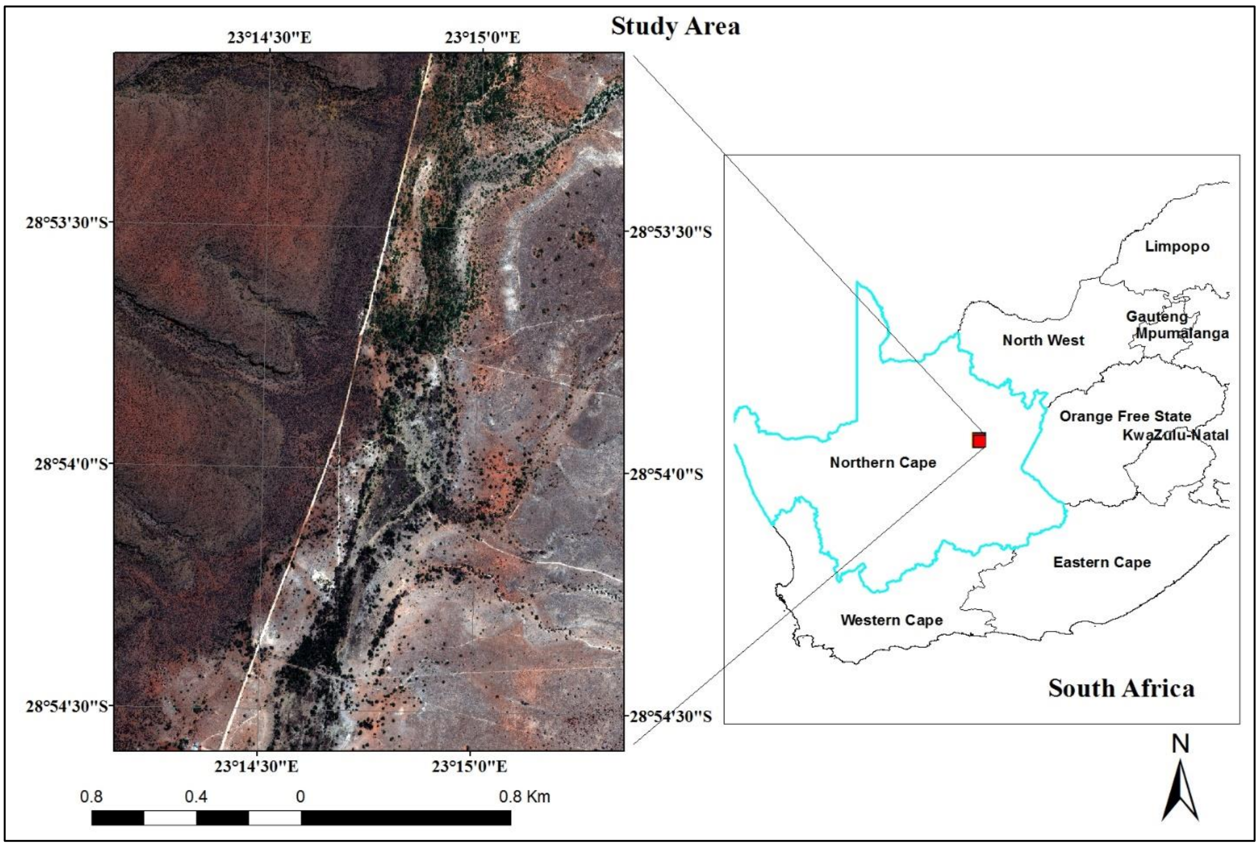


Figure 4.1: A true-colour WorldView-2 image showing the location of study area

4.3.2. Image acquisition and pre-processing

A 2 m spatial resolution WorldView-2 (WV-2) image captured on the 12th of January 2015 under cloudless conditions was used for this study. Digital Globe partnered with Boeing Commercial Launch Services to deliver the WorldView-2 satellite into a sun-synchronous orbit on October 8, 2009 as the first high-resolution 8-band multispectral commercial satellite. The satellite is capable of collecting up to 1 million square kilometres of 8-band imagery per day at 2 m spatial resolution. These eight multispectral bands are: coastal blue (400 to 450 nm), blue (450 to 510 nm), green (510 to 580 nm), yellow (585 to 625 nm), red (630 to 690 nm), red edge (705 to 745 nm), NIR1 (770 to 895 nm), and NIR2 (860 to 1040 nm) and a panchromatic band (450 to 800 nm). The satellite has a 16.4 km swath width and an average of 1.1 days revisit time, and the system offers unsurpassed accuracy, agility, capacity and spectral diversity, which is useful in mapping vegetation quality and quantity, coastal mapping, environmental monitoring, and physical infrastructure delineation (Digital Globe 2010).

The WV-2 image was ortho-rectified using a geo-referenced high resolution (0.5 m) aerial photograph of the study area. The orthorectification was done using 22 ground control points and a first-order polynomial transformation technique. An overall root-mean square error (RMSE) of 0.21% of a pixel was achieved. A visual assessment was carried out to ensure that the WV-2 image was perfectly aligned to the aerial photograph. The fast line-of-sight atmospheric analysis of spectral hyper cubes (FLAASH) algorithm was then used to atmospherically correct the image as described in the Environment for Visualizing Images (ENVI 5.2) 2014 software package.

4.3.3. Defining land-cover classes and reference data collection

Ground reference data were collected during the period of the 5th to the 8th of February 2015; this is about three weeks after the WV-2 image acquisition. Initially eight spectral land-cover classes were generated from the eight bands of WorldView-2 image using the IsoData unsupervised classification tool in ENVI 5.2 to identify the common land cover types and to guide the field data collection. These classes were then regrouped into six broad classes and ground points were randomly generated across the different land cover types, which were used in a GPS to navigate to the field sites. Purposive sampling was also adapted when a random point was not accessible, or to increase the variation of ground data for *Prosopis* and other co-existing species (Adam and Mutanga 2009). The WV-2 false colour composites and the GPS points were used in the field to directly locate and delineate *Prosopis* and the other land cover type classes. The ground reference data was then overlaid on WV-2 image to create regions of interest (ROIs) to train and validate the classifiers (Table 4.1) by randomly splitting the ground reference data into 70% training and 30% validation data sets (Table 4.1).

Table 4.1: Training and validation datasets collected for *Prosopis* and other land-cover classes in the study area.

Land-cover class	**Code**	**Training dataset**	**Validation dataset**	**Total**
Prosopis glandulosa	PRS	58	25	83
Acacia mellifera	AMF	72	31	103
Acacia karoo	AK	71	30	101
White calcareous soil	WS	69	30	99
Red apedal sand	SS	71	30	101
Grassland	GL	50	21	71

4.3.4. Image classification

4.3.4.1 Random Forest classifier

Decision learning trees such as classification and regression trees (CART) are commonly used for data mining and have been one of the most successful methods for supervised classification (Olshen and Stone 1984). To improve the accuracy of CART, Breiman (2001) developed an ensemble learning technique called Random Forest (RF) by introducing the idea of bagging (bootstrap aggregating) to the decision trees. This involves combining multiple decision trees and each tree contributes a single vote for the assignment of the most frequent class to the input data. Many binary classification trees (*ntree*) are built by RF using several bootstrap samples with replacements drawn from the original observations. Samples not in this bootstrap sample are called out-of-bag (OOB) samples. These OOB samples, which are about a third of the total data, can be used to estimate the misclassification error and to measure the importance of each variable in the final model (Breiman 2001; Lin et al. 2010). A given number of input variables (*mtry*) at each node were randomly chosen from a random subset of the features and the best split was calculated by utilizing only this subset of features. No pruning was performed and all trees in the forest are maximally grown trees so as to ensure low bias (Genuer et al. 2010). *Mtry* in this study is defined as the square root of the total number of spectral bands. In order to improve the classification accuracy, RF parameters (i.e. *mtry* and *ntree*) have to be optimized (Breiman 2001; Mutanga et al. 2012b) and the default number of trees (*ntree*) is 500, while the default value for the number of variables (*mtry*) is the square root of the total number of spectral bands used in the study (Breiman 2001). A 10-fold grid-search approach based on the OOB estimate of error was used in this study to find the optimal combination for these two parameters with the *mtry* value being varied from 1 to 5 and the *ntree* parameter varied from 500 to 10,000. The Image RF tool in EnMAP-Box was used to perform the RF classification.

4.3.4.1 Support Vector Machines

Support vector machines (SVM) is a nonparametric supervised machine learning classifier (Cortes and Vapnik 1995a). The algorithm was originally proposed by Vapnik (1979) as a binary linear classifier where the distance of each class from the data points in the training data to the optimal hyperplane or decision boundary is maximized. This in turn minimizes the misclassifications obtained during the training step (Mashao 2003). On the boundaries of the hyperplane there are two support hyperplanes that have data points on their edges called support vectors and these are the ones that define the optimal hyperplane (Mountrakis et al. 2011). In practice it has been found that data of different classes tends to overlap so that a non-linear polynomial is applied to improve on this limitation of linear separability and increase classification accuracy. SVM optimizes the non-linear algorithm through the use of a number of methods with one being the kernel method using the radial basis, which is the most common method used on remotely sensed data to date (Huang et al. 2002; Oommen et al. 2008). Two parameters are required for tuning in the radial basis method, namely, the cost 'sigma (C)', defined as a plenty value that is used for adjusting the error of misclassifying instants of the training dataset, and the kernel width 'gamma (γ)' (Karatzoglou et al. 2006; Waske and Benediktsson 2010). Hsu and Lin (2002) have described how studies have shown that when considering class size, the one-against-one procedure is more consistent than one-against-all and is used to implement multiclass-based SVM model. The Supervised Support Vector Machines classification tool in ENVI 5.2 was used to perform the SVM classification.

4.3.5. Accuracy assessment

In order to evaluate the predictive map of *Prosopis* and the other acacia species developed by RF and SVM algorithm classifiers on WorldView-2 imagery, an independent test dataset (Table 4.1) was used and confusion matrices were then generated to compare the true class with the class assigned by the classifiers by obtaining the overall accuracy, user and producer accuracies, and the Kappa statistic (Congalton and Green 2008). The overall, user and producer accuracies were calculated using the confusion matrix. The producer's accuracy shows the probability that specific vegetation species and land cover types of an area on the ground is classified as such, while the user's accuracy refers to the probability that a pixel labelled as specific vegetation species and land cover types in the map is the actual class. The overall accuracy was calculated based on the number of pixels correctly classified divided by the total number of pixels. In addition, the kappa coefficient was also calculated to provide a measure of the difference between the actual agreement, reference data and the classifier used to perform the classification versus the likelihood of agreement between the reference data and a random classifier (Congalton and Green 1999). A Kappa coefficient equal or close to 1, indicates strong agreement.

4.4. Results

4.4.1. Tuning of Random Forest parameters

In order to determine the best input parameters to train the random forest algorithm to classify the six land-cover classes, RF parameters were optimized. The lowest OOB error rate of 13.5% was produced from the combination of *ntree* value 500 and *mtry* value 5 (Figure 4.2). The combination of *mtry* value of 2 and *ntree* value of 500 produced the highest OOB error rate of 15.5%.

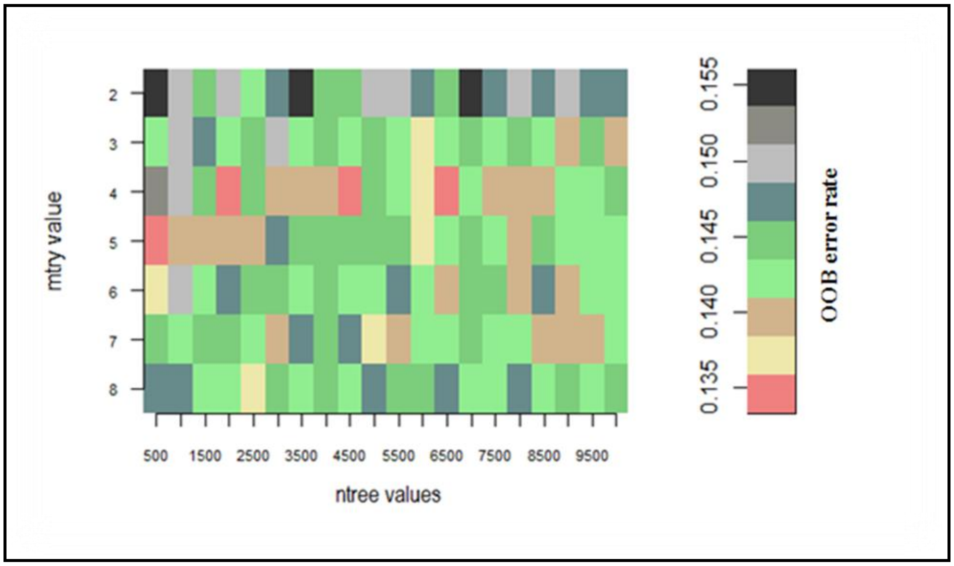


Figure 4.2: Random Forest Optimization of parameters (*ntree* and *mtry*) using the 10-fold grid search method. The Out-of-Bag sample was used to determine the error rate for all the different combinations.

4.4.2. Performance of Random Forest and Support Vector Machines in land-cover classification

The RF and SVM classifiers were able to provide the spatial distribution of *Prosopis* and other vegetation species (Figure 4.3). From both classifiers, it is clear that the most common vegetation species are *Acacia mellifera* and *Prosopis glandulosa*. The *Prosopis glandulosa* species are dominant in low land, while *Acacia mellifera* are dominant in the high land. Figure 4.3 also shows that clear ecotones exist between the vegetation species.

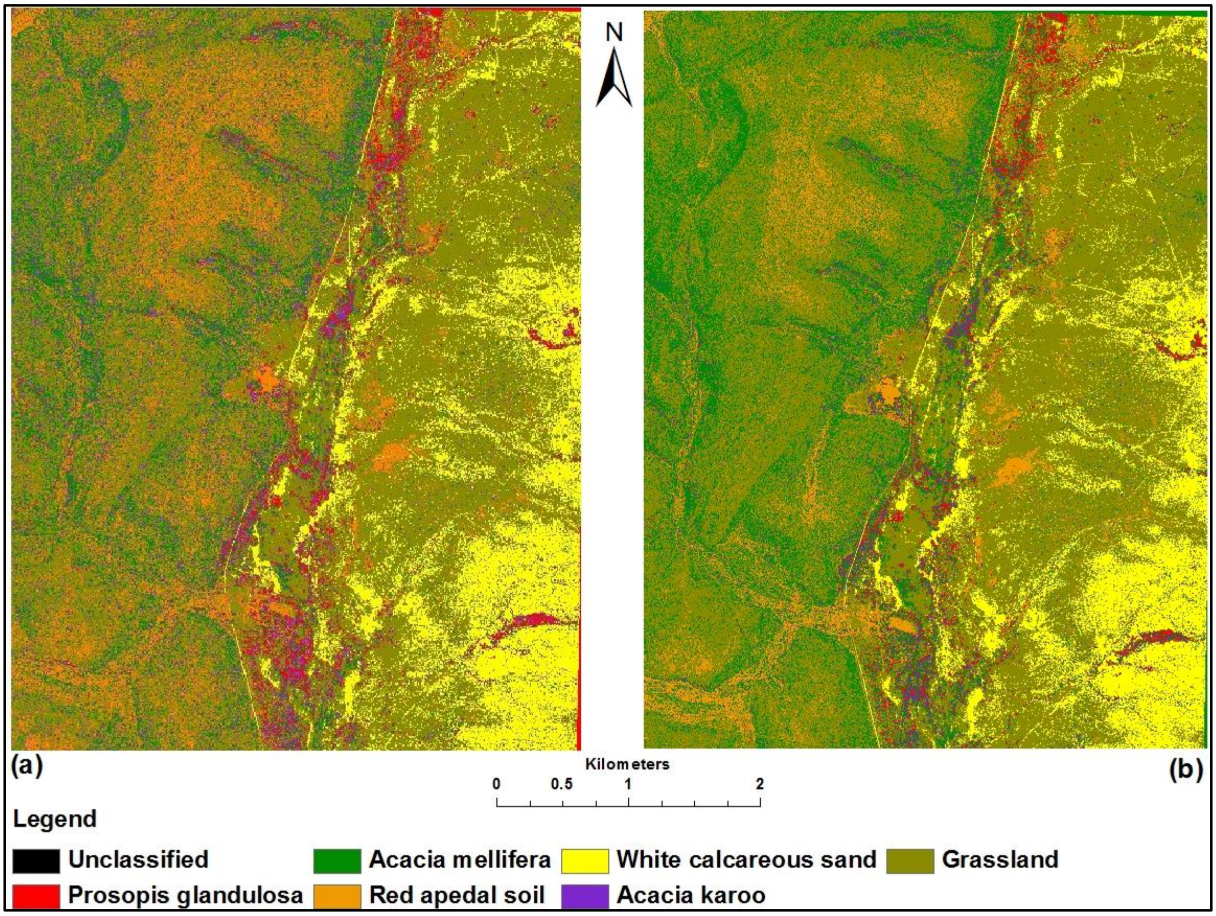


Figure 4.3: Land Use and Land Cover classification using Random Forest (a) and Support Vector Machines (b) classification algorithms.

The RF also provided a variable importance measurement to indicate the role of each band in the classification process. The most important bands are those with the highest mean decrease in accuracy, which in this classification were allocated at the red, yellow and blue bands (Figure 4.4). We further evaluated the utility of each band in mapping particular land-cover types, the red and blue bands were most important bands for classifying *Prosopis glandulosa* and other species (Figure 4.5). Likewise, areas covered in vegetation mainly fall in the red and near infrared and coastal regions of WorldView-2, while the non-vegetated areas mainly containing sandy soil and white soil lie in the green and yellow and blue bands.

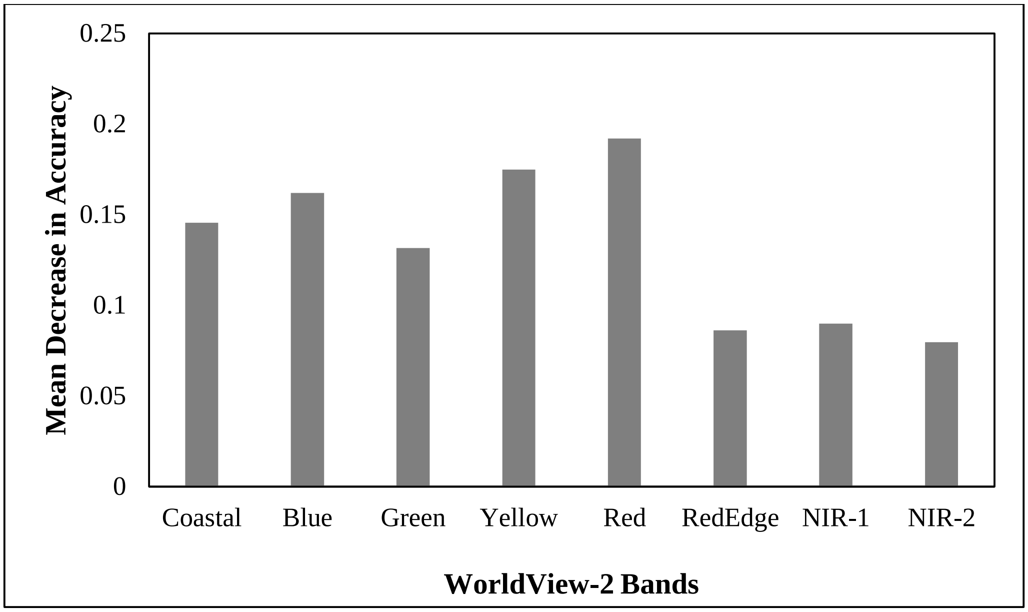


Figure 4.4: Variable importance of the WorldView-2 bands in classification for the entire vegetation species and other land-cover classes.

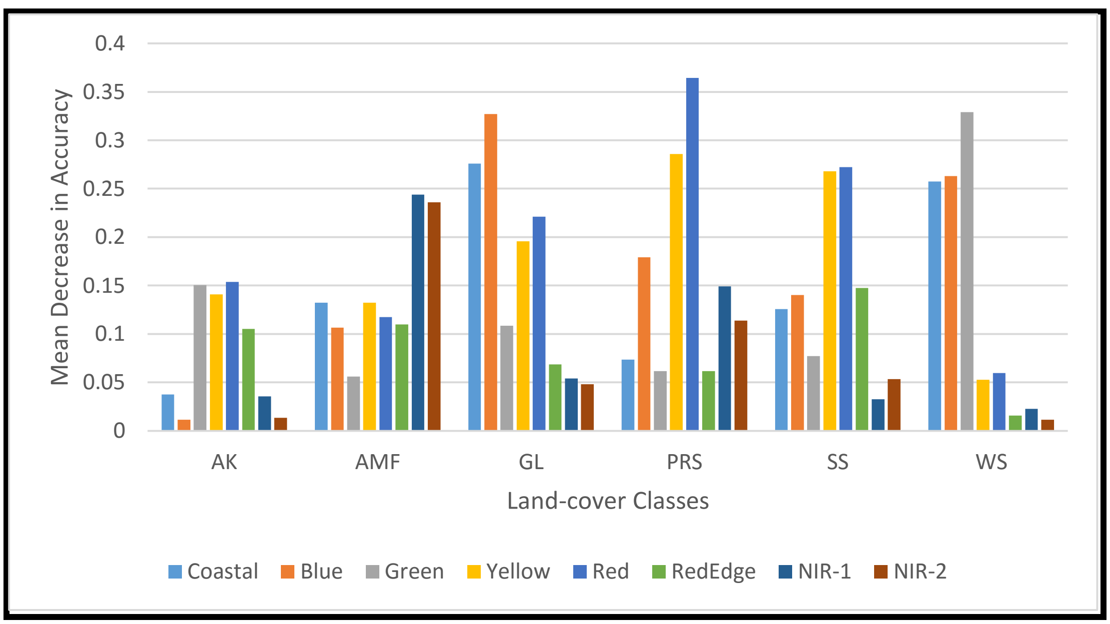


Figure 4.5: The relationship between each individual land-cover class and the importance of the WorldView-2 bands. The highest mean decrease in accuracy shows the most important band.

4.4.3. Accuracy assessment

The independent test dataset was used to evaluate the prediction performance of both RF (Table 4.2) and SVM (Table 4.3) as classifiers. The RF classifier produced an overall accuracy of 86.59% with a Kappa value of 0.84. From the user's accuracies it can be noted that generally all land-cover classes achieved a user's accuracy of over 90%. Spectral confusion was noted between *Acacia karoo* (AK) and *Acacia mellifera* (AMF) and therefore resulted in lower user accuracies of 76% and 65.71%, respectively. The random forest classifier also generated less than 90% for the producer's accuracy for other land cover types and lower for the producer's accuracy *Acacia karoo* (63.33%) and *Acacia mellifera* (76.67%) (Table 4.2). Conversely, the SVM classifier generated a slightly lower overall accuracy of 85.98% with a Kappa value of 0.83. Similar to the RF classifier, the SVM classifier also produced user accuracies of lower than 90%, or 72% and 64.86% for *Acacia karoo* and *Acacia mellifera*, respectively (Table 4.3). Likewise, these two classes had the relatively lower producer's accuracies and had the most spectral confusion.

Table 4.2: Confusion matrix using Random Forest classifier for *Acacia karoo* (AK), *Acacia mellifera* (AMF), grassland (GL) *Prosopis glandulosa* (PRS), red apedal soil (SS) and white calcareous sands (WS). The overall accuracy (OA); user's accuracy (UA); and producer's accuracy (PA) were developed on the test dataset using the EnMAP-Box *ImageRF* Accuracy Assessment tool.

Class	**Using Random Forest**								
	AK	**AMF**	**GL**	**PRS**	**SS**	**WS**	**Total**	**UA%**	**PA%**
AK	19	5	1	0	0	0	25	76.00	63.33
AMF	10	23	1	0	1	0	35	65.71	76.67
GL	0	0	19	0	1	0	20	95.00	90.48
PRS	1	0	0	24	0	0	25	96.00	100.00
SS	0	2	0	0	28	0	30	93.33	93.33
WS	0	0	0	0	0	29	29	100.00	100.00
Total	30	30	21	24	30	29	164		
OA = 86.59%; Kappa = 0.84									

Table 4.3: Confusion matrix using the Support Vector Machines classifier for *Acacia karoo* (AK), *Acacia mellifera* (AMF), grassland (GL) *Prosopis glandulosa* (PRS), red apedal soil (SS) and white calcareous sands (WS). The overall accuracy (OA); user's accuracy (UA); and producer's accuracy (PA) were developed on the test dataset using the ENVI-5.2 Confusion Matrix Workflow.

Class	**Using Support Vector Machines**								
	AK	**AMF**	**GL**	**PRS**	**SS**	**WS**	**Total**	**UA%**	**PA%**
AK	18	5	2	0	0	0	25	72.00	60.00
AMF	11	24	0	0	2	0	37	64.86	80.00
GL	0	0	19	0	1	0	20	95.00	90.48
PRS	1	0	0	24	0	0	25	96.00	100.00
SS	0	1	0	0	27	0	28	96.43	90.00
WS	0	0	0	0	0	29	29	100.00	100.00
Total	30	30	21	24	30	29	164		
OA = 85.98%; Kappa = 0.83									

Table 4.4 shows the area under each *Prosopis* and other classes obtained by RF and SVM classification algorithms. The comparable areas obtained by the two algorithms also confirm the similar performance of the two algorithms. The study area is dominated by grassland and bare soil.

Table 4.4: Area for each vegetation species and other land cover class obtained by Support Vector Machines and Random Forest classifiers.

Land-cover class	**Random Forest Area (ha^{-1})**	**%**	**Support Vector Machines Area (ha^{-1})**	**%**
Acacia karoo	443.9	9.3	467.0	9.7
Acacia mellifera	417.1	8.7	435.9	9.0
Grassland	2 657.0	55.4	2630.0	54.5
Prosopis glandulosa	104.9	2.2	108.9	2.3
Red Apedal sand	685.3	14.3	687.4	14.2
White Calcareous soil	486.4	10.1	495.4	10.3

4.5. Discussion

The rapid spread of the *Prosopis* species has caused considerable negative impacts to biodiversity and ecosystems across different landscapes. To better understand the status and to support researchers and decision makers to develop effective management for this problem, it is essential to obtain reliable and accurate information about the spatial distribution and the level of invasive species dynamism into the native eco-community. With developing technologies, remote sensing methods are increasingly being employed for monitoring a range of remotely detectable properties of invasive plant species, and there is now a growing demand to test the ability of different remotely sensed data in mapping and monitoring invasion status of alien plants accurately across a range of scales. The availability of high resolution satellite data provides a great potential to achieve better performance and results in studies of such alien plants.

The main objective of this study was to investigate the performance of the new high spatial resolution WorldView-2 sensor in the detection and mapping of *Prosopis glandulosa* and other co-existing species in the arid environment of the Northern Cape Province of South Africa. Random forest and support vector machines were used as classifiers and results demonstrated that the *Prosopis glandulosa* can be detected and distinguished accurately from the co-existing three indigenous acacia species.

The relatively high overall and individual classification accuracy obtained in this study demonstrates the capability of the high spatial and spectral resolutions of the WorldView-2 sensor to detect *Prosopis* and its co-existing species. Previous studies on mapping *Prosopis* were not able to discriminate it from the native background shrubs and trees using aerial photography (Robinson et al. 2008; van Klinken et al. 2007) and Landsat (Van den Berg et al. 2014) due to the lack of fine spatial and spectral resolution. While remote sensing has considerable potential to provide information on spatial and temporal dynamics of the invasive plant species, there are many uncertainties with maps of invasive species obtained from commonly used medium-spatial and spectral resolutions such as Landsat and SPOT (Van den Berg et al. 2014). This study however, indicates that the eight-band multispectral sensor of WV-2 is suitable to provide species maps with an overall accuracy of 86% at 2 m spatial resolution for *Prosopis* and other co- existing plant species in the arid environment of the site under study. Such high accuracies in mapping vegetation species have been largely restricted to hyperspectral platforms that have higher spatial and spectral resolutions (Artigas and Yang 2005; Belluco et al. 2006; Lawrence et al. 2006). However, the associated problems with the use of hyperspectral sensors such as high cost and high dimensionality or redundancy of data exist. The new generation multispectral sensors, the WV-2, can save money and time, while providing a high level of accuracy in mapping and monitoring of invasive alien plant species. Our results are also comparable to those found from a recent study where the ability of the WV2 satellite sensor was used to detect the invasive shrub mesquite in the north-west Pilbara region of Australia (Robinson et al. 2016).

Since RF and SVM algorithms were run using equivalent training and test data points in the present study, the ability of the RF and SVM classifiers to detect and discriminate *Prosopis* from other co-existing indigenous plant species were investigated, and both of them yielded comparable overall accuracies. RF achieved higher classification accuracy than SVM by about 1%. The performance similarity of the two classifiers was, however, not a surprise and it agrees with results of other studies from the literature, where hyperspectral data were used (Abdel-Rahman et al. 2014; Waske et al. 2010) and multispectral (Pal 2005; Priyanka et al. 2015; Taylor and Barker 2012).

The relatively high accuracies achieved by RF and SVM were expected. This is due to the fact that the combination of tree classifiers in RF is such that each classifier depends on the values of a random vector sampled independently (Breiman 2001). These random vectors have the same distribution for all classifiers in the forest, and each tree casts a unit vote for the most popular class input (Breiman 2001). This distribution and casting makes RF algorithms more robust to noise and outliers (Abdel-Rahman et al. 2014). SVM, on the other hand, is a known versatile classification algorithm that constructs models based on small data instances (support vectors) from different classes (Abedi et al. 2012; Mountrakis et al. 2011; Vapnik 1995). The classification error can therefore be considerably minimized by using a nonlinear kernel function to perform SVM classification. A nonlinear kernel is an efficient method to solve inseparability problems that may be found in the mapping of the vegetation species. The classification error is minimised by increasing the margin between data points and the hyperplane (Abedi et al. 2012; Cortes and Vapnik 1995a; Mountrakis et al. 2011; Vapnik 1995; Yu et al. 2012). A radial basis (non-linear) kernel function was used in this study because it solves the inseparability issues that could be associated with vegetation species mapping (Mountrakis et al. 2011).

The high spatial and spectral resolutions of the WV-2 dataset allowed us to detect small invaded areas of four square meters. Such high resolutions may support the early detection and eradication program and therefore eliminate any new invasion, minimizing the long-term damages and/or the control costs (Hunt 2009). Figure 4.3 clearly indicates that *Prosopis* showed a strong preference for riparian and floodplains as reflected by a higher rate of initial colonisation by patches and increase in canopy cover (Figure 4.3). Many previous studies have also shown that in the semi-arid and arid rangelands *Prosopis* species are frequently found in a flood zone and soils that have good moisture retention capacity (Lowe et al. 2000; Robinson et al. 2008; Zachariades et al. 2011). Unfortunately, these are the only areas characterised by year-round water supply in the arid region of the province and therefore, the local communities depend on these sites for farming and animal production. Thus, the invasion of the *Prosopis* in the only cultivable lands of the province (Northern Cape Province) results in socio-economic and ecological havoc such as destruction in biodiversity and potential grazing or rangelands (Hagos and Smit 2005; Lowe et al. 2000). The red apedal soil has a high density of *Acacia mellifera*, but low grass cover. It is therefore, unlikely for grazing animals to spend long time in these areas and consequently the spread of *Prosopis* seeds is low in these sites.

Both the RF and SVM models showed that the current *Prosopis* distribution covers about 3% of the study area. However, their extent of invasion especially in the riparian zones is expected to increase by ~ 27.5% annually (Lowe et al. 2000; Smit et al. 1999). Goats and horses are considered to be responsible for the rapid spread of the *Prosopis* within its native range from riparian zones into uplands.

4.6. Conclusion

Prosopis species are among the most widespread and damaging invasive woody plants in the Northern Cape of South Africa and there is much potential for the species to spread further. The negative impacts on the environment and the livelihood of the local communities are escalating rapidly and there is an urgent need for more effective management approaches to drastically reduce adverse impacts and enhance benefits. However, there are still critical gaps in our knowledge of its spatial distribution and the dynamic invasion impacts on the ecosystem.

This study obtained relatively high accuracies in mapping *Prosopis* and therefore provides reliable spatial information on the extent and the dynamic of *Prosopis* invasion as well as a number of other land cover classes. Results in this study provide new insights on the performance of WorldView-2 imagery in mapping vegetation cover at species level. This would help environmental managers to focus their existing monitoring and control efforts on areas of priority. Such monitoring efforts allow rapid assessment and proactive adoption of the most appropriate intervention in the control of the invasive alien plants.

4.7. Acknowledgements

Gratitude is given to the R Development Team and for their open source package that was used for data analysis. Additionally, a great thank-you to the Southern Mapping Company for supplying the WorldView-2 imagery. Finally, a thank you to the ENVI and EnMAP-Box development team for the open source software used to carry out the classification of the imagery.

CHAPTER FIVE

Cost effective approach for mapping *Prosopis* invasion in arid South Africa

This chapter based on:

Mureriwa, N., and Adam E., (in Preparation). Cost effective approach for mapping P*rosopis* invasion in arid South Africa using Spot-6 Imagery and machine learning classifiers

5.1. Abstract

This study evaluates the use of SPOT-6 imagery in conjunction with two machine learning classifiers, namely Random Forest (RF) and Support Vector Machines (SVM) to map *Prosopis glandulosa*, its co-existing acacia species and other land-cover types in an arid South African environment. *Prosopis glandulosa* is one of the 44 species of *Prosopis* which are rated the world's top-100 unwanted species by the International Union of Conservation and Nature (IUCN). This highly invasive species has been difficult to control using physical, chemical and biological methods because of insufficient knowledge of the species dynamic and lack of spatial data. Results show that it is possible to distinguish *Prosopis glandulosa* from its co-existing species of *Acacia karoo* and *Acacia mellifera* as well as three other general land cover types (grassland, red apedal soils and white calcareous sands). Classification using RF obtained a higher overall accuracy of 78.46% with a Kappa value of 0.7524. SVM classification on the other hand obtained a lower classification accuracy of 77.62% with a Kappa value of 0.7428. The high accuracies obtained from the use of the new-generation SPOT-6 sensor and two advanced classification algorithms show the potential to map the invasive species spread on a large scale. This data is useful to aid the current control methods so as to assist farmers, environmental managers and other affected parties to monitor and plan against future invasion.

Keywords: SPOT 6, *Prosopis glandulosa*, Random Forest, Support Vector Machines, cost effectiveness

5.2 Introduction

Prosopis glandulosa is one of the 44 species of *Prosopis* (Pasiecznik et al. 2004) that is native to North and South America. It was introduced intentionally to different countries across the world for purposes such as to provide shade and fodder in the arid areas of Australia and South Africa; for sand-dune stabilization, afforestation as well as fuelwood supply in Sudan and for local greening, ornamental cultivation and soil stabilization in many Middle Eastern countries (Chikuni et al. 2005; Ghazanfar 1996; Mwangi and Swallow 2005; Zimmermann 1991). Despite these useful characteristics, over the years the species was shown to have negative impacts such as forming dense impenetrable thickets associated with unfavourable impacts on human economic activities by overrunning grazing land (Pasiecznik et al. 2001); negatively affecting biodiversity and excessively consumes surface and ground water (Pasiecznik 1999). Studies have shown that the negative impacts associated with *Prosopis* were perceived to exceed benefits and consequently, an effective management of intervention to control the existing *Prosopis* invasion and to reduce its negative impacts is extremely important (Shackleton et al. 2015)..

Methods that include mechanical techniques such as burning and the removal of the plant; chemical control methods such as herbicide treatment of cut stumps and foliar spraying of saplings and burning (Harding 1987; van Klinken et al. 2009); and biological control using seed-eating beetles have been tested (Zachariades et al. 2011) in different countries all over the World. Unfortunately, these measures have not been as successful as intended due to the high costs and lack of knowledge on key aspects of *Prosopis* species such as spatial dynamic, scale of the invasion and its benefits. Therefore, as described by Nie et al. (2012) up to date temporal and spatial information on the distribution of mesquite invasion and its negative impacts on the ecosystem services is crucial for effective management. Remote sensing methods and field surveys are the two common methods of obtaining this kind of information.

Remote sensing techniques offer an economic and cost-effective technique that produces timely and accurate information for mapping vegetation species. Cost-effectiveness mainly depends on the overall accuracy of the thematic map generated. The main costs are associated with remotely sensed data can be grouped as field-based costs, set-up costs, image acquisition costs and data analysis costs (both field data and imagery processing) (Mumby et al. 1999) of these, field survey costs are usually higher than any of the other costs due to accessibility and time taken to collect the data (Hoshino et al. 2012). The most cost effective satellite for carrying out vegetative mapping depends on the size and how detailed the thematic map needs to be.

Multispectral and hyperspectral data have been used in mapping vegetation species in different landscapes (Akasheh et al. 2008; Harvey and Hill 2001; Lawrence et al. 2006; Peerbhay et al. 2013; Saatchi et al. 2008). It has been noted that the utility of commonly used multispectral data such as Landsat and SPOT has been limited by the lack of spectral and spatial resolutions. (Harvey and Hill 2001; Li et al. 2005). Conversely, data acquisition via the narrow bands of hyperspectral sensors allows the detection of vegetation at species level which would otherwise be masked by the broad bands of multispectral sensors (Adam et al. 2010; Goetz 2009). It should be noted however that the use of hyperspectral data has its own limitations such as cost, time, availability, processing and the inherent high dimensionality of the data (Goetz 2009).

As a compromise between the multispectral and hyperspectral imagery benefits and limitations, a suite of new-generation imagery such as RapidEye, Sentinel series and World-View series have emerged over the last decade and provide more detail on land cover mapping due to their high spatial and spectral resolutions (Cho et al. 2012; Schuster et al. 2012). The SPOT-6 and 7 twin satellites, a focus of this study are also part of this new generation of satellites. Numerous reasons led to the selection of SPOT-6 imagery for this particular study. Firstly, in South Africa the use of SPOT data is advantageous because it is free due to an agreement that stands between the South African National Space Agency (SANSA) and the Airbus Defence and Space (ADS) since November 2013 (sansa.org 2014). This new Spot Data Direct Receiving Station Supply, Reception and Distribution (DRS) Agreement is a continuation of the 2006 Spot Image Data Reception and Distribution Agreement that had allowed acquisition of SPOT 1-5 data. The entire 1 221 000 square kilometres of South Africa is covered bi-annually (3 months for each coverage) for a seamless country mosaic (sansa.org 2014; Web 2015). This frequency in monitoring of invasive species makes data acquisition using remote sensing methods cost-effective (Kokaly et al. 2003a; Müllerová et al. 2013). Furthermore, SPOT-6 imagery consists of four multispectral bands at 6 m resolution, and a panchromatic band at 1m resolution. This allows pan-sharpening of images obtained to an even higher 1.5 m resolution which makes the species-level discrimination between (*Prosopis*) mesquite and other acacia a possibility (Pohl and Van Genderen 1998). Lastly, similar research carried out using higher resolution (2 m) WorldView-2 imagery (Chapter 4) identified that the most important bands for mapping *Prosopis glandulosa* are the bands that intersect with the SPOT-6 multispectral bands of red, blue, green and NIR. This substantiates the possible ability of SPOT 6 to map *Prosopis* spp. at species level.

In land-cover mapping, image classification succeeds image acquisition. The results obtained from classification are believed to be dependent on factors such as test sample collection, image data available, pre-processing of data (feature extraction and selection), training sample selection, validation methods post-processing techniques and the classification scheme (Gong and Howarth 1990). For supervised image classification via pixel-based methods, conventional classifiers such as Maximum Likelihood classifier (MLC) have been used in remote sensing for many years with successful results of high accuracies (Li et al. 2014). However, this method has the disadvantages of being very dependent on the quality of the training data and classification outputs cannot be improved by including expert knowledge to the imagery (Srivastava et al. 2012). The development in computer and mathematical sciences in the past 10 years has led to more advanced algorithms such as random forest (RF), support vector machines (SVM), classification and regression trees and artificial neural networks (ANN), which has enhanced digital image processing (Ham et al. 2005; Kumar et al. 2015; Omer et al. 2015a; Petropoulos et al. 2012). Ability to handle unbalanced datasets as well as to synthesize regression and having insensitivity to over-training are some of the superior image-processing abilities that have ranked RF and SVM classifiers considerably higher than the other advanced algorithms (Breiman 2001; Li et al. 2014).

Consequently, the aims of this study are to test the use of SPOT-6 imagery as a cost effective method of vegetative mapping as well as to evaluate the robustness of Random Forest (RF) and Support Vector Machines (SVM) as machine learning algorithms in mapping *Prosopis glandulosa* (mesquite) invasion and its co-existing indigenous species in the semi-arid region of the Northern Cape Province of South Africa.

5.3. Materials and methods

5.3.1 Study area

The study area shown in Figure 5.1 below is located in the Northern Cape Province of South Africa (Figure 5.1). This Province covers about 363 203 km^2 and is an arid region that takes up nearly a third of South Africa's land area. It is a dry region that is heterogeneous with fluctuating temperatures, varying topographies and comprises of six biomes namely, the Savanna, Desert, Succulent Karoo, Grassland, Fynbos and Nama Karoo biomes (Mucina and Rutherford 2006). The study area is situated in the north-western part of the province and is about 5km from the small town of Griekwastad and 170km from the city of Kimberley. It covers plains with a variety of acacia, such as *Acacia erioloba*, *Acacia karoo* and *Acacia mellifera.* It also consists of a mixture of grasses such as *Stipagrostis amabilis*, *Aristida meridionalis* and *Eragrostis lehmanniana* that dominate the grassy plains (Van den Berg et al. 2014). In addition to these land-cover types there is also a range of soil types in the area such as the deep-grey calcareous sands, yellow sands and red-yellow apedal soils just to mention a few (Group 1991).

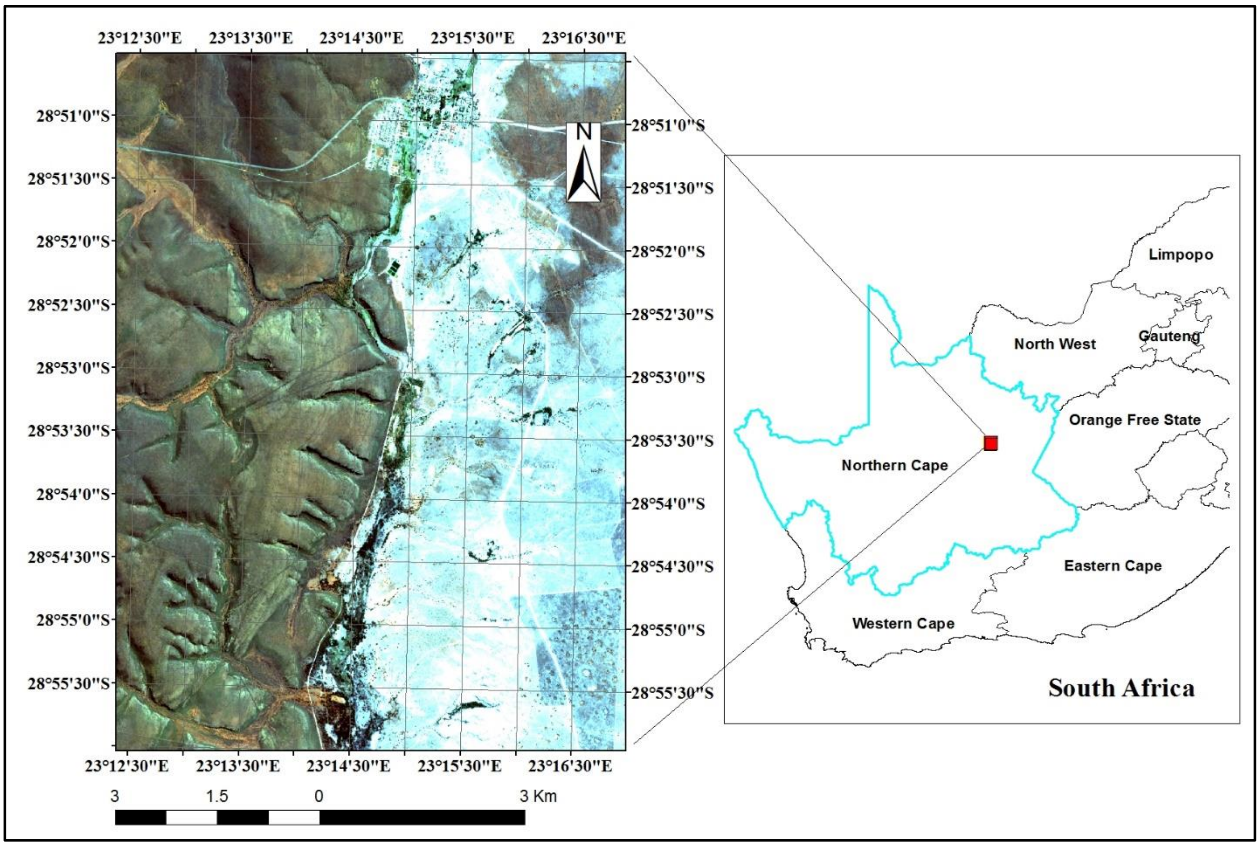


Figure 5.1: A true-colour composite SPOT-6 image showing the location of study area.

5.3.2 Image acquisition and pre-processing

A SPOT 6 image captured on the 24th of July 2015 under cloudless conditions was used for this study. This is a new generation optical satellite launched by Airbus Defence and Space (DS) on 9 September 2012. SPOT 6 has a large swath capacity of 60 km at nadir that enables a 6 million square kilometre daily acquisition at 1.5 m spatial resolution of 4-band imagery. The panchromatic band has the spatial resolution of 1.5 m and ranges from 450–745 nm. It has 4 multispectral bands with a spatial resolution of 6 m and are: Red (625-695nm), Green (530-590 nm), Blue (450-520 nm), and NIR (760-890 nm). There is a 1 to 3-day revisit time. Due to the high spatial and temporal resolutions the satellite offers an even wider range of remote sensing applications in agriculture, deforestation, environmental monitoring, mining and coastal surveillance (DefenceWeb 2015).

The SPOT-6 image was acquired ortho-rectified and geo-referenced in WGS84 UTM zone 34S from the South African Space Agency (SANSA). The Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm was then used to atmospherically correct the image as described in the Environment for Visualizing Images (ENVI 5.2) 2014 software package.

5.3.3 Defining land-cover classes and reference data collection

A week after the SPOT-6 image acquisition and pre-processing ground reference data was collected from the 1st to the 4th of August 2015. Unsupervised classification was carried out on the four band SPOT-6 image using the IsoData unsupervised classification tool in ENVI 5.2 so as to identify the most common land-cover types. This process initially identified nine main classes which were then regrouped into six broad classes and various ground points were randomly generated across the different land cover types. Random points were input into a GPS to navigate to the field sites. Whenever a random point was not accessible or as an attempt to increase the variation of ground data for *Prosopis* and other co-existing species, purposive sampling was adopted (Adam and Mutanga 2009). The SPOT-6 false colour composites and the GPS points were used in the field to directly locate and delineate *Prosopis* and the other land cover type classes. Regions of interest were then created by overlaying the ground reference data over the SPOT-6 image so as to train and test the classifiers (Table 5.1) by randomly splitting the ground reference data into 70% training and 30% test data sets (Table 5.1).

Table 5.1: Training and validation datasets collected for *Prosopis* and other land-cover classes in study area

Land-cover class	Code	Training dataset	Test dataset	Total
Prosopis glandulosa	PRS	58	25	83
Acacia mellifera	AMF	72	31	103
Acacia karoo	AK	71	30	101
White calcareous soil	WS	68	29	97
Red apedal sand	SS	71	30	101
Grassland	GL	49	21	70

5.3.4 Image classification

5.3.4.1 Random Forest classifier

Breiman (2001) developed an ensemble learning technique called Random Forest (RF) by introducing the idea of bootstrap aggregating ('bagging') to decision trees. In RF, multiple decision trees are combined and each tree contributes by a single vote towards the plural vote of the class assignment of the input data. From the original observations several bootstrap samples are drawn with replacement and many binary classification trees (*ntree*) are built. Any samples not in this bootstrap sample are called out-of-bag (OOB) samples which are usually about a third of the total data and can be used to estimate the misclassification error and to measure the importance of each variable in the final model (Breiman 2001; Lin et al. 2010). A given number of input variables at each node (*mtry*) are randomly chosen from a random subset of the features and the best split is calculated by utilizing only this subset of features. By definition *mtry* is the square root of the total number of spectral bands in the study. To ensure low bias, pruning is not performed and all trees in the forest are maximally grown (Genuer et al. 2010).

Moreover, in order to improve the classification accuracy, RF parameters (i.e. *mtry* and *ntree*) have to be optimized (Breiman 2001; Mutanga et al. 2012a). A 10-fold grid-search approach based on the OOB estimate of error was used in this study to find the optimal combination for these two parameters with the *mtry* value being varied from 1 to 5 and the *ntree* parameter varied from 500 to 10,000. By default *ntree* is 500, while the default value for *mtry* is the square root of the total number of spectral bands used in the study (Breiman 2001). The *ImageRF* tool in EnMAP-Box as well as R were used to perform the RF classification.

5.3.4.2 Support Vector Machines

Originally proposed by Vapnik in 1979, support vector machines (SVM) are defined as a nonparametric binary linear classifier where the distance of each class from the training data points to the optimal hyperplane or decision boundary is maximized (Cortes and Vapnik 1995a). Misclassifications obtained during the training step are thus minimised (Anthony et al. 2007). On the boundaries of the hyperplane are two support hyperplanes that have data points on their edges called support vectors and these are the ones that define the optimal hyperplane (Mountrakis et al. 2011). A drawback has been found in practice when using this linear approach of hyperplanes, which is that data of different classes tends to overlap. Therefore, to improve on this linear-separability limitation and increase classification accuracy, a non-linear polynomial is applied. This non-linear algorithm is optimised by using a number of different methods. To date, for remotely sensed data, the most commonly used method is the kernel method via the radial basis (Huang et al. 2002; Oommen et al. 2008). Two parameters are required for tuning in the radial basis method, namely, the cost 'sigma (C)', defined as a plenty value that is used for adjusting the error of misclassifying instants of the training data set, and the kernel width 'gamma (γ)' (Karatzoglou et al. 2006; Waske and Benediktsson 2010). Hsu and Lin (2002) have described how studies have shown that when considering class size, the one-against-one procedure is more consistent than one-against-all and is used to implement multiclass-based SVM model. The Supervised Support Vector Machines classification tool in ENVI 5.2 was used to perform the SVM classification.

5.3.5 Accuracy assessment

An independent test data set (Table 5.1) was used to assess the classification maps for *Prosopis* and its co-existing species developed by RF and SVM algorithms on SPOT-6 imagery. Confusion matrices were then generated to compare the true class with the class assigned by the classifiers by obtaining the overall accuracy, user and producer accuracies, and the kappa statistic (Congalton and Green 2008). The overall accuracy calculation is established by dividing the number of pixels correctly classified by the total number of pixels; the producer's accuracy shows the probability that specific vegetation species and land cover types of an area on the ground is correctly classified; while the user's accuracy refers to the probability that a pixel labelled as specific vegetation species and land cover type in the map is the actual class.

Moreover, the Kappa coefficient, which is defined as a measure of the difference between the actual agreement between reference data and the classifier used to perform the classification versus the likelihood of agreement between the reference data and a random classifier was also calculated (Congalton and Green 1999). If the Kappa coefficient is equal or close to 1, then there is strong agreement between the two.

5.4. Results

5.4.1 Tuning of Random Forest parameters

RF parameters were optimized so as to determine the best input parameters to train the algorithm to classify the six land-cover classes. The lowest OOB error rate of 25.5% was produced from the combination of *ntree* value 3000 and *mtry* value 3 (Figure 5.2). The combination of *mtry* value of 2 and *ntree* value of 8500 produced the highest OOB error rate of 27.5%.

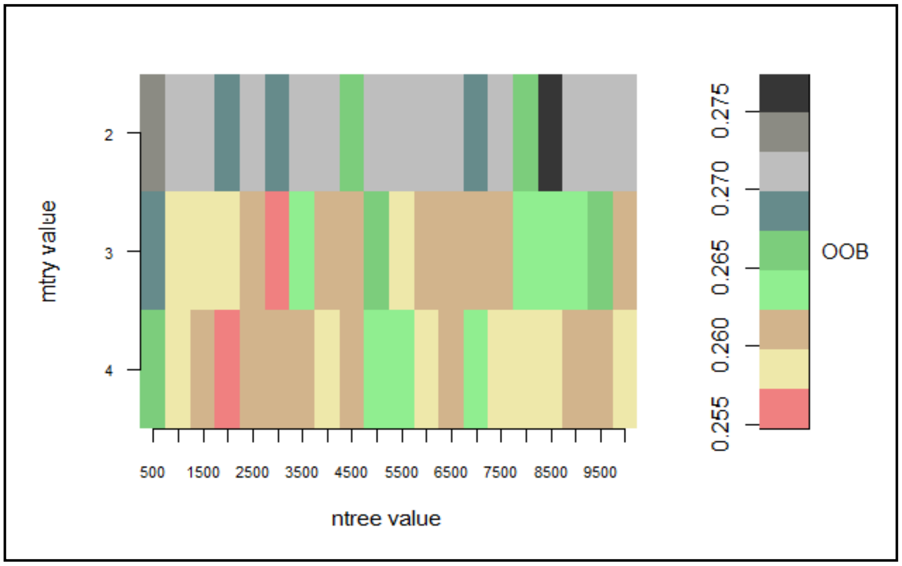


Figure 5.2: Random Forest Optimization of parameters (*ntree* and *mtry*) using the 10-fold grid search method. The Out-of-Bag (OOB) sample was used to determine the error rate for all the different combinations.

5.4.2 Tuning of SVM parameters

SVM parameters for classification via a radial basis kernel function were optimized to define the best input parameters to train the algorithm to classify the six land-cover classes. Using a 10-fold cross validation, the lowest error was produced from the combination of *g*amma (γ) value of 0.1 and cost (C) value of 100 (Figure 5.3).

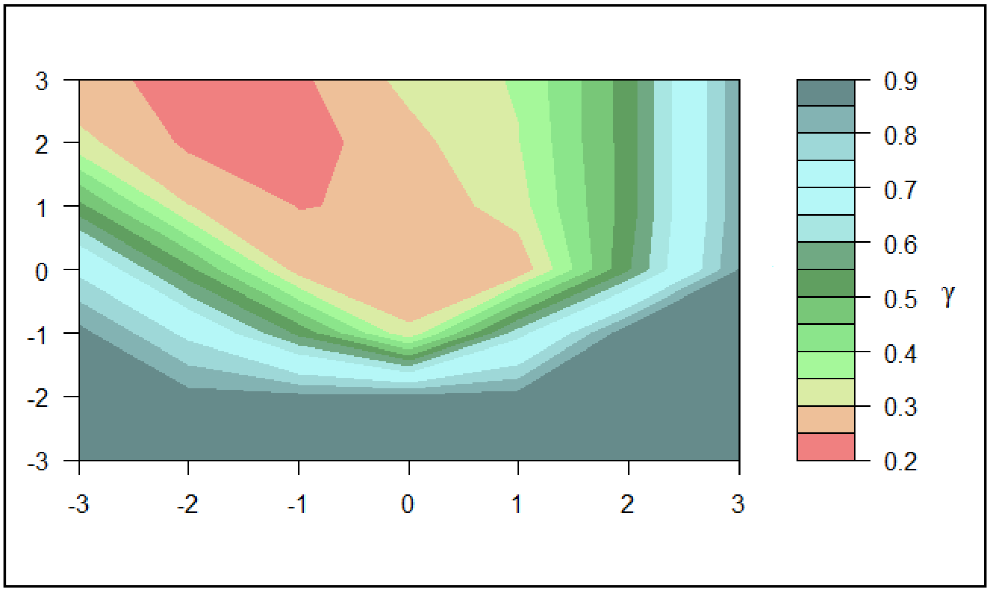


Figure 5.3: Support Vector Machines optimization of parameters (*C* and γ) using the 10-fold grid search method. The Out-of-Bag (OOB) sample was used to determine the error rate for all the different combinations.

5.4.3 Performance of RF and SVM in land-cover classification

The RF and SVM classifiers were able to classify the spatial distribution of *Prosopis glandulosa* and other vegetation species (Figure 5.4). Clear ecotones that exist between the vegetation species are shown in both classification images. For the Random Forest classification image, *Prosopis glandulosa*, *grassland* and *Acacia mellifera* are the most dominant species. Grassland and *Prosopis glandulosa* mainly occupying the lowlands whilst *Acacia mellifera* occupies the high lands. The Support Vector Machines classification has three dominant classes, namely, grassland, red apedal soils and *Acacia mellifera*. The higher lands are mainly occupied by *Acacia mellifera* and red apedal soils whilst grassland occupy the lower lands.

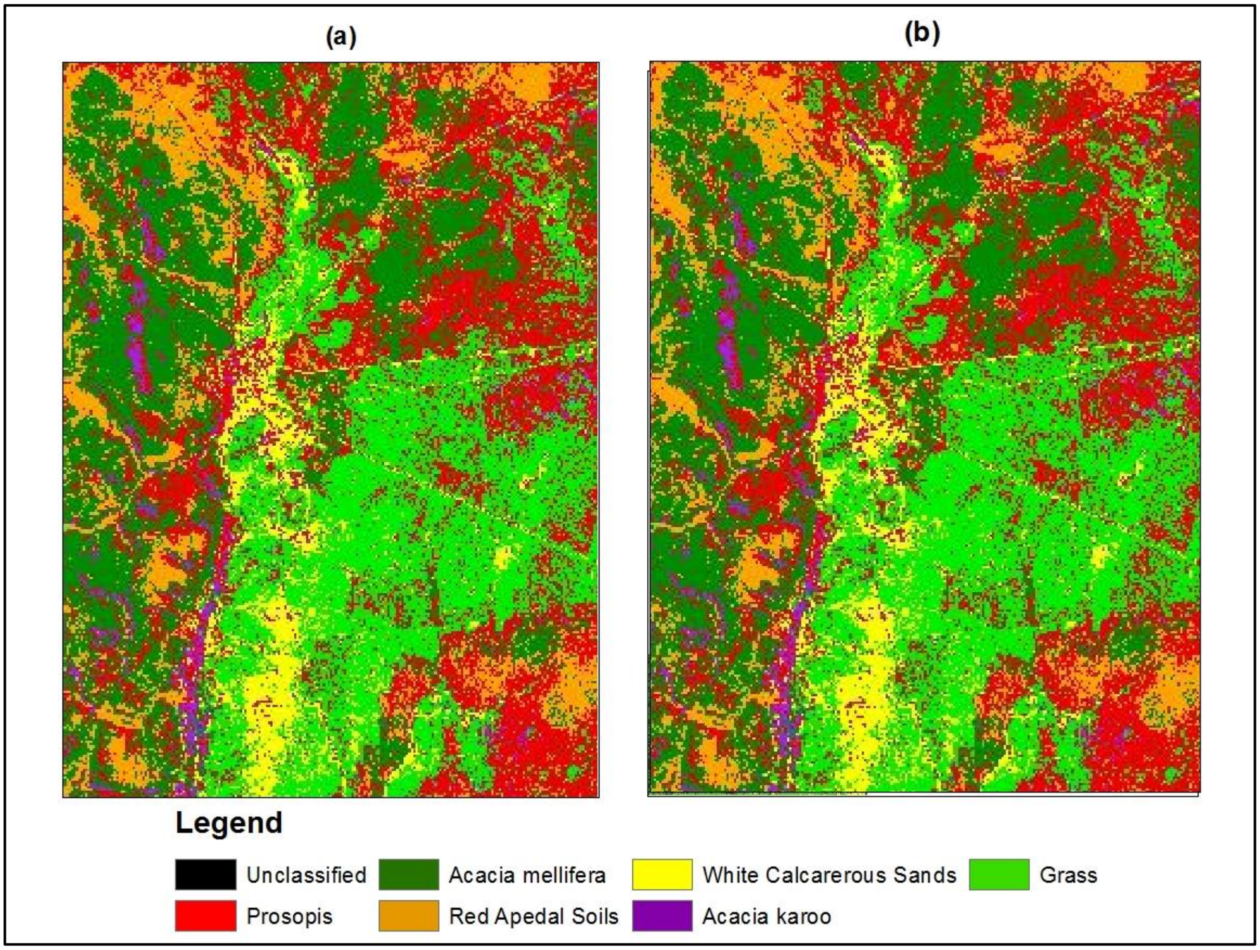


Figure 5.4: Classifications of SPOT-6 image: (a) Random Forest classification (b) Support Vector Machines classification

The role of each band in the random forest classification was provided by the inherent variable importance measurement of the classifier. The most important bands are those with the highest mean decrease in accuracy (Figure 5.5) which in this classification are allocated at the red and blue bands (Figure 5.6). Moreover, the effectiveness of each band in mapping the different land-cover types is investigated. *Prosopis glandulosa* and other species are best classified by the red and blue bands (Figure 5.6 and 5.7). Vegetated areas, namely those covered by *Prosopis glandulosa*, grassland, *Acacia karoo* and *Acacia mellifera* mainly fall in the red and blue regions of SPOT-6 while the non-vegetated areas mainly containing red apedal soil and white calcareous sand fall in the red and near-infrared bands (Figure 5.7).

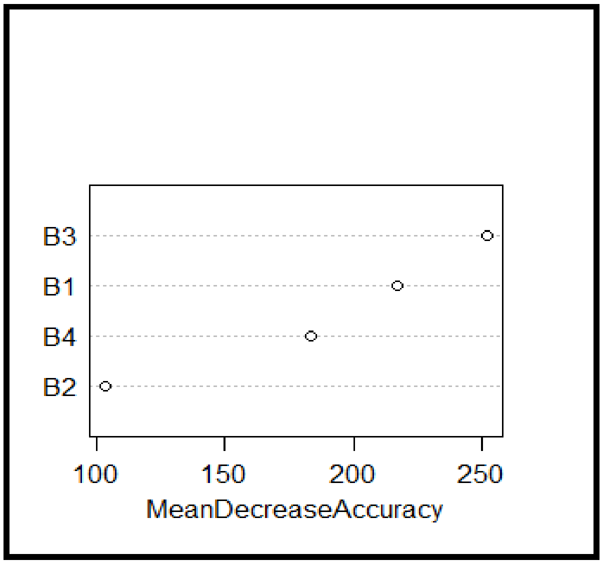


Figure 5.5: Ranking of band importance of SPOT-6 bands using Random Forest. The most important band has the highest mean decrease in accuracy.

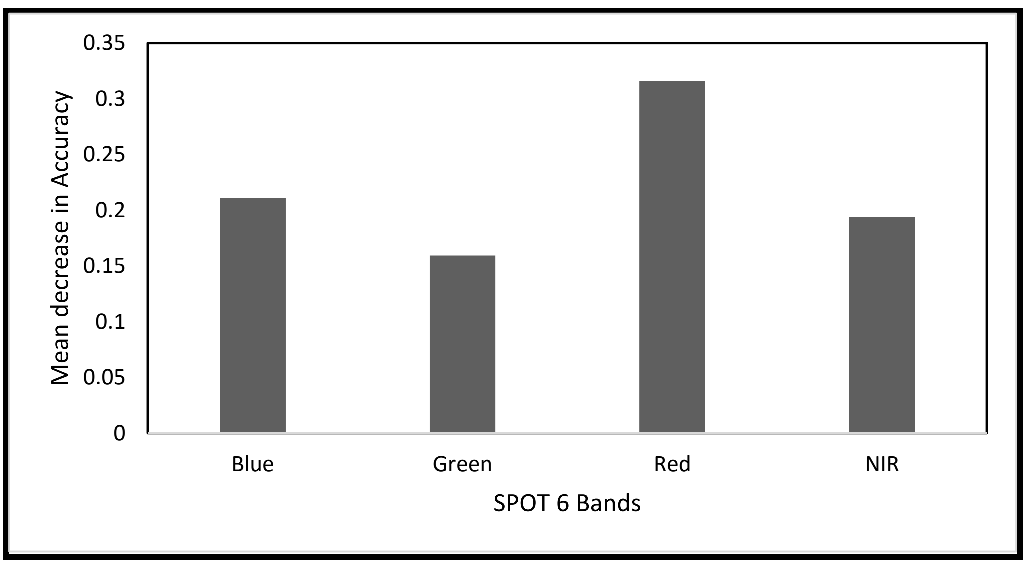


Figure 5.6: Variable importance of the SPOT-6 bands in classification for entire vegetation species and other land-cover classes.

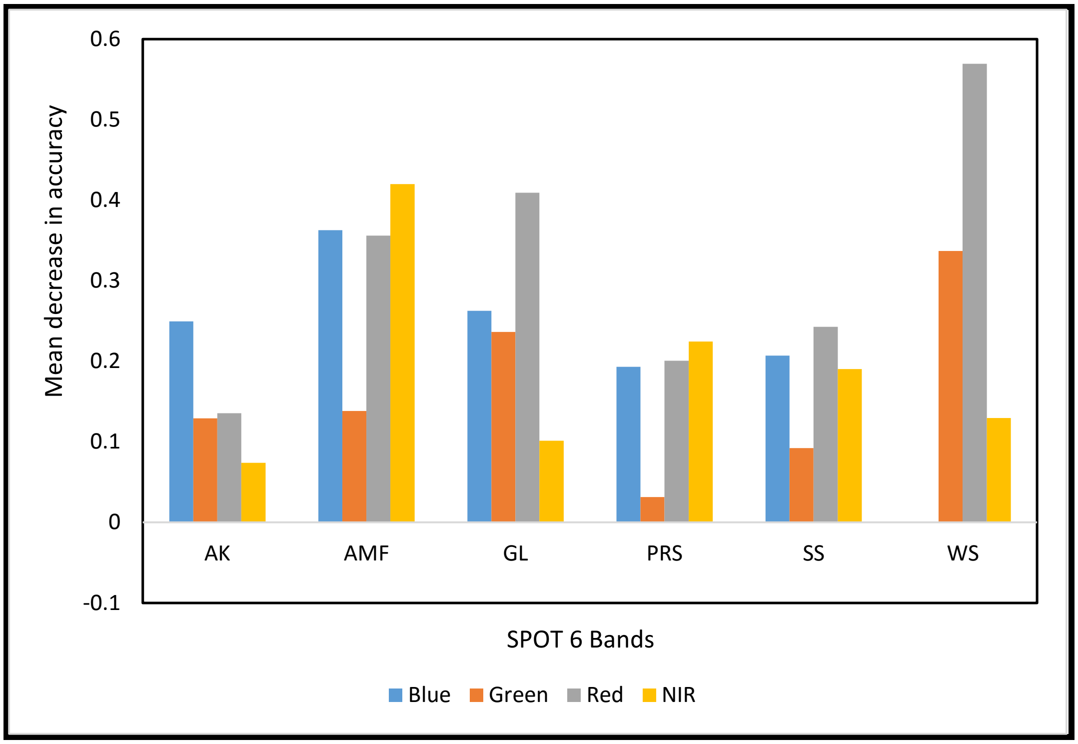


Figure 5.7: The relationship between each individual land-cover class and the importance of the SPOT-6 bands. The highest mean decrease in accuracy shows the most important band.

5.4.4 Accuracy assessment

The performance of both the RF and SVM as classifiers was assessed by using the test dataset (Table 5.1). The RF classifier produced an overall accuracy of 78.46% with a Kappa value of 0.7524 (Table 5.2). Spectral confusion was noted between *Acacia karoo* (AK) and *Prosopis glandulosa* (PRS) and therefore the lowest user accuracy for *Prosopis glandulosa* of 48.15% and a low producer's accuracy of 54.17% whilst *Acacia karoo* obtained a user's accuracy of 72.00% and a producer's accuracy of 60.00% (Table 5.2). The class separation of the six land-cover types (Figure 5.8) shows how there is a great overlap between classes and the main classes that are noticeably separable are *Prosopis glandulosa* and *Acacia mellifera* and white calcareous sands.

Table 5.2: Confusion matrix using Random Forest classifier for *Prosopis glandulosa* (PRS), *Acacia karoo* (AK), *Acacia mellifera* (AMF), grassland (GL), red apedal soil (SS) and white calcareous sands (WS). The overall accuracy (OA); user's accuracy (UA); and producer's accuracy (PA) were developed on the test dataset using the EnMAP-Box *ImageRF* Accuracy Assessment tool.

Class	**Using Random Forest**								
	AK	**AMF**	**GL**	**PRS**	**SS**	**WS**	**Total**	**UA%**	**PA%**
AK	18	1	0	4	0	2	25	72.00	60.00
AMF	1	27	2	1	3	0	34	79.41	90.00
GL	0	0	11	4	0	5	19	55.00	52.38
PRS	8	1	1	13	2	2	27	48.15	54.17
SS	0	1	0	1	25	0	27	92.59	83.33
WS	3	0	7	1	0	20	31	64.52	68.97
Total	30	30	20	24	30	29	163		
OA = 78.46%; Kappa = 0.7524									

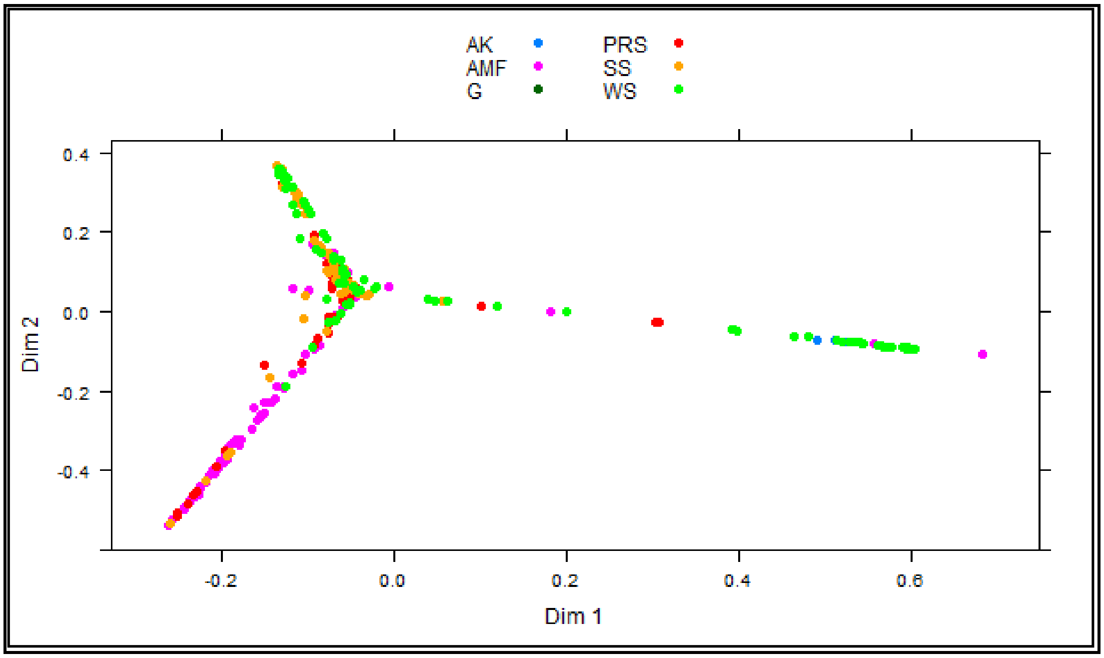


Figure 5.8: Class separation using Random Forest classification for *Prosopis glandulosa* (PRS) *Acacia karoo* (AK), *Acacia mellifera* (AMF), grassland (GL), red apedal soil (SS) and white calcareous sands (WS).

Unlike with RF, the SVM classifier generated a slightly lower overall accuracy of 77.62% with a Kappa value of 0.7428 (Table 5.3). In the same way as the RF classifier, due to spectral confusion, the SVM classifier obtained lower user accuracies for *Acacia karoo* (70.00%) and *Prosopis glandulosa* (72.73%) and producer's accuracies of 70.00% and 66.67% respectively (Table 5.3) The class separation shown in Figure 5.9 further substantiates the major confusion occurring almost species in this classification method. *Prosopis glandulosa* and *Acacia karoo* are greatly confused with almost every other class (Figure 5.9) and thus have the lowest user and producer accuracies (Table 5.3).

Table 5.3: Confusion matrix using the Support Vector Machines classifier for *Prosopis glandulosa* (PRS), *Acacia karoo* (AK), *Acacia mellifera* (AMF), grassland (GL), red apedal soil (SS) and white calcareous sands (WS). The overall accuracy (OA); user's accuracy (UA); and producer's accuracy (PA) were developed on the test dataset using the ENVI-5.2 Confusion Matrix Workflow.

Class	**Using Support Vector Machines**								
	AK	**AMF**	**GL**	**PRS**	**SS**	**WS**	**Total**	**UA%**	**PA%**
AK	21	1	1	3	0	4	30	70.00	70.00
AMF	0	28	1	1	2	0	32	87.50	93.33
GL	1	0	15	2	0	2	20	75.00	71.43
PRS	3	0	1	16	1	1	22	72.73	66.67
SS	1	1	0	1	27	0	30	90.00	90.00
WS	4	0	3	1	0	22	30	73.33	75.86
Total	30	30	21	24	30	29	164		
OA = 77.62%; Kappa = 0.7428									

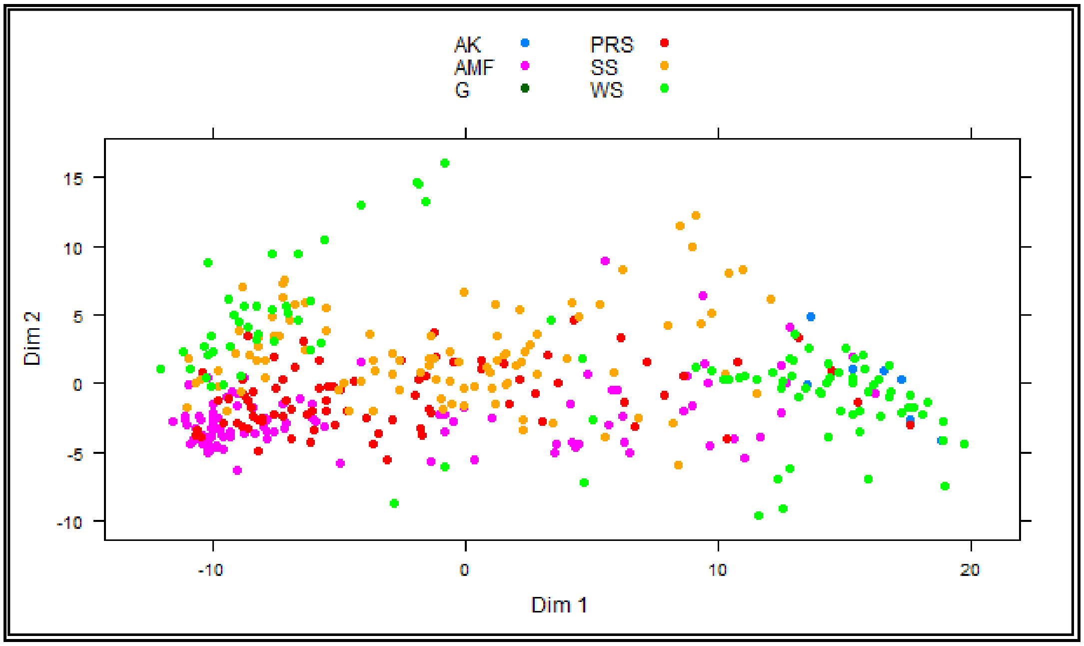


Figure 5.9: Class separation using Support Vector Machine classification for *Prosopis glandulosa* (PRS), *Acacia karoo* (AK), *Acacia mellifera* (AMF), grassland (GL), red apedal soil (SS) and white calcareous sands (WS).

5.5. Discussion

The taxa of *Prosopis* has proven to be an invasive species world-wide (Zimmermann 1991). It has negative effects socio-economically as well as posing a threat on biodiversity since its accidental and intentional introduction began in the 1800s (Van den Berg 2010; Zeila 2011). Invasion control approaches that include biological, physical and chemical methods have been tried and tested with little success over the years (Zachariades et al. 2011). One of the reasons for this has been the lack of timely spatial data as well as lack of knowledge into the dynamic of mesquite invasion (Le Maitre et al. 2011; Wise et al. 2012b). This study explored the performance of the new-generation SPOT-6 spectral sensor to map *Prosopis glandulosa*, other co-existing species and land-cover types. The study area is located in an arid environment thus there is low species diversity which makes species structure easier to distinguish. Results show that *Prosopis glandulosa* was accurately detected from its co-existing acacia species when Random Forest and Support Vector Machines classifications were used.

Both classifiers achieved high accuracies with RF obtaining a slightly higher accuracy of 78.46% as opposed to SVM's accuracy of 77.62% (Table 5.2 and 5.3). The support vector machines classifier (Cortes and Vapnik 1995b; Vapnik 1995) constructs models based on the separation of data points from an optimal hyperplane. In order to minimise the classification error the margin between the hyperplane and data points is increased (Cortes and Vapnik 1995a; Yu et al. 2012). For this study a non-linear (radial-basis) kernel function is used with the optimisation of two parameters (the cost 'sigma (C)' and the kernel width 'gamma (γ)') because it solves inseparability issues that could be associated with LULC classes (Karatzoglou et al. 2006; Mountrakis et al. 2011). As a classifier, SVM has the advantage of having more flexibility to be used for specific data sets and to choose specific data set parameters as well as kernel methods. Conversely, the random forest classifier (Breiman 2001) is better at dealing with outliers and noise associated with classification algorithms because it works on the basis of each tree contributing towards a plural vote of the most popular class input in a random sample. It also only uses two parameters (*ntree* and *mtry*) which make the classifier a lot easier to use once optimized.

Another advantage of using the RF classifier is its inherent ability to provide the importance of each of the four SPOT-6 bands to map each of the six land-cover types in the arid environment (Figure 5.5, 5.6 and 5.7). The most important band for the overall classification is the red band (Figure 5.5 and 5.6). The most important bands for classifying vegetative species *Acacia karoo*, *Acacia mellifera*, *Prosopis glandulosa* and grassland are the red and blue bands (Figure 5.7). The biochemical make-up of vegetation is greatly affected by these bands (Hansen and Schjoerring 2003; Kokaly et al. 2003a) – chlorophyll content, leaf area index and structure as well as canopy structure greatly affect photosynthesis and thus are dependent on this visible region of the electromagnetic spectrum (400 -700 nm) where the red and blue bands are located (Adjorlolo et al. 2012; Ceccato et al. 2001b; Ghulam et al. 2007).

Unlike other multispectral sensors such as Landsat that have the limitation of mixed pixels, the new-generation sensors of SPOT-6 reduce this limitation and have a higher accuracy when it comes to the discrimination of similar species due to their higher resolution (6 m for SPOT versus 30 m for Landsat) (Lu and Weng 2007; Mutanga and Skidmore 2004).

It should be noted that calculation errors were obtained for both RF and SVM due to the high spectral variation in the heterogeneous environment when using pixel-based landscape classification (Duro et al. 2012). No post-classification processing was carried out on the RF and SVM images to smooth the results. Performing this action may improve accuracy of results substantially. Moreover, as seen in other similar studies, misclassifications produced in this study can be attributed to mainly the high spectral variation within the associated indigenous land cover classes (namely *Acacia karoo* and *Acacia mellifera*) as well as the high spatial resolution of the SPOT-6 image. In addition, the misclassification error associated with remotely sensed data obtained from multispectral sensors is reduced by the inherent variable importance feature of the random forest algorithm (Congalton and Green 2008).

5.6. Conclusion

The research conducted in this study set to assess the utility of the advanced classification algorithms RF and SVM on a new-generation SPOT-6 image to delineate invasive *Prosopis glandulosa* from its co-existing species and other land-cover types in semi-arid South Africa. Support Vector Machines classification outperformed Random Forest classification as it was more flexible for parameter and kernel function selection for data sets that are as specific as this one. With only two parameters to optimize, RF provided variable importance ranking for each of the SPOT-6 four bands as well as land-cover type classification. Due to the images high resolution and the high spectral variation of land-cover types, misclassifications were noted for this study as in similar studies. The results provide valuable information on mesquite invasion, change in farm productivity, ecosystem balance and the potential areas of future invasion. Such free data can be used by farmers and environmental managers to evaluate the extent and dynamic of *Prosopis* invasion at a larger scale than would be possible with the use of more expensive imagery such as WorldView-2 or GeoEye (costs ranging from USD$16 – 25 per square kilometer) for example. This has been one of the main reasons why few farmers and environmental managers opt to not use remote sensing as an aid to the control and management methods already in place.

However, more studies have to be carried out to collect more datasets containing test and training samples of high quality to evaluate the performance of the RF and SVM algorithms on similar environments. Spectral-based analysis of *Prosopis* and its co-existing species is another aspect that has not yet been assessed in this current study that could potentially increase the accuracy of training the samples. Moreover, other classification methods that are object-based as opposed to pixel-based should be explored to compare the accuracy of the classifications. These aspects will be considered for future research.

5.7 Acknowledgements

Gratitude is given to the R Development Team and for their open source package that was used for data analysis. Thank you to the ENVI and EnMAP-Box development team for the open source software used to carry out the classification of the imagery. Moreover, much appreciation is given to the South African National Space Agency (SANSA) for providing the SPOT-6 imagery.

CHAPTER SIX

Conclusion

6.1. Introduction

For this research, the invasive species called *Prosopis glandulosa* (mesquite) was investigated with the focus on mapping it in arid areas of South Africa. *Prosopis glandulosa* was rated one of the World's top 100 worst invasive alien species in 2004 by the International Union of Conservation of Nature (IUCN)(Baillie et al. 2004). Mesquite is one of the 44 species of *Prosopis* that were originally introduced to 129 countries and islands intentionally and accidentally for benefits such as providing fodder, timber for furniture production and firewood (thus a source of income for some communities), stabilizing sand dunes, providing shade to livestock and curbing desertification (Pasiecznik 1999; Shackleton et al. 2014b). Over time, however, as with other invasive species, its spread has shown to have negative impacts on ecosystems and biodiversity so that there are reductions in the populations of other species as they are unable to compete; and human socio-economical practices such as overtaking grazing land, community boundaries, and overtaking farmland (Mwangi and Swallow 2005; Pasiecznik et al. 2001; Zimmermann and Pasiecznik 2005). In order to control this species, various methods involving biological, chemical, physical and integrated approaches have been tried and tested with little success when compared to the rapid spread and regeneration of the species (Geesingis et al. 2004; Mwangi and Swallow 2008; Pasiecznik et al. 2006; Shackleton et al. 2014b; Wise et al. 2012a; Zachariades et al. 2011). One of the reasons why control and management techniques have not been as successful as they should be, is that the species exists with very structurally similar co-existing species and there is lack of knowledge on the species invasion dynamic in relation to its co-existing species. The monitoring, detection and mapping of the invasive mesquite has been highlighted as a vital means to overcoming this gap in knowledge.

Consequently, remote sensing applications at different scales have been employed for this study. Both spectral and spatial analysis were conducted and evaluated to provide an overview of the extent of invasion and possible recommendations to enhance the control and management measures already in place. The aim of this research was to examine the possibility of mapping and spectrally discriminating *Prosopis glandulosa* from its native co-existing species in semi-arid South Africa. The specific objectives of the study were: (i) to investigate the usefulness of *in situ* spectroscopic data in discriminating *Prosopis glandulosa* from three other co-existing species (ii) to test the utility of the newly developed guided regularized random forest (GRRF) to accurately discriminate amongst mesquite and its co-existing species (multiclass classification) (iii) to examine if WorldView-2 imagery and two machine learning algorithms (RF and SVM) can map *Prosopis glandulosa* invasion and its co-existing species (iv) to explore the possibility of using SPOT-6 imagery to map mesquite invasion and its co-existing indigenous species using machine learning algorithms. The sections below evaluate each objective.

6.2. Investigating the usefulness of in situ spectroscopic data in discriminating *Prosopis glandulosa* from three other co-existing species.

Objective two of the study aimed to examine the possibility of spectrally discriminating *Prosopis glandulosa* from its native co-existing species. Field spectroscopy was applied by using the Spectral Evolution® RS-3500 Remote Sensing Portable Spectroradiometer Bundle to collect reflectance measurements from four vegetative species namely, *Prosopis glandulosa*, *Acacia karoo*, *Acacia mellifera* and *Ziziphus mucronata*. Each species spectra produced a distinctly different signature which made it easier to distinguish between the similar species (Figure 3.2). The random forest algorithm was then applied to the hyperspectral data obtained in the field to spectrally discriminate between the species and provide variable importance. From the total 2150 wavelengths of the spectral range between 350 nm to 2500 nm of the Spectroradiometer only 1825 wavelengths were used for analysis after removing noisy wavelengths from the spectra. The traditional RF classifier provided the measure of the importance of each wavelength across the 1825 wavelengths in discriminating between the four different vegetative species (Figure 3.4). A high overall accuracy level of 79.19% and a Kappa value of 0.7201 was achieved (Table 3.2). Inherently, hyperspectral data has the problem of high dimensionality which occurs when the number of training samples (n) is limited as compared to the large number of hyperspectral spectral bands (p) (Hsu 2007a). This has been shown through studies to be greatly reduced by using the RF classifier whilst retaining good accuracy levels (Abdel-Rahman et al. 2012, 2013; Adam et al. 2012a; Vincenzi et al. 2011).

6.3. Testing the utility of the newly developed guided regularized random forest (GRRF) to accurately discriminate amongst mesquite and its co-existing species (multiclass classification).

This objective was achieved as a continuation of objective two. A newly developed algorithm called guided regularized random forest (GRRF) that would reduce high dimensionality even further and increase the accuracy of the traditional random forest by identifying the key number of variables that could yield the lowest error rate was employed (Adam et al. 2012a). From the 1825 wavelengths used by the RF classifier to identify the most important wavelengths, 11 wavelengths were identified as key wavelengths by applying the GRRF algorithm and eliminating irrelevant and redundant wavelengths (Figure 3.5). These key wavelengths were identified to lie in three main regions of the electromagnetic spectrum: the visible region greatly affects absorption of photosynthetic pigments of vegetation (Ceccato et al. 2001b); the red edge region greatly affects the biochemical make-up of the vegetative species (Adjorlolo et al. 2013a) and; the short-wave-infrared region which affect the water properties associated with vegetation such as leaf area index, water absorption and macronutrient absorption (Carter 1994; Ceccato et al. 2001b; Ghulam et al. 2007). GRRF greatly increased the overall accuracy of data classification to 88.59% and a Kappa value of 0.8524 (Table 3.2). These results showed how the newly developed GRRF algorithm was a robust method for reducing high dimensionality and could be used to improve results of species discrimination between spectrally similar species.

6.4. Examining if WorldView-2 imagery and two machine learning algorithms (RF and SVM) can map *Prosopis glandulosa* invasion and its co-existing species.

The first approach of the study was to map *Prosopis glandulosa* from its co-existing acacia species and other land cover types using high resolution new-generation WorldView-2 imagery. Two advanced classification algorithms were applied to the image namely Random Forest (Breiman 2001) and Support Vector Machines (Cortes and Vapnik 1995a). An overall classification of 86.59% with a Kappa value of 0.84 (Table 4.2) was found using the random forest classifier, whilst the support vector machines classification obtained on overall value of 85.98% and a Kappa of 0.83 (Table 4.3). These high accuracies show how the advanced algorithms are robust methods for classification and species discrimination. The random forest classifier also provided a measure of variable importance in terms of the WorldView-2 bands (Figure 4.4 and Figure 4.5) which in this case were the red, blue, yellow and coastal bands of the visible region of the electromagnetic spectrum that are vital to the biochemical make-up of vegetative species (Adjorlolo et al. 2012; Ceccato et al. 2001b; Kokaly et al. 2003a). Differences in plant characteristics along these selected bands of the WV-2 sensors helped to successfully discriminate *Prosopis glandulosa* from its co-existing *Acacia karoo* and *Acacia mellifera* as well as other general land-cover types.

6.5. Exploring the cost-effectiveness of using SPOT-6 imagery to map mesquite invasion and its co-existing indigenous species using machine learning algorithms.

It has been noted that obtaining high resolution remote sensing data such as WorldView-2 imagery (2 m) is expensive (with costs ranging between USD$16 – 25 per square kilometre) for most farmers and other environmental organisations and thus not often applied to obtain useful information to enhance control measures. Free satellite data is preferred for remote sensing methods. However, when it comes to vegetation discrimination at species level for a species like *Prosopis*, higher spatial and spectral resolution is needed than is provided by freely available sensors such as Landsat (30 m) (Foody et al. 2005; Robinson et al. 2016). The need for a cost-effective way of mapping and monitoring the invasion of *Prosopis* at a higher resolution provided the necessity for the last chapter of this study. New-generation high resolution SPOT-6 (6 m) data was utilised as it is freely available in South Africa according to an agreement signed in November 2013 between the South African National Space Agency and the Airbus Defence and Space (Web 2015). Similar classification using advanced random forest and support vector machines algorithms as in the WorldView-2 imagery was applied. Variable importance measurement of the random forest classifier showed that the red band followed by the blue band (Figure 5.6 and Figure 5.7) were the most important bands for mapping and discriminating amongst the species of *Prosopis glandulosa*, *Acacia karoo* and *Acacia mellifera*. These bands of

SPOT-6 greatly affect the biochemical make-up of vegetation which include chlorophyll content, leaf area index and canopy structure (Hansen and Schjoerring 2003; Kokaly et al. 2003b). Image classification of the imagery using random forest yielded an overall accuracy of 78.46% with a Kappa value of 0.7524 (Table 5.2). Support vector machines classification on the other hand yielded an overall accuracy of 77.62% and a Kappa value of 0.7428 (Table 5.3). These accuracies are slightly lower than the ones obtained from the WorldView-2 image (86.59% for RF and 85.98% for SVM). This is mainly due to a much lower spatial resolution of SPOT-6 as compared to WV-2 which thus means an increase in spectral confusion amongst the species (Figure 5.8 and Figure 5.9). This thus increases the classification error of the two classifiers. Despite this, however, the classification accuracies are still high enough to provide useful information on the mapping and monitoring of mesquite invasion on an even larger scale than would be possible with the costly and higher-resolution methods.

6.6. Recommendations for future studies

A few recommendations for future studies have been identified from this study as discussed below:

- It is recommended that these studies have to be replicated using similar techniques and data as follows: (i) over larger areas in the country (ii) in other similarly invaded semi-arid and arid areas (iii) in invaded areas under different climatic condition. This will help establish *Prosopis* invasion on a country-scale as well as test the accuracy and reproducibility of these methods.
- Research into the increased utilisation of *Prosopis glandulosa* for its benefits is needed as this could be an added approach to the management and reduction of its spread. Kenya for example is in the process of establishing the utilisation of mesquite's biomass as a source of fuel for power plants and this in turn also creates jobs for community members (Shackleton et al. 2015).
- According to Kohavi et al. (1997), no single machine learning algorithm is superior in all applications, thus for future studies in order to test the RF and SVM classifier's robustness, other methods should also be tested such as artificial neural networks or

variable band set combination methods that have been used in previous invasive species discrimination.

- More research needs to be conducted on the dynamics of the spread of mesquite by considering the impacts of the environmental variables on the invasion such soil analysis studies, favourable habitats, its biology and effectiveness of control practices already in place.
- Testing the use of remote sensing data to estimate the biomass in order to establish an integrative method for controlling cost (physical control) and benefits.
- Object-based classification methods need to be considered for future study on a comparative method to the pixel-based classification methods used in this research.

6.7. Conclusion

The aim of this study was to examine the possibility of mapping and spectrally discriminating *Prosopis glandulosa* from its native co-existing species in semi-arid South Africa. The results from the research conducted showed that it is possible to spatially and spectrally discriminate *Prosopis glandulosa* from its co-existing species. This final conclusion is justified based on the following:

1. New-generation multispectral WorldView-2 and SPOT-6 data were able to accurately discriminate between *Prosopis glandulosa* and its co-existing acacia trees namely, *Acacia karoo*, *Acacia mellifera* and other general land cover types.
2. The random forest (RF) classification algorithm has proven to have great potential in accurately discriminating *Prosopis glandulosa* from its co-existing species. In more detail, (i) it provided high classification accuracies as shown by the accuracy of 86.59% with a Kappa value of 0.84 (Table 4.2) for the WorldView-2 study and 78.46% with a Kappa value of 0.75 (Table 5.8) for the SPOT-6 study. (ii) RF was a good predictor of variable importance by predicting the exact bands in the WV-2 sensor that accurately discriminated *Prosopis glandulosa* from its co-existing species (Figure 4.4 and Figure 4.5) which were identified as the red, blue, yellow and coastal bands. Similarly, the red and blue bands of the SPOT-6 sensor (Figure 5.5, Figure 5.6 and 5.7) were identified as

most important in accurately discriminating *Prosopis glandulosa* from its co-existing species.

3. The support vector machines (SVM) classification algorithm also has great potential in accurately discriminating *Prosopis glandulosa* from its co-existing species. More specifically, (i) SVM classification yielded high accuracies of 85.98% and a Kappa of 0.83 (Table 3) for the WV-2 study and 77.62% and a Kappa value of 0.7428 (Table 5.3) for the SPOT-6 study.
4. The RF ensemble can reduce the problem of high dimensionality associated with hyperspectral data by selecting the most optimal bands in a hyperspectral dataset to improve accuracy of classification. With n=1825, the high accuracy achieved for classification was 79.19% and a Kappa value of 0.7201 (Table 3.2).
5. The newly developed guided regularized random forest algorithm (GRRF) is an even more effective method of reducing high dimensionality associated with hyperspectral data. It uses the variables of importance identified by the traditional RF ensemble and removes all the redundant and irrelevant wavelengths so that only the key wavelengths for accurate discrimination are used for analysis. GRRF greatly increased the overall accuracy of data classification with n=11 identified as key wavelengths. Overall accuracy increased to 88.59% and a Kappa value of 0.8524 (Table 3.2) from 79.19% and a Kappa value of 0.7201 (Table 3.2) in traditional RF.
6. Overall accuracy obtained using spectral discrimination and GRRF was higher (88.59%) than the accuracies of the imagery data in which WV-2 obtained 86.59% and the SPOT-6 data obtained 79.70%. This can be explained by the difference in the numbers of species identified in each study as well as the spectral resolutions. The spectral discrimination data obtained showed the difference between four different tree species (*Prosopis glandulosa*, *Acacia karoo*, *Acacia mellifera* and *Ziziphus mucronata*) whilst the WV-2 and SPOT-6 data discriminated amongst only three species (*Prosopis glandulosa*, *Acacia karoo* and *Acacia mellifera)*. With fewer species identified there is a higher chance of pixel-mixing between species in the high resolution multispectral data so that the accuracies obtained are slightly lower than the hyperspectral data.

REFERENCES

Abdel-Rahman, E.M., Ahmed, F.B., & Ismail, R. (2012). Random forest regression for sugarcane yield prediction in umfolozi, South Africa based on Landsat TM and ETM+ derived spectral vegetation indices. In J.F. Goncalves, & K.D. Correia (Eds.), *Sugarcane: Production, Cultivation and Uses* (pp. 257-284). Hauppauge NY, USA: NOVA Science Publishers, Inc

Abdel-Rahman, E.M., Ahmed, F.B., & Ismail, R. (2013). Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing, 34*, 712-728

Abdel-Rahman, E.M., Mutanga, O., Adam, E., & Ismail, R. (2014). Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing, 88*, 48-59

Abedi, M., Norouzi, G.-H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences, 46*, 272-283

Adam, E., & Mutanga, O. (2009). Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 612-620

Adam, E., Mutanga, O., Abdel-Rahman, E.M., & Ismail, R. (2014). Estimating standing biomass in papyrus (Cyperus papyrus L.) swamp: exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing, 35*, 693-714

Adam, E., Mutanga, O., & Ismail, R. (2013). Determining the susceptibility of Eucalyptus nitens forests to Coryphodema tristis (cossid moth) occurrence in Mpumalanga, South Africa. *International Journal of Geographical Information Science*, 1-15

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

Adam, E.M., Mutanga, O., Rugege, D., & Ismail, R. (2009). Field spectrometry of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands of St Lucia, South Africa. In, *Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009* (pp. IV-260-IV-263)

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

Adjorlolo, C., Cho, M.A., Mutanga, O., & Ismail, R. (2012). Optimizing spectral resolutions for the classification of C3 and C4 grass species, using wavelengths of known absorption features. *Journal of Applied Remote Sensing, 6*, 063560-063561

Adjorlolo, C., Mutanga, O., Cho, M.A., & Ismail, R. (2013a). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 535-544

Adjorlolo, C., Mutanga, O., Cho, M.A., & Ismail, R. (2013b). Spectral resampling based on user-defined inter-band correlation filter: C 3 and C 4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation, 21*, 535-544

Akasheh, O.Z., Neale, C.M.U., & Jayanthi, H. (2008). Detailed mapping of riparian vegetation in the middle Rio Grande River using high resolution multi-spectral airborne remote sensing. *Journal of Arid Environments, 72*, 1734-1744

Albright, T.P., Moorhouse, T., & McNabb, T. (2004). The rise and fall of water hyacinth in Lake Victoria and the Kagera River Basin, 1989-2001. *Journal of Aquatic Plant Management, 42*, 73-84

Amaral, C.H., Roberts, D.A., Almeida, T.I., & Souza Filho, C.R. (2015). Mapping invasive species and spectral mixture relationships with neotropical woody formations in southeastern Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing, 108*, 80-93

Anthony, G., Gregg, H., & Tshilidzi, M. (2007). Image classification using SVMs: one-against-one vs one-against-all. *arXiv preprint arXiv:0711.2914*

Artigas, F., & Yang, J. (2005). Hyperspectral remote sensing of marsh species and plant vigour gradient in the New Jersey Meadowlands. *International Journal of Remote Sensing, 26*, 5209-5220

Asner, G.P., Jones, M.O., Martin, R.E., Knapp, D.E., & Hughes, R.F. (2008). Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment, 112*, 1912-1926

Awale, A.I., & Sugule, A.J. (2006). Proliferation of honey mesquite (prosopis juliflora) in Somaliland: Opportunities and challenges

Baillie, J., Hilton-Taylor, C., & Stuart, S.N. (2004). *2004 IUCN red list of threatened species: a global species assessment*. IUCN

Bajcsy, P., & Groves, P. (2004). Methodology for hyperspectral band selection. *Photogrammetric engineering and remote sensing, 70*, 793-802

Barry, P., Mendenhall, J., Jarecke, P., Folkman, M., Pearlman, J., & Markham, B. (2002). EO-1 Hyperion hyperspectral aggregation and comparison with EO-1 Advanced Land Imager and Landsat 7 ETM+.

In, *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International* (pp. 1648-1651): IEEE

Belluco, E., Camuffo, M., Ferrari, S., Modenese, L., Silvestri, S., Marani, A., & Marani, M. (2006). Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. *Remote Sensing of Environment, 105,* 54-67

Bian, M., Skidmore, A.K., Schlerf, M., Wang, T., Liu, Y., Zeng, R., & Fei, T. (2013). Predicting foliar biochemistry of tea (Camellia sinensis) using reflectance spectra measured at powder, leaf and canopy levels. *ISPRS Journal of Photogrammetry and Remote Sensing, 78*, 148-156

Binggeli, P. (2001). The human dimensions of invasive woody plants. *The Great Reshuffling: Human Dimensions of Invasive Alien Species. Gland (Switzerland): IUCN*, 145-159

Borges, J.S., Marcal, A.R.S., & Dias, J.M.B. (2007). Evaluation of feature extraction and reduction methods for hyperspectral images. In Z. Bochenek (Ed.), *New Developments and Challenges in Remote Sensing* (pp. 255-264). Poland

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

Bromilow, C. (2010). *Problem plants and alien weeds of South Africa*. Briza

Carter, G. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing, 15*, 697-703

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J. (2001a). Detecting vegetation leaf content using reflectance in the optical domain. *Remote Sensing of Environment, 77*, 22-33

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J. (2001b). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, *77*, 22-33

Chan, J.C.-W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment, 112*, 2999-3011

Chikuni, M., Dudley, C., & Sambo, E. (2005). Prosopis glandulosa Torrey (Leguminosae-Mimosoidae) at Swang'oma, Lake Chilwa plain: a blessing in disguise? *Malawi Journal of Science and Technology, 7*, 10-16

Chikuni, M., Dudley, C., & Sambo, E. (2014). Prosopis glandulosa Torrey (Leguminosae-Mimosoidae) at Swang'oma, Lake Chilwa plain: A blessing in disguise? *Malawi Journal of Science and Technology, 7*, 10-16

Cho, M.A., Debba, P., Mutanga, O., Dudeni-Tlhone, N., Magadla, T., & 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., Sobhan, I., Skidmore, A.K., & de Leeuw, J. (2008). Discriminating species using hyperspectral indices at leaf and canopy scales. In, *the XXI congress:Silk road for information from imagery:* (pp. 369-376). Beijing, China: the International Society for Photogrammetry and Remote Sensing

Choge, S., & Chikamai, B. (2004). Experiences of Prosopis utilization and management from outside Kenya. In, *Proceedings of the Workshop on Integrated Management of Prosopis Species in Kenya*

Choge, S., Clement, N., Gitonga, M., & Okuye, J. (2012). Status report on commercialization of Prosopis tree resources in Kenya. *Nairobi: KEFRI*

Coetzer, W., & Hoffmann, J. (1997). Establishment ofNeltumius arizonensis (Coleoptera: Bruchidae) on Mesquite (ProsopisSpecies: Mimosaceae) in South Africa. *Biological Control*, *10*, 187-192

Committee, A.W. (2012). Mesquite (Prosopis spp.) strategic plan 2012–17. Weeds of National Significance, Australian Governmental Department of Agriculture. *Fisheries and Forestry, Canberra*

Congalton, R., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. Lewis Publishers .

Congalton, R.G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data : principles and practices*. Boca Raton etc.: Lewis Publishers

Cortes, C., & Vapnik, V. (1995a). Support-vector networks. *Machine learning, 20,* 273-297

Cortes, C., & Vapnik, V. (1995b). Support-vector networks. *Machine learning*, *20*, 273−297

Dean, W., Anderson, M., Milton, S., & Anderson, T. (2002). Avian assemblages in native Acacia and alien Prosopis drainage line woodland in the Kalahari, South Africa. *Journal of Arid Environments, 51*, 1-19

Dean, W.R.J., & Milton, S. (1999). *The Karoo: ecological patterns and processes*. Cambridge University Press

DefenceWeb (2015). <r12317_9_spot6-7_technical_sheet.pdf>. In

Deng, H. (2013). Guided Random Forest in the RRF Package. *arXiv preprint arXiv:1306.0237*

Deng, H., & Runger, G. (2012). Feature selection via regularized trees. In, *Neural Networks (IJCNN), The 2012 International Joint Conference on* (pp. 1-8): IEEE

Deng, H., & Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition*, *46*, 3483-3489

Diaz-Uriarte, R., & Alvarez de Andres, S. (2006). Gene selection and classification of microarray data using random forest *BMC Bioinformatics, 7*, 1-13

Digital Globe (2010). *The Benefits of the 8 Spectral Bands of WorldView-2.White Paper*. Digital Globe®, USA

Djoudi, H., & Brockhaus, M. (2011). Is adaptation to climate change gender neutral? Lessons from communities dependent on livestock and forests in northern Mali. *International Forestry Review, 13*, 123-135

Duro, D.C., Franklin, S.E., & Dubé, M.G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment, 118*, 259-272

Dzikiti, S., Schachtschneider, K., Naiken, V., Gush, M., & Le Maitre, D. (2013a). Comparison of water-use by alien invasive pine trees growing in riparian and non-riparian zones in the Western Cape Province, South Africa. *Forest Ecology and Management, 293*, 92-102

Dzikiti, S., Schachtschneider, K., Naiken, V., Gush, M., Moses, G., & Le Maitre, D. (2013b). Water relations and the effects of clearing invasive Prosopis trees on groundwater in an arid environment in the Northern Cape, South Africa. *Journal of Arid Environments, 90*, 103-113

El-Keblawy, A., & Al-Rawai, A. (2006). Effects of seed maturation time and dry storage on light and temperature requirements during germination in invasive Prosopis juliflora. *Flora-Morphology, Distribution, Functional Ecology of Plants, 201*, 135-143

Elfadl, M.A., & Luukkanen, O. (2006). Field studies on the ecological strategies of Prosopis juliflora in a dryland ecosystem: 1. A leaf gas exchange approach. *Journal of Arid Environments, 66*, 1-15

Evolution, S. (2012). Spectral Evolution RS-3500_User's Manual. In (p. 2)

Felker, P., Clark, P.R., Laag, A., & Pratt, P. (1981). Salinity tolerance of the tree legumes: Mesquite (Prosopis glandulosa var. torreyana, P. velutina andP. articulata) Algarrobo (P. chilensis), Kiawe (P. pallida) and Tamarugo (P. tamarugo) grown in sand culture on nitrogen-free media. *Plant and Soil, 61*, 311-317

Foody, G.M., Atkinson, P.M., Gething, P.W., Ravenhill, N.A., & Kelly, C.K. (2005). Identification of specific tree species in ancient semi-natural woodland from digital aerial sensor imagery. *Ecological Applications, 15*, 1233-1244

Fung, T., Ma, F.Y., & Sui, W.L. (1999). Hyperspectral Data Analysis for Subtropical Tree Species Identification. . In, *Proceedings: 1999 ASPRS Annual Conference*. Portland Oregon

Geesingis, D., Al Khawlani, M., & Abba, M.L. (2004). Management of introduced Prosopisspecies: can economic exploitation control an invasive species? *Unasylva, 217*

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*, 2225-2236

Ghazanfar, S. (1996). Invasive Prosopis in sultanate of oman. *Aliens, 3*, 10

Ghulam, A., Qin, Q., Teyip, T., & Li, Z.-L. (2007). Modified perpendicular drought index (MPDI): a real-time drought monitoring method. *ISPRS Journal of Photogrammetry and Remote Sensing, 62*, 150-164

Goetz, A.F.H. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing of Environment, 113, Supplement 1*, S5-S16

Gong, P., & Howarth, P. (1990). An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric engineering and remote sensing, 56*, 597-603

Groom, M.J., Meffe, G.K., & Carroll, C.R. (2006). *Principles of conservation biology*. Sinauer Associates Sunderland

Group, S.C.W. (1991). *Soil Classification: A Taxonomic System for South Africa: a Report on a Research Project Conducted Under the Auspices of the Soil and Irrigation Research Institute*. Department of Agricultural Development

Hagos, M., & Smit, G. (2005). Soil enrichment by Acacia mellifera subsp. detinens on nutrient poor sandy soil in a semi-arid southern African savanna. *Journal of Arid Environments, 61*, 47-59

Ham, J., Chen, Y., Crawford, M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 492 - 501.

Hansen, P., & 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

Harding, G. (1987). status of Prosopis as a weed. *Applied plant science= Toegepaste plantwetenskap*

Harvey, K., & Hill, G. (2001). Vegetation mapping of a tropical freshwater swamp in the Northern Territory, Australia: a comparison of aerial photography, Landsat TM and SPOT satellite imagery. *International Journal of Remote Sensing, 22*, 2911-2925

Henderson, L. (2001). *Alien Weeds and Invasive Plants. A complete guide to declared weeds and invaders in South Africa*. ARC, Plant Protection Research Inst.

Hoffman, M.T., S, T., Ntshona, Z., & Turner, S. (1999). Land degradation in South Africa. In D.o.E.A.a. Tourism (Ed.). Pretoria

Hoshino, B., Karamalla, A., Manayeva, K., Yoda, K., Suliman, M., Elgamri, M., Nawata, H., & Yasuda, H. (2012). Evaluating the invasion strategic of Mesquite (Prosopis juliflora) in eastern Sudan using remotely sensed technique

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

Hsu, P. (2007a). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing, 62*, 78-92

Hsu, P.H. (2007b). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing, 62*, 78-92

Huang, C.-y., & Asner, G.P. (2009). Applications of remote sensing to alien invasive plant studies. *Sensors*, *9*, 4869-4889

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

Hunt, E. (2009). <Remote-Sensing-of-Invasive-Leafy-Spurge.pdf>

Ismail, R., & Mutanga, O. (2011). Discriminating the early stages of Sirex noctilio infestation using classification tree ensembles and shortwave infrared bands. *International Journal of Remote Sensing, 32*, 4249-4266

Itholeng, K.B.I. (2007). The indigenous knowledge of the local community towards weeds and alien invasive plants in the Dinokana area, North-West Province, South Africa/by Kebalepile Benedict Itholeng. In: North-West University

Joshi, C., de Leeuw, J., & van Duren, I.C. (2004). Remote sensing and GIS applications for mapping and spatial modelling of invasive species. In, *Proceedings of ISPRS* (p. B7)

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support Vector Machines in R. *Journal of Statistical Software, 15*

Kaur, R., Gonzales, W.L., Llambi, L.D., Soriano, P.J., Callaway, R.M., Rout, M.E., & Gallaher, T.J. (2012). Community impacts of Prosopis juliflora invasion: biogeographic and congeneric comparisons. *PloS one*, *7*, e44966

Kavzoglu, T., & Mather, P.M. (2002). The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing, 23*, 2919-2937

Kent, M., & Coker, P. (1994). *Vegetation Description and Analysis: A practical Approach*. London: John Wiley and Sons

Klein, H., Hill, M., Zachariades, C., & Zimmermann, H. (2011). Regulation and risk assessment for importations and releases of biological control agents against invasive alien plants in South Africa. *African Entomology: Biological control of invasive alien plants in South Africa (1999-2010): Special Issue 2, 19,* 488-497

Kohavi, R., Sommerfield, D., & Dougherty, J. (1997). Data Mining Using a Machine Learning Library in C++. *International Journal on Artificial Intelligence Tools, 6*, 537-566

Kokaly, R., Despain, D., Clark, R., & Livo, K. (2003a). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment, 84*, 437-456

Kokaly, R.F., Despain, D.G., Clark, R.N., & Livo, K.E. (2003b). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment, 84*, 437 - 456

Kolar, C.S., & Lodge, D.M. (2001). Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution, 16*, 199-204

Kumar, L., Dury, S.J., Schmidt, K., & Skidmore, A. (2003). Imaging spectrometry and vegetation science. In F.D.v.d. Meer, & S.M.d. Jong (Eds.), *In Image Spectrometry* (pp. 111–156). London, UK: Kluwer Academic Publishers

Kumar, L., & Skidmore, A.K. (1998). Use of derivative spectroscopy to identify regions of difference between some Australian Eucalypt species. In, *9th Australasian Remote Sensing and Photogrammetry Conference*. Sydney, Australia

Kumar, P., Gupta, D.K., Mishra, V.N., & Prasad, R. (2015). Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data. *International Journal of Remote Sensing, 36*, 1604-1617

Kursa, M.B., & Rudnicki, W.R. (2010). Feature selection with the Boruta package. In: Journal

Lass, L.W., Prather, T.S., Glenn, N.F., Weber, K.T., Mundt, J.T., & Pettingill, J. (2009). A review of remote sensing of invasive weeds and example of the early detection of spotted knapweed (Centaurea maculosa) and babysbreath (Gypsophila paniculata) with a hyperspectral sensor

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

Laxén, J. (2007). Is prosopis a curse or a blessing. *An ecological-economic analysis of an invasive alien tree species in Sudan (Dissertation). Helsinki: University of HelsinkiViikki, Tropical Resources Institute.(VTR Institute)*

Le Maitre, D., De Lange, W., Richardson, D., Wise, R., & Van Wilgen, B. (2011). The economic consequences of the environmental impacts of alien plant invasions in South Africa. *Biological invasions: economic and environmental costs of alien plant, animal and microbe species (2nd edn). CRC Press, Taylor & Francis Group, Boca Raton, USA*, 295-323

Le Maitre, D.C., Richardson, D.M., & Chapman, R.A. (2004). Alien plant invasions in South Africa: driving forces and the human dimension: working for water. *South African Journal of Science, 100*, p. 103-112

Lee, K., & Lunetta, R. (1995). Wetland detection methods. In J.G. Lyon, McCarthy, J. (Ed.), *Wetland and environmental applications of GIS.* . New York

Lewis Publishers

Li, C., Wang, J., Wang, L., Hu, L., & Gong, P. (2014). Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sensing, 6*, 964-983

Li, F., Kustas, W.P., Prueger, J.H., Neale, C.M., & Jackson, T.J. (2005). Utility of remote sensing-based two-source energy balance model under low-and high-vegetation cover conditions. *Journal of Hydrometeorology, 6*, 878-891

Li, Z., & Fox, J.M. (2012). Mapping rubber tree growth in mainland Southeast Asia using time-series MODIS 250 m NDVI and statistical data. *Applied Geography, 32*, 420-432

Liew, S.C., Chang, C.W., & Lim, K.H. (2002). Hyperspectral land cover classification of EO-1 Hyperion data by principal component analysis and pixel unmixing. In, *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International* (pp. 3111-3113): IEEE

Lin, X., Sun, L., Li, Y., Guo, Z., Li, Y., Zhong, K., Wang, Q., Lu, X., Yang, Y., & Xu, G. (2010). A random forest of combined features in the classification of cut tobacco based on gas chromatography fingerprinting. *Talanta, 82*, 1571-1575

Lin, X., Sun, L., Li, Y., Guo, Z., Li, Y., Zhong, K., Wang, Q., Lu, X., Yang, Y., & Xu, G. (2011). A random forest of combined features in the classification of cut tobacco based on gas chromatography fingerprinting. *Talanta, 82*, 1571-1575

Lloyd, J., Van den Berg, E., & Badenhorst, N. (2002). Mapping the spatial distribution and biomass of Prosopis in the Northern Cape Province, South Africa, with the aid of remote sensing and geographic information systems. *Agricultural Research Council-Institute for Soil, Climate and Water, Pretoria, South Africa, Report No. GW/A/98/68*

Lowe, S., Browne, M., Boudjelas, S., & De Poorter, M. (2000). *100 of the world's worst invasive alien species: a selection from the global invasive species database*. Invasive Species Specialist Group Auckland,, New Zealand

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

Mack, R.N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., & Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications, 10*, 689-710

Mampholo, R.K. (2006). To determine the extent of bush encroachment with focus on Prosopis species on selected farms in the Vryburg district of North West Province/by Ramakgwale Klaas Mampholo. In: North-West University

Mansour, K., Mutanga, O., Everson, T., & Adam, E. (2012). Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. *ISPRS Journal of Photogrammetry and Remote Sensing, 70*, 56-65

Mashao, D.J. (2003). Comparing SVM and GMM on parametric feature-sets. In, *Proceedings of the 14th Annual Symposium of the Pattern Recognition Association of South Africa*: Citeseer

Masilamani, P., & Vadivelu, K. (1997). Tracing the seed development and maturation in kassod tree. *Madras Agricultural Journal, 84*, 19-21

Mather, P., & Tso, B. (2003). *Classification methods for remotely sensed data*. (Second ed.). CRC press/Taylor & Francis Group, London.

Matthews, S., & Brand, K. (2004). Africa invaded: the growing danger of invasive alien species. *Africa invaded: the growing danger of invasive alien species*

Maundu, P., Kibet, S., Morimoto, Y., Imbumi, M., & Adeka, R. (2009). Impact of Prosopis juliflora on Kenya's semi-arid and arid ecosystems and local livelihoods. *Biodiversity, 10*, 33-50

May, A., Pinder, J., & Kroh, G. (1997). A comparison of Landsat Thematic Mapper and SPOT multi-spectral imagery for the classification of shrub and meadow vegetation in northern California, USA. *International Journal of Remote Sensing, 18*, 3719-3728

McGlynn, I.O., & Okin, G.S. (2006). Characterization of shrub distribution using high spatial resolution remote sensing: Ecosystem implications for a former Chihuahuan Desert grassland. *Remote Sensing of Environment, 101*, 554-566

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing Letters, 42*, 1778-1790

Milton, E., Schaepman, M., Anderson, K., Kneubühler, M., & Fox, N. (2009). Progress in field spectroscopy. *Remote Sensing of Environment, 113*, S92-S109

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

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

Müllerová, J., Pergl, J., & Pyšek, P. (2013). Remote sensing as a tool for monitoring plant invasions: Testing the effects of data resolution and image classification approach on the detection of a model plant species Heracleum mantegazzianum (giant hogweed). *International Journal of Applied Earth Observation and Geoinformation, 25*, 55-65

Mumby, P., Green, E., Edwards, A., & Clark, C. (1999). The cost-effectiveness of remote sensing for tropical coastal resources assessment and management. *Journal of Environmental Management, 55*, 157-166

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

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

Mwangi, E., & Swallow, B. (2005). Invasion of Prosopis juliflora and local livelihoods: Case study from the lake Baringo area of Kenya. *Nairobi, Kenya: World Agroforestry Centre*

Mwangi, E., & Swallow, B. (2008). Prosopis juliflora invasion and rural livelihoods in the Lake Baringo area of Kenya. *Conservation and Society, 6*, 130

Nie, W., Yuan, Y., Kepner, W., Erickson, C., & Jackson, M. (2012). Hydrological impacts of mesquite encroachment in the upper San Pedro watershed. *Journal of Arid Environments, 82*, 147-155

Olshen, L., & Stone, C.J. (1984). Classification and regression trees. *Wadsworth International Group, 93*, 101

Omer, G., Mutanga, O., Abdel-Rahman, E.M., & 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

Omer, G., Mutanga, O., Abdel-Rahman, E.M., & Adam, E. (2015b). Performance of Support Vector Machines and Artificial Neural Network for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, PP*, 1-16

Oommen, T., Misra, D., Twarakavi, N.K., Prakash, A., Sahoo, B., & Bandopadhyay, S. (2008). An objective analysis of support vector machine based classification for remote sensing. *Mathematical geosciences, 40*, 409-424

Osmond, R. (2003). Mesquite best practice manual: control and management options for mesquite (Prosopis spp.) in Australia

Ozesmi, S., & Bauer, M. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management, 10*, 381-402

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*, 217-222

Pasiecznik, N. (1999). Prosopis-pest or providence, weed or wonder tree?

Pasiecznik, N., Harris, P., Smith, S., & Association, H.D.R. (2004). *Identifying tropical Prosopis species: a field guide*. HDRA Publishing

Pasiecznik, N., Harris, P., Trenchard, E., & Smith, S. (2006). Implications of uncertain Prosopis taxonomy for biocontrol. *Biocontrol News and Information, 27*

Pasiecznik, N.M., Felker, P., & Association, H.D.R. (2001). *The'Prosopis Juliflora'-'Prosopis Pallida'Complex: A Monograph*. HDRA Coventry

Peerbhay, K.Y., Mutanga, O., & Ismail, R. (2013). Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing, 79*, 19-28

Petropoulos, G.P., Kalaitzidis, C., & Prasad Vadrevu, K. (2012). Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences, 41*, 99-107

Pohl, C., & Van Genderen, J.L. (1998). Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing, 19,* 823-854

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

Priyanka, C., Kumar, P., Bankar, S.P., & Karthik, L. (2015). In vitro antibacterial activity and gas chromatography–mass spectroscopy analysis of Acacia karoo and Ziziphus mauritiana extracts. *Journal of Taibah University for Science, 9*, 13-19

Rejmánek, M., & Richardson, D.M. (2013). Trees and shrubs as invasive alien species–2013 update of the global database. *Diversity and distributions, 19*, 1093-1094

Robinson, T., Van Klinken, R., & Metternicht, G. (2008). Spatial and temporal rates and patterns of mesquite (Prosopis species) invasion in Western Australia. *Journal of Arid Environments, 72*, 175-188

Robinson, T., Wardell-Johnson, G., Pracilio, G., Brown, C., Corner, R., & van Klinken, R. (2016). Testing the discrimination and detection limits of WorldView-2 imagery on a challenging invasive plant target. *International Journal of Applied Earth Observation and Geoinformation, 44*, 23-30

Saatchi, S., Buermann, W., ter Steege, H., Mori, S., & Smith, T.B. (2008). Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sensing of Environment, 112*, 2000-2017

sansa.org (2014). In

Schlesinger, W., Reynolds, J., Cunningham, G., Huenneke, L., Jarrell, W., Virginia, R., & Whitford, W. (1990). Biological feedbacks in global desertification. Science 247, 1043e1048. In

Schmidt, K., & Skidmore, A. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, *85*, 92-108

Schuster, C., Förster, M., & Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing, 33,* 5583-5599

Shackleton, R.T., Le Maitre, D.C., Pasiecznik, N.M., & Richardson, D.M. (2014a). Prosopis: a global assessment of the biogeography, benefits, impacts and management of one of the world's worst woody invasive plant taxa. *AoB plants*, *6*, plu027

Shackleton, R.T., Le Maitre, D.C., Pasiecznik, N.M., & Richardson, D.M. (2014b). Prosopis: a global assessment of the biogeography, benefits, impacts and management of one of the world's worst woody invasive plant taxa. *AoB Plants, 6*

Shackleton, R.T., Le Maitre, D.C., & Richardson, D.M. (2015). Stakeholder perceptions and practices regarding Prosopis (mesquite) invasions and management in South Africa. *Ambio*, 1-13

Shaikh, M., Green, D., & Cross, H. (2001). A remote sensing approach to determine environmental flows for wetlands of the Lower Darling River, New South Wales, Australia. *International Journal of Remote Sensing, 22*, 1737-1751

Shaw, G., & Manolakis, D. (2002). Signal processing for hyperspectral image exploitation. *IEEE Signal Processing Magazine*, *19*, 12-16

Shiferaw, H., Teketay, D., Nemomissa, S., & Assefa, F. (2004). Some biological characteristics that foster the invasion of Prosopis juliflora (Sw.) DC. at Middle Awash Rift Valley Area, north-eastern Ethiopia. *Journal of Arid Environments*, *58*, 135-154

Shouse, M., Liang, L., & Fei, S. (2013). Identification of understory invasive exotic plants with remote sensing in urban forests. *International Journal of Applied Earth Observation and Geoinformation, 21*, 525-534

Smit, G., Richter, C., & Aucamp, A. (1999). Bush encroachment: an approach to understanding and managing the problem. *Veld management in South Africa. University of Natal Press, Scottsville*, 246-260

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research, 50*, 1250-1265

Steenkamp, H., & Chown, S. (1996). Influence of dense stands of an exotic tree, Prosopis glandulosa Benson, on a savanna dung beetle (Coleoptera: Scarabaeinae) assemblage in southern Africa. *Biological Conservation*, *78*, 305-311

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9,* 307

Strobl, C., Boulesteix, A., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*, 25

Strobl, C., & Zeileis, A. (2008). Exploring the statistical properties of a test for random forest variable importance. *COMPSTAT 2008 - Proceedings in Computational Statistics, II*, 59 - 66

Taylor, C., & Barker, N. (2012). Species limits in Vachellia (Acacia) karroo (Mimosoideae: Leguminoseae): Evidence from automated ISSR DNA "fingerprinting". *South African Journal of Botany, 83*, 36-43

Tian, F., Yang, L., Lv, F., & Zhou, P. (2009). Predicting liquid chromatographic retention times of peptides from the Drosophila melanogaster proteome by machine learning approaches. *Analytica Chimica Acta, 644*, 10-16

Update, S. (1997). Desert tree is a victim of its own success. *Sudan Update, 8*, 11

Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., & Green, R.O. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience, 54*, 523-534

Van den Berg, E.C. (2010). Detection, quantification and monitoring Prosopis spp. in the Northern Cape Province of South Africa using remote sensing and GIS/EC van den Berg. In: North-West University

Van den Berg, E.C., Kotze, I., & Beukes, H. (2014). Detection, quantification and monitoring of Prosopis in the Northern Cape Province of South Africa using remote sensing and GIS. *South African Journal of Geomatics, 2*, 68-81

van Klinken, R., Shepherd, D., Parr, R., Robinson, T., & Anderson, L. (2007). Mapping mesquite (Prosopis) distribution and density using visual aerial surveys. *Rangeland Ecology & Management, 60*, 408-416

van Klinken, R.D., Hoffmann, J.H., Zimmermann, H.G., Roberts, A.P., Muniappan, R., Reddy, G., & Raman, A. (2009). Prosopis species (Leguminosae). *Biological Control of Tropical Weeds using Arthropods*, 353-377

van Wilgen, B.W., Forsyth, G.G., Le Maitre, D.C., Wannenburgh, A., Kotzé, J.D., van den Berg, E., & Henderson, L. (2012). An assessment of the effectiveness of a large, national-scale invasive alien plant control strategy in South Africa. *Biological Conservation, 148*, 28-38

Vapnik, V.N. (1995). *The nature of statistical learning theory* New York: Springer-Verlag

Versfeld, D., Le Maitre, D., & Chapman, R. (1998). *Alien invading plants and water resources in South Africa: a preliminary assessment*. The Commission

Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G.A., & Torricelli, P. (2011). Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecological Modelling, 222*, 1471-1478

Waske, B., & Benediktsson, J.A. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 2880 - 2889

Waske, B., van der Linden, S., Benediktsson, J.A., Rabe, A., & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on, 48*, 2880-2889

Web, D. (2015). <SPOT 6 and 7 satellite imagery now available at SANSA _ defenceWeb.pdf>

Wise, R.M., van Wilgen, B.W., & Le Maitre, D.C. (2012a). Costs, benefits and management options for an invasive alien tree species: The case of mesquite in the Northern Cape, South Africa. *Journal of Arid Environments, 84*, 80-90

Wise, R.M., van Wilgen, B.W., & Le Maitre, D.C. (2012b). Costs, benefits and management options for an invasive alien tree species: The case of mesquite in the Northern Cape, South Africa. *Journal of Arid Environments, 84*, 80-90

Witt, A. (2010). Impacts of invasive plants and their sustainable management in agro-ecosystems in Africa: a review. *CABI Africa*, 1102-1109

Woodborne, S. (2004). A water stress history of an Acacia erioloba (Kameeldoring) tree from the Koichab. *CSIR Environmentek, Pretoria.[Links]*

Yu, L., Porwal, A., Holden, E.-J., & Dentith, M.C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences, 45*, 229-239

Zachariades, C., Hoffmann, J., & Roberts, A. (2011). Biological control of mesquite (Prosopis species)(Fabaceae) in South Africa. *African Entomology: Biological control of invasive alien plants in South Africa (1999-2010): Special Issue 2, 19*, 402-415

Zeila, A. (2011). Mapping and managing the spread of Prosopis Juliflora in Garissa County, Kenya. In: Kenyatta University

Zhang, G., Li, H., & Fang, B. (2009). Discriminating acidic and alkaline enzymes using a random forest model with secondary structure amino acid composition. *Process Biochemistry, 44*, 654-660

Zhao, D., Huang, L., Li, J., & Qi, J. (2007). A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. *ISPRS Journal of Photogrammetry and Remote Sensing, 62*, 25-33

Zimmermann, H. (1991). Biological control of mesquite, Prosopis spp.(Fabaceae), in South Africa. *Agriculture, Ecosystems & Environment, 37*, 175-186

Zimmermann, H., Hoffmann, J., & Witt, A. (2006). A South African perspective on prosopis. *Biocontrol News and Information, 27*, 6-9

Zimmermann, H., & Pasiecznik, N. (2005). Realistic approaches to the management of Prosopis species in South Africa. *HDRA*, *Coventry*