Mapping and modelling spatial distribution of *Prosopis Juliflora* invasion in the Northern Cape Province of South Africa

Research report submitted to the Faculty of Science, University of the Witwatersrand. In partial fulfilment of the requirements for Master of Science (MSc) in Geographic Information Systems and Remote Sensing

Leolin Ntsika Qegu 1445225

3/20/2019

Supervisor: Dr Elhadi Adam

DECLARATION

I, Leolin Ntsika Qegu, declare that this research report is my own unaided work. It is being submitted to the Degree of Master of Science in Geographical Information Systems and Remote Sensing to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature of candidate

20th day of March 2019 in Johannesburg

Abstract

The negative effects of *Prosopis* on the ecological infrastructure and economies of invaded areas is well documented. One of the factors hindering invasion management of eradication is lack of reliable spatial data. Policy makers, ecologists and biodiversity planners are seeking efficient and cost friendly methods of evaluating spatial dynamics such as invasion rate and hot spots. Inventory methods such as field surveys are expensive and labour-intensive thus unsustainable while Remote Sensing (RS) and Geographic Information Systems (GIS) methods are efficient and cost effective. Moreover, field surveys are often limited to small geographic extents while RS and GIS techniques have the ability to analyse vast geographic extents.

The aim of this study was to evaluate potential of mapping and modelling *Prosopis Juliflora* using Sentinel – 2A and Species Distribution Modelling (SDM) in arid regions of the Northern Cape Province of South Africa. It was achieved by three specific objectives; using Random Forest (RF) and Support Vector Machine (SVM) classifiers to map *Prosopis Juliflora* and other land cover; evaluation of the distribution of *Prosopis Juliflora* using spatial point pattern analysis; predicting areas susceptible to invasion using binomial logistic regression. Results obtained from the first objective illustrated the potential of 10 m resolution Sentinel – 2A in classifying *Prosopis Juliflora* from other land cover, RF and SVM reported overall accuracies of 88.86 % and 91.42 % respectively. Spatial point pattern analysis illustrated that *Prosopis Juliflora* exhibited clustering which was evaluated with respect to various geophysical factors such as geology, soil, elevation and slope. The binomial regression successfully predicted areas that are susceptible to invasion. The final model had a pseudo R^2 of 0.78 and Akaike Information Criterion (AIC) of 223.35.

List of Tables

List of Figures

Contents

Acronyms

AIC - Akaike's Information Criterion

CART - Classification and Regression Trees

CBS - Citrus Black Spot

DEM - Digital Elevation Model

EDA - Exploratory Data Analysis

ESDA - Exploratory Spatial Data Analysis

ESA - European Space Agency

GIS - Geographic Information Systems

IUCN - International Union for the Conservation of Nature

MODIS - Moderate Resolution Imaging Spectro-radiometer

NDVI - Normalised Difference Vegetation Index

NDII - Normalised Difference Infrared Index

LULC - Land Use-Land Cover

OOB - out of bag

OA - Overall Accuracy

PA - Producer's Accuracy

RF - Random Forest

RS – Remote Sensing

SDM - Species Distribution Model

SVM - Support Vector Machine

SPPA - Spatial Point Pattern Analysis

UA - User's Accuracy

Chapter One

1. General introduction

1.1 Background

It is imperative to have reliable spatial information on the distribution of alien invasive species over large geographic areas, this is crucial for biodiversity spatial planning (Wakie et al. 2014). Biodiversity plays a critical role to the welfare of humans thus sustainable coexistence is highly imperative (Wakie et al. 2014b). Maintaining a harmonious relationship between humans and biodiversity is becoming increasingly difficult due to the occurrence of alien invasive species (Kumar & Mathur 2014). Alien invasive species refers to species accidentally or deliberately introduced to their nonindigenous ranges and disturb the ecology (Van den Berg et al. 2013). Alien invasive species have negative impacts on biodiversity, water resources and agriculture. Some alien invasive species are intentionally distributed to provide positive services but subsequently impact negatively on the environment (Berhanu et al. 2016). One example of such species is *Prosopis Juliflora (P. juliflora).* In South Africa, *P. juliflora* was intentionally planted to provide fodder and shade however recent literature shows its negative environmental impacts.

The negative impacts of *P. juliflora* have received significant attention across the world; the Invasive Species Specialist Group of the International Union for the Conservation of Nature (IUCN) listed it in the top 100 least wanted species. As a result many countries have developed policy framework to eradicate *Prosopis*. However, quantifying the spatial distribution remains a challenge. Field based inventory methods of tracking and recording spatial distribution of *P. juliflora* are expensive, time-consuming and tedious (Kumar & Mathur 2014). Remote Sensing (RS) and Species Distribution Modelling (SDM) offer an opportunity to analyse the distribution of *P. juliflora* over large geographic extents, which might not be accessible through field surveys (Van den Berg et al. 2013).

Quantifying the spatial distribution and monitoring invasion rate of an alien invasive species are fundamental towards building invasive species management strategies. However, there is an increasing interest in predicting invasive species distribution (Campbell 2002; Chuvieco & Huete 2010).

Many scientists are currently combining statistical methods, environmental data and advanced machine learning techniques to produce invasive species predictive models (Brunel et al. 2010; Li et al. 2005; Ustin et al. 2004). Repetitive field mapping and remote sensing may be used to measure invasion rate however, they often fall short in predicting future invasion trends (Mather 2004; Blaschke 2010). On the other hand, SDM offer insights into future trends by analysing environmental dynamics in invaded areas and subsequently predicting future ranges.

Various studies have employed RS and SDM to analyse the spatial dynamics of alien invasions (Malahlela et al, 2016; Van den Berg et al. 2013; Lillesand et al., 2004). However costs associated with high quality data limits studies to small geographic extents. In contrast, low quality data studies cover large geographic extents however they are prone to inaccuracies (Van den Berg et al. 2013). The moderate to high specifications of Sentinel 2 presents an opportunity improve accuracies at no cost (van der Werff & van der Meer, 2015). Sentinel's capacity for analysing *P. juliflora* has not yet been tested let alone in arid environments. This research will assess the ability of Sentinel 2 in mapping *P. juliflora* in Northern Cape (NC), South Africa.

1.2 Research Problem Statement

Prosopis was previously planted to provide fodder, shade, and fuel in arid areas across the world, however it was later listed in the top 100 least wanted species by the IUCN (Bromilow, 2010; Baillie et al. 2004). *Prosopis* invasion often result in socio-economic challenges particularly for developing countries which are largely dependent on natural resources (Berhanu et al. 2016; Van den Berg et al. 2014). *Prosopis* has invaded approximately 4 % of NC and it is spreading at an estimated annual rate of 15 % (Van den Berg et al. 2014). The deep root systems of *Prosopis* lower the water table in semi-arid to arid regions of the Northern Cape which result in *Acacia erioloba* fatalities (Wise et al. 2012b). *Prosopis* invasion reduced biodiversity of the Nama-Karoo biome affected species include birds, dung beetles and raptors (Steenkamp & Chown, 1996).

Mapping and analysing the spatial distribution of *P. juliflora* over large areas using traditional field survey methods is costly and labour-intensive (Abebe & Corresponding, 2012; Ndhlovu et al., 2011). Remote Sensing and species distribution modelling offer efficient methods to map and analyse *P. juliflora.* Freely available imagery such as Landsat and MODIS can map *P. juliflora* over large areas they are often prone to spectral confusion and omission errors respectively (Van den Berg et al. 2014; Hoshino et al., 2013) On the other hand commercial imagery with high spectral and spatial resolution produce higher accuracies however studies are often limited to small areas due to imagery costs (Mureriwa et al. 2015).

The launch of Sentinel-2A equipped with moderate to high spectral and spatial resolution offers an opportunity to map *P. juliflora* with less spectral confusion and omission error (Vanthof & Kelly. 2017). Moreover, the freely distributed imagery presents an opportunity to map *P. juliflora* invasion over large areas (van der Werff & van der Meer, 2015). The rapid development of species distribution modelling presents biodiversity planners with an opportunity of predicting areas susceptible to *P. juliflora* invasion (Colwell & Rangel, 2009). Risk and susceptibility maps have the potential to assist in the construction of *P. juliflora* invasion management plans.

1.3 Research Question

This research was guided by the following question: "How can Sentinel 2 and species distribution modelling analyse *P juliflora* in Hantam Local Municipality? "

1.4 Aim

The aim of this research was to map and model the spatial invasion of *P. juliflora* using Sentinel – 2A and species distribution modelling.

1.5 Objectives

The specific objectives of this research were to:

1.5.1 Map *P. juliflora* using Sentinel 2 imagery and advanced machine learning classifiers (Random Forest and Support Vector Machine Classifiers).

1.5.2 Evaluate the spatial distribution *P. juliflora* and the coexisting land cover classes across Hantam Local Municipality.

1.5.3 Map and evaluate susceptibility of Hantam Local Municipality to *P. juliflora* invasion through multivariate species distribution modelling.

1.6 Research Report Outline

The research report is composed of six chapters i.e chapter 1 (Introduction), chapter 2 (Literature review), chapter 3 (Methods and Materials), chapter 4 (Results), chapter 5 (Discussion) and chapter 6 (Conclusion). The content of the chapters is:

Chapter 1 sets the scene by providing an overview of the study and briefly discussing concepts and challenges of mapping and analysing *Prosopis* in South Africa. The research problem statement, research question, aim and objectives are also outlined.

Chapter 2 is the literature review that details the arrival of *Prosopis* in Africa and its biological characteristics, ecological dynamics i.e the negative impacts and benefits of *Prosopis* in arid environments. Furthermore, the chapter details the use of remote sensing and species distribution modelling and mapping *Prosopis.*

Chapter 3 provides a brief background on the geophysical settings of Hantam Local Municipality and the characteristics of Sentinel – 2A imagery and other datasets which are used as variables in the binomial logistic regression model. The chapter details sampling methods employed to select the dataset used to train and test Random Forest and Support Vector Machine classifiers. Additionally, accuracy assessment of the classifiers, spatial point pattern analysis and binomial logistic regression are discussed.

Chapter 4 reports and visualises the findings of the methods employed to achieve the aim and objectives of the study. Some of the key results include confusion matrices of both Random Forest and Support Vector Machine classifiers, *P. juliflora* density bar graphs obtained from the spatial point pattern analysis and a probability map obtained from the binomial logistic regression model.

Chapter 5 is a discussion that compares the findings of the study to previous studies which employed similar methods. It is noted that the results of this study compare well with studies carried out in other arid areas.

Chapter 6 draw conclusions, highlights limitations and makes recommendations for future studies.

Chapter Two

2. Literature Review

2.1 Prosopis Juliflora in Africa

P. juliflora is a perennial multipurpose dry land tree that originates from South America and the Caribbean (Abebe & Corresponding 2012). It is an evergreen, fast-growing and drought resistant tree that is mostly found in semi-arid environments. It is the descendent of the *Leguminosae* family, subfamily *Mimisoideae* and genus *Prosopis.* It has an average trunk diameter of 1, 2 m and its height ranges from 3 - 12 m (Tesema et al. 2013). The drought resistant tree or shrub was introduced to other parts of the world 100 – 150 years ago. In Africa the species has invaded arid environments in countries such as Kenya, Sudan, Ethiopia and South Africa (Abebe & Corresponding 2012). The introduction of *P. juliflora* in South Africa is documented to have occurred in the late 1800s. Deliberate plantation and distribution of *Prosopis* continued up to 1960, it was largely planted for shade and fodder during severe droughts (Harding & Bate, 1991; Zimmermann, 1991; Poynton, 2009). A study by Van Der Berg (2010) discovered that invasion covered about 1 million hectares from 2002 to 2007, this equals to 27% invasion rate. Thus far the invasion is prominent in arid and semiarid parts of Northern Cape and Western Cape (Shackleton et al. 2015).

2.2 Ecological dynamics of *Prosopis Juliflora*

A series of studies conducted through extensive surveys of small areas revealed some key findings about the invasion of *Prosopis* in South Africa (Ndhlovu et al. 2011; Wise et al. 2012; Schachtschneider & February 2013; Shackleton et al. 2015). *P. juliflora* primarily threatens the semi-arid and arid savannah and grassland ecosystems of the Northern and Western Cape provinces of South Africa (Shackleton et al. 2015). The threat also extends to ecosystem services such as water supply and pastures (Dzikiti et al. 2013). Furthermore, disruptions are reported on insect and bird species abundance and disturbances to the composition of the Kalahari (Steenkamp & Chown, 1996; Dean et al. 2002). Moreover, Schachtschneider and February (2013) noted that *P. juliflora* invasion is a driving factor responsible for the mortality of *Acacia erioloba* tree in the Kalahari Desert. The afore-discussed literature provides micro scale dynamics about *Prosopis* invasions which might be investigated at a macro scale using remote sensing.

Contrary to the well documented negative ecological effects of *Prosopis*, some studies explored positive ecological effects. *Prosopis* plays a critical role in reducing soil loss as a result of wind and water induced erosion. Furthermore, it is instrumental in improving soil salinity. These benefits have led to a policy dilemma in many parts of the world with some countries opting to control the species rather than eradication (Abebe & Corresponding 2012).

2.3 Monitoring and managing the spatial distribution of *Prosopis*

In South Africa, monitoring and controlling invasive alien plants is assigned to the Working for Water division, which is part of the Extended Public Works Programme. The main objective of the programme is to reduce the impacts of alien trees on water resources (Marais & Wannenburgh 2008). Mechanical, biological and chemical control measures have been used to eliminate *P. juliflora,* however they have failed to reduce overall invasion extents and impacts (Zachariades et al., 2011; Van Wilgen et al., 2012). One of the leading factors hindering effective monitoring and management is the absence of information about the vastness of *Prosopis* (Van Der Berg et al., 2013). Mechanical control measures include among others cutting, burning, uprooting and felling trees (Geesingis et al. 2004; Harding, 1987). These methods require remedial action as they often involve land and soil degradation (Van den Berg 2010). Chemical control measures include spraying and applying herbicides on cleared areas to prevent regrowth, precise dosage and constant monitoring are required to prevent negative environmental side effects (Van Klinken et al. 2009; Zeila 2011). Biological control measures consist of deploying species that feed on seeds of *Prosopis* to reduce the invasion rate. *Algarobius prosopis, Algarobius bottimeri* and *Neltumius arizonensis* are some the beetle species that were deployed at invaded areas in South Africa. Despite the heavy investment in biological control measures, they have achieved very little success (Coetzer & Hoffmann 1997; Klein et al. 2011; Zimmermann 1991). Other control measures involve cutting *Prosopis* for firewood, furniture and charcoal production (Shackleton et al. 2014b).

2.4 Use of remote sensing in mapping *Prosopis*

Field based inventory methods of mapping the spatial distribution of alien invasive species are laborious, time consuming, costly and often limited to small physically accessible areas (Kent & Coker 1994; Lee & Lunetta 1995; Kokaljl et al., 2007). In contrast, remote sensing technology allows professionals in various fields to observe and measure objects of interest without making contact (Adam & Mutanga 2009).

Recent studies have explored the capacity of remote sensing technology in mapping spatial distribution *Prosopis* over vast regions. Van den Berg et al (2014) combined terrain analysis and remote sensing techniques to monitor *Prosopis* in Northern Cape, South Africa. Here, thresh-holding of the Normalised Difference Vegetation Index (NDVI) was used to separate *Prosopis* from other species based on greenness. Classification accuracy of 72 % was achieved. However, *Prosopis* was spectrally confused with other tree species. Hoshino et al (2013) applied Normalised Difference Infrared Index (NDII) thresholds on Landsat imagery to map *Prosopis* in arid areas of Sudan. Accuracy assessment revealed that *Prosopis* clustered along rivers were detected, however sparse *Prosopis* trees away from the river were omitted. Both works portray the potential of using remote sensing to monitor *Prosopis.* However omission errors and spectral mixing are some of the limitations. Zeila (2011) integrated Landsat classification and citizen geographic information systems to investigate the spatial distribution and socio-economic impacts of *Prosopis* in Garissa County in Kenya. The study was carried out in the dry season to exploit the evergreen characteristic of *Prosopis* with the aim to improve chances of detecting *Prosopis* from surrounding species. Classifying imagery acquired during the dry season did not improve the spectral confusion associated with Landsat. Mureriwa et al (2015) used random forest and support vector machine to classify World-View 2 imagery to map the spatial distribution of *Prosopis* and its co-existing species near Griekwastad, a small town in Northern Cape. In this study, random forest and support vector machine classifiers produced overall accuracies of 86.59 % and 85.98 % respectively. This study proved that good spectral and spatial resolution has the potential to discriminate and reduce spectral confusion associated with moderate to poor spatial and spectral resolution. Vanthof and Kelly (2017) employed SVM, RF and CART classifiers on Sentinel-2A and Landsat 8 to detect and map *P. juliflora* in Tamil Nadu, India. Here, SVM classier produced the highest accuracies and Sentinel-2A performed better than Landsat 8. The work affirms increases in accuracy associated with improvements in spectral and spatial resolution. Vijayaraghavan and Sundaramoorthy (2012) derived vegetation indices and texture measures from SPOT-XS and IRS Linear Imaging Self-Scanning Sensor to map *P.*

julflora and co-existing species in Jodhpur, India. Overall accuracies of 86.98 % and 89.86 % were obtained from SPOT-XS and IRS-LISS3 respectively. User's accuracy of 62.24 % and 70.15 % for *P. juliflora* class were obtained from the SPOT and IRS datasets respectively. Here, high overall accuracies were achieved however moderate *P. juliflora* user's accuracy were obtained. Laliberte et al (2004) employed object based image classification on historic aerial photographs and QuickBird to quantify *Prosopis glandulosa* encroachment in grasslands of south western America. The results also reflected the shape and compactness of *Prosopis glandulosa* and other classified species. Eva et al (2005) evaluated the potential of wavelet analysis for detecting *Prosopis* invasion from aerial photographs. Their accuracy assessment reported commission and omission errors of 5 % and 8 % respectively. Robinson et al (2007) classified aerial photographs to detect and analyse spatial dynamics of *Prosopis* in Western Australia. Here, only canopies greater than 3 m^2 were detected even though the image spatial resolution was 1.4 m. Both studies show the potential of analysing *Prosopis* using aerial photographs, however costs associated with flights remains a limiting factor. The integration of machine learning and high resolution imagery presents a great opportunity to map *Prosopis* however data acquisition remains a hindrance to large scale studies.

Table 1: Comparison of applications of remote sensing on invasive species

Country	Type of data	Methods used	Accuracy	Author
South Africa	Landsat, NOAA, MODIS, SPOT-5	NDVI classification, EVI classification	72 %	Van den Berg et al. (2014)
South Africa	WorldView-2	RF and SVM	RF = 86.59 % SVM = 85.98 %	Mureriwa et al. (2015)
Sudan	Landsat	NDII thresh-holds	70 %	Hoshino et al. (2013)
Kenya	Landsat	Sub-pixel	84 %	Zeila (2011)

		classification in ERDAS imagine; change detection and surveys		
India	Landsat-8, Sentinel-2A	RF, SVM and CART	RF + Sentinel-2A: 87 % SVM +Sentinel-2A: 89.42 RF + Landsat-8: 71.03 % SVM + Landsat-8: 78.54 %	Vanthof and Kelly (2017)
India	SPOT-XS and IRS-LISS3	Vegetation indices and texture measures	SPOT-XS 86.98 % RS-LISS3 89.86 %	Vijayaraghavan and Sundaramoorthy (2012)
United States of America	Aerial photographs and QuickBird	OBIA	12 % omission and 9% commission error	Laliberte et al. (2004)
Australia	Aerial photographs	RF	76.32 %	Robinson et al. (2007)

2.5 Spatial point pattern analysis of vegetation

The degree of density of individual plants has an effect of how they use resources, thus quantifying spatial patterns plays a critical role in developing environmental management plans (Richardson & Thuiller., 2007). Jose et al (2010) used spatial point pattern analysis to assess the distribution of *Pyrus bourgaeana* trees and their seeds dispersed by three mammals (badger, fox, and wild boar) within a 72 ha plot. A total of nine plots exhibited clustering while three exhibited complete spatial randomness. Point pattern analysis with respect to the three mammals revealed that badger feces and dispersed seeds were clustered while boar and fox dispersed feces were scattered (Jose et al. 2010). Lan et al (2009) employed Ripley's function to assess distribution pattern of 131 tree species across different growth stages within a 20 ha plot of a tropical rain forest in China. Here, 83 % of the species exhibited clustering at sapling stage while 96 % of the species exhibited randomness. Sposito et al (2007) used binomial dispersion, Ripley's K function and Monte Carlo test to assess the distribution of Citrus Black Spot (CBS) across three commercial farms in Brazil. Binomial dispersal index reported clustering for all quadrat sizes, Ripley's K function indicated that two or three trees neighbouring were affected with CBS and disease dispersal occurred at distances less than 24 m. Pillay and Ward, (2012) employed nearest neighbour and Ripley's L function to assess the distribution of *Accacia karoo* from savannas in Kwa Zulu Natal, South Africa. The spatial distribution of trees was aggregated and the nearest neighbour analysis discovered positive correlations between the sum of the distances to the four nearest neighbours and the sum of the canopy diameters of the central tree and its four nearest neighbours, indicating the presence of competition (Pillay & Ward., 2012). The above discussed literature illustrated the potential of employing spatial point pattern analysis to evaluate distribution of vegetation. However, little work has been done to assess the spatial dynamics of alien invasive species using pattern analysis.

Dowhower and Teague, (2017) used the nearest neighbour to determine the distribution pattern of juvenile and mature *Prosopis glandulosa* in relation to soil depth in Texas. Juveniles were aggregated in both deep and shallow soils while mature *Prosopis glandulosa* was random in deep soils and aggregated shallow soils. Browning et al, (2014) compared the spatial distribution of *Prosopis velutina* between grazed and protected grasslands over 74 years. The *Prosopis velutina* density increased on both protected and grazed grasslands but the clustering increased with time on the former while the pattern remained random on the latter. Khosravi and Erfanifard, (2017) used quadrat analysis to investigate the spatial distribution of *Prosopis cineraria* in Barchah protected area, Iran. The analysis reported homogeneity of the species over the 700m x 700m plot. A spatial-temporal study by Robinson et al, (2007) discovered that *Prosopis* invasion rates were high across all land types in Western Australia however coalescence was noted in red loamy soils and riparian areas. These studies positively exploited the benefits of spatial point pattern analysis however they were all carried out at small scales. The extraction of species points from an image classification presents an opportunity to evaluate spatial patterns of invasive species at large scales.

2.6 Species Distribution Modelling and invasive species predictions

Traditionally, predicting future ranges for alien invasive species was challenging and often impossible due to lack of accurate models and poor data (Brunel et al., 2010 Cook et al. 2007; Richardson & Thuiller., 2007). Technological advances and access to improved data led to the inception of SDMs which are used to predict future ranges for invasive species. SDMs use statistical modelling, machine learning and environmental data to produce high risk maps (Colwell & Rangel, 2009). The two main advantages that make SDMs preferable over predictive models are that they use a small sample of the invaded area to identify areas at high of invasion and SDMs do not need repeat measures to make a prediction (Buckley et al., 2010; Kearney & Porter., 2009)

Lemke and Brown (2012) used logistic regression and maximum entropy to predict habitats of three alien invasive species in the Cumberland Plateau and Mountain Range in North America. Their logistic regression used information on both occurrence and absence while the Maxent used occurrence information only. The models used various input variables (anthropogenic, climate, hydrology, Landsat and land use land cover). The logistic regression used Akaike's Information Criterion (AIC) to select significant input variables (Lemke & Brown, 2012). An assessment of their models yielded an agreement of 93 % for the most prevalent species, 67 % and 87 % for the other species respectively (Lemke & Brown, 2012). The combination they used proved to be cost effective and efficient for aiding conservation plans.

Miyamoto *et al*, (2004) in a study on the negative impacts of *Callosciurus erythraeus thaiwanensis* on the native Japanese squirrel (*Sciuruslis*) habitat in Kamakura City, Kanagawa Prefecture introduced ten environmental variables to the model to determine which variables were effective in predicting habitats. Their model produced statistical significance at p-value < 0,001 and it correctly predicted 87.9 %

of the woods with *Callosciurus erythraeus thaiwanensis* presence. Malahlela et al, (2016) integrated environmental and remote sensing variables in a stepwise logistic regression to predict the occurrence of *Chromolaena odorata* in Dukuduku Forest, Kwa Zulu Natal, South Africa. Their model produced statistical significance at p-value < 0, 05 and it correctly predicted 87 % of the forest gaps with *Chromolaena odorata* presence however it was dominated by World-View derived variables which accounted for 6 of the 8 input variable of the model. The studies discussed above were useful in predicting the occurrence of alien invasive species however none of them carried out in arid regions.

Wakie et al, (2104) integrated species occurrence points, Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation indices, and topo-climatic predictors to map and predict areas susceptible to *P. juliflora* invasion in Afar, Ethiopia. Maxent models based on non-correlated variables managed to map current and future infestations of *P. juliflora* however the coarse spatial resolution of MODIS omitted small dispersed patches.

The above discussed literature outlines the origins, socio-economic and ecological dynamics of *P. juliflora*. Over the last three decades *P. juliflora* has been listed as one of the least wanted invasive species thus leading to crafting of strategies to manage its spread and impact. One of the factors that hinder effective control of *P. juliflora* is lack of spatial data quantifying the invasion rate. RS, GIS and SDM have proven to be effective in representing *P. juliflora* spatial dynamics efficiently. SPPA provides great insights about the spatial distribution of invasive species and potential factors driving invasion. However, cost associated with high quality data often limit studies to small geographic areas.

Chapter Three

3. Materials and Methods

3.1 Study Area

3.1.1 Background

Hantam Local Municipality is located of the southern part of Northern Cape and it borders the Western Cape Province in the south. It is a vast and arid area covering 36 128km², this amounts to 28 % of the Namakwa District Municipality's surface area (Figure 1). It is characterised by fluctuating temperatures and diverse topographies (Van Der Berg et al., 2013). The municipality is made up of four small towns namely: Calvinia and Newtown in the centre, Bradvlie in the north, Loeriesfontein and Nieuwoudtvillein the west and Middelpos in the south (Fig 1).

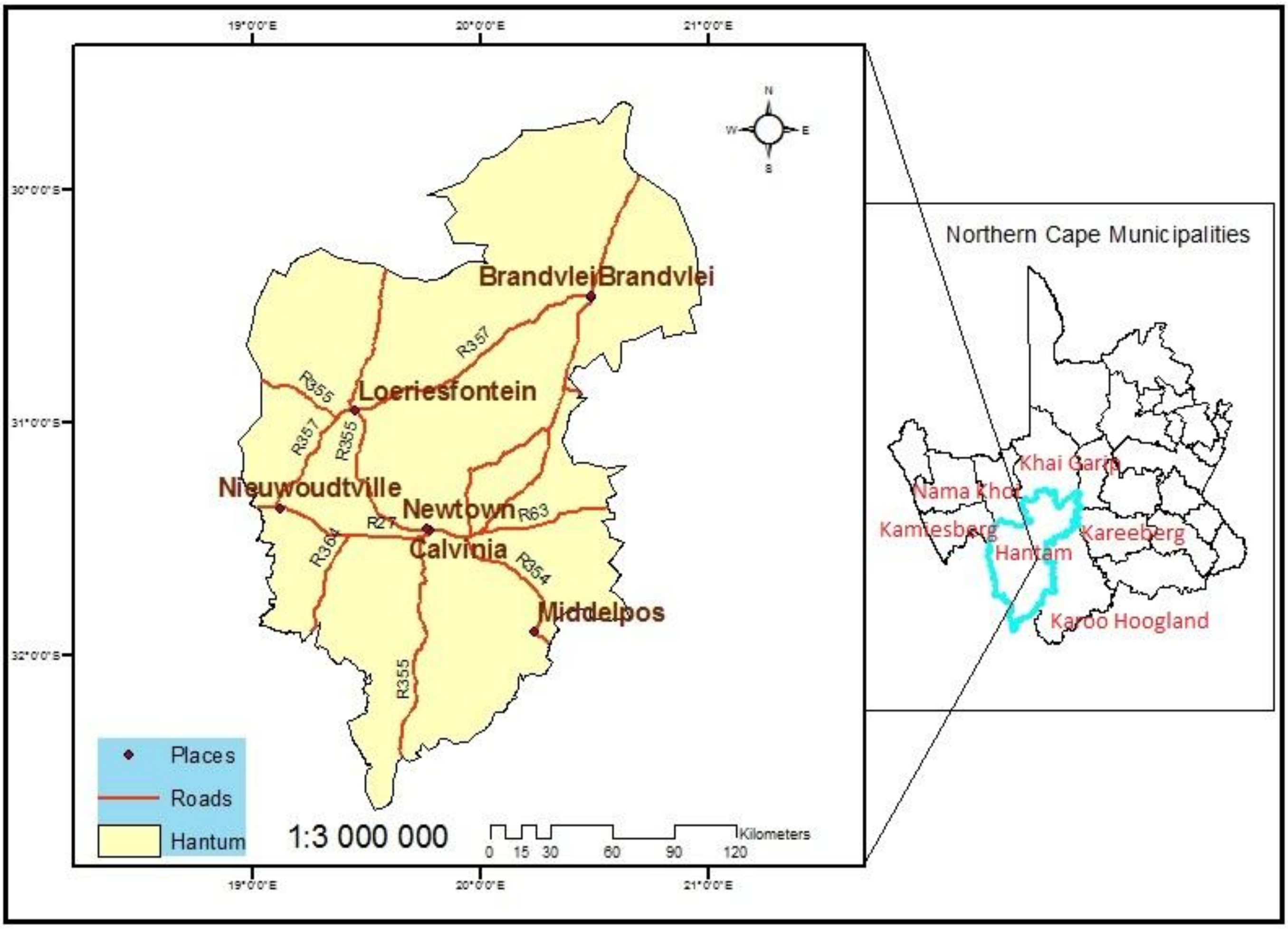

Figure 1: Map of the study area, Hantam Local Municipality Northern Cape in yellow

3.1.2 Geology

The underlying rocks of the Hantam Local Municipality include shales of the Dywaka and Ecca Group, granites of the Namaqua and shales of the Transvaal Supergroup in the east (Fig 2). The central and northern regions are dominated by sands of the Khalahari Group. The southern region is dominated by shales of the Beaufort Group (Voster, 2003).

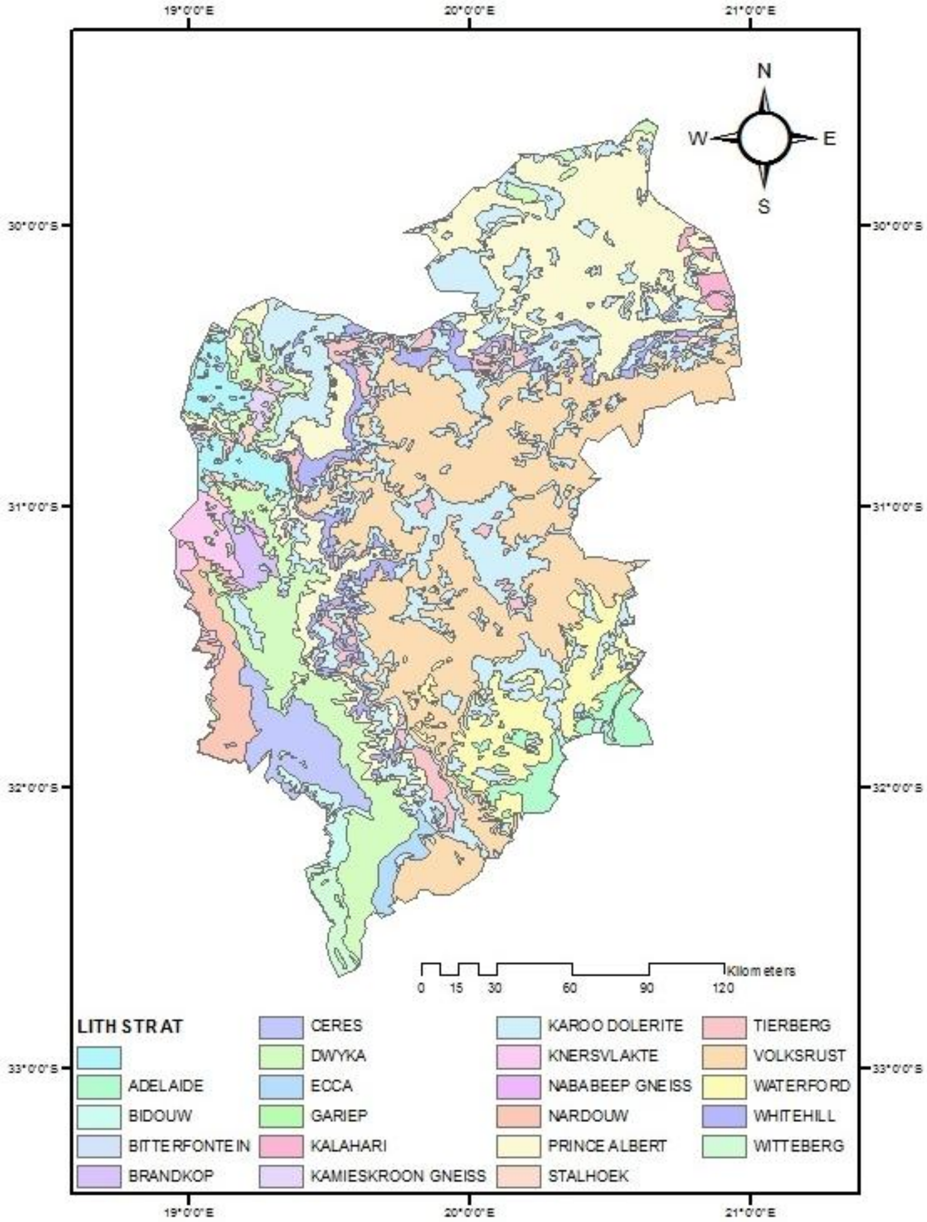

Figure 2: Hantam Local Municipality geological formations

3.1.3 Vegetation

The Hantam Local Municipality has approximately 2300 plant species that are grouped into three biomes (Fig 3). The Nama-Karoo is dominant in the northern part and the Succulent Karoo is dominant on the west. The Fynbos Biome is dominant on the central region (Mucina & Rutherford, 2006).

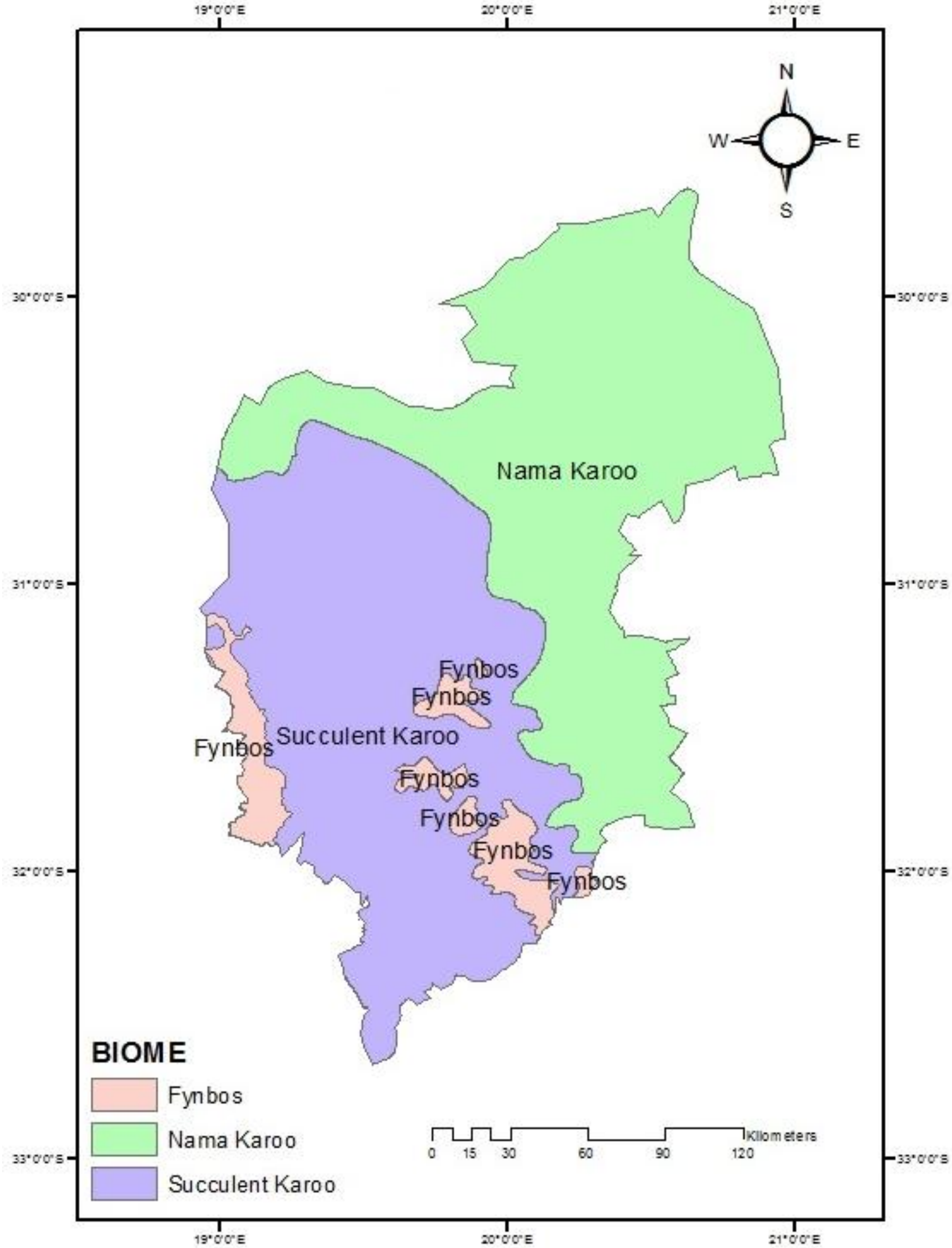

Figure 3: A map of Hantam Local Municipality showing the biomes

3.1.4 Climate

Hantam Local Municipality is characterised by semi-arid to arid climate, a low rainfall and high temperatures. In winter, temperatures range from a minimum of - 8 Degree

Celsius to a maximum of 25 Degrees Celsius. In summer, temperatures range from a minimum of 11 Degree Celsius to a maximum of 36 Degrees Celsius. In the Northern Cape summer runs in December to February and winter starts in May and ends in August (South African Weather Services, 2016).

Precipitation is rare, the mean annual rainfall increases eastwards. The annual rainfall ranges from 6 mm to 23 mm with an annual average of 146 mm (South African Weather Services, 2016).

3.2 Materials

3.2.1 Sentinel – 2A Imagery

Copernicus is an earth observation program that was launched by the European Space Agency (ESA) in association with European Commission. Its space component is made up of 5 missions, Sentinel 1 to 5. The main aim of Copernicus is providing environmental monitoring data at no cost (van der Werff & van der Meer, 2015).

In line with the Copernicus program the European Space Agency launched Sentinel 2 that is envisaged to provide data with better spectral and spatial resolution than Landsat. Sentinel 2 is a sun-synchronised multi spectral imager with 13 bands with spatial resolution ranging from 10m to 60m (Table 1). These bands are situated in Visible and Near Infrared and Shortwave Infrared regions of the spectrum (van der Werff & van der Meer, 2015). Moreover, Sentinel 2 has a revisit time of 5 days and red edge bands, these specifications makes it ideal for monitoring vegetation (van der Meer et al., 2014).

Table 2: Sentinel 2-A sensor characteristics

Band Number		**Central wavelength (um)**	**Bandwidth (nm)**	**Spatial Resolution (m)**
1	Coastal aerosol	0.443	20	60
2	Blue	0.490	65	10
3	Green	0.560	35	10
4	Red	0.665	30	10

5	Vegetation Red Edge	0.705	15	20
6	Vegetation Red Edge	0.740	15	20
7	Vegetation Red Edge	0.783	20	20
8a	NIR	0.842	115	10
8b	Vegetation Red Edge	0.865	20	20
9	Water vapour	0.945	20	60
10	SWIR- Cirrus	1.375	30	60
11	SWIR	1.610	90	20
12	SWIR	2.190	180	20

3.3 Methods

3.3.1 Sentinel 2 imagery acquisition and pre-processing

Sentinel 2 imagery was acquired from Sentinel Copernicus Scientific Data Hub (https://scihub.copernicus.eu/) at no cost. The imagery was acquired for the remote sensing date of 6th June 2016. The study area was covered by six separate tiles which were stitched using the Mosaic tool within the Raster toolbox in ArcGIS. Sentinel-2A imagery is distributed at Level 1 – C i.e Top of Atmosphere, imagery was corrected using Sen2Cor extension within the Sentinel Application Platform (SNAP) (van der Werff & van der Meer, 2015). The imagery bands were corrected at a spatial resolution of 10 m and stacked on ENVI 5.4.

3.3.2 Defining land use land cover classes and reference data sampling

ISO unsupervised classification tool in SNAP 6.0 was used to group pixels with common reflection characteristics into land cover classes. Initially, the number of classes was set at 6 which later extended to 12 for purposes of separating vegetation species. Polygons representing *Prosopis* invasion obtained from the Northern Cape Department of Environmental Affairs were used to guide the selection of reference data for the *P. juliflora* class. The polygons inaccurately represented the spatial distribution of *P. juliflora* because they also included other land cover types (Department of Environmental Affairs, 2015). Fine spatial resolution imagery from Google Earth Pro was used to select the rest of the classes. Purposive and stratified random sampling techniques were used to create reference data. Purposive sampling is particularly useful when there is limited number of elements. In this project, it was used to select reference points for the *P. juliflora* and *Fynbos open bush* classes. Stratified random sampling was used to select reference data for classes that have big populations. The reference data was digitised in ArGIS 10.3 then the regions of interest were created in ENVI 5.4 by overlaying the reference data over Sentinel-2A image to train and test SVM and RF classifiers (Table 2). Subsequently, the stratified sampling tool in EnMAP-BOX was used each class of the reference data was randomly divided into 70 % training and 30 % test sets.

Table 3: Training and test samples of *P. juliflora* and other land cover land use.

Class	**Code**	**Training Dataset**	**Test Dataset**	**Total**
Dark Brown Soil	DBS	280	120	400
Karoo Nama Karoo grassland	KBG	378	162	540
Karoo low shrub	KLS	280	120	400
Prosopis Juliflora	PSJ	237	101	338
Nama Karoo grassland	NKG	332	142	474
Fynbos open bush	FOB	231	99	330
Settlements	S	238	102	340
Nama Karoo open bush	NKOB	287	123	410
Succulent karoo open shrub	GL	280	120	400
Cultivated commercial annual crops non-pivot	CCACNP	280	120	400
Fynbos low shrub	FLS	145	61	206
Water bodies	W	154	66	220

3.3.3 Image Classification

In the last decade, remote sensing image classification has progressed beyond the traditional unsupervised and supervised classifiers. This shift is due to the unrealistic assumptions and limitations of both supervised and unsupervised classifiers. Several works recorded significant accuracy improvements where machine learning classifiers were used (Liu, Skidmore & Van Oosten 2002, Ghosh et al., 2014). There are many machine learning classifiers used in remote sensing however this study will only discuss random forest and Support Vector Machine (Fig 4).

3.3.4 Random Forest Classifier

Random forest (RF) is an ensemble classifier that is less sensitive to over fitting; it is combination of methods grouped to make a prediction regardless of sample size and the number of variables (Lin et al. 2011). RF uses the bagging approach and Classification and Regression Trees (CART) to make a prediction (Breiman, 2001). A portion of the total samples i.e in-bag samples are used to train the decision trees and the remainder of the samples i.e out-of-the-bag samples is used to validate the predictions (Genuer et al. 2010). The estimate of the internal performance is reported using the out of bag (OOB) error. The decision on how the samples are split for training and validation is user dependent and is often influenced by the size of the sample (Breiman, 2001). There are two parameters which must be set by the analyst i.e the number of decision trees and the number of important bands (Cho et al, 2015). The number of decision trees is informed by the stability of errors; the default number of trees is 500 (Ghosh et al. 2014). One of the impressive features of RF is the variable importance measure which employs three methods namely: the Gini importance, the number of times each variable is selected and the permutation accuracy importance measure (Strobl et al. 2007). For this study, a 10-fold grid-search approach based on the OOB error was employed to determine the best *mtry* and *ntree* parameter combination. The *mtry* value ranged from 1 to 10 and the *ntree* value ranged from 500 to 10 000. The image classification was carried out using the training dataset and RF tool in EnMAP-Box. The Gini importance measure was used to evaluate the effectiveness of bands in classifying the image and detecting each LULC class.

3.3.5 Support Vector Machines

Support Vector Machine (SVM) is a nonparametric machine learning technique with regression and classification capacity initially used in the medical sphere (Oommen et al. 2008). SVM is based on binary linear system that maximises the distance between training data points and the optimal hyperplane (Cortes & Vapnik 1995a).

The maximisation of distances between hyperplanes and data points reduces chances of misclassification during the training stage (Cortes & Vapnik 1995b). The support vectors are outlines situated of the edges of hyperplanes that predict or separate entities into classes specified in the training data (Mashao 2003). In application, these boundaries are often bridged when data points from different classes overlap thus, a non-linear polynomial is imposed to reduce misclassification induced by overlaps (Cho et al, 2015). SVM optimization is carried out using kernel methods such as linear, polynomial and sigmoid, however radial function is the often used in remote sensing (Huang et al. 2002; Oommen et al. 2008; Mureriwa et al, 2015). The kernel width i.e gamma (ɣ) and the cost sigma (C) i.e the value used to reduce misclassification in the training data (Mountrakis et al. 2011). In this study the radial function was used to optimize the kernel and the image classification was carried out using the training and SVM tool in EnMAP-Box.

3.3.4 Accuracy assessment

The performance of both SVM and RF classifiers was assessed using a traditional confusion matrix which employed an independent sample (Table 3). The traditional error matrix and its associated statistics such as overall accuracy (AO), user's accuracy assessment (UA) and producer's accuracy (PA) were computed. The accuracy assessment tool in EnMAP-Box used the test data to evaluate the performance of RF and SVM classifiers. AO represents the percentage of correctly classified elements from the test samples. UA is estimation that the classifier has accurately assigned an element within a sample to a correct class. PA represents the probability of accurately identifying the classes. Cohen's Kappa coefficient, a robust statistic used for testing interrater and intrarater reliability was used to calculate the agreement between the reference data and the classification. It ranges from -1 to 1, where 0 represents the amount of agreement that can be from random chance and 1 represents a perfect agreement. Cohen's Kappa is calculated according to the following formula:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{1}$$

Where Pr(a) represents the actual observed agreement and Pr(e) represents chance agreement.

Two raters use a number of observations to assess the performance of the agreement using the following formula:

$$\text{Expected (Chance) Agreement} = \frac{\left(\frac{cm^1 \times rm^1}{n}\right) + \left(\frac{cm^2 \times rm^2}{n}\right)}{n} \quad (2)$$

where: cm^1 represents column 1 marginal, cm^2 represents column 2 marginal, rm^1 represents row 1 marginal, rm^2 represents row 2 marginal, and n represents the number of observations (not the number of raters).

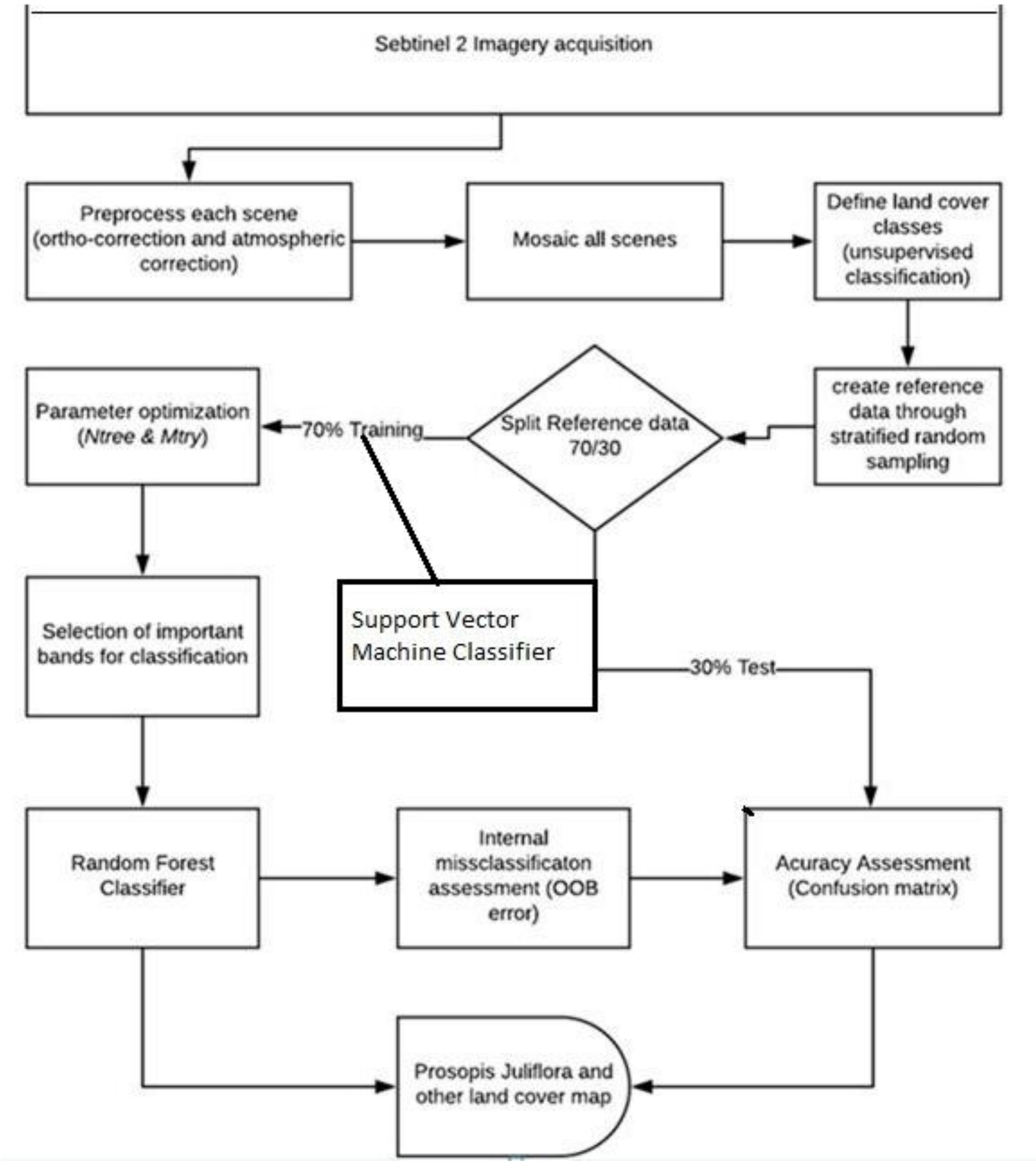

Figure 4: Process flow diagram showing the series of steps followed to classify the images using RF and SVM

3.4.1 *Prosopis Juliflora* spatial point pattern analysis

Classification pixels representing *P.juliflora* were converted to points using the Raster to Point tool in ArcGIS for Spatial Point Pattern Analysis (SPPA) in R. A spatial point pattern is defined as a set of locations can be irregularly distributed within a region of interest, which has been generated by random mechanisms

(Diggle, 1985). SPPA is a collection methods employed to evaluate how points are located with respect to each other. There are three main spatial point patterns: complete spatial randomness (CSR), regularity and clustering (Parmigiani, 2009). The initial step towards spatial point pattern analysis of *P. juliflora* within Hantam Local Municipality was to investigate the distribution and density of points; this was achieved using plots and basic summary statics. Various packages within R 3.3.1 were used to carry out the spatial point pattern analysis. Clark-Evans Chi Squared test was used to assess the data for complete spatial randomness. If the data shows complete spatial randomness, this means that there is little analysis to be carried out. Clark-Evans Test involves calculating the nearest-neighbour distances for each point, and analysing them with respect to their distribution under the hypothesis of Complete Spatial Randomness (CSR). If the nearest-neighbour distances are particularly small, this is an indication of clustering. Large nearest-neighbour distances are an indication of uniformity. The nearest-neighbour distances have a tendency to come in pairs, thus the nearest neighbour data is not independent (Akasapu et al. 2011)

3.4.2 Evaluating the density of *Prosopis Jiliflora* using quadrats and Kernel Density

Quadrat analysis examines the frequency of points occurring in various parts of an area. A set of stationery quadrats of cells (usually squares but not always) is superimposed on a study area and number of points in each cell is determined (Boots & Getis, 1998). The size of windows or quadrats affects the density of points. Larger quadrats yield smoother intensity maps; smaller quadrats yield “spiky” intensity maps with empty quadrats (Bailey & Gatrell, 1995). The primary objective of quadrat analysis is to determine whether a point pattern has been generated by a random (Poisson) process. The frequency of *P. juliflora* occurrence was examined using three window sizes i.e (10 x 10), (20 x 20) and (30 x 30). In Kernel Density, a moving window is employed to calculate the density of point features around each output raster cell (Grabarnik, 2002). The key objective is to find sigma which determines the bandwidth of the kernel. In K function (solid lines) deviates from the theoretical expected value assuming the points are completely random (dashed lines).

The spatial point pattern analysis revealed that *P. juliflora* was not randomly distributed thus its density was investigated with respect to geology, soil, and land cover land use (Equation 3).

$$Density = \frac{\sum Prosopis\ \ Patches}{Class\ Surface\ \ Area} \qquad (3)$$

3.3.3 Using SDM to access areas susceptible to *Prosopis Juliflora* invasion

Over the last decades, ecologists have been developing methods that represents biodiversity spatial dynamics (Cook et al., 2007). Statistical and technological developments have led to the rise of SDM for ecological modelling data (Brunel et al., 2010). Correlative species distribution modelling was used to predict areas which are susceptible to *P. juliflora* invasion. Multivariate binomial logistic regression used true-presence and true - absence data in addition to Digital Elevation Model (DEM) derivative such as slope, aspect, curvature and soil data, land cover–land use and hydrology.

3.3.4 Data pre-processing for the predictive model

The datasets were sourced at various departments and data portals (Table 4). In preparation for the predictive modelling all datasets was geo-referenced to one coordinate system and all vector data converted to raster. Rasterise-zing and assignment of a uniform coordinate system reduced errors associated with distortion. The last data pre-processing step was resampling all grid cells to 10 m^2 with the objective of standardizing spatial resolution differences and making layers conformable.

Table 4: Datasets to be included in the susceptibility model

#	Dataset	Source	Data Type	Scale/ Resolution
1	Field Survey	Department of Environmental Affairs	Vector	
2	Digital Elevation Model	Alaska Satellite Facility	Raster	12 m
3	Slope	DEM derivative	Raster	10 m

4	Aspect	DEM derivative	Raster	10 m
6	Land Cover	Department of Environmental Affairs	Raster	30 m
8	Soils	Council for Geosciences	Vector	1: 250 000
9	Geology	Council for Geosciences	Vector	1: 250 000
10	Climate	South African Weather Service	Raster	
12	Sentinel 2 *P.Juliflora* classification	Objective 1 outcome	Raster	10 m

Broad soil types was one of the determinant used to evaluate current and predict future treads of spatial distribution *P. juliflora*. The South African binomial soil classification system was adopted to differentiate and denote various soil types (Table 5). The higher level of the system categorized soils into 41 general types (MacVicar et al, 1977)

Table 5: Higher level Land Type Survey: Broad Soil Patterns

A: Red and/or yellow, freely-drained soils (Ia, Kp, Ma, Hu, Gf, Cv) dominant (>40%)	
Aa	**Humic** topsoils (**Ia, Kp, Ma** >40%), red and/or yellow
Ab	**Red** (yellow soils <10%); **dystrophic/mesotrophic** > eutrophic
Ac	**Yellow/red** (yellow & red soils each >10%); **dystrophic/mesotrophic** > eutrophic
Ad	**Yellow** (red soils <10%); **dystrophic/mesotrophic** > eutrophic
Ae	**Red** (yellow soils <10%); **eutrophic** > dystrophic/mesotrophic
Af	As for **Ae**, but with regular **dunes**. Mostly Northern Cape
Ag	**Red** (yellow soils < 10%), **<300 mm** soil depth; **eutrophic** > dystrophic/mesotrophic

Ah	**Yellow/red** (yellow & red soils each >10%), **sandy** (<15% clay); **eutrophic** > dys/meso
Ai	**Yellow** (red soils <10%), **sandy** (<15% clay); **eutrophic** > dystrophic/mesotrophic
B: Plinthic catena (Bv, Av, Gc, Wa, We, Ms11) >10%; upland duplex and margalitic soils (Ar, Bo, Tk, My, Mw, Es, Ss, Sw, Va, Kd) <10%	
Ba	**Red** (Hu, Bv >33%); **dystrophic/mesotrophic** > eutrophic
Bb	**Non-red** (Hu, Bv <33%); **dystrophic/mesotrophic** > eutrophic
Bc	**Red** (Hu, Bv >33%); **eutrophic** > dystrophic/mesotrophic
Bd	**Non-red** (Hu, Bv <33%); **eutrophic** > dystrophic/mesotrophic
C: Plinthic catena (Bv, Av, Gc, Wa, We, Ms11) >10%; upland duplex and margalitic soils (Ar, Bo, Tk, My, Mw, Es, Ss, Sw, Va, Kd) >10%	
Ca	As for **Ba-Bd**, but with >10% **clay soils** (***not*** in valley bottoms)
D: Duplex soils (Es, Ss, Sw, Va, Kd) >50%	
Da	**Red** subsoils **>50%** of duplex component
Db	**Non-red** subsoils **>50%** of duplex component
Dc	As for **Da/Db**, but also with **>10% Ea** soils
E: One or more of: vertic (Ar, Rg), melanic (Mw, My, Bo, Ik, Wo) and/or red structured (Sd) soils >50%	
Ea	Dark, blocky **clay** topsoils (often swelling clays) and/or red, structured clays
F: Mainly Glenrosa and/or Mispah forms (other soils may occur as long as land type does not qualify elsewhere)	
Fa	**Shallow**, and/or rocky, often steep, highly leached (very **little lime**)

Fb	**Shallow**, and/or rocky, often steep, moderately leached (**some lime**, mainly in valleys)
Fc	**Shallow**, and/or rocky, often steep, slightly leached (**lime is common** throughout)
G: Podzol (Lt, Hh) soils >10%	
Ga	Moderately deep to **deep** (Lt form), bleached sands with **podzol** horizon.
Gb	Usually **shallow** (Hh form), bleached sands with **podzol** horizon, over rock.
H: Grey regic sands (Fw, Ct, Sp, Vf)	
Ha	**Dominantly** (>80%) deep, **grey sands** (usually near coast)
Hb	**Some** (20-80%) deep, **grey sands** (usually near coast). Other soils may occur
I: Miscellaneous land classes	
Ia	Deep **alluvial** deposits (>60%), usually on river floodplains (**Du, Oa** forms)
Ib	Much **rock** (60-80%), usually with shallow and/or rocky soils on steep slopes
Ic	Mainly **rock** (>80%), with little soil (usually steep to very steep slopes)
Other units	
WA	**Water** bodies (dams and/or lakes)

3.3.5 Exploratory data analysis and variable selection

Regression was initiated with exploratory data analysis (EDA) and exploratory spatial data analysis (ESDA) to check for normality and dependent variable significance. Graphic and statistical packages in R global version 3.3.1 were used to carry out EDA and ESDA. EDA is categorised into graphical and non-graphical methods. These methods are further sub divided into two techniques i.e univariate and multivariate. Graphical methods use diagrams to summarize data whilst non-graphical methods use calculations and tables. ESDA pictorially summarizes data through maps and show the locations of outliers and spatial autocorrelation (McHugh, 2013).

A Logistic regression model (LRM) can be used to estimate the probability of the presence of an event, given information about predictors that can potentially influence the outcome (Richardson & Thuiller., 2007). As a class of generalized linear models, LRMs are distinguished from ordinary linear regression (OLS) models by the range of their predicted values, the assumption of the variance of the predicted response and the distribution of the prediction errors (Hosmer & Lemeshow 1989). Logistic regression is a variation of the regression model. It is used when the dependent response variable is binary in nature (Buckley et al., 2010), (in this case 1 representing “susceptible” and 0 representing “insusceptible”). Logistic regression predicts the probability of the dependent response, rather than the value of the response (as in simple linear regression).

Malahlela et al, (2016) introduced variables individually in a step-wise binomial logistic regression to produce a model that estimated areas susceptible to *Chromolaena odorata* occurrence. This project employed a predictive methodology based on logistic regression (LR) analysis to estimate the likelihood of areas susceptible to *P. juliflora* invasion. The subset feature tool in ArcGIS was used to randomly select 25 % of both the presence and absence sample to access the performance of the binomial logistic regression.

Chapter Four

4. Results

4.1 Tuning of Support Vector Machine parameters

SVM parameters were optimized to find the best input parameters to train the SVM classifier to separate the Sentinel-2A image pixels into 12 land cover-land use classes. The 10-fold cross validation method revealed that the best performance was produced at a parameter combination of gamma = 0.1 and cost 100 (Fig 5).

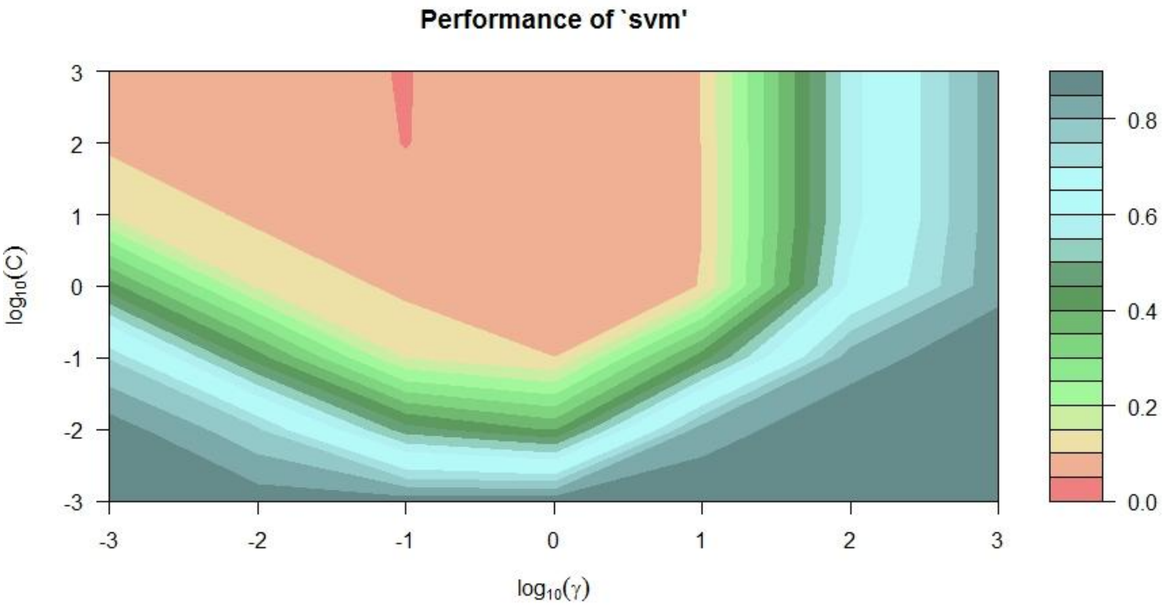

Figure 5: Support Vector Machines optimization of parameters (C and γ) using the 10 fold grid search method

4.2 Tuning of Random Forest parameters

RF parameters were optimized to find the best input parameters to train the RF classifier to separate the Sentinel-2A image pixels into 12 land cover-land use classes. The 10-fold cross validation method revealed that the best performance (78 %) was produced at a parameter combination of *mtry* = 4 and *ntree* 7500. The internal assessment of the RF optimization produced the lowest OOB error of 6.0 % at a combination of *mtry* = 3 and *ntree* = 1500 and the highest OOB error of 7.0 % was obtained at a combination of *mtry* = 10 and *ntree* = 500 (Fig 6).

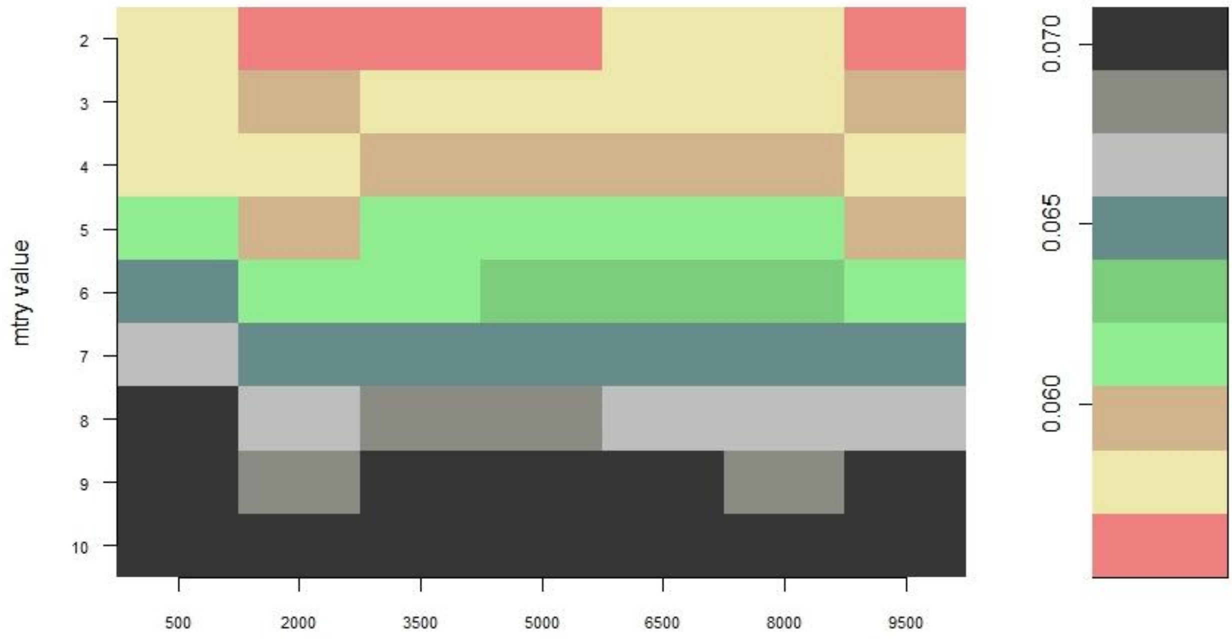

Figure 6: Random Forest optimization of parameters (*mtry and ntree*) used the 10 fold grid search method where the Out-of-Bag (OOB) sample was employed to evaluate the error rate for all combinations.

4.3 Performance of RF and SVM classifiers in mapping *Prosopis Juliflora* and other land cover-land use classes

The RF and SVM classifiers managed to separate image pixels into 12 land cover-land use classes thus depicting the spatial distribution of *P. juliflora* across Hantam Local Municipality. The classifications were dominated by the Nama Karoo grassland class with Succulent Karoo low shrub and *P. juliflora* being the most common vegetation classes. The grasslands were dominant in highlands while *P. juliflora* frequently occurred along water courses. The water courses were often composed of dark brown or Nama Karoo open bush (Fig 7 & 8).

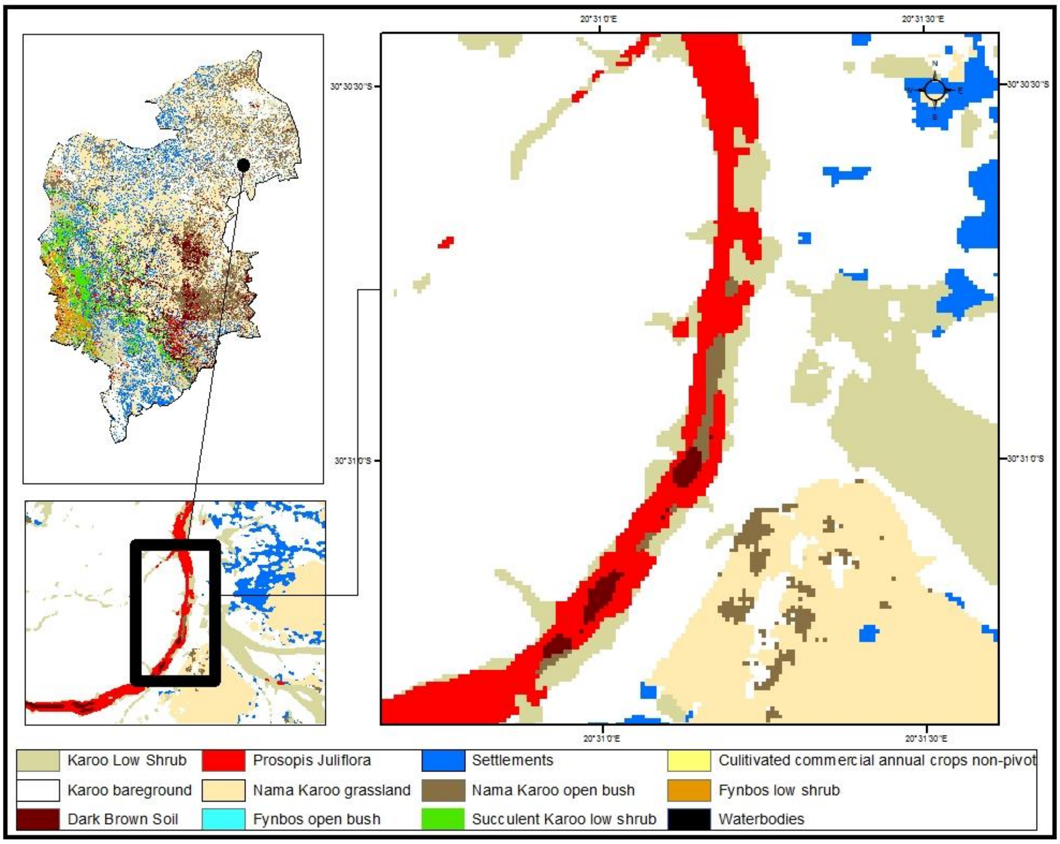

Figure 7: Land Use and Land Cover classification produced by Support Vector Machines classifier.

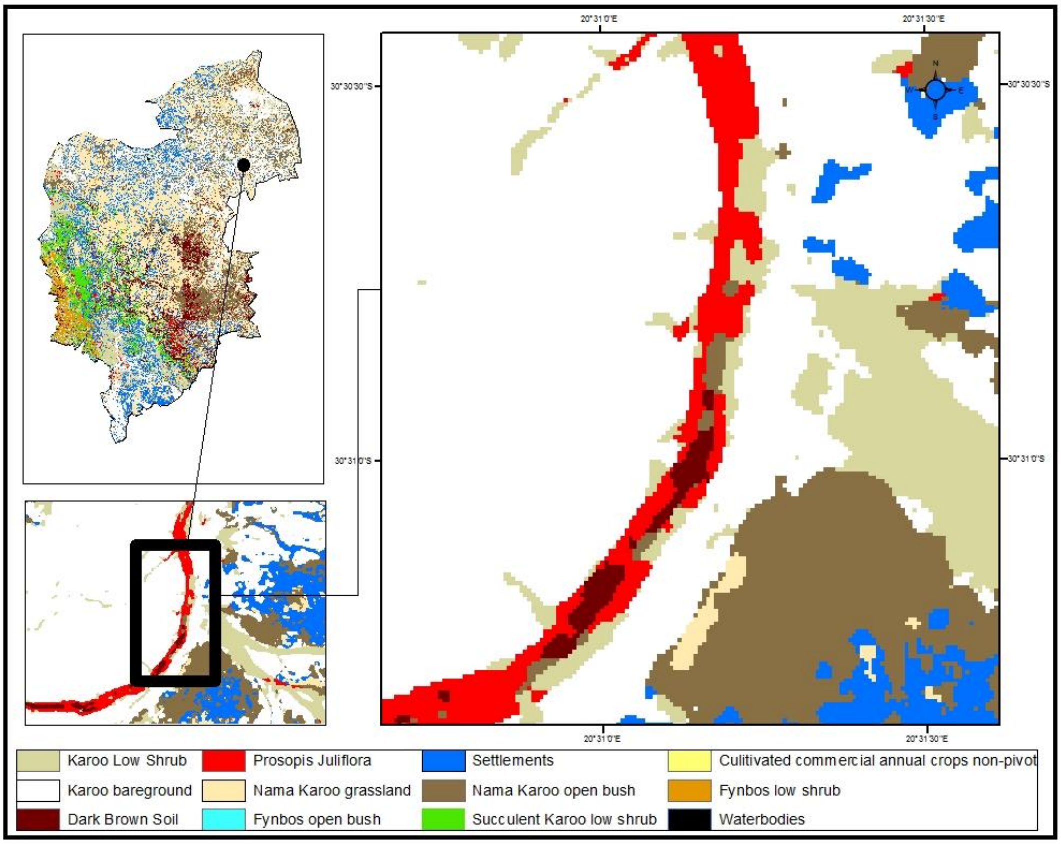

Figure 8: Land Use and Land Cover classification produced by Support Vector Machines classifier.

The RF variable importance measurement feature investigated and reported the effectiveness of bands in *P. juliflora* and other LULC classification. The most effective or important bands were assigned high mean decrease accuracy and mean decrease gini, for this classification the visible bands i.e red (B4), green (B3) and blue (B2) reported high mean accuracy decrease (Fig 8). Subsequently, an investigation into the effectiveness of each band in classifying different land cover types was carried out, red, green and blue were the most important bands for classifying vegetation (Fynbos open bush, *P. juliflora*, Cultivated commercial annual crops non-pivot and grassland) while the green blue and narrow infra-red bands were the most effective in classifying soils (dark brown soil, Nama Karoo open bush, Karoo Nama Karoo grassland and Karoo Low Shrub). The narrow near infra-red, near infra-red and blue bands were the most important for classifying Nama Karoo grassland, settlements and Fynbos low shrub while the narrow near infra-red and near infra-red bands were effective in classifying water bodies (Fig 9).

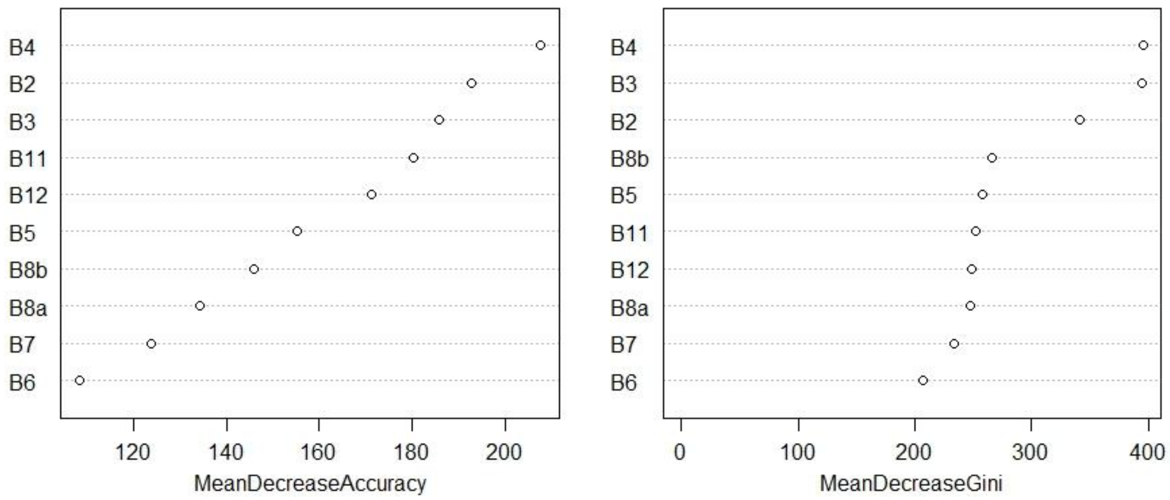

Figure 9: Band importance ranking of Sentinel-2A bands using Random Forest. The most important band has the highest mean decrease in accuracy and mean decrease Gini.

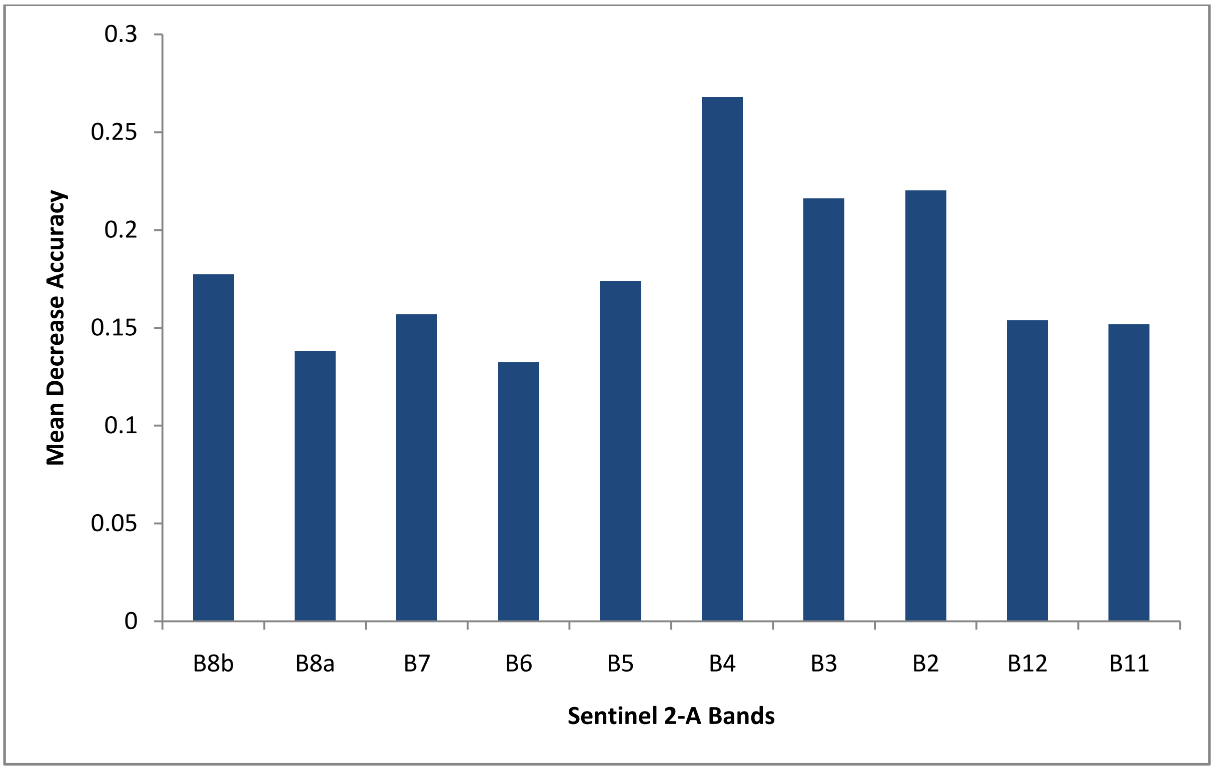

Figure 10: Sentinel -2A bands variable importance of the classification for *P. juliflora* and other LULC classes.

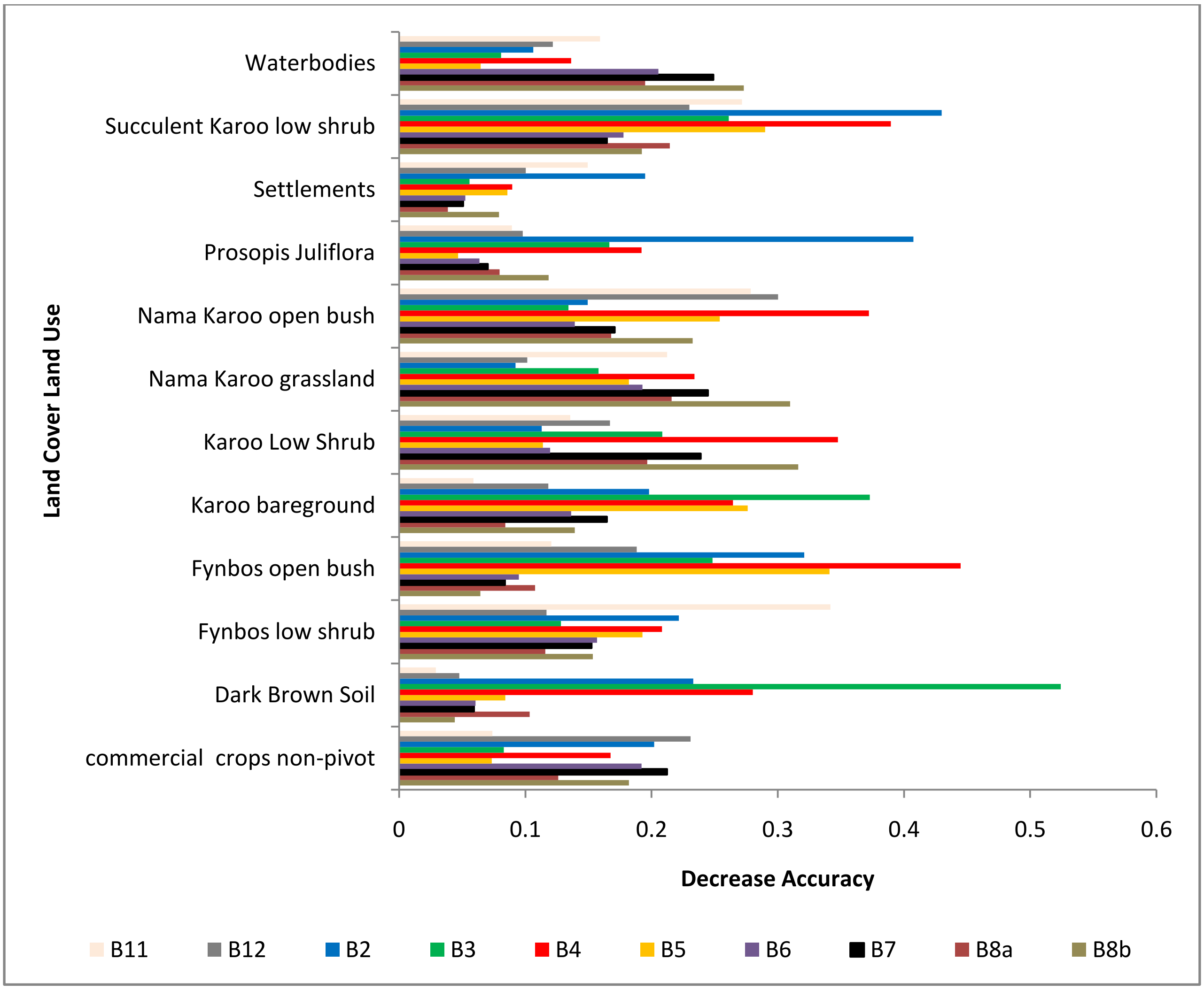

Figure 11: Sentinel – 2A band importance for each LULC class, the highest decrease in accuracy shows the most important band.

4.4 Accuracy assessment

The performance of the RF and SVM classifiers was assessed using an independent test dataset (Table 2). SVM classifier produced an overall accuracy of 91.42 % with a kappa coefficient of 0.90. Eight out of twelve land cover land use classes produced a user's accuracy above 90% namely: dark brown soil (90.48 %), Karoo Nama Karoo Succulent Karoo low shrub (96.86 %), Nama Karoo open bush (95.31 %), *Fynbos open bush* (90.63 %), Succulent Karoo low shrub, Cultivated commercial annual crops non-pivot (97.50 %), Fynbos low shrub (95.00 %) and water-bodies (100 %) while settlements, Karoo Low Shrub and Nama Karoo grassland recorded a user's accuracy of 85.53 %, 85.12 % and 85.12 % respectively. The SVM classifier confusion matrix reported spectral confusion between *Fynbos open bush* and *P. juliflora,* as a result *P. juliflora* produced the lowest user's accuracy and producer's accuracy, 80.42 % and 77.23 % respectively (Table 6). RF classifier produced an overall accuracy of 88.86 % i.e 2.56 % lower than SVM and a kappa coefficient of 0.87. At least six of the twelve classes produced a user's accuracy above 90 % namely: dark brown soil (91.67 %), Karoo Nama Karoo Succulent Karoo low shrub (95.00 %), Nama Karoo open bush (92.31 %), Succulent Karoo low shrub (92.17 %), Cultivated commercial annual crops non-pivot (94.35 %) and water (100 %) while Nama Karoo grassland, Fynbos low shrub, Karoo Low Shrub, *Fynbos open bush*, settlements produced user's accuracy ranging between 80 % - 88 % . Similarly, the RF classifier confusion matrix revealed spectral confusion between *P. juliflora* and *Fynbos open bush* and once again *P. juliflora* recorded the lowest user's accuracy and producer's accuracy, 76.42 % and 80.2 % respectively (Table 6). The SVM classifier class separation measurement showed moderate overlaps among land cover types however. *P. juliflora*, cultivated commercial annual crops non-pivot, dark brown soil and Succulent Karoo low shrub are clearly separable (Fig 11). Contrary, RF classifier class separation method showed great overlap between land cover types however *P. juliflora* and Succulent Karoo low shrub are slightly separable (Fig 12).

Table 6:SVM Classifier confusion matrix for *P. juliflora* and other LULC classes. The overall accuracy; user's accuracy; and producer's accuracy were developed on the test dataset using the EnMAP-Box *ImageSVM* Accuracy Assessment tool.

Using Support Vector Machine															
Map class	DBS	KBG	KLS	PSJ	NKG	FOB	S	NKOB	GL	CCACNP	FLS	W	**Sum**	UA %	PA %
DBS	114	4	0	4	0	0	4	0	0	0	0	0	126	90.48	95
KBG	1	154	0	0	0	0	4	0	0	0	0	0	159	96.86	95.1
KLS	0	0	103	14	0	3	0	0	0	0	1	0	121	85.12	85.1
PSJ	2	0	7	78	0	8	2	0	0	0	0	0	97	80.41	77.2
NKG	0	0	2	0	141	0	21	0	0	0	0	0	164	85.98	98.6
FOB	0	0	4	4	0	87	0	0	0	1	0	0	96	90.63	88.8
S	0	4	1	0	1	0	65	1	0	1	3	0	76	85.53	64.4
NKOB	3	0	1	1	1	0	0	122	0	0	0	0	128	95.31	99.2
GL	0	0	3	0	0	0	2	0	116	0	0	0	121	95.87	97.5
CCACNP	0	0	0	0	0	0	0	0	3	117	0	0	120	97.5	98.3
FLS	0	0	0	0	0	0	3	0	0	0	57	0	60	95	93.4
Sum	120	162	121	101	143	98	101	123	119	119	61	61	1329		
Overall Accuracy (OA) = 91.42 %												0.90Kappa Coefficient = 0.90			

Table 7:RF Classifier confusion matrix for *P. juliflora* and other LULC classes. The overall accuracy; user's accuracy; and producer's accuracy were developed on the test dataset using the EnMAP-Box *ImageRF* Accuracy Assessment tool.

Using Random Forest															
Map class	DBS	KBG	KLS	PSJ	NKG	FOB	S	NKOB	GL	CCACNP	FLS	W	**Sum**	UA %	PA %
DBS	110	4	0	4	0	0	2	0	0	0	0	0	120	91.7	85.4
KBG	3	152	0	0	0	0	5	0	0	0	0	0	160	95	89.5
KLS	0	1	100	9	0	2	0	0	1	1	0	0	114	87.7	82.6
PSJ	2	0	11	81	0	8	1	0	1	1	1	0	106	76.4	62.9
NKG	0	0	2	1	133	0	20	3	0	0	7	0	166	80.1	89
FOB	0	0	7	5	0	83	0	0	2	0	0	0	97	85.6	6.23
S	1	4	0	0	4	0	68	0	1	0	1	0	79	86.1	61.2
NKOB	4	1	0	1	2	0	2	120	0	0	0	0	130	92.3	95.1
GL	0	0	1	0	1	5	0	0	106	0	2	0	115	92.2	81.7
CCACNP	0	0	0	0	0	0	0	0	7	117	0	0	124	94.4	84.7
FLS	0	0	0	0	3	0	3	0	1	0	50	0	57	87.7	70.4
W	0	0	0	0	0	0	0	0	0	0	0	61	61	100	100
Sum	120	162	121	101	143	98	101	123	109	119	61	61	1329		
Overall Accuracy (OA) = 88.86 %											Kappa Coefficient = 0.87				

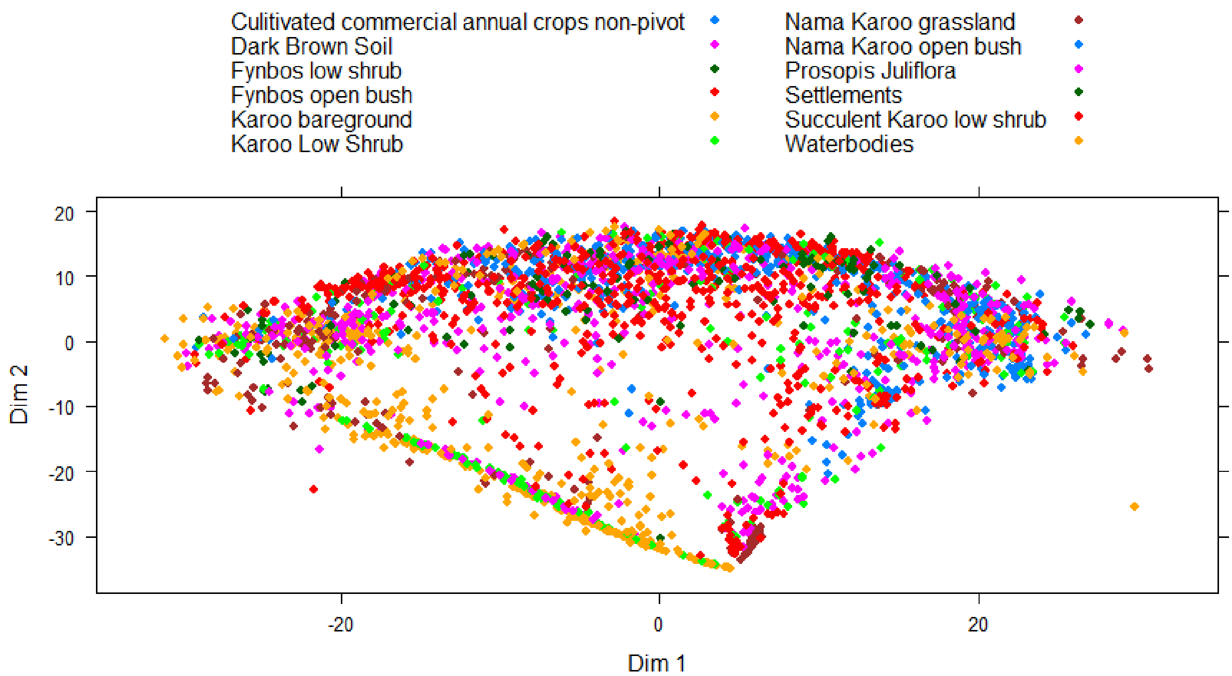

Figure 12 : Separation of classes using Support classification for *P. juliflora* and other LULC classes

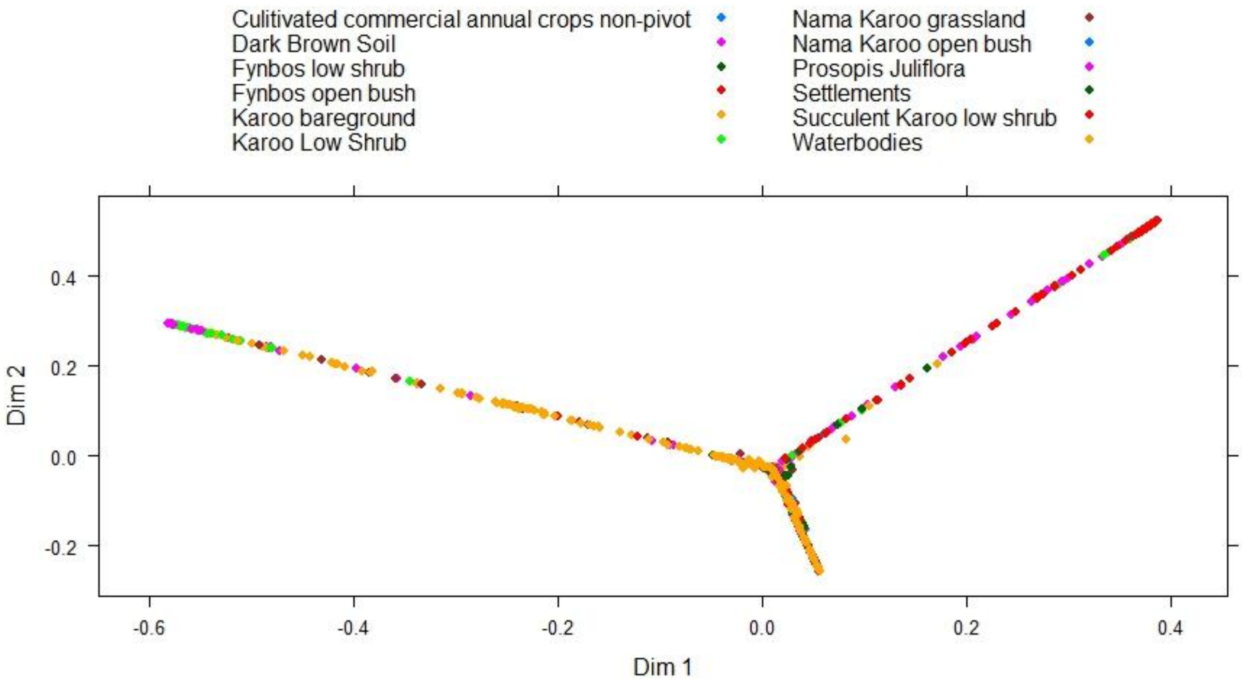

Figure 13 : Separation of classes using Random Forest classification for *P. juliflora* and other LULC classes

4.5 *Prosopis Juliflora* spatial point pattern analysis

Graphic and nongraphic descriptive techniques were used to determine the overall distribution and average *P. juliflora* intesity. The summary statistics revealed an average intensisity of 1.4 e^{-5} trees per square unit. A graphic summary technique displayed the general distribution of trees across Hantam Local Municipality, patches of high concentration are situated on the southern and western parts of the region.

The Clark Evans Test was used to calculate the nearest-neighbour distance of each point and analysed their distribution for the entire study area. The following hypothesis was created to test Complete Spatial Randomness (CSR):

H_0: There is complete spatial randomness (*P. juliflora* is not clustered)

H_A: There is no complete spatial randomness (*P. juliflora* is clustered)

In this case the p-value < 2.2 e $^{(-16)}$ which is less than 0.05, thus we reject the null hypothesis; there is sufficient evidence to conclude in favour of the alternative hypothesis i.e *P. juliflora* is clustered.

The Clark Evans test revealed absence of complete spatial randomness to further understand the distribution and intensity of *P. julifora* quadrant analysis and Kernel Density was employed. The quadrant analysis used various window sizes ranging from (10X10) to (30X30) and the Chi squared test to determine randomness per window (Table 7). The following hypothesis was developed to evaluate the randomness of the distribution all window sizes:

H_0: *P. juliflora* pattern is random (there is no clustering).

H_A: *P. juliflora* pattern is not random (there is clustering).

Table 7: Chi-square test p-values for quadrat analysis

Window size	**DF**	**P-value**
10X10	21	$< 2.2\ e^{-16}$
20X20	399	$< 2.2\ e^{-16}$
30X30	899	$< 2.2\ e^{-16}$

The Chi squared test reported a p-value is $< 2.2\ e^{(-16)}$ for all window sizes which is less than 0.05, thus we reject the null hypothesis; there is sufficient evidence to conclude in favour of the alternative hypothesis i.e *P. juliflora* is clustered. Kernel density used a moving window to reveal high intensity and density in the south-eastern and western parts of the study area (Fig 13). This confirms the results obtained from the summary graphic.

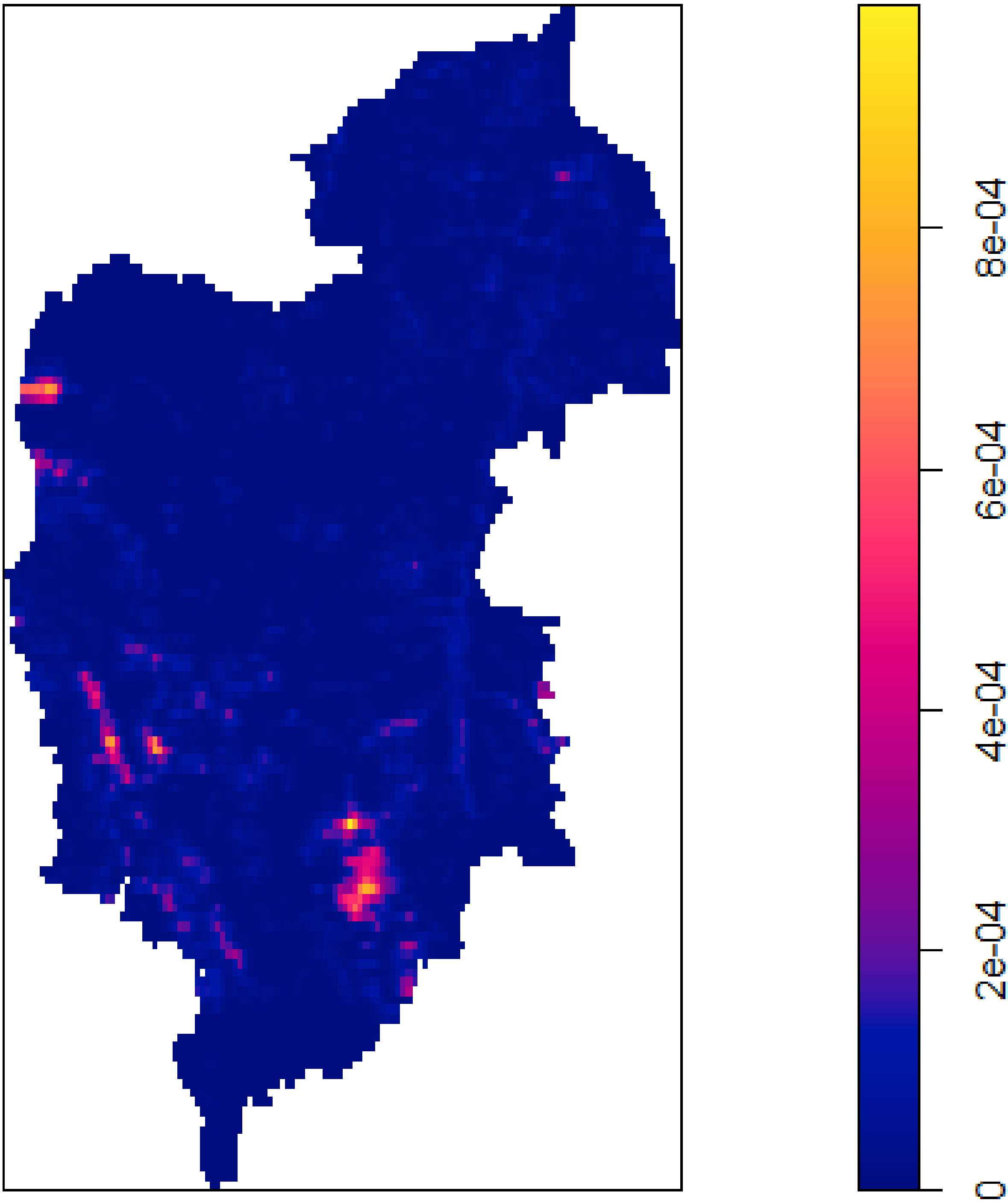

Figure 14: The density of *P. juliflora*

4.7 Evaluating the *Prosopis Juliflora* spatial distribution per geological lithology type

The results of quadrat analysis and kernel density paved a way for further evaluation of *P. juliflora* with respect to geology. *P. juliflora* density per geological lithology type was calculated; Waterford lithology stratigraphy reported the highest *P.* density. The

Kamieskroon-Gneiss Cares and Bidouw lithology types also reported high *P. juliflora* density. Dywaka, Kalahari, Stalhoek and Nardouw reported moderate *Prosopis* density and other lithology stratigraphy types reported low *P. juliflora* density (Fig 14). Karoo Dolorite recorded the highest *P. julifora* infestation with approximately 19 %, Volksrust and Waterford came in second and third with an infestation of approximately 18% and 16 % respectively. Dywaka, Prince Albert recorded approximately 11 % and other lithology stratigraphy types reported infestation less than 10% (Table 8).

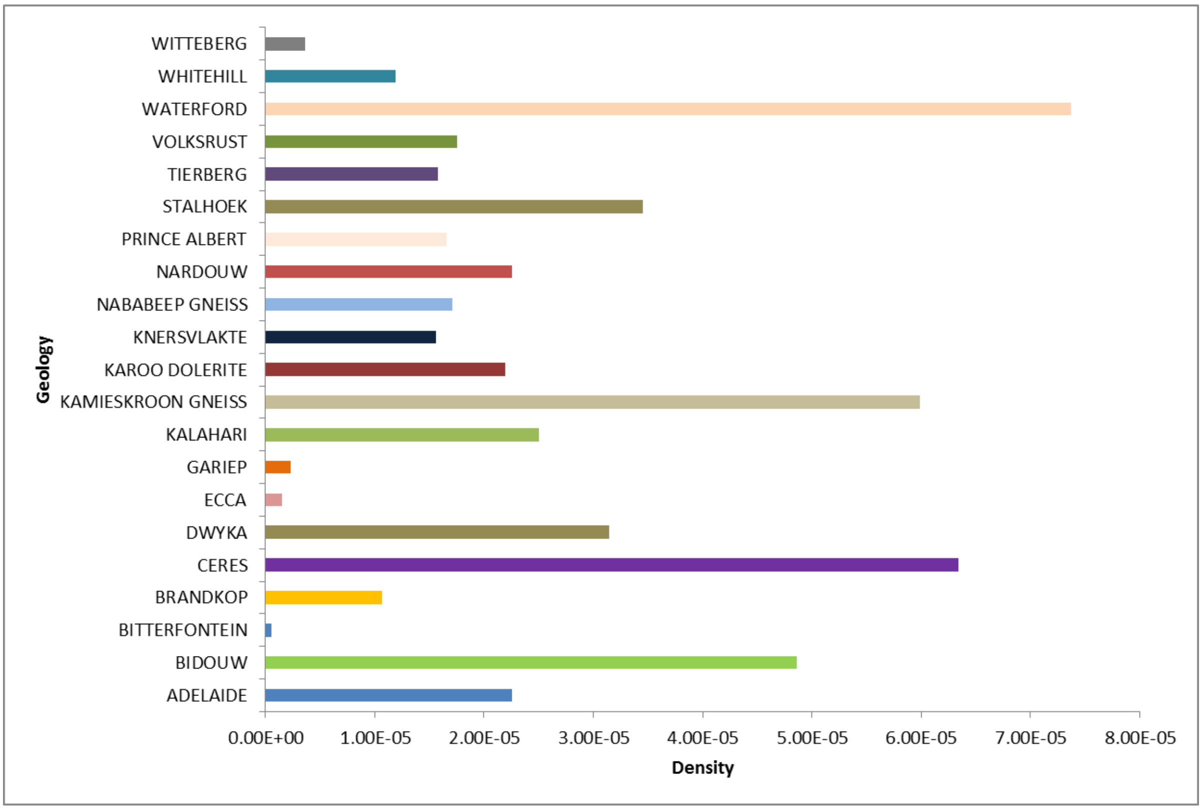

Figure 15 : Distribution of *P.juliflora* per rock formation

Table 8 : Distribution and percentage cover of *P.juliflora* with respect to rock formation

Lithology Stratigraphy	**Density**	***Prosopis* Patches**	**% *Prosopis* Patches**
ADELAIDE	2.26E-05	17185	2.0270
BIDOUW	4.86E-05	12465	1.4703
BITTERFONTEIN	5.11E-07	2	0.0002

BRANDKOP	1.07E-05	2727	0.3217
CERES	6.34E-05	67722	7.9881
DWYKA	3.15E-05	100078	11.8046
ECCA	1.51E-06	407	0.0480
GARIEP	2.29E-06	22	0.0026
KALAHARI	2.50E-05	5430	0.6405
KAMIESKROON GNEISS	5.99E-05	13558	1.5992
KAROO DOLERITE	2.20E-05	161742	19.0781
KNERSVLAKTE	1.56E-05	8444	0.9960
NABABEEP GNEISS	1.71E-05	394	0.0465
NARDOUW	2.25E-05	19859	2.3424
PRINCE ALBERT	1.65E-05	101337	11.9531
STALHOEK	3.45E-05	7358	0.8679
TIERBERG	1.58E-05	14195	1.6744
VOLKSRUST	1.75E-05	158130	18.6520
WATERFORD	7.37E-05	138431	16.3285
WHITEHILL	1.19E-05	17186	2.0272
WITTEBERG	3.64E-06	1117	0.1318

4.8 Evaluating the *Prosopis Juliflora* spatial distribution per soil type and soil depth class

P. juliflora density per soil type was calculated; Fc754 soils reported the highest *P. juliflora* density and other soil types recorded low *P. juliflora* density (Fig 15). It is worth noting that figure 15 only represents the top 14 densities which were corrected to 3 significant figures other densities are recorded on (Table 9). The percentage computation revealed that Fa650 accounted for the highest number of *P. juliflora* patches with approximately 6 % followed Ag91 and Ea365 accounting for approximately 4 % each. *P. juliflora* density per soil depth class was calculated (Fig 16); D4 (903.2 m - 1097.2 m) reported the highest density followed by D3 (605.1 m - 879.3 m) and D1 (118.8 m - 299.9 m). The percentage computation revealed that D2 (300.1 m - 584.7 m) accounted for the highest number of *P. juliflora* patches with approximately 45 % followed D1 (118.1 m - 299.9 m) with D4 (903.2 m - 1097.2 m) and D3 (605.1 m - 879.3 m) accounting for approximately 7.5 % and 6.0 % respectively.

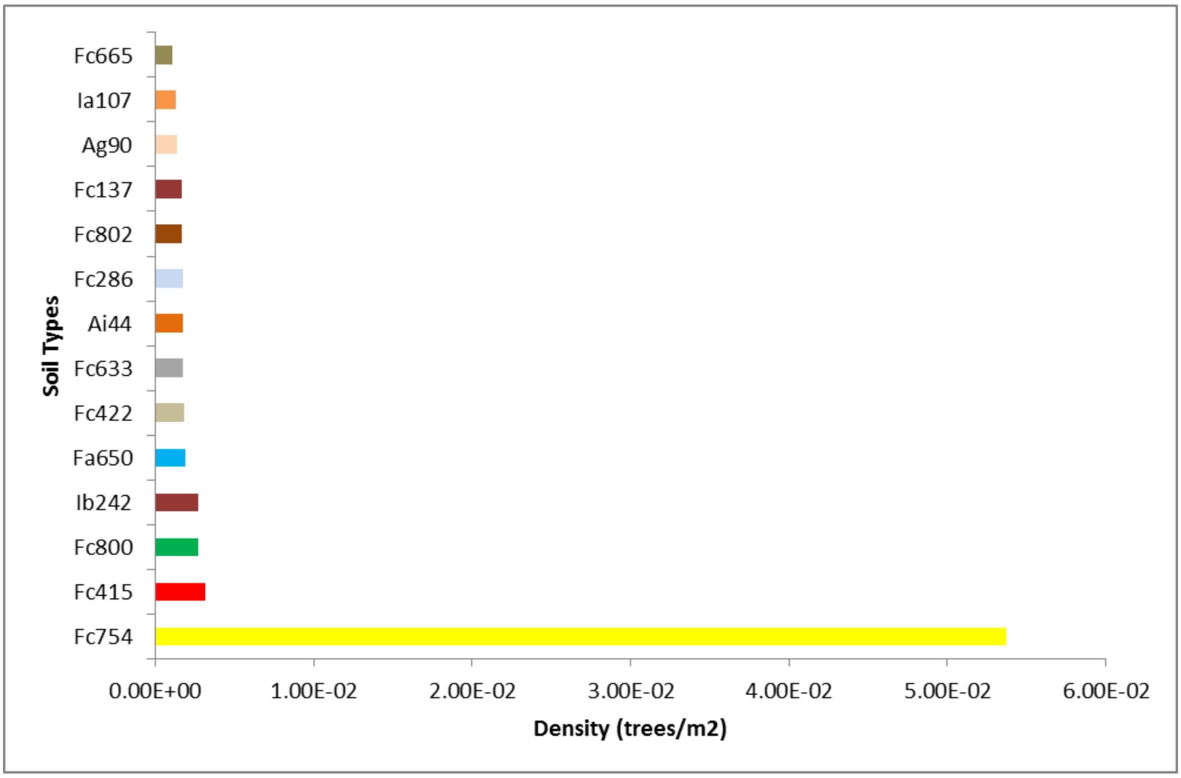

Figure 16: Distribution of *P. juliflora* per soil type

Table 9: Distribution and percentage cover of *P. juliflora* with respect to soil type

LANDTYPE	**Density**	***Prosopis* Patches**	**% *Prosopis* Patches**
Fa650	1.94E-03	62263	6.86
Ag91	3.03E-04	43187	4.76
Ea365	1.75E-04	41053	4.52
Ia103	5.49E-04	40405	4.45
Fc284	2.26E-05	37066	4.08

Ag199	1.30E-04	32361	3.56
Fc137	1.71E-03	32056	3.53
Ia16	1.61E-04	25016	2.76
Fc764	1.44E-04	23688	2.61
Fc10	1.58E-04	21867	2.41
Fb564	6.83E-04	19380	2.13
Da69	3.47E-04	18645	2.05
Ag189	3.66E-04	16704	1.84
Fc633	1.76E-03	16676	1.84
Da173	4.03E-04	14751	1.63
Dc66	2.38E-04	14668	1.62
Ia161	3.09E-04	14597	1.61
Ag90	1.37E-03	13940	1.54
Fc421	4.42E-05	13619	1.50
Ia108	5.26E-04	13503	1.49
Ag200	9.38E-05	11776	1.30
Fc760	1.69E-04	11461	1.26
Ia102	9.21E-05	10352	1.14
Ai44	1.76E-03	10296	1.13
Ia55	5.49E-05	9751	1.07
Da70	7.40E-04	9671	1.07
Fc762	2.21E-04	9348	1.03
Fc805	1.37E-04	9012	0.99
Fc754	5.37E-02	8945	0.99
Ag196	2.60E-05	7561	0.83

Fc672	3.16E-05	7224	0.80
Fc422	1.84E-03	6867	0.76
Ae314	4.13E-05	6635	0.73
Ae378	3.84E-05	6446	0.71
Fc426	6.97E-06	6367	0.70
Fb558	2.28E-04	6251	0.69
Ib275	3.32E-04	6191	0.68
Fc434	1.85E-05	6035	0.66
Ib475	5.42E-05	5968	0.66

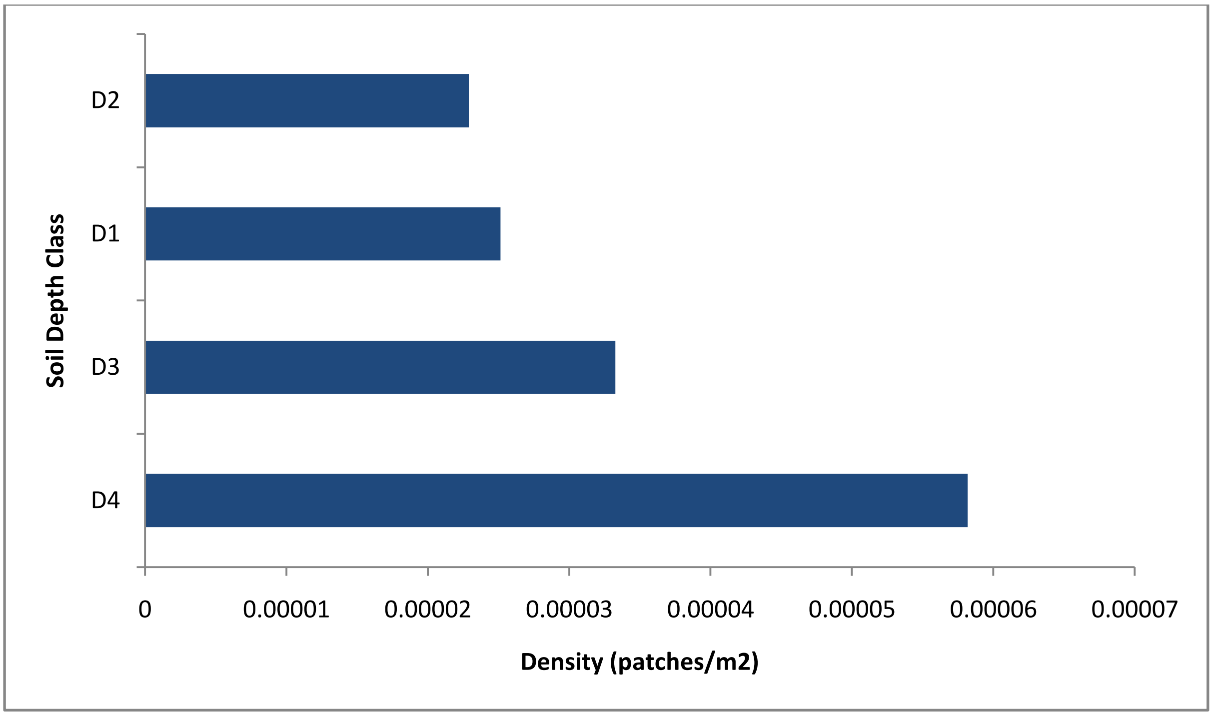

Figure 17: Distribution of *P. juliflora* per soil depth class

Table 10: Distribution and percentage cover of *P. juliflora* with respect to soil depth class

Soil Depth Class (m)	**Density (patches/m2)**	**Percentage (%)**
D4 (903.2 - 1097.2)	5.81807E-05	6.04
D3 (605.1 - 879.3)	3.32661E-05	7.54
D1 (118.8 - 299.9)	2.51471E-05	41.30
D2 (300.1 - 584.7)	2.29114E-05	45.11

4.9 Evaluating the *Prosopis Juliflora* spatial distribution per land use land cover class

P. juliflora density per class was calculated using the 2013 South African land use land cover map; Succulent Karoo: grassland reported the highest *P. juliflora* density followed by Succulent Karoo: thicket, Nama Karoo: thicket and Succulent Karoo: open bush (Fig 17). Nama Karoo: open bush, *Fynbos*: low shrub and *Fynbos*: open bush reported moderate density other classes reported low density. The percentage computation revealed that Succulent Karoo: low shrub accounted for the highest number of *P. juliflora* patches with approximately 28 % followed by *Fynbos*: low shrub and cultivated commercial annual crops non-pivot accounting for 17.64 % and 10.93 % respectively (Table 11).

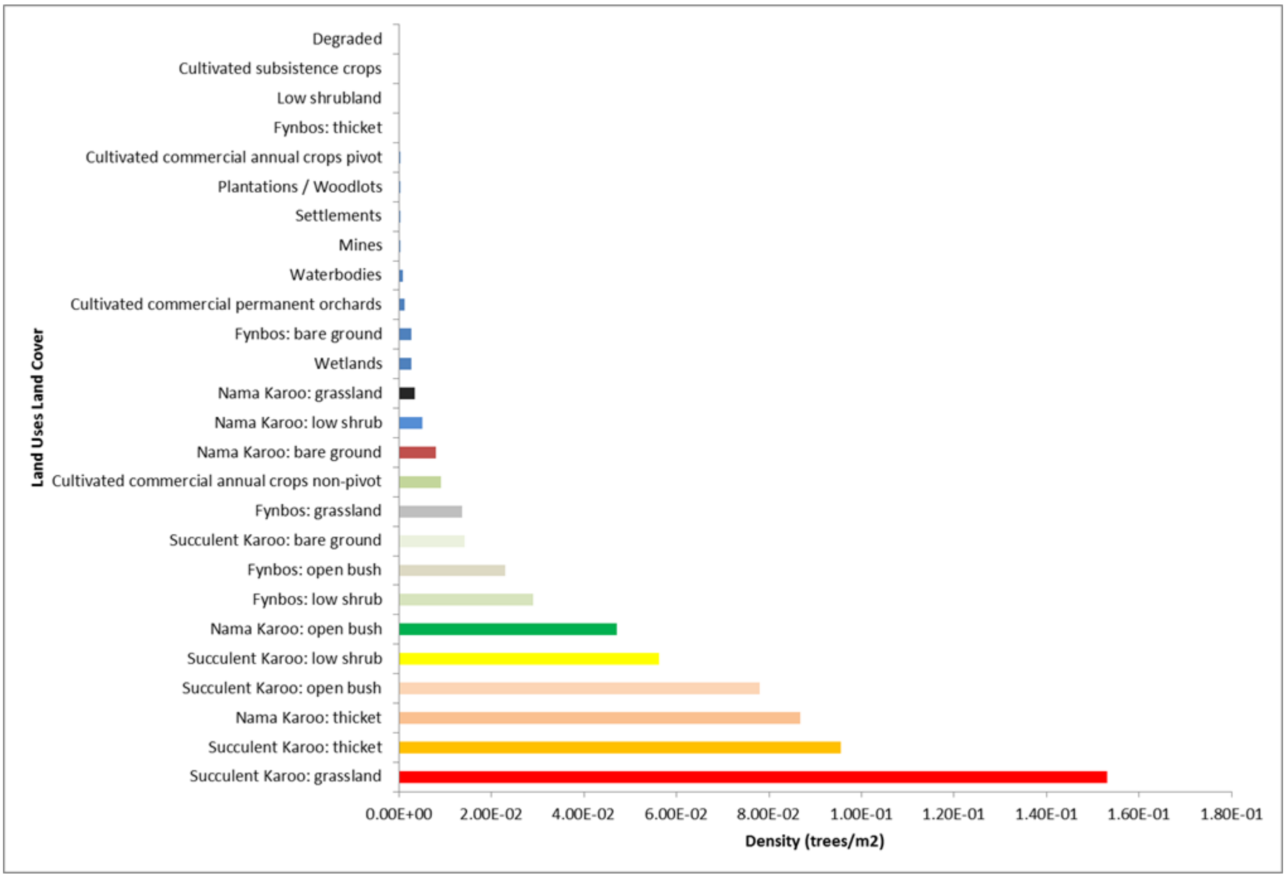

Figure 18: Distribution of *P. juliflora* per land cover land use

Table 11 : Distribution and percentage cover of *P. juliflora* with respect to land cover land use

LULC_13	***Prosopis* Patches**	**% *Prospis* Patches**
Succulent Karoo: low shrub	249724	28.57
Fynbos: low shrub	154249	17.64
Cultivated commercial annual crops non-pivot	95535	10.93
Nama Karoo: bare ground	74205	8.49
Nama Karoo: low shrub	72084	8.25
Succulent Karoo: open bush	52855	6.05
Succulent Karoo: grassland	51343	5.87
Nama Karoo: thicket	28513	3.26

Succulent Karoo: bare ground	27982	3.20
Nama Karoo: open bush	25509	2.92
Succulent Karoo: thicket	16689	1.91
Fynbos: open bush	7057	0.81
Fynbos: grassland	5237	0.60
Nama Karoo: grassland	4426	0.51
Wetlands	2756	0.32
Waterbodies	1733	0.20
Low shrubland	1479	0.17
Settlements	902	0.10
Fynbos: bare ground	552	0.06
Plantations / Woodlots	481	0.06
Cultivated commercial permanent orchards	444	0.05
Cultivated commercial annual crops pivot	193	0.02
Mines	128	0.01
Fynbos: thicket	66	0.01
Cultivated subsistence crops	50	0.01
Degraded	1	0.00

4.10 Evaluating the ***Prosopis Juliflora*** spatial distribution with respect to slope and aspect

East, North, West, South facing slopes and flat surfaces account for approximately 76 % of *P. juliflora* patches while North East, South West, South East and North West facing slopes account for the remaining 24 % (Fig 18). Flat surfaces (0° – 6°) account for 51 % of the *P. juliflora* patches with gentle (7° - 15°) and steep slopes (16° - 30°) account for 34 % and 14 % respectively.

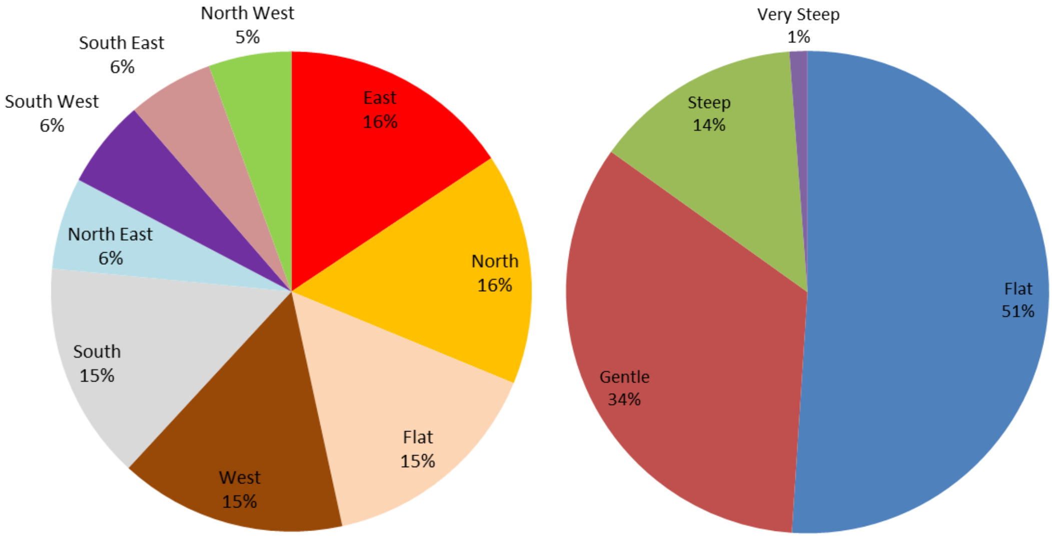

Figure 19: Distribution of *P. juliflora* with respect to aspect and slope

4.11 Exploratory data analysis and variable selection

Exploratory data analysis was used to investigate variables that would be effective in building the binomial logistic regression based predictive model. The dependent variable was *P. juliflora* presence (1) and absence (0) locations while land use land cover, soil type, soil depth, aspect, slope, geology and average temperature and average precipitation were independent variables. A stepwise statistical variable significance validation was assessed at 3 confidence levels i.e ($\alpha = 0.05$), ($\alpha = 0.01$) and ($\alpha = 0.001$). Exploratory data analysis yielded 5 statistically significant variables while 4 were not significant. Interior assessment of the insignificant variables was carried to investigate possible sub variable significance. The thorough assessment yielded negative results which confirm the insignificance of independent variables such as precipitation, slope and aspect (Table 12).

Table 12: Exploratory data analysis for all variables

Deviance Residuals:					
Min	1Q	Median	3Q	Max	
-2.94358	-0.10960	-0.01052	0.38066	3.08913	
Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-5.193e+01	1.339e+01	-3.880	0.000105	***
average_temperature	1.795e+00	5.191e-01	3.458	0.000545	***
precipation	-2.382e-01	1.523e-01	-1.565	0.117653	
land_cover_land_use	-7.772e-02	1.103e-02	-7.043	1.88e-12	***
slope	1.630e-02	2.141e-02	0.761	0.446399	
aspect	-9.217e-04	1.278e-03	-0.721	0.470921	
elevation	8.954e-03	2.817e-03	3.178	0.001481	**
geology	1.843e-01	4.479e-02	4.115	3.87e-05	***
soil_type	-1.484e-02	1.249e-02	-1.187	0.235049	
average_soil_depth	8.866e-03	1.391e-03	6.376	1.82e-10	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
(Dispersion parameter for binomial family taken to be 1)					
Null deviance: 1017.87 on 737 degrees of freedom					
Residual deviance: 346.94 on 728 degrees of freedom					
AIC: 366.94					

4.12 Selecting final variables for the predictive model

All variables statistically significant at α = 0.05 were included as covariates in the predictive binomail logistic regression and the AIC was monitored to assess the performance of the model. The AIC for the model that included al variables was 366.9 and 728 degreess of freedom with residual deviance. In a stepwise process variables that were not statiscally significant were individually remoeved. The AIC evevtually dropped to 223.35 where elevation, soil depth and average temperature were the most effective and signicant variables to predict future invasion ranges (Table 13). The following expression was used to produce the susciptibility raster:

P. juliflora^ = *-29.453827 + (average temperature x 0.908076) + (elevation x 0.002861) + (average soil depth x 0.008828)*

Table13: Variables included in the best performing model

Deviance Residuals:					
Min	1Q	Median	3Q	Max	
-2.39223	-0.23550	-0.05887	0.40719	2.23744	
Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-29.453827	2.756626	-10.685	<2e-16	***
average_temperature	0.908076	0.110279	8.234	<2e-16	***
elevation	0.002861	0.001291	2.217	0.0266	*
average_soil_depth	0.008828	0.000872	10.124	<2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
(Dispersion parameter for binomial family taken to be 1)					
Null deviance: 1017.87 on 737 degrees of freedom					
Residual deviance: 428.09 on 734 degrees of freedom					
AIC: 223.26					

4.13 Performance of the predictive model

The assessment was in two-fold, firstly the performace of the model was assessed based on AIC and pseudo R^2 . Secondly the model’s predictive perfomance was assessed using the 25 % of the presence and absence sample that was excluded from the model. The AIC of the first model that included all variables was 366.95 while the AIC of the final model was 223.26. The pseudo R^2 for the for the first model was 0.62 while the pseudo R^2 for the final model was 0.78. The above narrated results indicated continous improvement of the model. The probability was divided into two, a threshold of 0.5 was used to distinguish low and high probabality. A total of 188 points were used to validate the model and 78 % of them were located in moderate to high probability areas (Fig 19).

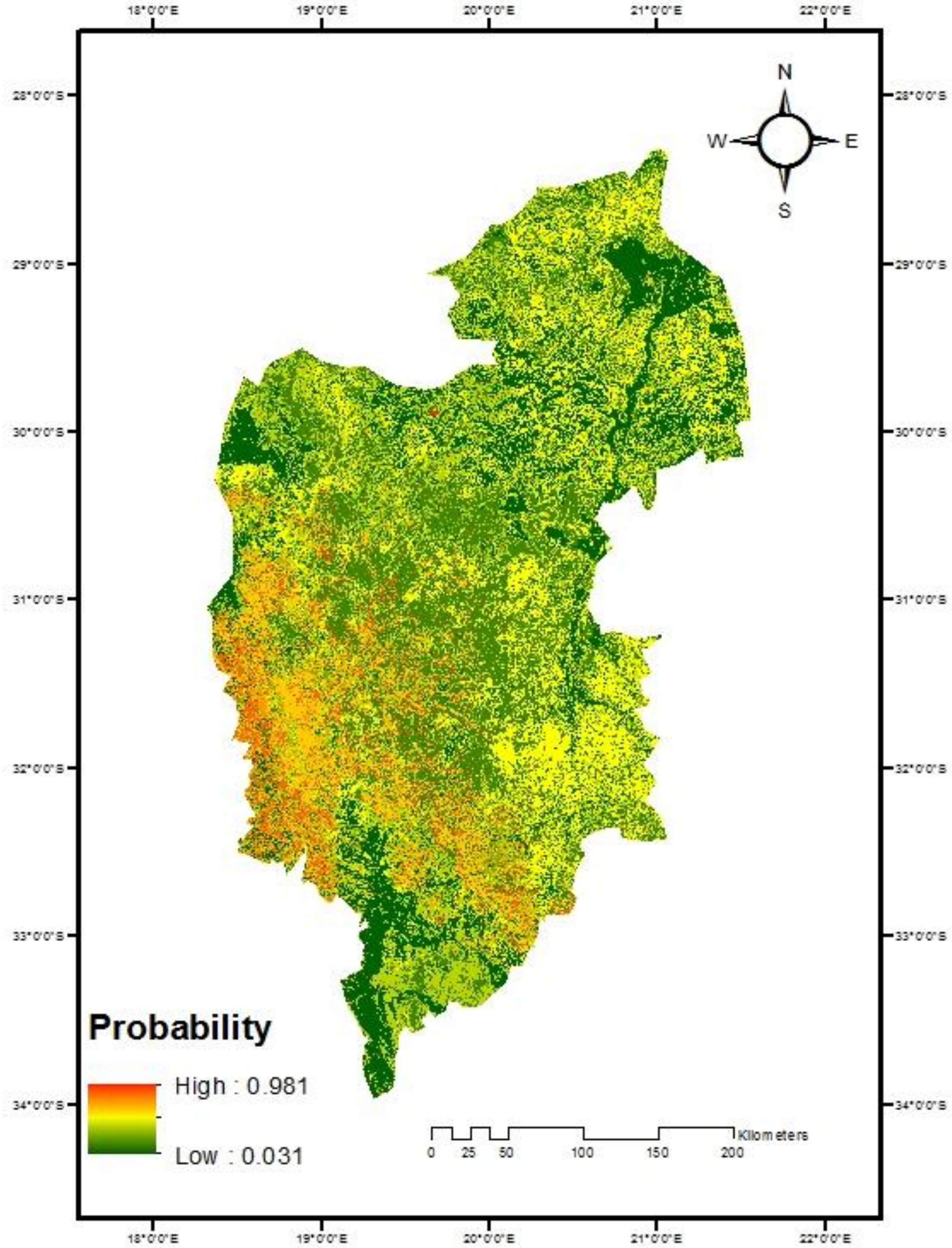

Figure 20: A probability map showing areas susceptible to *P. juliflora* invasion

Chapter Five

5. Discussion

In light of the negative impacts of *Prosopis* on ecological infrastructure and critical natural resources, policy makers are seeking methods to control or eradicate it (Wakie et al. 2014). The main impediment hindering development of *Prosopis* control management strategies is lack of credible spatial data (Berhanu et al. 2016). Small scale studies used field surveys to understand the dynamics such as spread rate and infestation density, these proved to be labour intensive and costly (Le Maitre et al. 2004; Versfeld et al. 1998; Robinson et al. 2007). Furthermore, monitoring *Prosopis* invasion rate requires repeat field measurements which would be unsustainable whilst remote sensing offers repeat measurements sustainably. This study investigated the potential of using Sentinel - 2A imagery classification, spatial point pattern analysis and step-wise multivariate binomial logistic regression to map, evaluate the density and predict *Prosopis* dynamics.

The first objective was to map the spatial distribution of *P. juliflora* and other land cover land use classes. This was achieved by employing random forest and support vector machine classifiers to assign Sentinel – 2A pixels to 12 classes. The satisfactory OA, UA and PA illustrate the potential of using Sentinel – 2A imagery however visual validation of the classification revealed spatial and spectral resolution limitations. Omission of *Prosopis* clusters less than the 10 m spatial resolution and spectral confusion between *Fynbos open bush* and *P. juliflora* had a negative impact on the overall accuracy and individual accuracies. Spectral similarities between some rock outcrops and settlements also led to spectral confusion thus impacting negatively on the accuracies of the classifiers. The spectral confusion between the above mentioned classes was confirmed by the class separation plots of both RF and SVM classifiers. In this study SVM outperformed RF by (3 %). Similar results were obtained in a study by Cho et al, (2015) where both classifiers' ability in tree species classification was tested. The RF classifier variable importance measure revealed that the red and blue bands were the most effective for the overall classification, this compares well with Mureriwa et al, (2015). Mapping *Prosopis* with remotely sensed data has gained a lot of traction, although it has proved to be an efficient method it still has some limitations. Studies that integrated geographic information systems and remote sensing methods to map *Prosopis* showed great potential however they were subjected to conflicting resolutions between GIS and RS datasets (Van der Berg, 2014; Zeila, 2011). Previous attempts of mapping *Prosopis* using Landsat were prone to omission/commission and spectral confusion due to poor spatial and spectral resolution respectively (Van der Berg, 2014; Hoshino et al 2013). Researchers attempted mapping *Prosopis* using aerial photographs with high spatial resolution managed to detect individual trees however they were unable to separate species due to poor resolution and costs associated with data acquisition remain a limiting factor (Laliberte et al 2004; Eva et al. 2005; Robinson et al. 2007). Studies that employed commercial remotely sensed data with high spectral and spatial results produced high accuracies and able to map small patches of *Prosopis* however they were limited to small geographic areas due to data acquisition costs (Robinson et al. 2016; Mureriwa et al. 2015; Laliberte et al. 2004). The results obtained from this study indicate that moderate to high spectral and spatial resolution of Sentinel – 2A has the potential to separate *Prosopis* from other LULC. Results from this study have some similarities with a study carried out by Vanthof and Kelly (2017) in Tamil Nadu, India where SVM and RF classifiers were employed to map *P. juliflora* using Sentinel – 2A. The studies discussed above suggest that spectral and spatial resolution plays a critical role in determining the accuracies of mapping using remotely sensed data. Freely available data with low spectral and spatial resolution is associated with commission errors and spectral confusion while commercial data with high spectral and spatial resolution is often limited to small scale studies (Van der Berg. 2014; and Mureriwa et al. 2015). The results from this study presents Sentinel – 2A data as a viable option for mapping and monitoring *Prosopis* invasion at large scales cost effectively.

The degree of density of invasive alien species has an effect of how they deplete resources for native species thus quantifying spatial patterns plays a critical role in crafting environmental management plans (Richardson & Thuiller., 2007). The second objective was to evaluate the spatial distribution and density of *P. juliflora* across the study area. This was achieved by turning *P. juliflora* pixels from the best performing classifier (SVM) to points in ArcGIS. Approximately 800 000 points were imported into R studio to carry out spatial point pattern analysis. The Clark Evans test for complete randomness produced a p-value < 2.2 e $^{(-16)}$ which indicated clustering. This was also confirmed by results obtained from quadrat analysis and kernel density evaluation. *P. juliflora* density is relatively high on the south eastern and western parts. Spatial point pattern analysis with respect to geology illustrated that Cares, Waterford and Bidouw had the highest density of *P. juliflora.* An evaluation of spatial distribution of *P. juliflora* with respect to soil depth show that deep soils have the highest density, this is similar to findings in a study by Browning et al, (2014). Approximately 28 % and 17 % of *P. juliflora* coexisted with Succulent Karoo: low shrub and Fynbos: low shrub respectively however the highest density was within Succulent Karoo grassland. *P. juliflora* density analysis with respect to slope indicated that *P. juliflora* was clustered on flat areas, river banks and riparian areas. This mirrors findings of the image classification in the first objective and studies by Hoshino et al. (2013). *P. juliflora* density evaluation with respect to soil types reported high densities in loamy and clay soils. This is comparable to results obtained by Robinson et al. (2007) where they discovered that *Prosopis* was highly clustered in red loamy soils in South Western Australia.

The third objective was to produce a predictive model based on logistic regression. The model was expected to produce location-specific probability and associated error. Moreover, it estimated the probability of determinants that are conducive for *P. juliflora* invasion i.e to investigate environmental settings favourable for invasion. The statistical determination of significant variable and model performance assessment was carried out in R global version 3.3.1 and the raster calculator was used to create the probability surface. Only three of the 9 independent were dimmed statistically significant and effective in predicting future ranges of *P. juliflora* invasion. The pseudo R^2 and AIC were used to evaluate the performance of the logistic regression model, the final model had a pseudo R^2 of 0.78, AIC of 223.26 and three significant independent variable.

Limitations

All objectives were completed as per the proposal and satisfactory results were obtained however there were some hindrances that limited attainment of better results. These were some of the challenges:

- The average canopy diameter of *Prosopis Juliflora* is 2 m and the spatial resolution of Sentinel – 2A is 10 m this implies that only *Prosopis Juliflora* clusters greater than the sensor spatial resolution were detected.

- The field data obtained from the Northern Cape Department of Environmental Affairs was too coarse and had no metadata thus there was uncertainty about date of acquisition and methods employed.
- Google Earth imagery used to validate the image classification was captured for different seasons.
- The climatic data (average temperature and precipitation) was too coarse.
- The predictive model's transferability to a different region was not carried out due to time limitations

Chapter Six

6. Conclusions

A thorough analysis of the results of the study led to the following conclusions:

- The of Sentinel – 2A imagery shows great potential for classifying *Prosopis Juliflora* and its surrounding land cover classes however it omits individual trees with canopy size less than its spatial resolution. Moreover, the classification is prone to spectral confusion amongst surfaces with similar reflection characteristics.
- The Random Forest classifier variable importance selector detected the visible and near infrared bands as the most effective in the classification of *Prosopis Juliflora* and other vegetation species. Support Vector Machine classifiers yielded better accuracies than Random Forest classifier however the *Mcnemar Test* was not carried out to validate the results. From both classification outputs it was noted that *Prosopis Juliflora* was concentrated along hydrological paths.
- The spatial point pattern analysis results from this study illustrated that the distribution of *Prosopis Juliflora* was not random and geophysical factors such as soil-depth, slope and rock formations have an influence on invasion density.
- Step-wise regression shows great potential as a tool for predicting future invasion ranges however lack of good quality data may limit its potential.

The successful integration of Remote Sensing and Geographical Information Systems for detecting and analysing spatial distribution of *Prosopis Juliflora* presents ecologists and biodiversity planners an efficient and cost friendly method of obtaining spatial dynamics of invasions.

6.1 Recommendations

- The potential of Sentinel – 2A for classifying *Prosopis Juliflora* in arid regions should be further investigated with the aid of field data that matches the image acquisition dates.
- Future studies should investigate the influence of anthropogenic drivers in spreading *Prosopis Juliflora*.

- Research should be carried out to investigate whether sub pixel classification reduce the omission of individual trees smaller than the spatial resolution of Sentinel – 2A.
- Supplementary studies should investigate the apparent clustering of *Prosopis Juliflora* along hydrological paths.
- Future studies should evaluate the performance of methods employed in this study in different regions or with a different invasive species.

7. References

Abebe, Y. and Corresponding, T., (2012)., Ecological and Economic Dimensions of the Paradoxical Invasive Species- *P.juliflora* and Policy Challenges in Ethiopia. Journal of Economics and Sustainable Development, 3(8), pp.62–71.

Adam, E. and Mutanga, O. (2009)., Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. ISPRS Journal of Photogrammetry and Remote Sensing,3(64), p.612-620

Akasapu, A.K. Rao, P.S. Sharma, L.K. and Satpathy, S.K., (2011)., Density based k-nearest neighbors clustering algorithm for trajectory data. International Journal of Advanced Science and Technology, 31(1), p.121-145

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

Bailey, T. C. and A. C. Gatrell., (1995) Interactive Spatial Data Analysis. Longman Scientific & Technical, Harlow, U.K. 395 pp..

Berhanu, A. et al., (2016)., The *Prosopis* dilemma , impacts on dryland biodiversity and some controlling methods. Journal of the drylands, 1(February).

Breiman, L., (2001)., Random forest. Mach. Learn. 45.

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

Brunel,S.,Branquart,E.,Fried,G.et al. (2010)., The EPPO prioritization process for invasive alien plants. EPPO Bulletin, 40(3), p.407–422.

Bethea, L., (1986)., Statistical methods for engineers and scientists, 2nd Edition, Marcel Dekker..

Boots, B. and A. Getis., (1988)Point Pattern Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07–001. Sage Publications..

Blaschke T., (2010)., Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 4(65), p.2-16.

Browning, D. M. S. R. Archer, G. P. Asner, M. P. M and Wessman, C. A., (2014)., Spatial patterns of grassland–shrubland state transitions: a 74-year record on grazed and protected areas. Ecological Applications 2 (45),p. 1421–1433

Buckley,L.B. Urban,M.C. Angilletta, M.J.et al., (2010)., Can mechanism inform species' distribution models? Ecology Letters, 13(8), p.1041–1054.

Campbell, J.B., (2002). Introduction to remote sensing. 3rd ed. New York: The Guildford Press.

Cochran , WG (1997). Sampling techniques Wiley.

Cook, D.C. Thomas, M.B. Cunningham, S.A. Anderson,D.L. and De Barro,P.J. (2007)., Predicting the economic impact of an invasive species on an ecosystem service. Ecological Applications, 17(6), p.1832–1840.

Cohen, J., (1997). Statistical power analysis for the behavioural sciences. Academic Press.

Coetzer, W. and Hoffmann, J., (1997)., Establishment of Neltumius arizonensis (Coleoptera: Bruchidae) on Mesquite (ProsopisSpecies: Mimosaceae) in South Africa. Biological Control, 3(10), p.187-192

Colwell,R.K. & Rangel,T.F., (2009)., Hutchinson's duality: The once and future niche. Proceedings of the National Academy of Sciences, 106(2), p.19651–19658.

Cortes, C. and Vapnik, V., (1995a)., Support-vector networks. Machine learning, 3(20), p.273-297

Cortes, C. and Vapnik, V. (1995b)., Support-vector networks. Machine learning, 3(20), p.273−297

Chuvieco, E. and Huete, A., (2010)., Fundamentals of satellite remote sensing. Boca Raton: Taylor & Francis, 5(2), p.78-83

Diggle, P.J., (1985)., A kernel method for smoothing point process data. Applied Statistics Journal of the Royal Statistical Society, Series 34(1),p.138–147.

Dowhower, SL and Teague, WR., (2017)., Spatial Patterns in a Prosopis – Juniperus Savannah. The Texas Journal of Agriculture and Natural Resources 2(30),p.63-77

Ghosh, A. Fassnacht, F.E. Joshi, P.K. and Koch, B., (2014)., A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. International Journal Applications. Earth Observations. Geoinformation. 4(26), p.49–63.

Geesingis, D. Al-Khawlani, M. and Abba, M.L., (2004)., Management of introduced Prosopisspecies: can economic exploitation control an invasive species? Unasylva, 5(3), p.209-217

Genuer, R. Poggi, J.M. and Tuleau-Malot, C., (2010)., Variable selection using random forests. Pattern Recognition Letters, 3(31), p.2225-2236

Grabarnik, P. and Chiu, S.N., (2002)., Goodness-of-fit test for complete spatial randomness against mixtures of regular and clustered spatial point processes. Biometrika, 89(2), p.411-421.

Harding, G., (1987). status of *Prosopis* as a weed. Applied plant science= Toegepaste plantwetenskap 23(4), p.34-41

Hoshino, B. Karamalla, A. Manayeva, K. Yoda, K. Suliman, M. Elgamri, M., Nawata, H. and Yasuda, H., (2012)., Evaluating the invasion strategic of Mesquite (Prosopis juliflora) in eastern Sudan using remotely sensed technique 53(4), p. 76-83

Huang, C. Davis, L. and Townshend, J. (2002)., An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23(2), p. 725-749

Jose, M. Fedriani. and T. Wiegand, M., (2010)., Spatial pattern of adult trees and the mammal-generated seed rain in the Iberian pear 32(3),p. 531-536

Kent, M. and Coker, P., (1994)., Vegetation Description and Analysis*:* A practical Approach. London: John Wiley and Sons.45(2),p. 412-418

Kearney,M. and Porter,W.(2009)., Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. Ecology Letters, 12(4),p. 334–350.

Kokalj1 Z. and Ostril K., (2007)., Land Cover Mapping Using Landsat Satellite Image Classification in the Classical Karst - Kras region. Institute of Anthropological and Spatial Studies ZRC SAZU

Klein, H. Hill, M. Zachariades, C. and 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, 19(2), p.488-

497

Kumar, S. and Mathur, M., (2014)., Impact of invasion by *P.juliflora* on plant communities in arid grazing lands. Tropical Ecology, 55(1), p.33–46.

MacVicar, C.N., et al, (1977)., Soil Classification, A Binomial System for South Africa. Department of Agricultural Technical Services, Pretoria, South Africa, 1(2), p.166-171.

Malahlela, OE, Cho M.A and Mutanga O., (2016)., Mapping canopy gaps in an indigenous subtropical coastal forest using high resolution WorldView-2 data. Internatinal Journal of Remote Sensing, 41(1), p. 113-118

Marais, C. and Wannenburgh, A.M., (2012)., Restoration of water resources (natural capital) through the clearing of invasive alien plants from riparian areas in South Africa — Costs and water benefits. South African Journal of Botany, 74(1), p.526–537.

Mureriwa, N. and Adam, E., (2016)., Mapping *Prosopis glandulosa* (mesquite) invasion in the semi-arid environment of South Africa using high resolution WorldView2 imagery and machine learning classifiers. Journal of Arid Environment.

Mucina, L and Rutherford MC (2006)., The vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria, South Africa, Strelitzia, vol. 19.

Mather PM (2004)., Computer processing of remotely-sensed images. 3rd ed. Chicester, West Sussex: Wiley.

McHugh, M.L., (2013)., The Chi-square test of independence. Biocheia Medica. 23 (2), p.143–149.

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

Mountrakis, G. Im, J. and Ogole, C. (2011)., Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), p.247-259

Miyamoto, A. Tamura, N. Sugimura. N and Yamada F., (2004)., Predicting Habitat Distribution of the Alien Formosan Squirrel Using Logistic Regression Model: Global Environmental Research ©2004 AIRIES 8(1), p.13-21

Oommen, T. Misra, D. Twarakavi, N.K. Prakash, A. Sahoo, B. and Bandopadhyay, S. (2008)., An objective analysis of support vector machine based classification for remote sensing. Mathematical geosciences, 40(3), p.409-424

Lan, G. Hua, Z. Cao, M. Wang, Y. Deng, X. and Song, Z., (2009) Spatial dispersion patterns of trees in a tropical rainforest in Xishuangbanna, southwest China. Ecological Research, 24(3), p.1117–1124 DOI 10.1007/s11284-009-0590-9

Le Maitre, D.C. Richardson, D.M. and 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(1),p. 103-112

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

Lemke, D. and Brown, J.A., (2011) Distribution modelling of Japanese honeysuckle (Lonicera japonica) invasion in the Cumberland Plateau and Mountain Region, USA. For. Ecoligical. Management., 262, p.139–149.

Li, F., Kustas, W.P. Prueger, J.H. Neale, C.M. and 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(3), p.878-891

Lillesand, M. Thomas and Ralph W., (2004) Remote Sensing and image Interpretation: 5th ed. John Wiley & Sons, Inc.

Lalibertea, A.R.K Havstada, M. Reldon, F. and Gonzaleza, L., (2004) Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico : Remote Sensing of Environment ,93(2), p.73-78

Liu X-H, Skidmore AK and Van Oosten H., (2002)., Integration of classification methods for improvement of land-cover map accuracy. ISPRS Journal of Photogrammetry and Remote Sensing 56(4),p. 257-268.

Lin, X. Sun, L. Li, Y. Guo, Z., Li, Y. Zhong, K. Wang, Q. Lu, X. Yang, Y. and Xu, G., (2011)., A random forest of combined features in the classification of cut tobacco based on gas chromatography fingerprinting. Talanta, 82(1), p.1571-1575

Robinson, T.P. van Klinken and R.D, Metternicht , G (2007)., Spatial and temporal rates and patterns of mesquite (*Prosopis* species) invasion in Western Australia.Journal of Arid Environments 72 (3), p. 112-117

Robinson, T. Wardell-Johnson, G. Pracilio, G. Brown, C. Corner, R. and 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(2), p.23-30

Richardson,D.M. and Thuiller,W. (2007)., Home away from home–objective mapping of high---risk source areas for plant introductions. Diversity and Distributions, 13(3), p.299–312.

Shackleton, R.T. et al., (2015)., South African Journal of Botany The impact of invasive alien Prosopis species (mesquite) on native plants in different environments in South Africa. South African Journal of Botany, 97(1), p.25–31. Available at: http://dx.doi.org/10.1016/j.sajb.2014.12.008.

Schmidt, K., and Skidmore, A. (2003)., Spectral discrimination of vegetation types in a coastal wetland. Remote Sensing of Environment, 85(4), p.92-108

South African Weather Services, (2016)., Climate of the Northern Cape. Pretoria: South African Weather Services.

Spósito, M. B. Amorim, L., Ribeiro, P. J., Jr. Bassanezi, R. B. and Krainski, E. T. (2007)., Spatial pattern of trees affected by black spot in citrus groves in Brazil. Plant Dis. 91(2), p.36-40.

Tesema, B. Animut, G. and Urge, M. (2013)., Effect of Green *P. juliflora* Pods and Noug Seed (Guizotia obissynica) Cake Supplementation on Digestibility and Performance of Blackhead Ogaden Sheep Fed Hay as a Basal Diet. Science, Technology and Arts R esearch Journal, 752(2), p.38–47.

van der Meer, F. D., van der Werff, H. M. and van Ruitenbeek, F. J., (2014). Potential of ESA's Sentinel-2 for geological applications. Remote Sensing of the Environment, Volume 148(3), p. 124-133.

van der Werff, H. and van der Meer, F., (2015). Sentinel-2 for Mapping Iron

Absorption Feature Parameters. Remote Sensing*,* 7(1), p. 12635-12653.

van Klinken, R.D. Hoffmann, J.H., Zimmermann, H.G. Roberts, A.P. Muniappan, R. Reddy, G. and Raman, A., (2009)., *Prosopis* species (Leguminosae). Biological Control of Tropical Weeds using Arthropods,31(2), p. 353-377

Vijayaraghavan,R. and Sundaramoorthy, S.,(2012)., Mapping *Prosopis juliflora* Using Satellite Data in a Part of Indian Thar Desert. International Journal of Ecology and Environmental Sciences, (38) 1,p.23-29

Ustin, S.L. Roberts, D.A. Gamon, J.A. Asner, G.P. and Green, R.O., (2004)., Using imaging spectroscopy to study ecosystem processes and properties. *BioScience,* 54(1), p.523-534

Shackleton, R.T. Le Maitre, D.C. Pasiecznik, N.M. and 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

Steenkamp, H. and 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(4), p.305-311

Strobl, C. and Zeileis, A., (2008)., Exploring the statistical properties of a test for random forest variable importance. Compastat 2008 - Proceedings in Computational Statistics, 23(2), p.59 - 66

Vanthof R. and Kelly R. E. J., (2017)., "Mapping *Prosopis juliflora* invasion within rainwater harvesting structures in India using Google Earth Engine," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, 67(1) p.1115-1118. doi: 10.1109/IGARSS.2017.8127152i: 10.1109/IGARSS.2017.8127152

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

Voster, C.J., (2003)., Simplified Geology, South Africa. Lesotho and Swaziland. Council for Geoscience, South Africa

Wakie, T.T. et al., (2014)., Mapping Current and Potential Distribution of Non- Native

P.juliflora in the Afar Region of Ethiopia. PLOS ONE, 9(11), p.3–11.

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

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