INTEGRATING REMOTE SENSING AND GEOSTATISTICS IN MAPPING SERIPHIUM PLUMOSUM (BANKRUPT BUSH) INVASION

Morwapula Jurrian Mashalane

Submitted in fulfilment of the academic requirements for the degree of
Master of Science in the Discipline of Geography in the School of
Agricultural, Earth and Environmental Sciences

University of KwaZulu-Natal
Pietermaritzburg, South Africa
December 2016

DECLARATION

I, Morwapula Jurrian Mashalane, declare that the research reported in this dissertation is my original work, unless indicated through references. This dissertation has not been previously/currently submitted for the attainment of a degree or examination purposes at any university apart from University of KwaZulu-Natal. This dissertation does not contain any data, graphics, and information from other persons whom have not been acknowledged. This dissertation does not contain other persons' writing, unless acknowledged as such, and in cases where written sources have been cited, their words have been paraphrased and general information attributed to them referenced, and where exact words have been used, they were placed inside quotation marks, referenced and italicized. Furthermore, this dissertation does not contain graphics, tables directly copied and pasted from the internet, unless the sources are acknowledged within the content of this research.

Signed:……………………………………….
Morwapula Jurrian Mashalane (214576458)

I, Dr John Odindi, being the candidate's supervisor, I certify that the above declaration is true to the best of my knowledge and have approved this dissertation for submission.

Signed:……………………………………….
Supervisor: Dr. John Odindi

I, Dr Clement Adjorlolo, being the candidate's co-supervisor, certify that the above declaration is true to the best of my knowledge and have approved this dissertation for submission.

Signed:………………………………………
Co-supervisor: Dr Clement Adjorlolo

Table of Contents

DECLARATION .... **i**

List of Figures .... **v**

List of Tables .... **vi**

List of Appendices .... **vii**

ABSTRACT **1**

DEDICATIONS **2**

ACKNOWLEDGEMENTS **3**

CHAPTER ONE **4**

INTRODUCTION **4**

1.1 Background 4

1.2 Aim and objectives of the study 8

1.3 Specific objectives 8

1.4 Key research questions 9

1.5 Organisation of the thesis 9

1.6 Study area 11

CHAPTER TWO **12**

LITERATURE REVIEW **12**

2.1 Applications of optical remote sensing for mapping shrub invasion 12

2.2 Potentials of mapping *S. plumosum* using remote sensing 13

2.3 Image transformation techniques used in remote sensing of vegetation 14

2.4 SPOT 6 data derived Getis statistics 15

2.5 Getis statistics and classification techniques 17

2.6 Geostatistics and remote sensing of vegetation 18

2.6.1 Kriging technique and application to *S. plumosum* mapping 19

2.6.2 Co-kriging techniques and application to *S. plumosum* mapping 20

2.7 Accuracy assessment 21

2.8 Lessons learnt from literature review 22

CHAPTER THREE **23**

MATERIALS AND METHODS **23**

3.1 Field sample plot survey and structural data collection 23

3.2. Data processing and analysis 26

3.3. Getis statistics methods 28

3.3.1. Local indicator of spatial association 28

3.3.2. Getis image transformations 28

3.4 Variable selection analysis (Assessment and comparison of the Getis statistics indices layers) 30

3.5 Classification methods 32

3.6 Regression methods 32

3.7 Geostatistics methods 33

3.7.1 Sampling image data using variogram 35

3.7.2 The Experimental Variogram 35

3.7.3 Kriging and cokriging technigues of interpolations 36

CHAPTER FOUR **42**

RESULTS **42**

4.1 Getis transformed indices layer variables. 42

4.2 Classification and accuracy assessment 45

4.3 Geostatistical analysis 47

4.3.1 Selecting the optimal variable for density estimation 47

4.3.2 Investigating directional influence in the data 48

4.3.3 Estimation of *S. plumosum* canopy density and percentage cover. 49

CHAPTER FIVE **54**

DISCUSSIONS AND CONCLUSIONS **54**

5.1 Assessing the ability of Getis statistics to discriminate pixels of *S. plumosum* 54

5.3 Geostatistics analysis results 56

5.3 Conclusions 58

5.3.1 Aims and objectives re-visited 59

5.3.2 Key research questions 60

5.4 Recommendation for future use of Getis and geostatistics 60

REFERENCES...**61**
APPENDICES...**69**

List of Figures

Figure 1: Flowchart for the methods followed and outputs in the study. 10

Figure 2: Study area demonstrating farms surveyed...11

Figure 3: Surveyed and high resolution image *S. plumosum* plots. 25

Figure 4: Variable importance reduction according to the performance of each Getis index layer for lag 5. 44

Figure 5: Variable importance reduction according to the performance of each Getis index layer for lag 3. 44

Figure 6: Demonstration of *S. plumosum* infested areas mapped using top 3 Getis indices layers. The area covered by *S. plumosum* is 13148.7 hectares. 45

Figure 7: Demonstration of the extent of *S. plumosum* patch mapped, a) SPOT 6 image used for the study, b) SPOT 6 image with *S. plumosum* patches mapped. 46

Figure 8 : Field collected canopy density histogram 47

Figure 9: Demonstration of the presence of directional influence to the semivariogram model. 48

Figure 10: Variograms for prediction models of *S. plumosum* canopy density, a) field canopy density data variogram; b) Variogram for Rook, Horizontal and positive Getis transformed layers; c) Cross-variogram for field canopy density data combined with Getis transformed layer variable for cokriging modelling. 51

Figure 11: Density cover map of *S. plumosum* created through Geostatistical cokriging of field canopy density information and Getis transformed layer variables. Filled contours depict different levels of density infestation per 400 sqm. 52

Figure 12 : Percentage cover map of *S. plumosum* created through Geostatistical cokriging of field canopy percentage information, SPOT 6 multispectral band 1, Positive and horizontal Getis transformed layer variables. Filled contours depict different levels of percentage infestation per 400 sqm. 53

List of Tables

Table 1: Plot canopy measurements and observations 25
Table 2: Specifications of the SPOT 6 multispectral imagery 27
Table 3: Sample confusion error matrix 39
Table 4 : Top 20 Getis indices transformed layers, 10 for each lag 5 and 3 respectively. "Mean" refers to layer in the naming. 43
Table 5 : Classification confusion matrix 46
Table 6 : Correlation (r) between field data parameters and SPOT 6 data (multispectral data and Getis transformed data). 48
Table 7: Kriging and cokriging model root mean square error results. The variables used in the model are Rook, Horizontal and Positive Getis transformed layer variables. 50
Table 8: Variogram parameters for predicted *S. plumosum* canopy density 52

List of Appendices

Appendix 1: Field sheet……………………………………………………………………66

Appendix 2: Field work schedule…..………………….………………………………..……….67

Appendix 3: Field pictures…..…………………………………………………..……....68

ABSTRACT

The impacts of plant species invasion in natural ecosystems have attracted geo-scientific studies globally. Several studies have demonstrated that the effects of invasive species can permanently alter an ecosystem structure and affect its provision of goods and services, e.g. the provision of food and fibre, aesthetics, recreation and tourism, and regulating the spread of diseases. Plant invasion causes transformation of ecosystems including replacement of native vegetation. This study focuses on invasive plant impacting on grasslands called *Seriphium plumosum*. The plant is known to have allelopathic effects, killing grass species and turning grazing lands into degraded shrublands. The major challenge in grassland management is the eradication and management of *S. plumosum.* Central to this challenge is locating, mapping and estimating the invasion status/cover over large areas. Remote sensing based earth observation approaches offer a viable method for invasion plants mapping. Moreover, mapping of vegetation requires robust statistical analysis to determine relationships between field and remotely sensed data. Such relationships can be achieved using spatial autocorrelation. In this study, Getis statistics transformed images and geostatistical techniques, which involve modelling the spatial autocorrelation of canopy variables have been used in mapping *S. plumosum*. Getis statistics was used to transform SPOT (Satellites Pour l'Observation de la Terre)-6 image bands into spatially dependent Getis indices layer variables for mapping *S. plumosum*. Stepwise multiple Regression, ordinary kriging and cokriging were used to evaluate the cross-correlated information between SPOT6-derived Getis indices transformed layer variables and field sampled *S. plumosum* canopy density and percentage. To select the best SPOT6-derived Getis indices to map *S. plumosum*, 308 spectral Getis indices transformed layer variables were statistically evaluated. Results indicated that Rook, Positive and Horizontal Getis indices are most suitable for mapping *S. plumosum* with 0.83, 0.828 and 0.828 importance. The most accurate Getis index obtained using 5x5 (Lag 5) moving window yielded 0.83 mapping importance. Cokriging with the most important Getis index yielded the best in *S. plumosum* density prediction with root mean square error (RMSE) of 25.8 compared to ordinary kriging with RMSE of 26.1 and regression with RMSE of 35.6. This study demonstrated that Getis statistics and geostatistics were successful in mapping and predicting *S. plumosum*. The current study provides insights critical for developing sound framework for planning and management of *S. plumosum* in agro-ecological systems.

DEDICATIONS

I dedicate this work to my Mother, Ngwakwana A. Mashalane, Grandmother Mosibudi E. Mashalane and Uncle Nkama J. Mashalane and Mabu D. Mashalane for their continued invaluable advice, support, prayers and believing in me. I could not have asked God for a better family, Love you all. I also dedicate this to my brother Sontaga Kolobe Mashalane, I know you will achieve more than me.

ACKNOWLEDGEMENTS

Above all, I thank Modimo wa thaba ya Sione for granting me the mind and the strength to carry on, for His grace, mercy and patience, preserving my life, for providing me with great people to learn from and for his guidance. I also would like to thank my Mother and Grandmother, who taught me to pray, to work hard, and all other principles of life.

Additional acknowledgement and thanks go to the South African National Space Agency (SANSA) for funding my studies, for all the opportunities in support of competence. I thank Dr Paida Mhangara and Dr Clement Adjorlolo for their guidance, support, and motivation. I would like to thank Mr. Oupa Malahlela, for assistance during fieldwork. I would also like to extend my sincere gratitude to all SANSA Earth observation team for their support with technical, scientific and social support. I further express my infinite appreciation and deepest gratitude to my supervisors, Dr John Odindi, and Dr Clement Adjorlolo for their constant support, invaluable critical comments and expert advice, wisdom, enthusiasm and determination to build me scientifically.

I would like to thank Department of Agriculture for their recognition and support in providing data and infestation database for the study. I further thank farm owners of Zyfer fontein; Hillyside; DeWildt; Rondebos; Nelville; Witsand; Witzand; Nieuwejaarsfontein; Aurora; Koekemoers rekwest; Jocador; Virginia; Virginia; Arundel; Bidsulphsberg; Canada; Cyferkuil; Liebenbergs bult; Astorea; Maclear; Frieden; Eenzaamheid; Groot Taaiboschfontein; Schurvekop; Melsetter; Gethsemane; Waterval; Dupreezpoort; Langverwacht; Weiveld; Kalkoenkrans; Stoffelina; De la Harpe; Kranskop; Deelfontein; Brakwater; Makwera; Liebenbergs bult in the Free Sate for allowing us access to their properties.

CHAPTER ONE

INTRODUCTION

1.1 Background

Invasive plant species are defined as plants occurring outside their natural environmental range (Turlings et al., 2001). Invasive plant species affect a wide range of ecosystem goods and services that underpin human wellbeing, e.g. provision of food, aesthetics, recreation and tourism (Brooks et al., 2004). Plant invasions also compromise ecosystem stability and threaten agricultural productivity (Devine and Fei, 2011). Several literature on the effects of plant invasions (Brooks et al., 2004, Richardson et al., 2000, Le Maitre et al., 2000, Dogra et al., 2010, Richardson and Van Wilgen, 2004) suggests that most invasive plant species transform ecosystems by excessive water, light and oxygen usage. Furthermore, plant invasion transform ecosystems through the addition of nitrogen to the soil, promote or suppress fires, induce soil erosion, or accumulate and redistribute salts (Richardson et al., 2000). Such changes may alter the flow, availability and/or quality of nutrient resources in biogeochemical cycles, modify trophic resources within food webs, and alter physical resources such as living space or habitat, sediment distribution, light and water (Vitousek, 1990). According to Pyšek and Richardson (2010), invasion causes a wide range of socio-economic and ecosystem impacts that include a decline in the population of threatened and endangered species, habitat alteration and loss, shifts in food webs and nutrient cycling and loss of agricultural crops and productive lands.

To date, mitigating impacts of plant species invasion remains a challenge (Richardson and Van Wilgen, 2004). In the United States of America for instance, the cost of alien plant species mitigation is estimated to be about USD137 billion per year, a cost that excludes the monetary value of native species extinctions, biodiversity reduction, ecosystem services and aesthetics (Pimentel et al., 2001). In other parts of the world, about 80% of the endangered species are threatened by pressure from invasive species (Pimentel et al., 2005). Pimentel et al. (2005) further noted that the global monetary losses accruing from invasive species amount to USD1.5 trillion per annum. In China, for instance, the total economic losses caused by invasive alien species are estimated to be USD14.45 billion, with direct and indirect economic losses accounting for 16.59% and 83.41% of total economic losses respectively (Xu et al., 2006).

According to Richardson and Van Wilgen (2004), South Africa has been identified as one of the countries most vulnerable to alien plant invasions in the world. A survey by the Southern African Plant Invader Atlas (SAPIA) project in South Africa, Lesotho and Swaziland, identified 548 alien plant species, with most invasion recorded in the fynbos, forests, the moist eastern grassland and savannah biomes (Henderson, 2007). The *Seriphium plumosum* (also known as Slangbos or Bankrupt bush), though indigenous to Western Cape Province of South Africa, has been identified as one of the biggest shrub that threatens the savanna and grassland biomes in Free State, North West, Mpumalanga, Eastern Cape and Gauteng Provinces. The shrub mainly invades bottom and mid slope terrains (Jordaan, 2009). It is a small multi-stemmed woody shrub that grows to an average height of 60 cm and a width of 60 cm. According to Snyman (2009), the shrub's light colour reflects sun light and it's wholly covering and small leaves reduce plant water loss, making it highly adaptive to the long dry summer seasons. The shrub is known to reduce grazing capacity by displacing grass species and excretes volatile oil which makes it unpalatable to livestock and wildlife (Jordaan, 2009). Furthermore, the plant is a strong competitor for soil moisture, light, space and nutrients and commonly out-competes grass species. It produces millions of seeds which contaminate wool on sheep and it is highly flammable during winter, altering fire regimes (Jordaan and Jordaan, 2007).

The *S. plumosum's* ecological and agricultural impacts necessitate a determination of its geographical extents for mitigation purposes. Studies have demonstrated that the effect of invasive species is multi-scale, in which invasive impact is a product of the potential geographical range of the invader, its density, and the measurable impacts at the smallest spatial scale (Richardson and Van Wilgen, 2004). This information is valuable in quantifying the impacts of invasion.

Traditionally, land surveys and analysis of aerial photographs (also a remote sensing technique) have been used in vegetation mapping. However, these survey techniques are labour intensive and time consuming, while aerial photograph acquisition is expensive for large areas, particularly when analysing and mapping large areas. However, satellite-based remote sensing approaches have become popular in vegetation mapping (Madden, 2004). Unlike traditional approaches, satellite data offer quick, reliable and relatively economical mapping, particularly for large areas. The selection of remote sensing data to be used is dependent on the spatial, spectral and radiometric characteristics, availability, cost, technical image interpretation and mapping objective and climatic conditions (Xie et al., 2008). Due to *S. plumosum's* large patch sizes, affordable remote sensing data from sensors with moderate spatial resolution (Ground Sampling Distance of 2.0 m to 30 m) for instance are useful for its modelling.

Whereas several studies (Hudak and Brockett, 2004, Hudak and Wessman, 1998, Mutanga and Skidmore, 2004, Wessels et al., 2006, Adjorlolo and Mutanga, 2013) have investigated the value of image spectral characteristics for estimating vegetation distribution and density, identification of image characteristics such as spatial dependence that best correlate with invasive shrubs canopy cover or density have not been well-established. It is, therefore, essential to explore the techniques accounting for additional image characteristics in estimating canopy cover and density. Spatial dependence is the spatial relationship of variable values or locations. Spatial dependence is measured as the existence of statistical dependence in a collection of random variables (Anselin, 1995). Spatial autocorrelation is the measure of the degree of spatial dependence (Bannari et al., 2005). Spatial autocorrelation within remotely sensed imagery occurs in terms of the variable (i.e. *S. plumosum*) location represented as the pixel location and variable information as the reflectance value within that pixel (Wulder and Boots, 1998). The remote sensing data represents continuous landscapes in a form of regularly spaced grid showing positive spatial autocorrelation (Wulder and Boots, 1998). Remote sensing data is inherently spatially autocorrelated and *S. plumosum*, like other natural vegetation, also depict spatial interdependency (Mutanga and Rugege, 2006, Adjorlolo and Mutanga, 2008). Autocorrelation represents information which can be exploited as image characteristic and can be integrated with spectral information to enhance mapping and estimating canopy density (Wulder and Boots, 1998). To date, various image transformation techniques have been tested for the purpose of estimating vegetation distribution and canopy volume (Hudak and Brockett, 2004, Hudak and Wessman, 1998, Mutanga and Skidmore, 2004, Wessels et al., 2006, Adjorlolo and Mutanga, 2013). However, few studies have used the spectral-spatial approach using spectrally derived and spatially dependent statistics in landscape mapping. On the other hand, geostatistical techniques have been adopted for vegetation estimation (Adjorlolo and Mutanga, 2013), however, there is paucity in the literature on integration of spatial dependence and geostatistics for vegetation mapping. In this study, we integrate spectral data with spatial dependence to map and quantify *S. plumosum* invasion.

Spatial dependence and the neighbourhood context can be used to supplement the spectral information in land-cover characterisation to reduce the high intra-class variability (Ghimire et al., 2010). In this study, Getis statistics was adopted as it takes into account the spatial dependence of remotely sensed image pixels in clustering pixels that belong to a particular feature. This characteristic provides information in addition to the spectral reflectance information of earth features with remote sensing data (Wulder and Boots, 1998). The Getis statistics used to describe spatial information has a distinct advantage over conventional contextual texture-based classification approaches because, unlike the standard contextual methods that consider only values at a given neighbourhood of each pixel, the Getis statistics is a ratio of values of the neighbourhood for each pixel versus values of the entire image (Ghimire et al., 2010). The selection of the most suitable remotely sensed vegetation variable is critical for the reliability of vegetation density modelling. Hence this study incorporates the identification of the best performing Getis statistics index in mapping *S. plumosum* geographic extent and establishes statistical relationships between mapping and density estimation using geostatistical kriging and cokriging methods of interpolations.

Geostatistics was selected for modelling canopy cover/density because of the robustness of its interpolation kriging and cokriging techniques. These techniques enable prediction using multiple input variables and incorporate spatial autocorrelation into the prediction model (Eldeiry and Garcia, 2010). Spatial autocorrelation is the degree of dependence between values of the same environmental variable associated with a location close to each other. It arises when the value of an environmental variable recorded at a location on the earth surface is related to values of the same environmental variable at nearby locations (Bannari et al., 2005) Geostatistical techniques which take spatial autocorrelation of sparsely (e.g. canopy cover and density) and intensively sampled variables into consideration can be used to combine field and remote sensing data and model their interdependence simultaneously through cokriging. The cokriging technique was selected because it allows for the integration of secondary (2^{nd}, 3^{rd}, and 4^{th}) input information into the interpolation model and applies spatial autocorrelation during modelling. Therefore, integrating remotely sensed data with geostatistics can improve our understanding of the spatial dynamics of vegetation spatial distribution in heterogeneous natural environments (Adjorlolo and Mutanga, 2008). Whereas the approach has been applied to model herbaceous biomass distribution in the African savannah woodland (Mutanga and Rugege, 2006, Adjorlolo and Mutanga, 2013), to date, the technique has not been tested for mapping shrub cover and density distribution.

The modelling of ecological systems requires the application of methods for identifying statistical relationships between field and remotely sensed data. Generally, regression is applied to field and remotely sensed data for the spatial estimation of canopy vegetation variables.

However, ordinary regression methods do not make maximum use of field and remotely sensed data because they ignore the spatial dependence of the two datasets and do not account for the interdependence of the field and remotely sensed data (Mutanga and Rugege, 2006). Since ordinary regressions do not consider the spatial autocorrelation in the vegetation and its radiation (Atkinson et al., 1994, Mutanga and Rugege, 2006), the technique commonly underestimates or overestimates vegetation cover (Mutanga and Rugege, 2006). It is thus important for vegetation modelling to consider the fundamental practical principle that vegetation natural groupings depict spatial distribution and spatial interdependence, and the radiation of vegetation derived from remote sensing data are also spatially correlated both to themselves and to one another (Mutanga and Rugege, 2006, Adjorlolo and Mutanga, 2008). Variogram or semivariance modelling is used to establish a relationship between field and remote sensing data demonstrating the spatial dependence of the data for use in kriging and cokriging interpolation.

This study, therefore, investigates the utility Getis statistic transformed variables from the SPOT-6 multispectral image and geostatistical techniques for mapping *S. plumosum* spatial extents and predict density.

1.2 Aim and objectives of the study

The aim of this study was to investigate the Getis based image transformations applied to high-resolution SPOT 6 multispectral imagery and integration with geostatistical analysis for mapping and estimating *S. plumosum* canopy density and percentage. The study focused on the use of geostatistical techniques: kriging (i.e. ordinary kriging of sampled field data) and cokriging (i.e. interpolation of field data combined with Getis statistics transformed SPOT6 data) to estimate the density and percentage cover of *S. plumosum*.

1.3 Specific objectives

- To evaluate the performance of different Getis-statistic indices in mapping spatial distribution of *S. plumosum* in grasslands.
- To estimate *S. plumosum* canopy density and percentage cover by integrating field data with best performing Getis statistics transformed index through geostatististical

cokriging technique, and compare ordinary kriging with cokriging, and simple linear regression.

1.4 Key research questions

- Which Getis image transformation index performs better in mapping spatial distribution of invasive plant species *S. plumosum*?
- To what level can geostatistical technique, cokriging, improve mapping and estimate canopy cover density and percentage of *S. plumosum*?

1.5 Organisation of the thesis

This study is organised in two major sections. The first section deals with SPOT 6 image processing to create Getis variables. These layers serve as input variables for the mapping of *S. plumosum.* The analysis of field sampled data and identification of the best performing Getis index layer variable are also established in this section. The identification of the best performing Getis index layer variable is evaluated through statistical variable selection. The second section of the study integrates the best Getis statistic index layer variable (identified in section one) with field sampled data parameters (i.e. canopy density and percentage cover) in a geostatistical technique cokriging to estimate canopy density of *S. plumosum*. An interpolation of field data through kriging, linear regression of field data combined with SPOT 6 band ratios, image bands was also executed. The section also dwells on modelling *S. plumosum* canopy density using regression, ordinary kriging and cokriging in the entire study area. The analysis in this section is done in three ways: 1) ordinary kriging using field data parameters alone for quantitative prediction of *S. plumosum* canopy density and percentage cover. 2) cokriging with the *S. plumosum* plant field data parameters serving as input primary variables and the best Getis indices layer variable as input secondary independent variable, 3) linear regression of field and SPOT 6 data and band ratios not accounting for spatial dependence. In order to draw conclusions, the results obtained from cokriging are compared to those obtained from kriging and regression. The conceptual workflow of the study is illustrated in Figure 2.

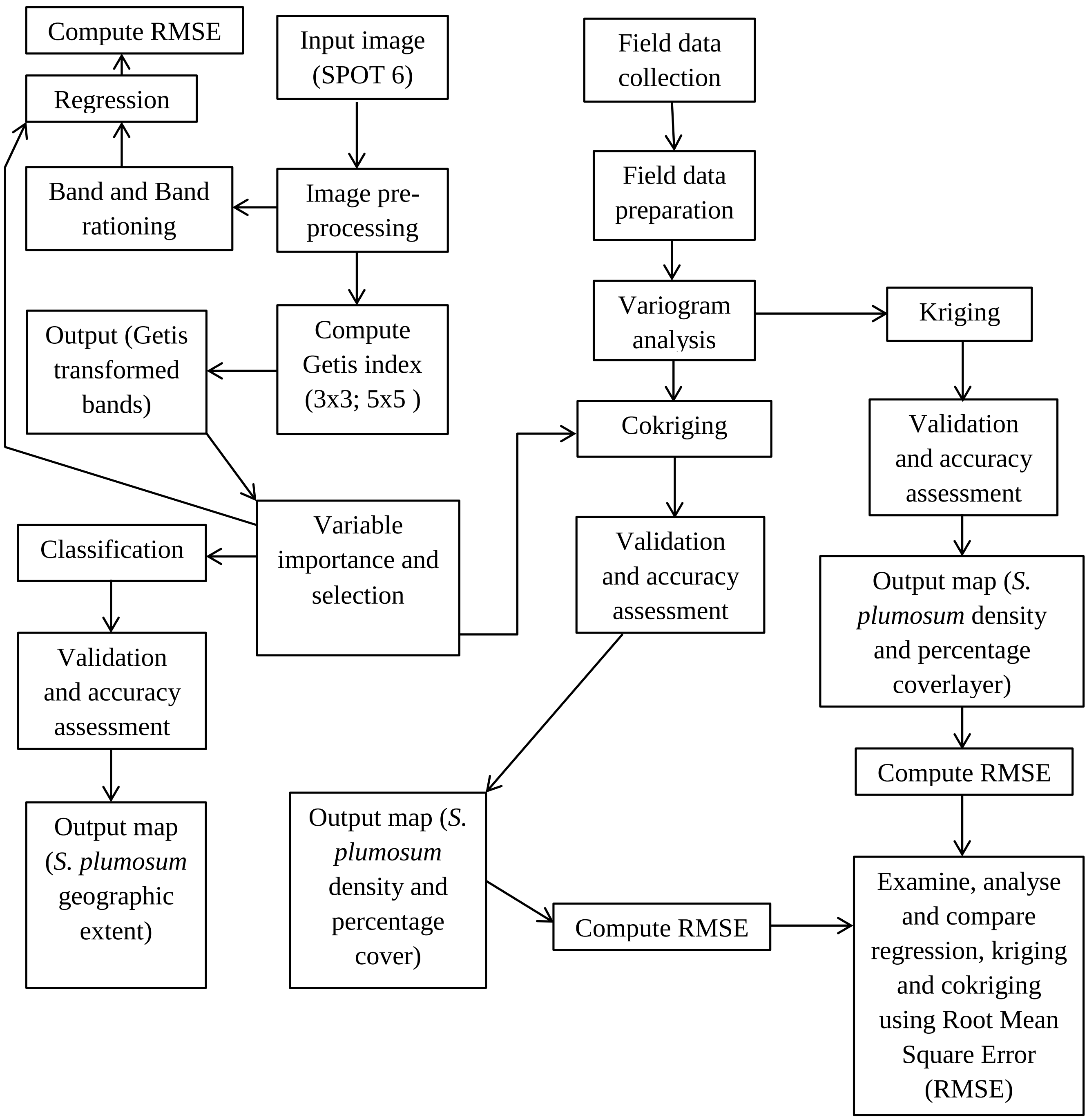


Figure 1: Flowchart for the methods followed and outputs in the study.

1.6 Study area

The study area is located in the Senekal area of the Free State province of South Africa. The area is located in the eastern parts of Free State province with heights exceeding 1000 meters above sea level. The area has rich soil and favourable climate, which allows for a thriving agricultural industry. Rainfall is received in summer with an annual rainfall of 477 mm and average daily temperatures range between 16 to 28 ^{o}C. The area consists of grasslands and shrub-tree natural ecosystems which are infested with *S. plumosum*. The landscape is characterised by both flat and mountainous surfaces and high mountains that dominate the eastern part of the study area (see Figure 3). The farms visited for the study lie east of Senekal town.

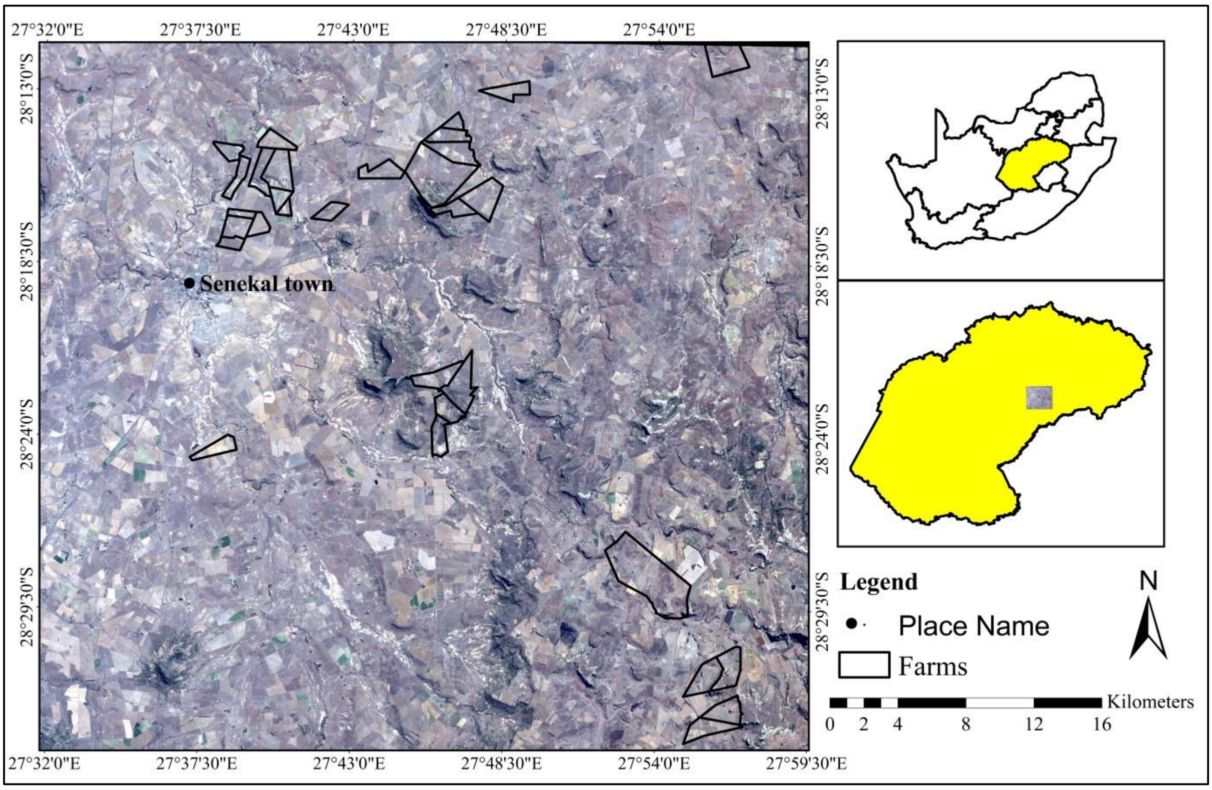


Figure 2: Study area demonstrating farms surveyed.

CHAPTER TWO

LITERATURE REVIEW

Plant species invasion into foreign environments has become a significant problem in South Africa. Among them is the *S. plumosum* which encroaches grasslands, turning them into less productive shrubby grasslands (Snyman, 2010). Jordaan (2009) notes that the encroachment of *S. plumosum* in the natural veld, grasslands and cultivated pastures is a serious problem in the North West, Mpumalanga, Gauteng, Free State and Eastern Cape provinces of South Africa. Consequently, it is essential to establish the geographical distributions of the species over large areas for eradication and management.

Remotely sensed imagery has increasingly become popular in vegetation mapping. Hence we utilised optical remotely sensed data in this study to investigate approaches for mapping and estimating *S. plumosum* cover and density. Remotely sensed data is inherently spatially auto-correlated; it is this spatial autocorrelation that we explore for the mapping of *S. plumosum*. Futhermore, we integrate the optical remote sensing data with field sampled physical canopy parameter data to estimate the quantity of in-situ canopy using geostatistical kriging and cokriging techniques. Therefore, this chapter reviews literature on three main approaches which are addressed in this study: 1) remote sensing of shrub vegetation and potential to map *S. plumosum*, 2) Image transformation techniques and Getis statistics, and 3) geostatistical techniques of interpolation.

2.1 Applications of optical remote sensing for mapping shrub invasion

Invasive plant species and their densities vary from field to field, hence the uniform application of invasive plant control measures over an entire field is neither cost effective nor environmentally friendly (Goel et al., 2002). Mapping invasive plants using ground surveys is often time consuming and labour intensive, however remotely sensed data offer a viable option for invasion mapping. Remotely sensed imagery covers a large geographical area and is not restricted to fence boundaries. Furthermore remotely sensed data offers access to areas which may otherwise not be accessible on the ground due to terrain (e.g. mountainous areas), authoritative (e.g. owner restrictions), political (e.g. differing political regimes) and safety (e.g. protected areas with wildlife) limitations. Although remotely sensed imagery has many advantages over traditional ground surveys, it is commonly limited to delivering information on vegetation structure in two dimensions (i.e. location and reflectance information) (Hantson et al., 2012). Furthermore, remote sensing mapping of specific plant species in forests, rangelands, natural landscapes and riparian areas has proved to be a challenge (Evangelista et al., 2009). For example, different invasive plants may have similar spectral reflectance to other vegetation or may be mixed with other vegetation within optical remote sensing data, resulting in incorrect classification and mapping (Shafii et al., 2004).

A number of studies have used remotely sensed data in vegetation mapping. Müllerová et al. (2013) for instance tested the effects of remotely sensed data resolution and image classification approaches for detection of *Heracleum mantegazzianum* (giant hogweed). In the study, high accuracies were achieved using high resolution data. Using high spatial resolution imagery and field data, Sanchez-Flores et al. (2008) used predictive skill of combined genetic algorithm ruleset-production and change analysis models to model plant invasion in dynamic desert landscapes. Their models identified areas vulnerable to *Brassica tournefortil* and *Schismus arabicus* invasion. The models representing most dynamic landscapes with high probability of invasion showed good spatial agreement with the distribution of invasions. Peters et al. (1992) identified infestations of broom snakeweed (*Gutierrezia sarothrae*) using advanced very high resolution radiometer (AVHRR) and normalised difference vegetation index (NDVI) data while Evangelista et al. (2009) used Landsat 7 ETM+ to map the invasive species tamarisk (*Tamarix*). Using varied approaches, studies above provide an indication of the potential of remotely sensed data in invasive species mapping.

2.2 Potentials of mapping *S. plumosum* using remote sensing

S. plumosum is a woody dwarf shrub that thrives during summer rainfall of approximately 620-750 mm (Snyman, 2009), thus dominant in mesic and semiarid grasslands. Grasslands are a major component of the natural vegetation, with the biome comprising about 295 233 km^2 of the central regions of South Africa and extending into most of the adjoining biomes such as forest, savannah, thicket, Nama-karoo. *S. plumosum* commonly occur in vast patches. It has previously been mapped through traditional field based survey methods which are limited to small scale, easily accessible areas. These methods require locality GPS coordinates, collecting encroachment information detailing extent and density of encroachment (Avenant, 2015). Remote sensing, conversely, provides an opportunity for mapping encroacher shrubs over large area.

Vegetation mapping using remotely sensed data is a process of extracting vegetation information by interpreting satellite image using features like image texture, colour, tone, pattern and association information (Xie et al., 2008). Successful delineation of vegetation components is dependent on the differences in reflectance properties (within the electromagnetic spectrum) of each vegetation species and their phonological differences at the time of image acquisition (Mirik and Ansley, 2012).

Furthermore, remote sensing sensors commonly used for vegetation mapping are those equipped with the green band (wavelength 530 nanometres (nm) – 590 nm), near-infrared band (wavelength 760 nm – 890 nm), red band (wavelength 625 nm – 695 nm) (Mutanga and Skidmore, 2004, Cho et al., 2013). Additionally, the general rule of thumb for mapping vegetation using remote sensing data is that the object mapped must at least be twice or three times the size of the pixel of used remote sensing data.

Whereas the adoption of remotely sensed data in vegetation assessment has been hugely successful, literature on canopy mapping and estimation, particularly in areas characterised by multiple vegetation structures remain limited. Consequently, a number of studies have adopted a number of approaches that include spectral assessments, rationing and texture techniques in vegetation mapping and modelling (Yang and Prince, 1997, Hudak and Brockett, 2004, Adjorlolo and Mutanga, 2008). These techniques are discussed in detail in section 2.3.

2.3 Image transformation techniques used in remote sensing of vegetation

Whereas remotely sensed data can be efficiently used to assess vegetation cover, it is characterised by inherent error as it relies on the regression of spectral responses of vegetation signal between ground data and the measured vegetation radiation (Adjorlolo and Mutanga, 2008). Remote sensing data is also subject to radiation scatter which influence spectral information of features recorded by sensors. Generally, methods for land cover characterisation using medium to high spatial resolution are well established, but the size of pixels is often larger than the land cover feature of interest, resulting in class mixing within pixels (Blaschke et al., 2004). Several studies have applied spectral assessments, rationing (vegetation indices) and texture techniques for vegetation mapping. Adjorlolo and Mutanga (2013), for instance, assessed vegetation indices, texture properties and geostatistics to quantify woody cover. Using probability occurrence of indicator values, Roelofsen et al. (2014) used remotely sensed vegetation characteristics from the visible near-infrared and shortwave infrared to map vegetation. Gil et al. (2013) used the very high resolution remote sensing IKONOS multispectral imagery to map the invasive woody plants through object based image techniques while Wang et al. (2015) applied a combination of spectral, spatial, texture and local statistical analysis Getis statistic to IKONOS multispectral imagery to map forest health. Therefore, integrating spatial information in vegetation mapping models could provide additional information for the discrimination of vegetation. These studies demonstrate the work done for the utilisation of remote sensing techniques in vegetation mapping. This study aims to integrate spatial dependence techniques (i.e. Getis statistics) with geostatistics to map and predict density and percentage cover of woody shrub *S. plumosum*, a methodology combination that have not been tested for woody shrub mapping. Bannari et al. (2005) in their study to determine the potential of Getis statistics to characterise radiometric uniformity and stability of test sites, found Getis statistics to provide an excellent spatial analysis for calibration. Getis showed to have good potential for the extraction of radiometric heterogeneities for surfaces. This potential application of extracting radiometric heterogeneity present advantages in image spatial analysis for all landcover, including vegetation.

2.4 SPOT 6 data derived Getis statistics

Remotely sensed imagery consists of digital numbers presented as pixels which represent earth's surface. This data is highly spatially auto-correlated. The characterisation and integration of this spatial autocorrelation into mapping models can provide valuable information for applied and theoretical remote sensing (Wulder and Boots, 1998). As a result, various techniques of local spatial association have been developed. These techniques focus on variations within regions of spatial dependence, known as Getis statistics. Spatial dependence and the neighbourhood context can be used to supplement spectral information in land-cover characterisation to reduce the high intra-class variability (Ghimire et al., 2010). It can also describe the local spatial structure and variability of land cover categories, hence can be used to improve mapping accuracies in heterogeneous landscapes. In a remotely sensed image, Getis statistics determines spatial dependence for each pixel while also indicating the relative magnitude of the digital numbers/reflectance in the neighbourhood of the pixel (Wulder and Boots, 2001). Typically, the Getis statistics measures the degree of association that results from the concentration of weighted points and all other weighted points included within a distance from the original weighted point and can be applied at different levels of spatial scales. On remotely sensed imagery, three types of relationships can be detected; 1) structural trend,

2) correlated variation, and 3) uncorrelated variation or noise. Although inclusion of spatial dependence in classification can be used to increase land cover classification accuracies, most existing classification approaches do not include spatial dependence to neighbouring pixels.

Spatial autocorrelation is the degree of dependence between values of the same variable associated with a location close to each other. Spatial autocorrelation arises when the value of a variable (*S. plumosum*) recorded at a location on the earth surface is related to values of the same variable at nearby locations (Wulder and Boots, 1998). In image processing, the locations are the pixel coordinates and the attributed data are the reflectance values within the image. It can be expected that pixels from similar land covers will generate clusters in image feature space that differ from other land cover types (Bannari et al., 2005). This clustering will translate into a positive spatial autocorrelation when there are similar reflectance values and a negative autocorrelation when there is a cluster of dissimilar values. Spatial autocorrelation can be measured by using global or local statistics, one of these being Getis statistics. Getis and Ord (1992) have demonstrated the potential of these statistics to identify significant spatial dependency in remotely sensed imagery. The Getis statistics describing spatial information has a distinct advantage over conventional contextual texture-based classification approaches because, unlike the standard contextual methods that consider only values at a given neighbourhood of each pixel, the Getis statistics is a ratio of values of the neighbourhood for each pixel versus values of the entire image resulting in a new layer. The measurement of spatial autocorrelation thus involves the simultaneous consideration of both the locational and attributes information (Wulder and Boots, 1998). The processing of image scenes allows for the use of the statistic in a more exploratory manner by providing each pixel with a spatial dependency value based upon processing pixel with a series of windows (3x3, 5x5), a new layer of information is derived (Wulder and Boots, 2001). This new layer of information is inherently spatially dependent and expected to give a proper examination of properties of local spatial dependence and provide insights into spatial autocorrelation characteristics of image data not revealed by traditional image analysis (Wulder and Boots, 2001). This means landcover types such as *S. plumosum* with similar image attributes will be grouped together and be easier to discriminate from other landcover types.

This study focuses on the evaluation of the Getis statistics for mapping *S. plumosum* and identifying the best Getis indices to achieve the highest accuracies. The Getis statistics is applicable in the study of *S. plumosum* mapping because the *S. plumosum* canopy grows in patches, and its weights in one spatial location are expected to be similar to those in nearby locations.

2.5 Getis statistics and classification techniques

Imagery covering large geographic areas with high temporal resolution offers an opportunity for mapping surfaces through image interpretation and classification. Image classification is a process of grouping pixels of similar values into meaningful categories (Abburu and Golla, 2015). To date, a variety of pixel based classification techniques categorised as supervised (i.e. Maximum likelihood, artificial neural network, support vector machine, random forest, and decision tree), unsupervised (i.e. k-means and ISODATA) and hybrid (semi-supervised, of supervised and unsupervised learning) have been developed (Alajlan et al., 2012). These classification techniques have limitations when applied to imagery with heterogeneous landscapes as the size of objects may be smaller than the pixel size, and the pixel may contain information about a mixture of land-cover types. Additionally, land-use and land-cover types are not effectively separated using spectral information, resulting in less accuracy (Li et al., 2014). Limitations of supervised classification include its unsuitability for big data as it is time-consuming and requires area expert knowledge. Limitation of unsupervised classification is that no training data can be incorporated. The supervised and unsupervised classification techniques utilize spectral variables and generally ignore spatial information which is inherent in real-world remote sensing (Li et al., 2014). This is problematic as with higher spatial resolutions, images are likely to have higher within-class spectral variability (Li et al., 2014), resulting in less than satisfactory results reached with spectral classifiers (Myint et al., 2011).

With the development of very high resolution remote sensing data, object base image analysis (OBIA) classification methods that offer new classification abilities have been developed. These object based methods group pixels with homogeneous properties into objects which are considered as a basic unit for analysis (Nussbaum and Menz, 2008). Object-based image classification techniques incorporate spectral and spatio-contextual information in the classification process and are considered superior when compared to traditional pixel-based techniques (Blaschke, 2010). This spatio-contextual information is incorporated in the image segmentation process grouping similar pixels into objects (i.e. *S. plumosum* patches). This method of classification was selected for classifying the SPOT 6 Getis transformed spatially dependent data. Recent studies (Su et al., 2008, Blaschke, 2010, Gianinetto et al., 2014) have achieved highly accurate classification results when applying OBIA to high-spatial-resolution for land use land cover mapping.

2.6 Geostatistics and remote sensing of vegetation

Geostatistics is the term applied to a group of spatial statistical techniques which describes the correlation of spatial data by exploration, modelling and surface generation of local variables and their estimation at un-sampled locations (Curran and Atkinson, 1998). According to Hengl (2007), geostatistics can be described as a collection of numerical techniques that are used for the characterisation of spatial attributes employing random models that may include spatial interpolation. Geostatistical techniques are centred on the regionalised variable theory which states that interpolation from points in space should not be based on a smooth continuous object, it should be based on a stochastic model that takes into consideration the various trends in the original set of points (Hengl, 2007). This approach offers a way of describing the spatial continuity of natural phenomenon and provides adaptations of classical regression techniques to take advantage of this continuity (Hengl, 2007). Studies on invasive plants and ecology assessment have investigated the applicability of integrating remote sensing and geostatistical techniques for spatial estimation of vegetation resources (Mutanga and Rugege, 2006, Adjorlolo and Mutanga, 2013).

The combination of remote sensing with geostatistics can improve our understanding of spatial dynamics of vegetation spatial distribution in heterogeneous natural environments (Adjorlolo and Mutanga, 2013). Optical remote sensing data application studies on vegetation need to take advantage of the spatial factors of vegetation density and distribution when making quantitative estimates of vegetation cover (Adjorlolo and Mutanga, 2008). It is important to consider the spatial aspects because the assessment of sample data from the patches of vegetation is spatially dependent on natural groupings of the vegetation species. *S. plumosum* for instance is assumed to have a spatial dependent characteristic.

Spatial interpolation is the process of using points with known values to estimate values at other points (Hengl, 2007). Spatial interpolation is, therefore, a means of creating continuous surface data from sample points so that the surface data from sample points can be used for analysis and modelling (Hengl, 2007). Fortunately, remote sensing allows for the integration of empirical and physical methods in which statistical methods of interpolation can combine limited field data with remote sensing data to estimate and map vegetation quantity (Ferwerda and Skidmore, 2007, Liang, 2005).

In geostatiscs, kriging techniques can spatially provide quantitative measures in estimating *S. plumosum* cover based on the regionalized variable theory. In this study, we use kriging and cokriging to interpolate the availability of *S. plumosum* in areas with no data including farms which we did not get authorization to survey. These interpolations are based on the assumption that *S. plumosum* is available everywhere within the study area and only attempts to estimate the density quantity. In this subsection literature on geostatistical kriging and cokriging are reviewed.

2.6.1 Kriging technique and application to *S. plumosum* mapping

Kriging is a technique for estimating the value of a regional variable from adjacent variable values while considering the dependence expressed in the variogram (Webster and Oliver, 2007). Kriging assumes that spatial variations of natural variables are spatially correlated, regionalised and represent a trend with an inherent error. Kriging differs from other conventional interpolation methods (e.g. Trend surface models, linear regression, Thiessen polygons, Inverse Distance Weighting) in that it can assess the quality of prediction and estimate prediction errors. For instance, the conventional approach to spatial predictions combines classical estimation with spatial information to overcome their weaknesses. *S. plumosum* is natural vegetation with its growth dependent on random natural processes making its occurrence random and spatially autocorrelated within landscapes. This motivated for the use of kriging and co-kriging techniques of interpolation. Kriging provides a solution to the problem of estimation based on a continuous model of stochastic spatial variation. It makes the best use of existing knowledge by accounting for the way that a property varies in space through the variogram model. There are linear and non-linear kriging techniques. Linear kriging estimates are weighted linear combinations of the data while non-linear kriging estimates whether or not variable estimates exceed or are below a particular set threshold (Webster and Oliver, 2007). In this study, we explored linear kriging estimates. The weights allocated to the sample data within the neighbourhood of the points or block to be estimated are in such a way as to minimise the estimation or kriging variance, and the estimates thus become unbiased. A semi-variogram is applied to measure spatially correlated components for estimation using kriging models (Webster and Oliver, 2007). Whereas there are many types of kriging methods

(Webster and Oliver, 2007), we chose ordinary kriging for this study due to its popularity. In this study, kriging was used to estimate the value of a random variable, *S. plumosum* at unsampled locations from sample field data and SPOT 6 Getis statistics transformed data.

Kriging has been applied in several studies to estimate vegetation. Adjorlolo and Mutanga (2013) for instance used kriging and co-kriging to estimate woody tree cover. Dwyer (2011) applied cokriging to estimate herbaceous biomass in savannah environments. Miller and Franklin (2002) used indicator kriging to model the distribution of vegetation alliances. While Valley et al. (2005) evaluated interpolation techniques for mapping the distribution of vegetation. Valley et al. (2005) found kriging to be interpolating vegetation better than ordinary interpolation methods which do not account for spatial dependence such as IDW and spline. Miller et al. (2007) also investigated the incorporation of spatial dependence in predictive vegetation models in which spatial clustering and geostatistics were applied and results demonstrated improvement in predictions.

2.6.2 Co-kriging techniques and application to *S. plumosum* mapping

Cokriging is an extension of kriging, it takes advantage of correlation that may exist between the variable of interest and other more easily measured variables. It allows for the addition of secondary variable information to assist in improving the model for estimating values at unsampled locations. Many studies (Odeh et al., 1995, Ver Hoef and Barry, 1998, Pan et al., 1993, Laurenceau and Sagaut, 2008) showed the superiority of cokriging to ordinary kriging. To ensure the validity of the estimates made by kriging and cokriging, the semivariogram and cross-semivariogram of the variables must accurately describe the spatial structures. Regression kriging involves spatially interpolating the residuals from a non-spatial model using kriging and adding the results to the prediction obtained from the non-spatial model. Cokriging equations for estimating a primary variable from a set of variables are extensions of those for kriging. Cokriging works best where the primary variable of interest is less densely sampled than the others (Eldeiry and Garcia, 2010).

2.7 Accuracy assessment

Any map is a model or generalized representation of reality and it contains some level of error (Brown et al., 1999, Dicks and Lo, 1990, Smits et al., 1999). Generally, thematic maps provide a simplification of reality and therefore often have flaws (Woodcock and Gopal, 2000). It is important that the quality of thematic maps derived from remotely sensed data be assessed and expressed in a meaningful way. This is important in providing a guide to the quality of a map and its relevance to a particular purpose, and also in understanding the error and its likely implications on map analyses (Arbia et al., 1998, Janssen and Vanderwel, 1994). Generally, classification accuracy assessment is widely accepted as a fundamental component of thematic mapping investigations (Cihlar, 2000, Cohen and Justice, 1999, Justice et al., 2000, Merchant et al., 1993).

In remote sensing thematic mapping, the term accuracy is used to express the degree of 'correctness' of a map or classification and typically means the degree to which the derived image classification agrees with reality (Janssen and Vanderwel, 1994, MALING, 1989, Smits et al., 1999). Accuracy is a difficult property to determine as it comprises bias and precision (Campbell, 1996, MALING, 1989). Commonly, a classification error is regarded as some discrepancy between the situation depicted on the thematic map and reality (Foody, 2002). In this research, accuracy assessment is used to understand the errors of thematic maps generated from classification and geostatistical regression.

2.8 Lessons learnt from literature review

This section reviewed the literature on the use of remote sensing, Getis statistics and geostatistics for mapping of invasive woody shrub *S. plumosum*. It has been outlined in literature that challenges in mapping *S. plumosum* can be resolved using remote sensing and field data. Remote sensing has been identified as an efficient method for obtaining spatial information about the location, extent and distribution of *S. plumosum*. As the remote sensing data is reliant on regression of spectral information, some misclassifications accrue from the application of traditional methods. This motivates for an exploration of other techniques in remote sensing data transformation for vegetation analysis.

Although several studies (Mutanga and Rugege, 2006, Adjorlolo and Mutanga, 2013, Miller et al., 2007, Valley et al., 2005) have demonstrated that vegetation and its radiation are spatially related and the spatial characteristics of vegetation traits can be estimated from its spectral reflectance properties, regression techniques e.g., (Gong et al., 2003, Wessels et al., 2006) have been the typical method used in evaluating the relationship between spectral data and vegetation parameters. Generally, whereas geostatistical techniques and spectral data contain information that captures spatial autocorrelation, which is significant in improving the accuracy of spatial estimation of vegetation resources, their potential remain largely unexplored. As a result, this study investigates the use of geostatistical techniques (kriging and co-kriging) to estimate the density and spatial distribution of *S. plumosum* cover.

CHAPTER THREE

MATERIALS AND METHODS

The methods section provides the theoretical considerations integrated into the techniques used for the study. The fundamental principle is that vegetation and its radiation at any given time or season are correlated (Adjorlolo and Mutanga, 2008). Therefore it is important that the time for field data collection must coincide with the dates in which images were acquired with minimum time lapses. Also taking into account that land-use changes rapidly in the farms surveyed, for example, when in the field, it was realised that some of the areas have been newly cleared of vegetation and others newly cultivated. This poses a threat that field data can be collected which indicate there is *S. plumosum* infestation in a particular location, however, by the time the satellite acquire an image the area might have been cleared off *S. plumosum.* It was therefore very desirable to have field data collection date to be as close as possible to the SPOT6 image acquisition date.

3.1 Field sample plot survey and structural data collection

Remote sensing and geographic information systems (GIS) applications for vegetation studies require that pre-fieldwork be carried out, the pre-fieldwork for this study consisted of GIS operations on auxiliary datasets. This involved analysis of shapefiles consisting of farm boundaries layers and 2014 SPOT5 satellite image were assessed. To ensure the best accuracy and avoid any spatial or geometric distortions in the GIS data, all shapefiles and imagery were geo-referenced. Assessment of appropriate sampling design (e.g. optimum sample size, plot size and the stratified random locations of the sample plots) before commencing fieldwork was also done. This assisted in addressing any foreseen sampling difficulties and established an unbiased fieldwork sampling criteria.

Field data for the study was collected from 23rd to 27th February 2015 and was used to process remotely sensed Getis transformed variables derived from SPOT6 multispectral data, acquired on 13th March 2015. Thirty-one farms previously surveyed (formally and informally) in Senekal indicated large *S. plumosum* infestation.

Since the *S. plumosum* is a small shrub (60 cm by 60 cm) which occurs in patches, plots of 20x20 meters (m) were used in the field. This plot size was selected because it can accommodate a minimum of 3x3 pixels of the SPOT6 multispectral image, essential for identifying a land-cover feature. Furthermore, observed field patches were generally larger than 20 metres in diameter. The size of the sampling plots conformed to the general guideline that the spatial scale of an object on the ground must be at least twice the spatial resolution of the remote sensing sensor (Cowen and Consortium, 2002). All *S. plumosum* canopy completely inside the plots were identified, counted and structurally assessed. Since a square plot has greater perimeter compared to circular plots, the chances that maximum canopy cover can fall completely within the plot was likely. Additionally, the square plots coincide with the square shaped remotely sensed image pixels, which enhance computational efficiency. The 20x20m plot was considered large enough to represent the surrounding *S. plumosum* plant properties, as well as optimum to retrieve spatial information contained in the SPOT6 6m multispectral spatial resolution images. The overall distribution of the sample plots is presented in Figure 3. Because the SPOT6 image coverage was over the entire study area, the selected image for cokriging was sampled at the predetermined lag size (2429 m) using a semivariogram model parameter (Detailed explanation on this is presented in the geostatistical analysis section of this chapter). The *S. plumosum* variables information on canopy diameter, canopy height, the number of canopy per 20x20m plot, the distance between canopies, and percentage observation were measured to evaluate how they relate to the remotely sensed SPOT6 derived Getis indices.

Georeferenced field data was collected at stratified randomly selected locations. This approach was followed to ensure spatial consistency. However, because the distance between plot sites in farms are large (Figure 3), they do not represent the range of spatial influence in the surrounding vegetation community, thus spatial interpolation of the field data alone will not provide reliable spatial *S. plumosum* density maps. The plot locations were loaded into a GPS and used to navigate to plot locations. Measurements are outlined in Table 1. Due to the survey team's inability to access some of the farms within the study area, high resolution data captured on a date closer to the field work date was utilised to collect more sample plot data to supplement field data. These sample plots filled the large gap distance among the sparse farms surveyed. Figure 3 demonstrate the distribution of the plots collected.

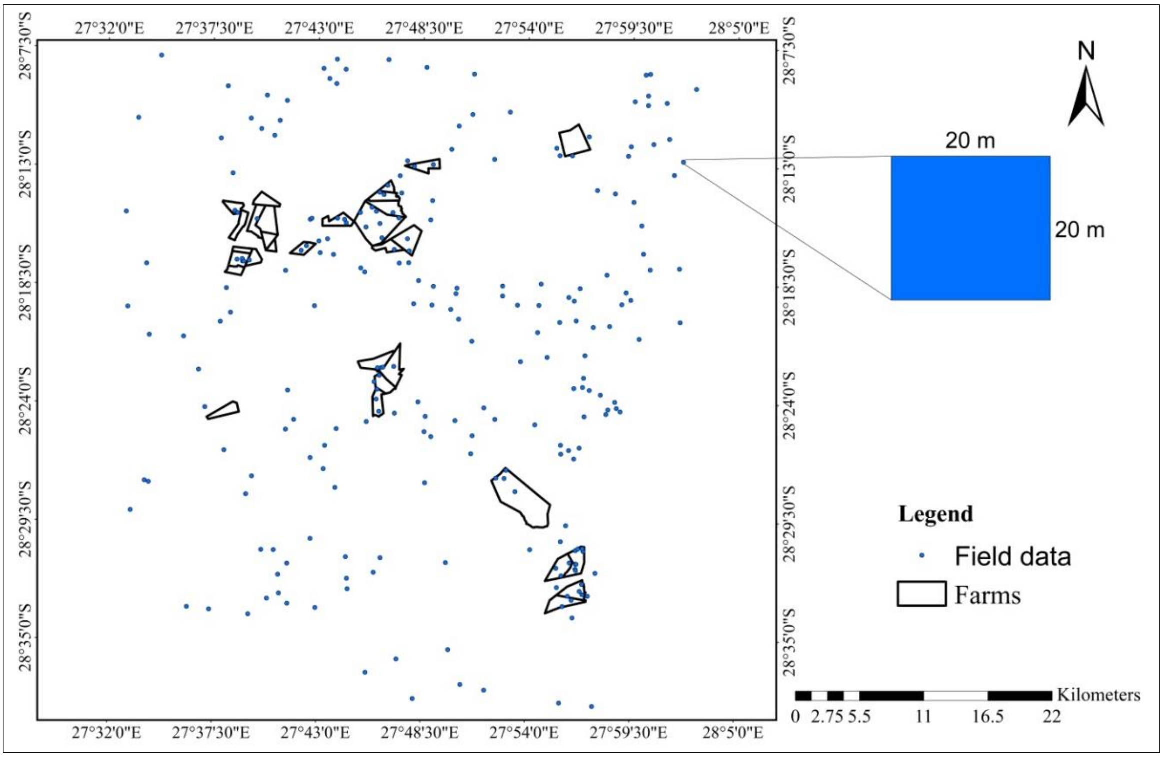


Figure 3: Surveyed and high resolution image *S. plumosum* plots.

The identification of *S. plumosum* species was done with assistance from local farmers. The data sheet (Appendix 1), was used to record *S. plumosum* canopy structural parameters and the total number of canopy per plot. The field work focused on *S. plumosum* canopy variables considered for spatial estimation of canopy cover and density mapping. But because canopy density and cover variables relate to other canopy parameters namely: canopy diameter, canopy height, and spacing, these variables were also measured (see Table 1).

Table 1: Plot canopy measurements and observations

Number.	Measurements of *S. plumosum* canopy per plot
1.	Count number of canopy per plot and estimate percentage cover
2.	Measure length, height and width of the canopy
3.	Measure the canopy spacing
4.	Take GPS coordinate at the centre of the plot
5.	Observe and record soil colour and soil type
6.	Observe and record surrounding land features and take picture of the plot

3.2. Data processing and analysis

3.2.1 Field data processing

Field variables (Table 1) measured at each sample plots were processed according to the study objectives. The field data comprised of 241 records of stratified random 20x20 meter plots, and additional point data records. The plot data was collected in the farms which the survey team had authorisation from landowners.

3.2.2. **SPOT 6 multispectral image processing**

- **SPOT 6 image pre-processing**

The information extracted through the remote sensing of plant patches is dependent on the spatial resolution of the data collected (Wulder and Boots, 1998). When the objects of interest are composed of a number of pixels, the imagery is considered to be high resolution, when multiple objects compose one pixel the imagery is considered a low resolution (Woodcock and Strahler, 1987). In the case of the *S. plumosum*, SPOT6 six meter spatial resolution was a low spatial resolution for individual 0.6x0.6 m canopy, however, the SPOT6 image was high resolution for the sensing of *S. plumosum* plant patches. In addition, SPOT 6 data was easily accessible to the author, covers large areas, satisfy cost, temporal, spatial and spectral requirements for use in the study. When considering imagery in terms of plant canopy patches, neighbouring image pixels may be expected to display some degree of spatial dependence which can serve as a source of information or a form of noise and error which must be accounted for when classical statistical analyses of dependence are applied to the data. This motivated for the use of SPOT 6 multispectral data as the data resolution is ideal to minimise noise and also provide high resolution information about the canopy patches of *S. plumosum.* The data consisted of nine (9) SPOT 6 scenes mosaicked together to cover an area of 27639 hectares. This mosaicked data was georeferenced and radiometrically corrected to top-of-atmospheric reflectance. Specifications of the remote sensing data are tabulated in Table 2.

Table 2: Specifications of the SPOT 6 multispectral imagery

Band	Wavelength range (um)	Spectral location	Spatial resolution (m)	Characteristics
1	0.455 - 0.525	Blue	6	Water penetration, differentiates between vegetated and non-vegetated
2	0.530 – 0.590	Green	6	Visible vegetation features
3	0.625 – 0.695	Red	6	Discerning between soil and vegetation
4	0.760 – 0.890	Near infrared	6	Discrimination of differing vegetation and varieties and conditions

- **SPOT 6 image Getis statistics transformations**

The mosaicked and radiometrically corrected SPOT6 multispectral image was used to compute Getis statistics indices. These indices are seven in total and are discussed in sub-section 3.3.2. Getis's local spatial statistics look for specific areas in an image that have clusters of similar or dissimilar values (Getis and Ord, 1992). The statistics output image layers for each index calculated each image layer contains a measure of autocorrelation around that pixel. This is useful for determining clusters of similar values, where concentrations of high values result in a high Gi value and concentrations of low values result in a low Gi value. This image processing was carried out using a 3x3 and 5x5 moving windows which generated twelve and twenty layers respectively per each Getis index. The moving window indicates that spatial dependency is confined to a very localised region while large distance indicates more spatially extensive spatial dependence (Bannari et al., 2005). A total of 308 layers were computed. From this total, 140 layers were computed through a 5x5 moving window and 84 from 3x3 moving window.

3.3. Getis statistics methods

3.3.1. Local indicator of spatial association

Local indicators of spatial association (LISA) Getis (Gi*) statistics measurements evaluate the extent and nature of concentration in the values of a particular variable in a local region within the imaged area. The Getis statistics achieve this by expressing the sum of the weighted variate values within a specified distance of a particular observation as a proportion of the sum of the variate values for the entire study area. The distance is called the “lag” which is defined by the ground sampling distance and quantity in a remote sensing data. This value can be compared with the statistics expected value under a hypothesis of no local spatial autocorrelation to indicate if the degree of clustering of variable values in the vicinity of distance is greater or less than chance would dictate (Anselin, 1995).

3.3.2. Getis image transformations

The Getis statistic is a local indicator of spatial dependence describing local variability in spatial dependence (Getis and Ord, 1992). The Getis (Gi) index identifies hot spots which are areas of very high or very low values that occur near one another. This is used in the clustering of similar values within an image where concentrations of high values result in a high Gi value and concentrations of low values result in a low Gi value. The Getis statistics is used to identify clusters of high values clusters of *S. plumosum* and low values which represent other land-cover types.

The improved approach to incorporate spatial dependence in land-cover classification is through the Getis statistics. This statistic behaves similarly to a moving filter in a remote sensing context by considering values within a local neighbourhood of the focus pixel. This helps to remove labelling errors caused by noisy data and or complex spectral measurements while simultaneously accommodating the values in the entire image reflecting global landscape heterogeneity characteristics (Wulder and Boots, 1998).

Conceptually, Getis behaves similar to a moving filter window and can be computed for each pixel of a given spectral band as the ratio of the sum of radiometric values in the image. The different moving windows are used to investigate the influence of different spatial scales on the classification accuracy. A window comprises of a specified block of pixel with a lag, for example, a lag3 window comprises 3x3 pixels. The application of Getis statistics result in the creation of a new image band representing the spatial structure of a given spectral image (Wulder and Boots, 1998). The Getis statistic provides a measure of spatial dependence for each pixel while also indicating the relative magnitudes of the reflectance value in the neighbouring pixels.

There are two versions of the Getis statistic, Gi and Gi* (Wulder and Boots, 2001). In the Gi, the value of the variable at a specified distance is excluded from the local sum, while in the Gi* it is included. Gi* is more appropriate for remote sensing applications as it allows for the computation of the statistics with a window of user-defined dimensions (3x3, 5x5,). The statistic *Gi*(d)* for variable *x* at specified distance *d* is defined as:

$$G_i^*(d) = \frac{\sum_j W_{ij}(d)x_j - W_i^*\bar{x}}{s[(w_i^*(n-w_i^*))/(n-1)]^{1/2}} \qquad \textit{equation 1}$$

Where $w_{ij}(d)$ is a spatial weights matrix with ones assigned to all locations within distance *d* of observation *i*, including at *i* (i.e. w_{ij}=1), and zero otherwise. In remote sensing data, a window can be specified around an observation *i* by specifying the value of *d*.

High Gi* values denote a cluster of high reflectance values, and low Gi* denote a cluster of low reflectance. It can be expected that pixels from similar land covers will generate clusters in image features space that differ in intensity compared to pixel clusters from other land-cover types, this clustering translates into positive spatial autocorrelation when we have a cluster of similar reflectance and a negative autocorrelation when we have a cluster of dissimilar values. Computing Gi* within a series of increasing windows and noting the distance at which the largest absolute Gi* values occur allows for an assessment of the size of the region of association around an individual pixel. A small window size (distance) indicates that spatial dependency is maximised within a localised region while a large distance value indicates more spatially extensive spatial dependence. A weakness of the Gi* statistic is that it cannot be used to identify clustering of medium values since mid-range values of Gi* (i.e values around zero) can result from either this situation or an absence of clustering of similar variate values.

There are several neighbourhood rules (also known as indices) that must be adhered to, these rules define which adjacent pixels to compare to the central pixel, the rules are:

- Rook’s Case (default): Selects the pixels on the top, bottom, left, and right.
- Bishop’s Case: Selects four diagonal neighbouring pixels.
- Queen’s Case: Selects all eight neighbouring pixels.
- Horizontal: Selects two neighbouring pixels in the same row.
- Vertical: Selects two neighbouring pixels in the same column.
- Positive Slope: Selects two neighbouring pixels in opposite corners in a positive diagonal.
- Negative Slope: Selects two neighbouring pixels in opposite corners in a negative diagonal. (www.exelisvis.com/docs/LocalSpatialStatisticcs.html)

The spatially auto-correlated SPOT6 Getis index layers created through Getis statistics are compared through a variable selection process. The SPOT6 Getis index layer variable that cluster areas of *S. plumosum* and discriminate these areas more accurately than other layers are classified according to their importance to mapping model. The Getis transformed layer variables with the highest importance are used to map *S. plumosum* and further integrated into geostatistical technique cokriging as input data for density estimation.

3.4 Variable selection analysis (Assessment and comparison of the Getis statistics indices layers)

Given the large volume of data obtained from the Getis statistics transformation process, it is impractical to use all the data for mapping *S. plumosum*. This section aims to identify Getis transformed layer variables optimal in clustering pixels representing *S. plumosum* from the SPOT6 image. Furthermore, it is vital to select variables that are significant for accurate classification of the *S. plumosum*. To carry out this task, we employed the variable selection technique.

Variable selection provides for the selection of an optimal set of predictors (i.e. SPOT 6 Getis transformed layers) essential for the inclusion in a model for mapping *S. plumosum*. Variable selection selected the best subset of predictors (i.e. Getis transformed indices layers) for modelling and mapping *S. plumosum* based on variable importance model. The variable importance model used was the “Recursive feature elimination” in which each Getis variable layer was ranked using its importance to the model of clustering *S. plumosum*. The variable importance used the second order model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_n x_n \qquad \textit{equation 1}$$

Getis transformed layer variables were selected and discarded based on their overall importance to the model indicating likelihood that a particular Getis transformed layer can cluster and discriminate *S. plumosum* from other vegetation. The Getis transformed layer variables with highest R-squared value demonstrated high clustering probability. The term "probability" refers to the likelihood that "mean layer values" of Getis transformed layer variables clustered pixel values represent *S. plumosum*. This was done with consideration of possible overfitting, i.e. a model adjusts to specific random features or noise of the training data but works poorly on other datasets (Packalén et al., 2012).

The Getis statistics transformed layer variables needed to be evaluated for their performance in clustering pixels. This is a criterion used to define the level to which an index layer is considered to be having a high probability in clustering pixels of *S. plumosum*. Field collected plot data was used in this selection of the most significant Getis index layer variable. Because the field data is applied at the plot level, remote sensing image's spatially dependent pixel values were averaged to one value per plot. The field data was split into 70% training data and 30% test data. The concept of the split criterion gives more weights to the training dataset which provides a reliable model building in vegetation studies (Kokaly and Clark, 1999). The training dataset was input into the variable selection model to train the model in selecting the Getis index layer with the best *S. plumosum* pixel clusters. The variable selection was executed in a binary environmental measurement, i.e. presence or absence of *S. plumosum* where a variable containing 1(presence) or 0 (absence) was integrated into the Getis transformed layer selection. The variable selection was executed in R environment using the caret packages. The results of this assessment are demonstrated in chapter 4.

3.5 Classification methods

The image transformation techniques applied to SPOT6 multispectral imagery in this study resulted in image layers containing clusters of pixels which are spatially dependent. Some of these pixel clusters are representations of an image feature of interest (i.e. *S. plumosum*) which we intend to discriminate and classify from the rest of the land-cover features. The ideal classification algorithm for this Getis transformed data is an algorithm that considers spatial information and accurately classifies the clusters within the transformed dataset. Image segmentation techniques segment clustered pixels into objects for classification making it ideal for classification (Li et al., 2014). Segmentation incorporates spatio-contextual information in finding an optimal partitioning of the data into subdivisions (Li et al., 2014). Each image object segmented contains spatially contiguous and homogenous pixels, and different regions have a high degree of heterogeneity. This object image analysis can be applied in both supervised and unsupervised classification technique which contributes to the ideals of automating features mapping within remote sensing. This method of classification allows for rectification of human error which may be introduced through supervised classification. The segmentation process measures pixel clusters compactness around the cluster centre and measurements of the compactness of the cluster can be taken as the set of standard deviation for the cluster measured separately for each object (Nussbaum and Menz, 2008). This ensures that the clustered pixels are well separated, a characteristic achieved through Getis transformation. The classification algorithm used was “assign class” to categorize pixel clusters based on thresholding method.

3.6 Regression methods

Stepwise linear regression is used to determine the relationships between remotely sensed data and ground surveyed vegetation variables (Mutanga and Skidmore, 2004, Mutanga and Rugege, 2006). In this study, stepwise linear regression was performed on the SPOT6 bands, and band ratios derived vegetation variables that demonstrated to have a significant correlation (significance level: $p < 0.01$) with *S. plumosum* canopy cover and density. Regression analysis was done for the band ratios and individual bands for SPOT6 multispectral. In line with specified criteria for the best fit model, stepwise linear regression analysis was used to find subsets of the predictor variables that best predict responses on a dependent variable by a regression equation (Mutanga and Rugege, 2006). However, to avoid overfitting a given stepwise regression model, the rule of thumb, as suggested by studies (Skidmore et al., 1997, Mutanga and Skidmore, 2004, Mutanga and Rugege, 2006) are that: (1) the number of predictor variables to enter the model be less than 1/3 the number of observations and (2) the number of steps in the stepwise linear regression analysis be 10 to 20 times less than the training data set. This was recommended to circumvent the problem of predictions being very unstable.

3.7 Geostatistics methods

Preliminary statistical tests and analyses were performed on sampled plots and selected Getis transformed layers. When constructing a linear model of co-regionalization, it is important that isotropic variograms be applied. Examining the different pairs of the sample (in this case *S. plumosum* cover) locations is an efficient means to assess the spatial structure in the regionalized variables. The set of actual values of *S. plumosum* that comprise the realization of random functions is known to be regionalized variable. A region is made of population units, so we can think of a random function as a super-population, with an infinite number of units in space and an infinite number of values of *S. plumosum* at each point in space. This is vital as exploring the data gives a good understanding of the spatial autocorrelation among measured values.

Spatial variation of irregular (randomly) sampled data is not necessarily the same in all directions, resulting in anisotropy. Anisotropy is a property of being directionally dependent. If the process is anisotropic then so are the variogram and the covariance function (Webster and Oliver, 2007). This helps to account for directional influences (i.e. anisotropy) in the sampled data. The presence of anisotropy affects geostatistical prediction methods by imposing directional influence on the predicted surfaces. It is thus important to investigate anisotropy so that if directional differences are detected in the autocorrelation, they can be accounted for by applying a transformation.

The environment, though continuous, vary from place to place with great complexities. Environmental variables (i.e. *S. plumosum*) are affected by factors (water, soil, slope, disturbances) which are largely unknown in detail and interact with a complexity we cannot disentangle, thus we can consider these environmental variables occurrence as random. If we adopt a stochastic view at each point in space, we regard the observed value as one drawn at random according to some probability distribution. This means at each point there is variation, thus at a particular point, an environmental variable is treated as a random variable with a mean, variance, and higher order moments, and a cumulative distribution function (Webster and

Oliver, 2007). The set of random variables (i.e. *S. plumosum*), constitute a random function, a random process or a stochastic process.

Through spatial interpolation, we estimate the quantity value of a land-cover feature (i.e. *S. plumosum*) at a location with no recorded data by using known observations of that land-cover feature at nearby locations. Therefore geostatistics conforms to the assumption of stationarity, in which we assume that the values of *S. plumosum* in all locations have some degree of similarity irrespective of their distance from each other. Places close to one another tend to have similar values, whereas ones that are farther apart differ. When there is a spatial correlation, then by stratifying we can estimate more precisely or sample more efficiently or both. If the strata are of different sizes then we might vary the weights attributable to their data in proportion. Geostatistical prediction differs from classical regression models estimation in that it relies on spatial models, whereas classical regression models methods do not (Webster and Oliver, 2007). Geostatistics requires the assumption that the variable is random, that the reality on the ground is the outcome of one or more random processes. Also, it assumes that the environmental variable (i.e. *S. plumosum*) to be predicted is present within the entire study area in varying quantity. The models on which predictions are based are of these random processes. Geostatistics treats a set of spatial data as a sample from the realization of a random process, therefore our summary data must include the spatial correlation. This will usually be the experimental or sample variogram in which the variance is estimated at increasing intervals of distance and several directions. Furthermore, we must recognize that spatial positions of the sampling points matter. This will show the extent to which the sample fills the region of interest, the clustering, and any mistakes in spatial position recording such as reversed coordinates. The methods are deterministic and to that extent accord with the understanding that the variation in the environment has physical causes, which means, it is physically determined.

3.7.1 Sampling image data using variogram

The field data surveyed are unevenly distributed and each pair of observation is separated by a unique lag in both distance and direction. To obtain averages containing directional information we must group the separations by direction and distance (Webster and Oliver, 2007). We choose a lag interval, the multiples of which will form a regular progression of nominal lag distance. The significant *S. plumosum* density measured in the field and best selected Getis statistic index layer data (identified from variable selection analysis of Getis index layers) were analysed using the basic geostatistical tool known as semivariogram/ variogram. The variogram is based on the theory of regionalization (Woodcock et al., 1988). The best selected Getis index layer data (i.e. pixels that correspond to ground sampled locations) was modelled using the variogram to determine the spatial range of influence in the samples. The range is highly related to the size of objects in an image and for this study, the range is related to *S. plumosum* density.

3.7.2 The Experimental Variogram

The variogram describes the variance in *S. plumosum* density within the region. Each calculated semivariance for a particular lag is only an estimate of a mean semivariance of that lag, as such it is subject to error. This error can give the experimental variogram an erratic appearance. The variogram is very important to geostatistics and is vital to estimate, interpret and model correctly. The variogram is a function of an underlying stochastic process. It may be thought of as the average of the variograms from all possible realizations of the process. The experimental variogram is computed from data which constitute a sample from a region. It applies to an actual realization and estimates the regional variogram for that realization. The shape of the points in the experimental variogram can reveal much at this stage about the way that properties change with distance and the adequacy of sampling. Variograms computed for different directions can show whether there is anisotropy and what form it takes. The variogram and estimates provide a basis for interpreting the causes of spatial variation and for identifying some of the controlling factors and processes (Webster and Oliver, 2007).

3.7.3 Kriging and cokriging techniques of interpolations

Two kriging methods were evaluated for this study with the main focus on cokriging method. However, ordinary kriging was assessed for the purposes of comparative analysis of the *S. plumosum* density prediction. Kriging and cokriging apply the same techniques to prediction models, however, cokriging allows for supplementing primary data with secondary and tertiary data into modelling taking advantage of correlation that may exist between the variable of interest and other more easily measured variables while kriging only use primary data. For both ordinary kriging and cokriging, all prediction models were evaluated for prediction of *S. plumosum* as no model was previously studied and found to excel more than the other in predicting *S. plumosum*. Kriging methods generally use the distance weighting factor and the spatial arrangement in the weights to quantify spatial autocorrelation in sample data. Ordinary kriging is thus based on the linear model of regionalization, which is fundamentally a weighting function approach which uses the variogram to estimate or predict variables.

Kriging provides a mechanism for combining global and local information in predictions, however, the ability of the variogram to describe spatial dependence is directly a function of the quantity and quality of the sampled data (Eldeiry and Garcia, 2010). The correlations among neighbouring values are modelled as a function of the geographic distance between the points across the study area, defined by a variogram (Eldeiry and Garcia, 2010). The variogram plays the significant role as a function that describes spatial dependence for a regional variable *S. plumosum*. The semivariance is a measure of the variance as distance increases from all points of recorded *S. plumosum*, it eventually reaches a value equal to the variance for the entire array of data locations (i.e. image), regardless of distance. That is, at recorded point of *S. plumosum*, the semivariance is zero, but the semivariance increases moving away from recorded point of *S. plumosum* until at a distance called the "range" where the semivariance value is equal to the variance, a point called the sill. The semivariance and covariance functions quantify the assumption that things nearby tend to be more similar than things that are farther apart by measuring the strength of statistical correlation as a function of distance.

The semivariogram is defined as:

$$\gamma(\mathbf{x}_i,\mathbf{x}_j) = \tfrac{1}{2}\, var(Z(\mathbf{x}_i) - Z(\mathbf{x}_j)), \qquad \textit{equation 2}$$

Where *var* is the variance. If two locations, $\mathbf{x}_i$ and $\mathbf{x}_j$, are close to each other in terms of the distance measure of d ($\mathbf{x}_i$, $\mathbf{x}_j$), we expect them to be similar, so the difference in their values, $Z(\mathbf{x}_i) - Z(\mathbf{x}_j)$, will be small. As $\mathbf{x}_i$ and $\mathbf{x}_j$ get farther apart, they become less similar, so the difference in their values, $Z(\mathbf{x}_i) - Z(\mathbf{x}_j)$, will become larger (Johnston et al., 2001).

Covariance is a function of the lag, it describes the dependence between values of a variable with changing lag. If the variable has a multivariate normal distribution for all positions than the mean and the covariance function completely characterise the process because all of the higher-order moments are constant. Satellite imagery has local erratic nature of the variation, wherever within an image there is some fluctuation, which emanates from variation in pixel values.

The variogram summarises the spatial relations in the data, however, we want the variogram to describe the variance of the data in the region. In other words the variogram analyses the spatial structure of the residuals from the multiple regression models (Eldeiry and Garcia, 2010). As a function of distance, this plot gives information on the spatial dependency of the variable. Regression kriging involves various combinations of linear regressions and kriging. The simplest model is based on a normal regression followed by ordinary kriging with the regression residuals.

Each calculated semivariance for a particular lag is only an estimate of a mean semivariance for that lag and errors may occur for the semivariogram. For the semivariogram, it is essential to choose a lag distance that best represents the distance among the points. The selection of a lag size has important effects on the empirical semivariogram, that is, if the lag size is too large, short-range autocorrelation may be masked. If the lag size is too small, there may be many empty bins, and sample sizes within bins will be too small to get representative averages for bins (Johnston et al., 2001). We used the “Average Nearest Neighbour” tool to determine the average distance between points and their neighbours, choosing the lag distance of 2429 metres. The true variogram representing the regional variation is continuous, and it is this variogram that we should know. Fitting the continuous model is required, this is done so that we can describe the spatial variation to predict values at unsampled locations and in larger blocks of land optimally by kriging. Based on the correlation theory, it is reasonable to expect that the density data are correlated along the transect, thus the density values of *S. plumosum* which are close together are likely to be more similar than density values which are distant from each other.

The semivariogram depicts the spatial autocorrelation of the measured sample points. It measures the strength of statistical correlation as a function of distance, the process of modelling semivariogram and covariance functions fit a semivariogram curve to the empirical data. The goal is to achieve the best fit which is ideal for the *S. plumosum* predictions. The empirical semivariogram and covariance provide information on the spatial autocorrelation of data. However they do not provide information for all possible directions and distances, for this reason, it is essential to fit a model to the empirical semivariogram to ensure that kriging predictions have positive kriging variances (Johnston et al., 2001).

Within the semivariogram plot, the points are paired and grouped so that they have a common distance and direction to reduce the number of points in the empirical semivariogram. This grouping process is known as binning. This property makes the empirical semivariogram symmetric. For each bin, you form the squared difference from the values for all pairs of locations that are linked, and these are then averaged and multiplied by 0.5 to give one empirical semivariogram value per bin.

3.8 Validation and accuracy assessment

Classification accuracy has various components for different user needs, it is, therefore, important to measure the desired properties (Lark, 1995). It is important that the component of accuracy measured is appropriate for the requirements of a particular study to avoid misinterpretation (Stehman, 1997). Consequently, seeking to optimize accuracy expressed by one metric may lead to a suboptimal classification when quantified with a different metric (Morisette et al., 1999). These coupled with other issues complicate the assessment and reporting of classification accuracy, which also limits the value of remote sensing as a source of land-cover data.

Many measures of classification accuracy can be derived from a confusion matrix. One of these is the percentage of classes correctly allocated. This is an easily interpretable guide to the overall accuracy of the classification. If attention focuses on the accuracy of individual classes, then the percentage of cases correctly allocated may be derived from the confusion matrix by relating the number of cases correctly allocated to the class to the total number of cases of that class. This may be achieved from two standpoints, giving rise to the user's and producer's accuracy, depending on whether the calculations are based on the matrix's row or column marginal (Campbell, 1996). The calculation of these and some other major indices is illustrated in Table 3 for data obtained via simple random sampling. The accuracy assessment is based on minimum mapping unit of 20m by 20m.

In the derivation of these indices, a number of fundamental assumptions are typically made. For example, it is generally assumed implicitly that each case pixel to be classified belongs fully to one of the classes in an exhaustively defined set of discrete and mutually exclusive classes (Congalton et al., 1998, Congalton and Green, 1999). To accommodate for the effects of chance agreement, kappa coefficient has often been used and some commentators argue that it should, in some circumstances, be adopted as a standard measure of classification accuracy (Smits et al., 1999). The kappa coefficient has many attractive features as an index of classification accuracy (see equation 2). It makes some compensation for chance agreement and a variance term may be calculated for it enabling the statistical testing of the significance of the difference between two coefficients (Rosenfield and Fitzpatrick-Lins, 1986). This is often important, as frequently, there is a desire to compare different classifications and so matrices. The rule of thumb is that there are at least test 10 points for accuracy assessment of a class.

Table 3: Sample confusion error matrix

		Reference		
Producer		*S. plumosum*	Other	Total
	S. plumosum	a	b	a+b
	Other	c	d	c+d
	Total	a+c	b+d	(a+c)+(b+d)

The promoted standard method of kappa coefficient is not always appropriate. There is nothing unique about the kappa coefficient in compensating for chance agreement or in allowing the significance of differences in accuracy to be evaluated as these are features shared with other accuracy metrics.

$$\hat{k} = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{i+})}{N^2 - \sum_{i=1}^{r} (x_{i+} * x_{+i})} \qquad \textit{equation 3}$$

where **r** is the number of rows, $\mathbf{x_{ii}}$ is the number of observations**,** $\mathbf{x_{i+}}$ is the marginal total of row **I,** $\mathbf{x_{+i}}$ is the marginal total of column **i, N** is the total number of observations.

The design of an accuracy assessment programme has several elements including the definition of an appropriate sample size and sampling design as well as the specification and use of a measure of accuracy appropriate to the application in-hand (Dicks and Lo, 1990, Stehman, 1999). The sample size, for example, must be selected with care and be sufficient to provide a representative and meaningful basis for accuracy assessment. An appropriately defined sample will aid the ability to infer the properties of the population from which it was drawn. The sampling design used to select the cases upon which the accuracy assessment is based is of major importance. If for example, a probability-based measure of classification accuracy is to be used, it is essential that the cases were acquired according to an appropriate sampling design (Stehman et al., 2000). How often this is achieved is open to question as typically, little information on the sampling design used in evaluating classification accuracy is provided with the accuracy statement. It is nonetheless important that the sampling design used is specified as it can significantly influence the results of an analysis (Friedl et al., 2000, Green et al., 1993, Stehman, 1995). Indeed, the confusion matrix cannot be properly interpreted without knowledge of the sampling design used in its construction (Stehman, 1995, MALING, 1989). Basic sampling designs, such as simple random sampling, can be appropriate if the sample size is large enough to ensure that all classes are adequately represented. The adoption of a simple sampling design is also valuable in helping to meet the requirements of a broad range of users (Stehman and Czaplewski, 1998) although the objectives of all users cannot be anticipated (Stehman et al., 2000). Frequently, ground data collection is constrained as physical access to some sites is impractical and restricted to sites of opportunity. Where possible ground data, or high-quality fine spatial resolution imagery acquired at an appropriate date can be used for validation (Edwards et al., 1998, Estes et al., 1999). However, it must be realized that the sampling design used to collect the sample of cases, upon which the accuracy assessment is based, has important implications for the estimation of classification accuracy. While there is an obvious desire to balance statistical requirements with practicalities (Edwards et al., 1998, Merchant et al., 1993), the choice of sampling design influences the reliability of an accuracy assessment (Stehman et al., 2000, Muller et al., 1998).

A method for assessing the accuracy of interpolated layers is the root mean square error (RMSE). The RMSE describe the global accuracy of the interpolation method, measuring the degree of deviation of the predicted values from the field observed data. For this method, samples of accuracy assessment must be adequately spaced apart or they will be spatially auto-correlated. If the user makes a mistake at a particular location then the chances of the user making the same mistake at a closer location are very high. Therefore the errors around a particular error will be highly correlated. The root mean square error is calculated using equation 3.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{P}_i - P_i)^2}{N}} \qquad \textit{equation 4}$$

where N is the number of observations, P_i is the population and $\acute{P}_i$ is the estimated population.

CHAPTER FOUR

RESULTS

This chapter provides the results obtained from analysis of the Getis transformed layer variable selection which selected the best Getis index at clustering pixels representing *S. plumosum* and the classification results. Furthermore, there is an outline of the geostatistical kriging and cokriging, and regression results of interpolation. The variable selection examined the 308 Getis transformed layer variables computed through the seven Getis statistic indices from SPOT 6 multispectral data. These variable selection results served to identify Getis transformed layers utilized to map *S. plumosum* and also for geostatistical predictions. Succeeding the variable selection process, the chapter presents the object based image analysis classification results for the top performing Getis transformed index layer variables. This classification is further improved through geostatistics to predict density quantity and percentage cover of *S. plumosum*.

Geostatistics allows for dealing with properties that vary in ways that are far from systematic and at all spatial scales. It also provides estimates of error. In this chapter, we also demonstrate model semivariograms which feed into kriging and cokriging techniques of interpolation. The results produced by different geostatistical tools and analysis, such as the variogram modelling for kriging and cokriging techniques of interpolation, as outlined earlier in chapter three are presented.

4.1 Getis transformed indices layer variables.

The SPOT 6 multispectral image Getis transformations resulted in 168 lag 3 bands and 140 lag5 bands for all the seven Getis indices. The lag 3 indices were computed in two ways; 1) to include intermediate lag 1, 2 and 3; and 2) to exclude the intermediate lags. This was motivated by the need to establish if there is any significant difference in the ability and probability of clustering pixels of *S. plumosum* by inclusion or exclusion of intermediate lags. The field plot data was used to extract 9 pixels per plot for each index for all indices and averaging the values of the 9 pixels within each 20x20m plot to a single mean value per plot. These mean layer values were used for the variable importance selecting the Getis transformed bands with a high probability of *S. plumosum* pixel clusters in variable selection process. The results from variable selection of Getis indices layers/bands for each lag are summarised in Table 4.

The model ranked the Getis transformed layers according to clustering of pixels in areas of presence of *S. plumosum*. The SPOT 6 bands most useful are the bands 2 (Green band) and 4(Infrared band) with their Getis transformed layers importance ranging from 0.79 to 0.83 for both lags. Table 4 shows the SPOT 6 bands, related Getis index variable, and their importance.

Table 4 : Top 20 Getis indices transformed layers, 10 for each lag 5 and 3 respectively. "Mean" refers to layer in the naming.

	Lag 5			Lag 3		
Layer No:	**SPOT band**	**Getis transformed layer variable**	**Variable importance value**	**SPOT band**	**Getis transformed layer variable**	**Variable importance value**
1	2	Rook mean.10	0.83045977	4	Positive mean.12	0.82758621
2	2	Horizontal mean.10	0.82758620	4	Horizontal mean.12	0.82327586
3	4	Positive mean.18	0.82758620	4	Positive mean.11	0.82183908
4	4	Horizontal mean.18	0.82327586	4	Positive intermediate mean.12	0.82183908
5	4	Positive mean.17	0.82183908	4	Queen mean.11	0.81609195
6	4	Positive mean.19	0.82040229	4	Bishop mean.11	0.81465517
7	2	Rook mean.9	0.82040229	4	Positive intermediate mean.11	0.81465517
8	2	Horizontal mean.9	0.81609195	4	Rook mean.11	0.81465517
9	2	Queen mean.9	0.81609195	2	Bishop mean.6	0.81321839
10	2	Queen mean.10	0.81609195	4	Rook intermediate mean.12	0.81321839

Figure 4 and 5 indicate the change in variable importance for Getis transformed layer variables. The layer variables displayed in each figure are the top 10 for each lag with variable 1 having the highest clustering importance.

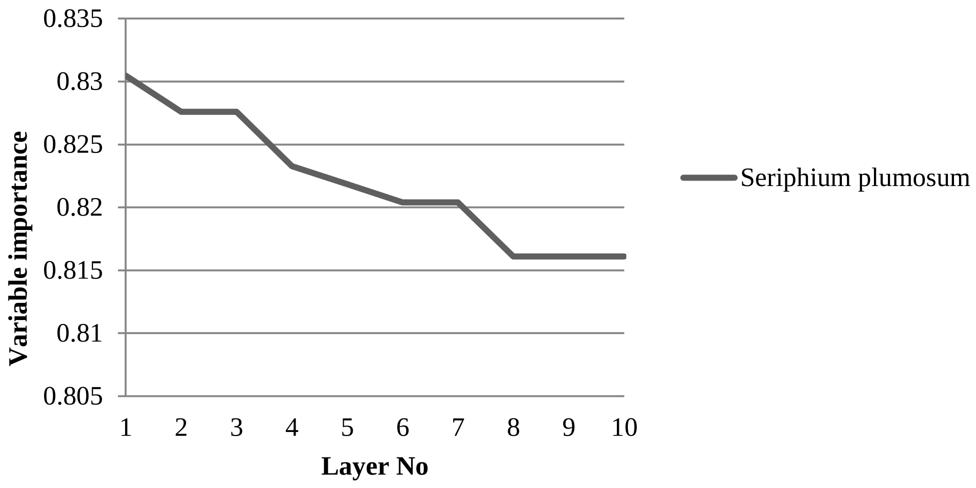


Figure 4: Variable importance reduction according to the performance of each Getis index layer for lag 5.

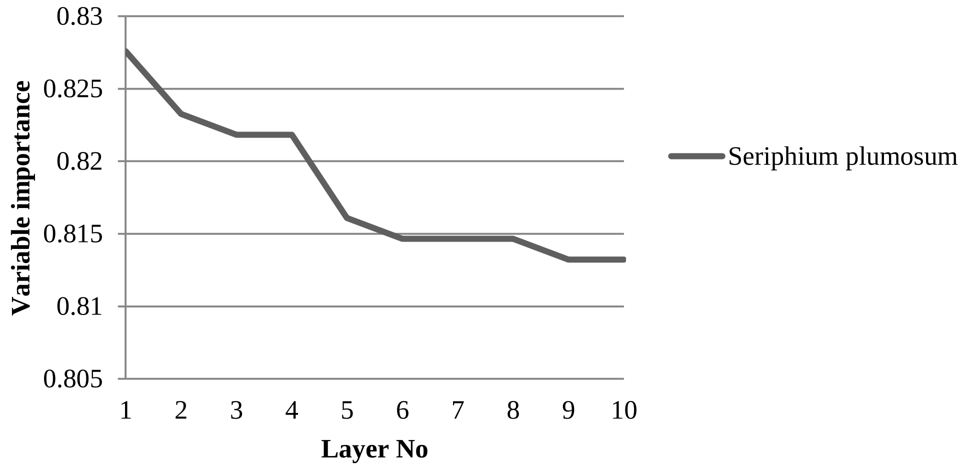


Figure 5: Variable importance reduction according to the performance of each Getis index layer for lag 3.

The Getis transformed layer variables selected with high importance to cluster pixels representing *S. plumosum* were utilized in the classification process of *S. plumosum*.

4.2 Classification and accuracy assessment

The Getis transformed layers identified in the variable selection process contained spatially dependent pixel cluster information. The classification algorithm suitable for the classification of this spatially dependent data must be able to classify clustered pixel. The object based image classification algorithm identified was successful in classifying the clusters of pixels representing the patches of *S. plumosum*. The classification was performed on the best top three performing indices bands namely Rook, Horizontal and Positive. These classification results are demonstrated in figure 6 and figure 7.

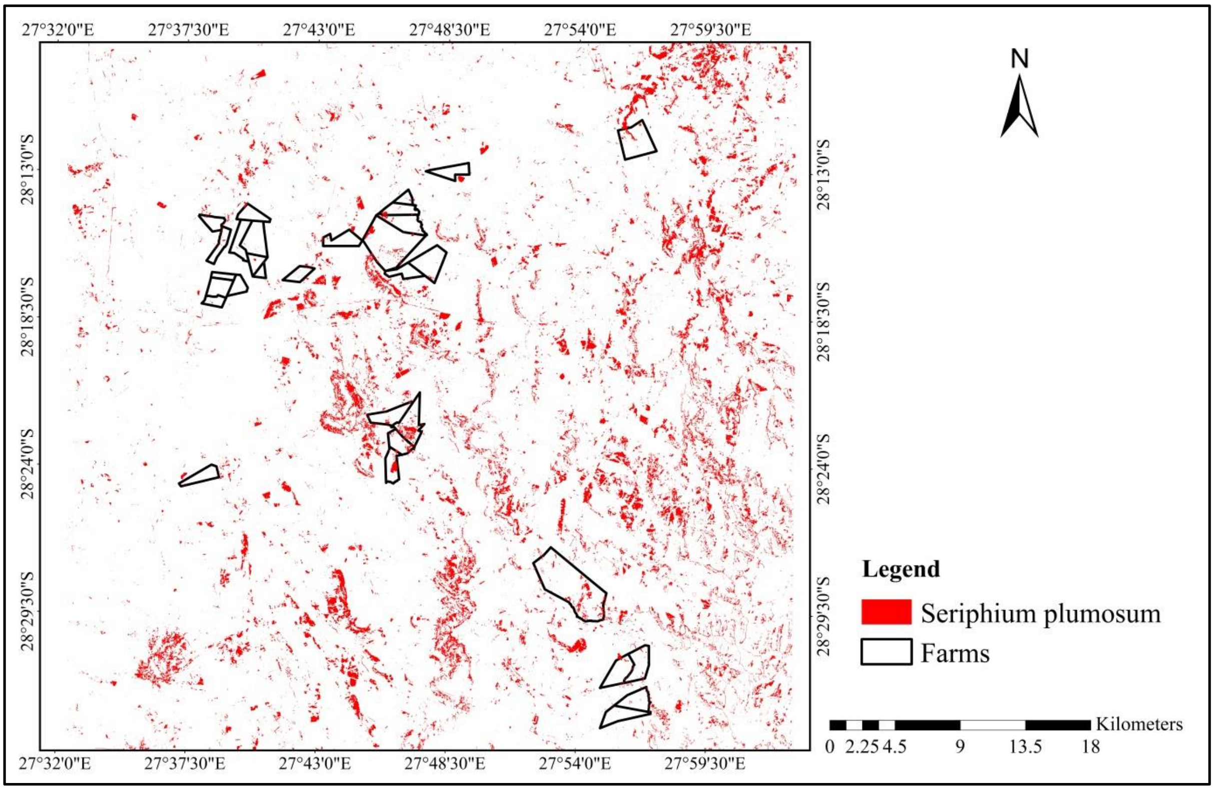


Figure 6: Demonstration of *S. plumosum* infested areas mapped using top 3 Getis indices layers. The area covered by *S. plumosum* is 13148.7 hectares.

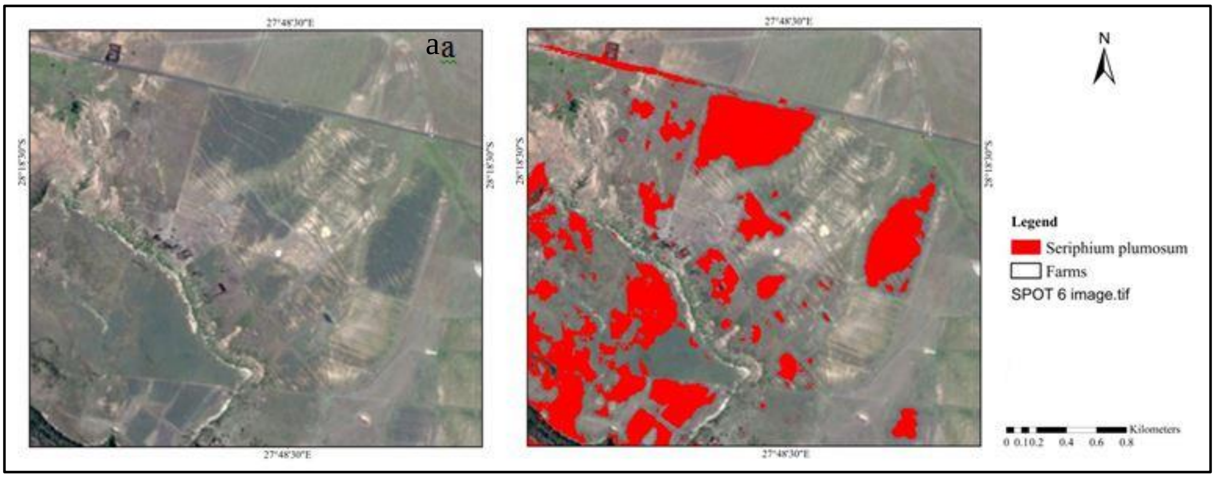


Figure 7: Demonstration of the extent of *S. plumosum* patch mapped, a) SPOT 6 image used for the study, b) SPOT 6 image with *S. plumosum* patches mapped.

Accuracy assessment

Table 5 : Classification confusion matrix

		Reference		
		S. plumosum	Other	Total
Producer	*S. plumosum*	37	0	37
	Other	26	9	35
	Total	63	9	72

The confusion matrix demonstrated the overall accuracy of 63.9%, Omission error of 41.3%, commission error of 0% and Kappa coefficient: 0.262411.

4.3 Geostatistical analysis

4.3.1 Selecting the optimal variable for density estimation

To apply geostatistics to the field and remote sensing data, a relationship needed to be determined between the Getis transformed layer variables and the field data collected. This relation is essential for the kriging and cokriging of the data to produce the most optimal and accurate density estimations (Mutanga and Skidmore, 2004, Mutanga and Rugege, 2006). The correlation of the SPOT 6 multispectral data, Getis transformed layer variables and field parameters data determined the degree of significance of the relationship between the data (Table 6). The data collected in the field included the canopy density, length, height, width, spacing, observed percentage estimation, and surrounding features. The canopy length, width, and spacing were used to calculate the reliable percentage area covered by *S. plumosum* per plot. Not all the field collected information was significantly useful for density prediction but was used to understand the results and analysis. The field data was found to be non-normally distributed when assessing “canopy density and canopy percentage” histograms, thus dictating that the data be transformed. It is important to have the data normally distributed for the selection of suitable geostatistical technique which assumes a normal distribution of the dataset. The transformation of this data was done through LogK transformation which produced satisfactory normal distribution in the data (Figure 8).

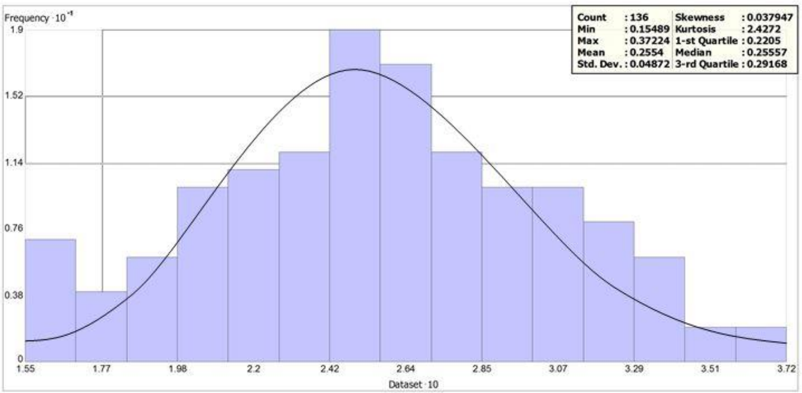


Figure 8 : Field collected canopy density histogram. The x-axis shows dataset, and y-axis shows frequency.

Table 6 : Correlation (r) between field data parameters and SPOT 6 data (multispectral data and Getis transformed data).

Variable	Canopy density	Percentage canopy cover
Canopy percentage	0.93496819	1
Band 1	0.405516157	0.579167269
Band 2	0.552321516	0.499910696
Band 3	0.336705719	0.320507736
Band 4	0.527769021	0.438021563
NDVI	0.470247477	0.376780205
Rook Getis layer variable	0.578564458	0.521730927
Horizontal Getis layer variable	0.578020357	0.520603033
Positive Getis layer variable	0.596441439	0.53331117

p-value < 0.01

4.3.2 Investigating directional influence in the data

The data used in the study was randomly sampled, indicating some level of anisotropy which affects the variogram in various ways (Figure 9). If the variogram has a sill, then variation in gradient will lead to variation in the range. The variogram cloud information of field data fields reflected a non-gradual increase in the semivariogram values from the centre to all directions indicating the presence of anisotropy. The presence of anisotropy in the variogram surfaces measured *S. plumosum* density datasets allowing for accounting for anisotropy for modelling a reliable linear model of coregionalization and better surface predictions. The significant principle is that when there is anisotropy for spatial correlation, the semi-variance depends on the magnitude of distance and its direction. The directional influence measured was at 36.4 degrees in the northeast direction.

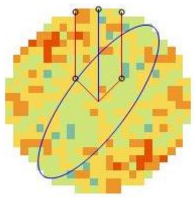


Figure 9: Demonstration of the presence of directional influence to the semivariogram model**.**

4.3.3 Estimation of *S. plumosum* canopy density and percentage cover.

The variogram is the cornerstone of geostatistics and very important to estimate, interpret and model correctly (Webster and Oliver, 2007). It is computed using the Matheron's method-of-moments estimator (moments is a method for estimation of population parameters). The values of the variance can then be plotted against lag distance as a scatter diagram in the variogram cloud. The variogram cloud shows the spread of values at each lag and assists in identifying outliers. This variogram cloud contains information on the spatial relations in the data to the lag. However, in practical terms, it is difficult to judge from a variogram cloud if there is any spatial autocorrelation present, the form it might have, and how we can model it (Webster and Oliver, 2007). As a result, we average the variance for each of the lags and examine the results. The dense distribution of points in the variogram cloud signifies stronger spatial continuity in the data. That is, the closer the points lie on the variogram line, the stronger is the correlation and the smaller the semivariance.

The autocorrelation coefficients and semivariogram are affected by the lag distance intervals, as the correlation between pairs of points decreases and the semivariance increases. The lag increments can affect the resulting variogram, and so distance should be chosen wisely. The variogram was computed using the data in kriging and cokriging tools. All the geostatistical models were used to model *S. plumosum* density and assessed through their root mean square error (RMSE). Kriging was applied to the field measured canopy density information whilst cokriging was applied to the field measured canopy density combined with Getis transformed layer information. Cokriging produced better results than kriging with a circular model having the smallest RMSE (Table 7). Further cokriging was performed for the percentage canopy cover using SPOT 6 band 1, Positive and Horizontal Getis transformed layer variables. Figure 12 demonstrate the map for percentage cover.

Table 7: Kriging and cokriging model root mean square error results. The variables used in the model are Rook, Horizontal and Positive Getis transformed layer variables.

Geostatistics model	**Cokriging RMSE**	**Kriging RMSE**
Circular	25.80420785	26.88747865
Spherical	26.99353244	26.74096792
Tetraspherical	26.94055773	26.37429003
Pentaspherical	26.8815745	26.09141439
Exponential	26.96823528	26.68273326
Gaussian	26.92228675	26.88629792
Rational Quadratic	26.8606043	26.84878289
Hole effect	26.91285171	26.4835978
K-Bessel	26.97529734	26.74660637
J-Bessel	27.2693793	26.44371081

Regression root mean square error: 35.61255352

Variograms were constructed for the sampled field measured canopy density, percentage cover, Getis transformed layer variables, and Cross-variogram for field canopy density data combined with Getis transformed layer variables (see Figure 10). The variography aimed to explore and quantify spatial dependence. The variogram models were used to predict *S. plumosum* density using ordinary kriging and cokriging. The cokriging circular model demonstrated to be performing better than all other geostatistical models, we, therefore, chose to display its variograms. With the support of the Getis transformed data, we integrated field canopy density measurements with Getis transformed layer variable in a cokriging technique, the results are demonstrated in Figure 10.

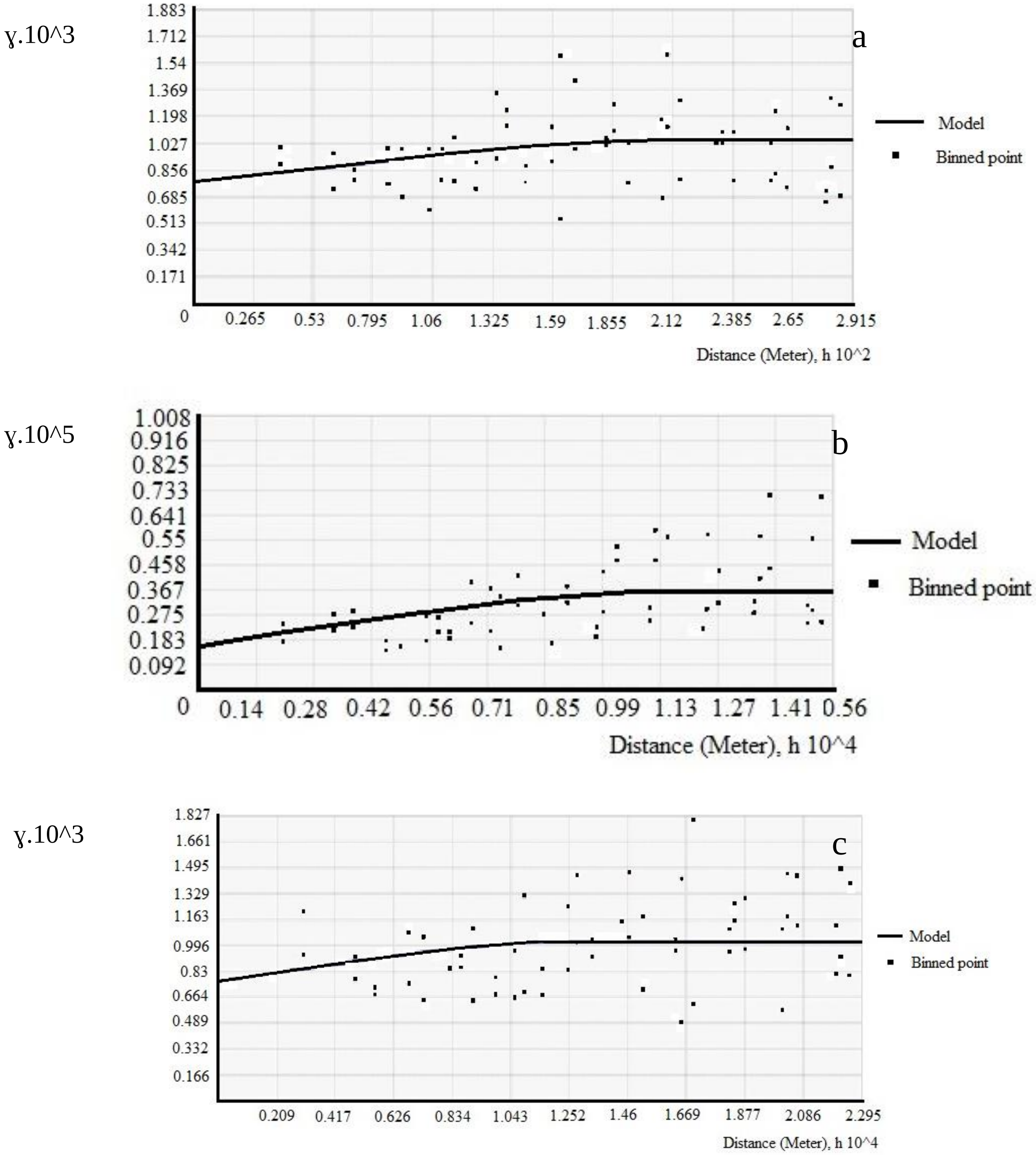


Figure 10: Variograms for prediction models of *S. plumosum* canopy density, a) field canopy density data variogram; b) Variogram for Rook, Horizontal and positive Getis transformed layers; c) Cross-variogram for field canopy density data combined with Getis transformed layer variable for cokriging modelling.

It is essential to notice that the variogram for the plot density has a nugget value whilst the Getis transformed layer variable does not have. This suggests that the measurement error can be reduced when cross-correlation factor between intensively sampled image data and Getis transformed layer variable are combined in semivariogram modelling. Therefore, it is important to understand the nugget effect for cokriging which can be exact or smoothed depending on the measurement error model calculated from the cross-semivariogram.

Table 8: Variogram parameters for predicted *S. plumosum* canopy density.

No.	Variable	Range (m)	Nugget (m)	Sill (m)
1	Field	29151	0.7705	1.2
4	Getis transformed layer variable	13025	0.17	0.367
5	Cross variogram	13914	0.73	1.1

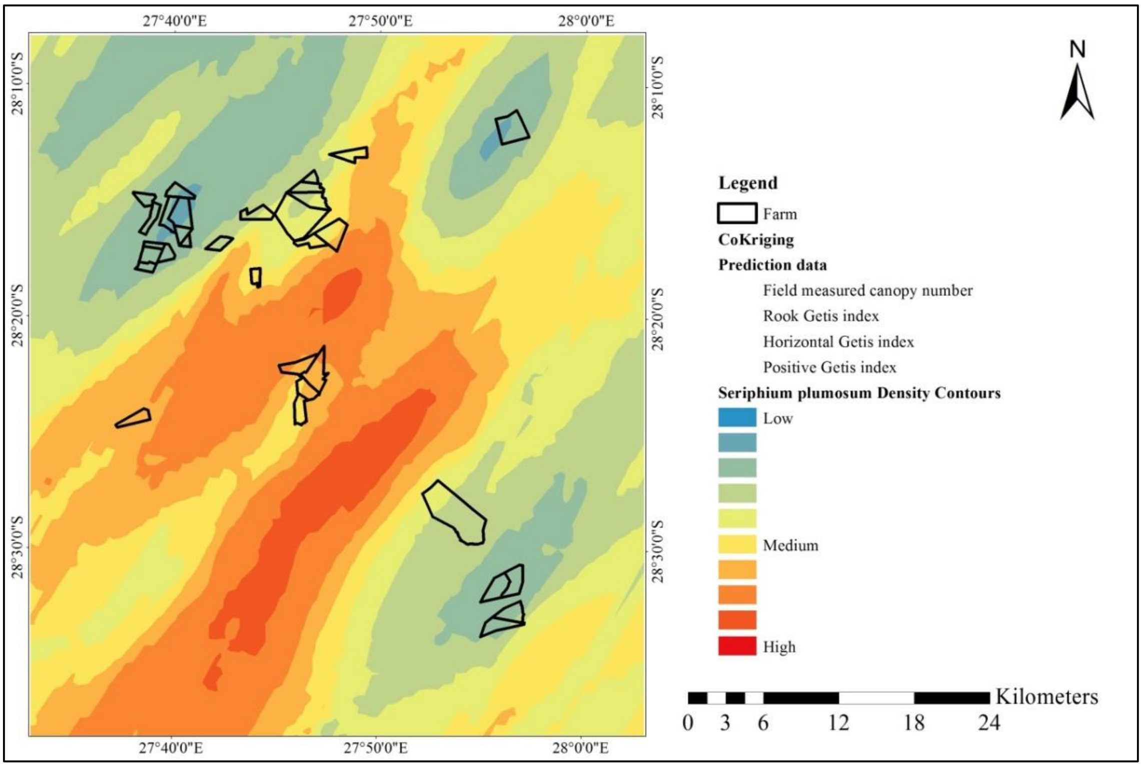


Figure 11: Density cover map of *S. plumosum* created through Geostatistical cokriging of field canopy density information and Getis transformed layer variables. Filled contours depict different levels of density infestation per 400 sqm.

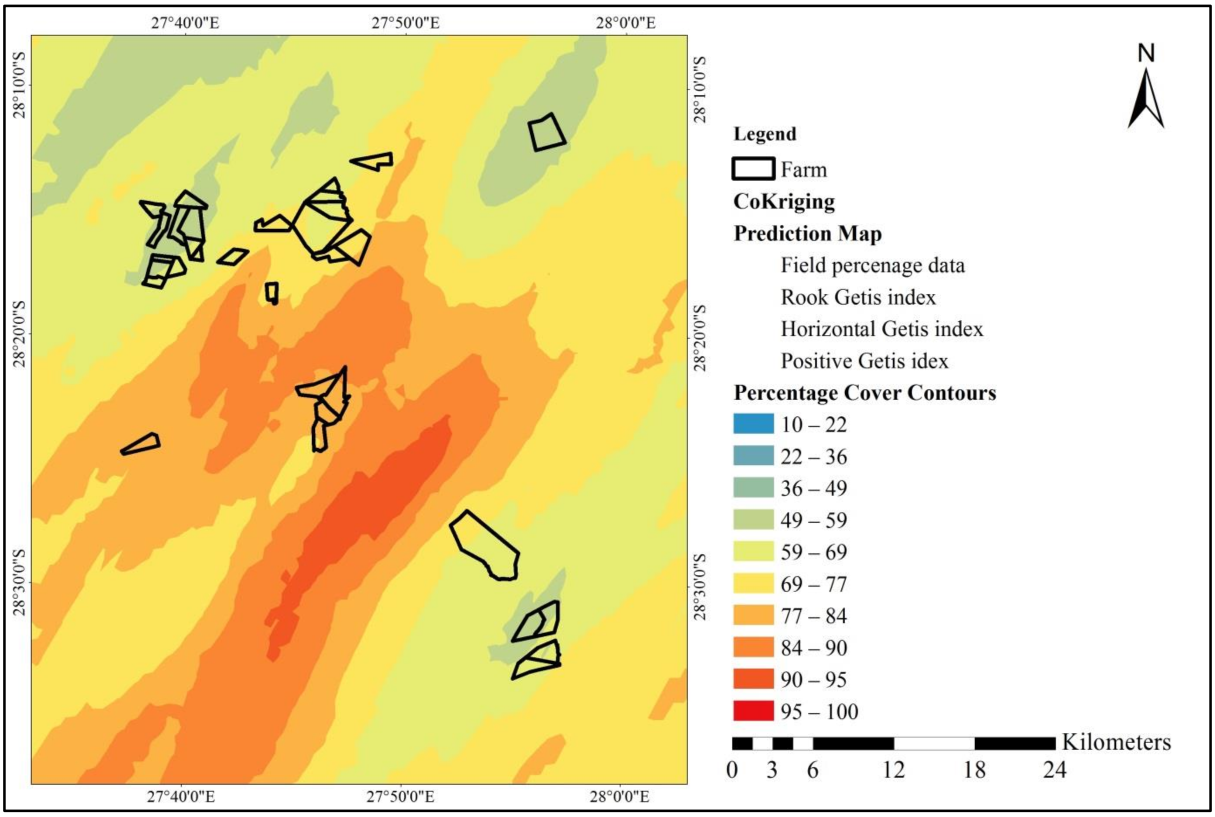


Figure 12 : Percentage cover map of *S. plumosum* created through Geostatistical cokriging of field canopy percentage information, SPOT 6 multispectral band 1, Positive and horizontal Getis transformed layer variables. Filled contours depict different levels of percentage infestation per 400 sqm.

CHAPTER FIVE

DISCUSSIONS AND CONCLUSIONS

In this chapter Getis statistics, classification, Geostatistics and regression results are discussed. This study had two objectives to: 1) assess if Getis statistics can be used to map *S. plumosum*; and 2) predict canopy density and percentage cover of *S. plumosum* using Geostatistics.

5.1 Assessing the ability of Getis statistics to discriminate pixels of *S. plumosum*

The potential of Getis statistics in discriminating pixels of *S. plumosum* was demonstrated using SPOT 6 multispectral imagery. The Getis statistics image transformations yielded layers of spatially dependent pixel clusters depicting a variety of landcover features. Among these features is the *S. plumosum* shrub species which was of interest to this study. The SPOT 6 Getis transformed data was found to have positive autocorrelation. The statistical test for the pixel clustering probability of the Getis transformed bands through variable selection demonstrated *S. plumosum* importance ranging from 0.79 to 0.83 for both 3 and 5 lags. Lag 5 produced the highest importance thus used as input into classification and geostatistical interpolation. The clusters of image spectral values and the strength of the association between neighbouring pixels are demonstrated by the magnitude of the *Gi** results (Wulder and Boots, 1998). These *Gi** values within a moving window create fuzzy boundaries between objects reflecting the subtle changes occurring in spectral values among image objects. The visualization of the *Gi** results illustrated that result conveyed meaningful spatial information, demonstrating the potential of Getis statistics in remote sensing context. Measures of spatial dependence such as variogram have proved to have value in remote sensing image processing. Local indicators of spatial association are complementary to variograms while providing information not detectable in variogram analysis allowing for an improved understanding of image spatial structure. This understanding of the magnitude of autocorrelated values is valuable supplementary information, providing insights to image spatial structure, which may allow for the creation of fuzzy boundaries around image objects (Wulder and Boots, 1998).

5.2 Object based image analysis classification results

Spatially dependent reflectance value clusters of varying intensities were distinguished, having low and high *Gi** values which translating into good spatial uniformity within the image. Of the 308 Getis index layers, top 20 indices layers were tested for the classification of *S. plumosum.* However, top 3 Lag 5 layers demonstrated better classifications compared to top 1, 5, 10, 15 and 20. The overall accuracy of the classification obtained was 63.9%. The classification error of omission was large (41.3%). This large error can be attributed to canopy spacing which, within an image, increases the reflectance of other vegetation and soil in the neighbourhood of *S. plumosum*, reducing reflectance purity of the *S. plumosum* canopy. Areas recorded to have large canopy spacing (average of 2.5 meters) in the field are not mapped due to limitations in spatial resolution. Additionally, some small patches of *S. plumosum* canopies recorded in the field were not clustered to be *S. plumosum* on the image introducing errors of omission. Other misclassifications and omission have been identified in areas with “young” plant canopies. These canopies have a height and width of fewer than 50 centimeters on average. There was misclassification of wet soil into *S. plumosum*, this behaviour can be attributed to the fact that water is known to absorb light, therefore, making soil appearing darker. The very low value of kappa coefficient of agreement can be attributed to the large error of omission indicating Getis statistics does not achieve very high accuracy for mapping *S. plumosum*.

Generally, Getis statistics were moderate in discriminating pixels representing *S. plumosum* from other vegetation types. The cluster of the image values succeeded in mapping *S. plumosum* patches with mature adult canopy averaging height of 70 centimetres and width averaging 1 metre with less than one meter average canopy spacing. This is very significant as the study area contained diverse landscapes comprising of different crops and natural vegetation. The Getis statistics weaknesses on creating clusters of medium values which do not belong to hot or cold spots are prevalent for mapping *S. plumosum*. Some of the features such as tar road and minor shadows were also misclassified to be *S. plumosum* introducing an error of commission. These minor shadows and tar had similar values to the peripheral mixed reflectance of *S. plumosum* and other vegetation.

The classification of spatially dependent image values therefore demonstrated that Getis statistics can be used to discriminate mature *S. plumosum* canopy and not "young" canopies. To compensate for the weaknesses in this classification, predictive techniques were used to predict density and percentage cover of *S. plumosum* in the entire study area. The density is predicted from field density and percentage measurements using the geostatistics analysis.

5.3 Geostatistics analysis results

Geostatistical techniques utilized field measured canopy density and percentage information, and Getis transformed layer variables and SPOT 6 multispectral band to predict density cover (Figure 11) and cover percentage (Figure 12). The identified best performing Getis transformed layer variable was used as input into Geostatistics instead of the classification mainly due to the error that may accrue from the classification. Irrespective of the error's magnitude, it would be introduced into the geostatistical cokriging emanating to further bias and uncertainties. Furthermore, the Getis transformed layer also had the advantage over classification because it provided continuous spatially dependent information throughout the study area.

The correlation results showed that certain canopy parameters (e.g. height) do not have a significant correlation to the Getis transformed data. The canopy density and percentage are parameters that had the highest level of significance of correlation (Table 6). Canopy density and percentage had the highest correlation with SPOT 6 Getis transformed layers and was thus selected for cokriging. The canopy density is the only parameter per plot which gives the canopy density of *S. plumosum*. Canopy size (width, length) and spacing help understand some of the reasons why in some field records, areas of infestation are not mapped on the SPOT image.

The variograms for the field canopy density values had a nugget value greater than zero at zero distance. Theoretically, we expected a zero nugget at zero distance, however, all the semivariograms depict a nugget value, and this nugget effect can be attributed to measurement errors. Measurement error occurs because of the error inherent in measuring devices, in this case, SPOT 6 sensor and the GPS. The nugget effect can further be attributed to spatial sources of variation at distances smaller than the sampling interval.

When assessing the model of a semivariogram, we noticed that at a certain range, the model levels out (i.e. reaches a sill). This means sample locations separated by distances closer than the range are spatially autocorrelated, yielding better density estimations reducing prediction error, whereas locations further apart than the range are not spatially autocorrelated. The partial sill is at the centre of data spatial autocorrelation, it is relatively small for Getis transformed layers indicating that only a few points in the dataset are spatially autocorrelated having low prediction error closer to them and larger prediction errors at distances farther from them.

The semivariogram for the *S. plumosum* density cover and the percentage was at lag 2429 meters. At this lag size, the experimental model with a nugget effect yielded the best estimation for the field dataset. The large error reflected the non-systematic sample field plots layout and lack of field information for large areas between surveyed farms. It is significant to notice that the large prediction errors occurred in areas that sparsely sampled. The Cokriging model had a root mean square error of 25.8 and the ordinary kriging model had a root mean square error of 26.1 and linear regression had a root mean square error of 35.6. Contrary to Eldeiry and Garcia (2010) whose study found ordinary kriging to perform more accurately than cokriging, we have found cokriging to have the lowest RSME compared to both ordinary kriging and ordinary regression.

The cokriging technique of density prediction resulted in the smallest root mean square error (RMSE) compared to kriging and regression prediction. This means the integration of Getis statistic and field data is significant for improving density prediction. Furthermore, the integration of spatial dependence information plays a significant role in improving the accuracy of predictions. This is indicated by the 9% RMSE difference between ordinary regression taking no account of spatial dependence and geostatistics which account for spatial dependence. The density values predicted on the density map in Figure 11 demonstrate low, medium and high infestations of *S. plumosum*. Areas of low infestations are areas in which *S. plumosum* canopy is dispersed and with large spacing among individual canopy.

5.3 Conclusions

This section reviews the aim and objectives of the study, determine what has been achieved regarding the set goals and provides a synthesis of the approaches adopted for this study. Conclusion, limitations, and recommendations for future applications of Getis statistics and geostatistical techniques for the estimation of *S. plumosum* canopy density and percentage cover, are also provided.

*Gi** computation is efficient in the assessment, generation and utilization of spatial dependence information present based upon neighbouring pixel relationships. LISA statistics, especially Getis statistics, provide values based on the spatial structure of digital images. The ability to assess the strength of inter-pixel relationships as well as the magnitude of autocorrelated data proved valuable when the values computed semivariance, as a positive valued function (Wulder and Boots, 2001). The positive spatial autocorrelation can be considered as positive additional of information source to image classification. Local spatial autocorrelation characteristics, the Getis statistics, generated to explore spatial autocorrelation characteristics present in remote sensing imagery is useable in classification of shrubs (Wulder and Boots, 2001). The Getis statistics was successful in discriminating mature *S. plumosum* from other vegetation types. More research into Getis statistics could result in large scale mapping of the invasive shrub considering the potential demonstrated in this study. The use of OBIA classification method also provides an interesting area of exploration as remote sensing classification and mapping are moving towards automation. Getis statistics, however, can be explored future in conjunction with spectral techniques of classification.

5.3.1 Aims and objectives re-visited

The aim of the study was to evaluate the different Getis-statistic indices in mapping invasive shrub *S. plumosum* in which local statistics LISA Getis was applied to SPOT 6 data to produce spatially autocorrelated Getis layers. The aim was also to estimate *S. plumosum* canopy density and cover by integrating field data with best performing Getis transformed indices through geostatistics technique called cokriging. The Getis indices which create clusters of pixels with similar reflectance proved to be useful in mapping spatial extent of the patches for invasive shrub *S. plumosum*. The mapping model importance of between 0.66 and 0.83 obtained is satisfactory for the application of Getis statistics in vegetation studies and allows for further refinement to optimize the ability of Getis to map vegetation. The study also indicated the variation in the correlation between field data and Getis transformed data. Findings established that the variation depends on the type of field sampled parameter and how it relates to both spectral and spatial characteristics of the remotely sensed image data.

The objective of this study was to investigate the use of Getis indices image transformations applied to high-resolution SPOT 6 multispectral imagery and geostatistic techniques in mapping and estimating canopy density and percentage cover of the invasive plant *S. plumosum* in the South African grasslands. This objective was achieved with Rook, Positive and Horizontal indices of Getis statistics yielding the highest mapping importance resulting in the *S. plumosum* map (Figure 6). The study also specifically focused on the use of geostatistical techniques known as kriging (i.e. ordinary kriging of sampled field data) and cokriging (i.e. interpolation of field and Getis transformed data) to estimate the density and percentage cover of *S. plumosum*. The results obtained demonstrate that SPOT 6 Getis transformed data utilised in conjunction with field data can be utilised to estimate the *S. plumosum* using Geostatistics technigues. This was demonstrated by cokriging which performed better than ordinary kriging and regression.

5.3.2 Key research questions

5.3.2.1 The research question "Which Getis image transformation index performs better in detecting and mapping invasive species *S. plumosum?"* was answered by the study with Lag 5 Rook Getis index performing better than all other indices for the Lags 3 and 5.

5.3.2.2 "To what level can geostatistical technique, cokriging, improve mapping and estimate canopy density and percentage over of *S. plumosum?"* Given that we were only able to take measurements at discrete locations and not over the entire, the utility of geostatistics has improved mapping through estimations in areas which were not field surveyed.

5.4 Recommendation for future use of Getis and geostatistics

Getis statistics demonstrated potential in local spatial studies in remote sensing. However, more research beyond the scope of this study should be carried out in which investigations attempt to eliminate the errors. The use of very high resolution imagery can also assist in the refinement of the Getis statistics classifications.

REFERENCES

ABBURU, S. & GOLLA, S. B. 2015. Satellite image classification methods and techniques: A review. *International journal of computer applications,* 119.

ADJORLOLO, C. & MUTANGA, O. 2008. *Estimating woody vegetation cover in an African savanna using remote sensing and geostatistics.* University of KwaZulu-Natal, Pietermaritzburg.

ADJORLOLO, C. & MUTANGA, O. 2013. Integrating remote sensing and geostatistics to estimate woody vegetation in an African savanna. *Journal of Spatial Science,* 58**,** 305-322.

ALAJLAN, N., BAZI, Y., MELGANI, F. & YAGER, R. R. 2012. Fusion of supervised and unsupervised learning for improved classification of hyperspectral images. *Information Sciences,* 217**,** 39-55.

ANSELIN, L. 1995. Local indicators of spatial association—LISA. *Geographical analysis,* 27**,** 93-115.

ARBIA, G., GRIFFITH, D. & HAINING, R. 1998. Error propagation modelling in raster GIS: overlay operations. *International Journal of Geographical Information Science,* 12**,** 145-167.

ATKINSON, P. M., WEBSTER, R. & CURRAN, P. J. 1994. Cokriging with airborne MSS imagery. *Remote Sensing of Environment,* 50**,** 335-345.

AVENANT, P. 2015. BANKRUPT BUSH (Seriphium plumosum).

BANNARI, A., OMARI, K., TEILLET, P. M. & FEDOSEJEVS, G. 2005. Potential of Getis statistics to characterize the radiometric uniformity and stability of test sites used for the calibration of Earth observation sensors. *Geoscience and Remote Sensing, IEEE Transactions on,* 43**,** 2918-2926.

BLASCHKE, T. 2010. Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing,* 65**,** 2-16.

BLASCHKE, T., BURNETT, C. & PEKKARINEN, A. 2004. Image segmentation methods for object-based analysis and classification. *Remote sensing image analysis: Including the spatial domain.* Springer.

BROOKS, M. L., D'ANTONIO, C. M., RICHARDSON, D. M., GRACE, J. B., KEELEY, J. E., DITOMASO, J. M., HOBBS, R. J., PELLANT, M. & PYKE, D. 2004. Effects of invasive alien plants on fire regimes. *BioScience,* 54**,** 677-688.

BROWN, J. F., LOVELAND, T. R., OHLEN, D. O. & ZHU, Z.-L. 1999. The global land-cover characteristics database: the users' perspective. *Photogrammetric Engineering and Remote Sensing,* 65**,** 1069-1074.

CAMPBELL, J. B. 1996. Introduction to Remote Sensing. Virginia Polytechnic Institute and State University. Taylor & Francis Ltd. London, England.

CHO, M. A., RAMOELO, A., DEBBA, P., MUTANGA, O., MATHIEU, R., VAN DEVENTER, H. & NDLOVU, N. 2013. Assessing the effects of subtropical forest fragmentation on leaf nitrogen distribution using remote sensing data. *Landscape ecology,* 28**,** 1479-1491.

CIHLAR, J. 2000. Land cover mapping of large areas from satellites: status and research priorities. *International journal of remote sensing,* 21**,** 1093-1114.

COHEN, W. B. & JUSTICE, C. O. 1999. Validating MODIS terrestrial ecology products: linking in situ and satellite measurements. *Remote Sensing of Environment,* 70**,** 1-3.

CONGALTON, R. & GREEN, K. 1999. Assessing the accuracy of remotely sensed data: principles and applications. *Lewis Pub—lishers, Boca Raton, Fla*.

CONGALTON, R. G., BALOGH, M., BELL, C., GREEN, K., MILLIKEN, J. A. & OTTMAN, R. 1998. Mapping and monitoring agricultural crops and other land cover in the Lower Colorado River Basin. *Photogrammetric Engineering and Remote Sensing,* 64**,** 1107-1114.

COWEN, D. J. & CONSORTIUM, S. C. S. G. 2002. *State of knowledge on GIS databases and land use/cover patterns: South Carolina*, LU-CES Program Office, SC Sea Grant Consortium.

CURRAN, P. J. & ATKINSON, P. M. 1998. Geostatistics and remote sensing. *Progress in Physical Geography,* 22**,** 61-78.

DEVINE, K. & FEI, S. 2011. A review of impacts by invasive exotic plants on forest ecosystem services.

DICKS, S. E. & LO, T. H. 1990. Evaluation of thematic map accuracy in a land-use and land-cover mapping program. *Photogrammetric Engineering and Remote Sensing,* 56**,** 1247-1252.

DOGRA, K. S., SOOD, S. K., DOBHAL, P. K. & SHARMA, S. 2010. Alien plant invasion and their impact on indigenous species diversity at global scale: A review. *Journal of ecology and the natural environment,* 2**,** 175-186.

DWYER, P. C. 2011. *Spatial estimation of herbaceous biomass using remote sensing in Southern African savannas.*

EDWARDS, T. C., MOISEN, G. G. & CUTLER, D. R. 1998. Assessing map accuracy in a remotely sensed, ecoregion-scale cover map. *Remote Sensing of Environment,* 63**,** 73-83.

ELDEIRY, A. A. & GARCIA, L. A. 2010. Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using LANDSAT images. *Journal of Irrigation and Drainage Engineering,* 136**,** 355-364.

ESTES, J., BELWARD, A., LOVELAND, T., SCEPAN, J., STRAHLER, A., TOWNSHEND, J. & JUSTICE, C. 1999. The way forward. *Photogrammetric Engineering and Remote Sensing,* 65**,** 1089-1093.

EVANGELISTA, P. H., STOHLGREN, T. J., MORISETTE, J. T. & KUMAR, S. 2009. Mapping invasive tamarisk (Tamarix): a comparison of single-scene and time-series analyses of remotely sensed data. *Remote Sensing,* 1**,** 519-533.

FERWERDA, J. G. & SKIDMORE, A. K. 2007. Can nutrient status of four woody plant species be predicted using field spectrometry? *ISPRS Journal of Photogrammetry and Remote Sensing,* 62**,** 406-414.

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

FRIEDL, M., WOODCOCK, C., GOPAL, S., MUCHONEY, D., STRAHLER, A. & BARKER-SCHAAF, C. 2000. A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data.

GETIS, A. & ORD, J. K. 1992. The analysis of spatial association by use of distance statistics. *Geographical analysis,* 24**,** 189-206.

GHIMIRE, B., ROGAN, J. & MILLER, J. 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters,* 1**,** 45-54.

GIANINETTO, M., RUSMINI, M., CANDIANI, G., DALLA VIA, G., FRASSY, F., MAIANTI, P., MARCHESI, A., NODARI, F. R. & DINI, L. 2014. Hierarchical classification of complex landscape with VHR pan-sharpened satellite data and OBIA techniques. *European Journal of Remote Sensing,* 47**,** 229-250.

GIL, A., LOBO, A., ABADI, M., SILVA, L. & CALADO, H. 2013. Mapping invasive woody plants in Azores protected areas by using very high-resolution multispectral imagery.

GOEL, P., PRASHER, S., PATEL, R., SMITH, D. & DITOMMASO, A. 2002. Use of airborne multi–spectral imagery for weed detection in field crops. *Transactions of the ASAE,* 45**,** 443.

GONG, P., PU, R., BIGING, G. S. & LARRIEU, M. R. 2003. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on,* 41**,** 1355-1362.

GREEN, E., STRAWDERMAN, W. & AIROLA, T. 1993. Assessing classification probabilities for thematic maps. *Photogrammetric Engineering and Remote Sensing,* 59**,** 635-639.

HANTSON, W., KOOISTRA, L. & SLIM, P. A. 2012. Mapping invasive woody species in coastal dunes in the Netherlands: a remote sensing approach using LIDAR and high-resolution aerial photographs. *Applied Vegetation Science,* 15**,** 536-547.

HENDERSON, L. 2007. Invasive, naturalized and casual alien plants in southern Africa: a summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia,* 37**,** 215-248.

HENGL, T. 2007. *A practical guide to geostatistical mapping of environmental variables*.

HUDAK, A. & BROCKETT, B. 2004. Mapping fire scars in a southern African savannah using Landsat imagery. *International Journal of Remote Sensing,* 25**,** 3231-3243.

HUDAK, A. T. & WESSMAN, C. A. 1998. Textural analysis of historical aerial photography to characterize woody plant encroachment in South African savanna. *Remote Sensing of Environment,* 66**,** 317-330.

JANSSEN, L. L. & VANDERWEL, F. J. 1994. Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing;(United States)*, 60.

JOHNSTON, K., VER HOEF, J. M., KRIVORUCHKO, K. & LUCAS, N. 2001. *Using ArcGIS geostatistical analyst*, Esri Redlands.

JORDAAN, D. 2009. Bankruptbush (Slangbos)–a silent threat to grasslands. *Grassroots,* 9**,** 40-42.

JORDAAN, D. & JORDAAN, F. 2007. Probleemplante in the Suidwesgebied. *Ons Eie,* 42**,** 38-40.

JUSTICE, C., BELWARD, A., MORISETTE, J., LEWIS, P., PRIVETTE, J. & BARET, F. 2000. Developments in the'validation'of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing,* 21**,** 3383-3390.

KOKALY, R. F. & CLARK, R. N. 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote sensing of environment,* 67**,** 267-287.

LARK, R. 1995. Components of accuracy of maps with special reference to discriminant analysis on remote sensor data. *International Journal of Remote Sensing,* 16**,** 1461-1480.

LAURENCEAU, J. & SAGAUT, P. 2008. Building efficient response surfaces of aerodynamic functions with kriging and cokriging. *AIAA journal,* 46**,** 498-507.

LE MAITRE, D., VERSFELD, D. & CHAPMAN, R. 2000. Impact of invading alien plants on surface water resources in South Africa: A preliminary assessment.

LI, M., ZANG, S., ZHANG, B., LI, S. & WU, C. 2014. A review of remote sensing image classification techniques: the role of spatio-contextual information. *European Journal of Remote Sensing,* 47**,** 389-411.

LIANG, S. 2005. *Quantitative remote sensing of land surfaces*, John Wiley & Sons.

MADDEN, M. 2004. Remote Sensing and Geographic Information System Operations for Vegetation Mapping of Invasive Exotics 1. *Weed Technology,* 18**,** 1457-1463.

MALING, D. 1989. . Oxford: Pergamon Measurements from maps. *Principles and methods of cartometry*.

MERCHANT, J. W., YANG, L. & YANG, W. Validation of continental-scale land cover data bases developed from AVHRR data. Proceedings of the Pecora 12 Symposium, L and Information f rom Space-Based Systems, 1993. 63-72.

MILLER, J. & FRANKLIN, J. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling,* 157**,** 227-247.

MILLER, J., FRANKLIN, J. & ASPINALL, R. 2007. Incorporating spatial dependence in predictive vegetation models. *ecological modelling,* 202**,** 225-242.

MIRIK, M. & ANSLEY, R. J. 2012. Utility of satellite and aerial images for quantification of canopy cover and infilling rates of the invasive woody species honey mesquite (Prosopis Glandulosa) on rangeland. *Remote Sensing,* 4**,** 1947-1962.

MORISETTE, J., KHORRAM, S. & MACE, T. 1999. Land-cover change detection enhanced with generalized linear models. *International Journal of Remote Sensing,* 20**,** 2703-2721.

MULLER, S., WALKER, D., NELSON, F., AUERBACK, N., BOCKHEIM, J., GUYER, S. & SHERBA, D. 1998. Accuracy assessment of a land-cover map of the Kuparuk river basin, Alaska: considerations for remote regions. *Photogrammetric Engineering and Remote Sensing,* 64**,** 619-628.

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

MUTANGA, O. & RUGEGE, D. 2006. Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing,* 27**,** 3499-3514.

MUTANGA, O. & SKIDMORE, A. K. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing,* 25**,** 3999-4014.

MYINT, S. W., GOBER, P., BRAZEL, A., GROSSMAN-CLARKE, S. & WENG, Q. 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment,* 115**,** 1145-1161.

NUSSBAUM, S. & MENZ, G. 2008. *Object-based image analysis and treaty verification: new approaches in remote sensing-applied to nuclear facilities in Iran*, Springer Science & Business Media.

ODEH, I. O., MCBRATNEY, A. & CHITTLEBOROUGH, D. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma,* 67**,** 215-226.

PACKALÉN, P., TEMESGEN, H. & MALTAMO, M. 2012. Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. *Canadian Journal of Remote Sensing,* 38**,** 557-569.

PAN, G., GAARD, D., MOSS, K. & HEINER, T. 1993. A comparison between cokriging and ordinary kriging: case study with a polymetallic deposit. *Mathematical Geology,* 25**,** 377-398.

PETERS, A. J., REED, B. C., EVE, M. D. & MCDANIEL, K. C. 1992. Remote sensing of broom snakeweed (Gutierrezia sarothrae) with NOAA-10 spectral image processing. *Weed Technology***,** 1015-1020.

PIMENTEL, D., MCNAIR, S., JANECKA, J., WIGHTMAN, J., SIMMONDS, C., O'CONNELL, C., WONG, E., RUSSEL, L., ZERN, J. & AQUINO, T. 2001. Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems & Environment,* 84**,** 1-20.

PIMENTEL, D., ZUNIGA, R. & MORRISON, D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological economics,* 52**,** 273-288.

PYŠEK, P. & RICHARDSON, D. M. 2010. Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources,* 35**,** 25-55.

RICHARDSON, D. M., PYŠEK, P., REJMÁNEK, M., BARBOUR, M. G., PANETTA, F. D. & WEST, C. J. 2000. Naturalization and invasion of alien plants: concepts and definitions. *Diversity and distributions,* 6**,** 93-107.

RICHARDSON, D. M. & VAN WILGEN, B. W. 2004. Invasive alien plants in South Africa: how well do we understand the ecological impacts?

ROELOFSEN, H. D., KOOISTRA, L., VAN BODEGOM, P. M., VERRELST, J., KROL, J. & WITTE, J.-P. M. 2014. Mapping a priori defined plant associations using remotely sensed vegetation characteristics. *Remote Sensing of Environment,* 140**,** 639-651.

ROSENFIELD, G. H. & FITZPATRICK-LINS, K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric engineering and remote sensing,* 52**,** 223-227.

SANCHEZ-FLORES, E., RODRIGUEZ-GALLEGOS, H. & YOOL, S. 2008. Plant invasions in dynamic desert landscapes. A field and remote sensing assessment of predictive and change modeling. *Journal of Arid Environments,* 72**,** 189-206.

SHAFII, B., PRICE, W. J., PRATHER, T. S., LASS, L. W. & THILL, D. C. 2004. Using landscape characteristics as prior information for Bayesian classification of yellow starthistle. *Weed science*, 52**,** 948-953.

SKIDMORE, A., VAREKAMP, C., WILSON, L., KNOWLES, E. & DELANEY, J. 1997. Remote sensing of soils in a eucalypt forest environment. *International Journal of Remote Sensing,* 18**,** 39-56.

SMITS, P., DELLEPIANE, S. & SCHOWENGERDT, R. 1999. Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost-based approach. *International journal of remote sensing,* 20**,** 1461-1486.

SNYMAN, H. 2009. A philosophical approach to the distribution and spread of Seriphium plumosum. *Grassroots,* 9**,** 29-37.

SNYMAN, H. 2010. Allelopathic potential, seed ecology and germination of the encroacher shrub Seriphium plumosum. *African Journal of Range & Forage Science,* 27**,** 29-37.

STEHMAN, S. 1995. Thematic map accuracy assessment from the perspective of finite population sampling. *Remote Sensing,* 16**,** 589-593.

STEHMAN, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment,* 62**,** 77-89.

STEHMAN, S. V. 1999. Basic probability sampling designs for thematic map accuracy assessment. *International Journal of remote sensing,* 20**,** 2423-2441.

STEHMAN, S. V. & CZAPLEWSKI, R. L. 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment,* 64**,** 331-344.

STEHMAN, S. V., WICKHAM, J. D., YANG, L. & SMITH, J. H. Assessing the accuracy of large-area land cover maps: experiences from the multi-resolution land-cover characteristics (MRLC) Project. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 2000. 601-608.

SU, W., LI, J., CHEN, Y., LIU, Z., ZHANG, J., LOW, T. M., SUPPIAH, I. & HASHIM, S. A. M. 2008. Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. *International Journal of Remote Sensing,* 29**,** 3105-3117.

TURLINGS, L., BERGMANS, W. & BLOM, E. Invasive plants and animals: is there a way out? Invasive plants and animals: is there a way out? Proceedings of a Conference on Alien Invasive Species on the 26th September, 2000 in the National Museum of Natural History Naturalis in Leiden, the Netherlands., 2001. Netherlands Committee for IUCN, 10-18.

VALLEY, R. D., DRAKE, M. T. & ANDERSON, C. S. 2005. Evaluation of alternative interpolation techniques for the mapping of remotely-sensed submersed vegetation abundance. *Aquatic botany,* 81**,** 13-25.

VER HOEF, J. M. & BARRY, R. P. 1998. Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference,* 69**,** 275-294.

VITOUSEK, P. M. 1990. Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. *Ecosystem Management.* Springer.

WANG, H., ZHAO, Y., PU, R. & ZHANG, Z. 2015. Mapping Robinia Pseudoacacia Forest Health Conditions by Using Combined Spectral, Spatial, and Textural Information Extracted from IKONOS Imagery and Random Forest Classifier. *Remote Sensing,* 7**,** 9020-9044.

WEBSTER, R. & OLIVER, M. A. 2007. *Geostatistics for environmental scientists*, John Wiley & Sons.

WESSELS, K., PRINCE, S., ZAMBATIS, N., MACFADYEN, S., FROST, P. & VAN ZYL, D. 2006. Relationship between herbaceous biomass and 1-km2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa. *International Journal of Remote Sensing,* 27**,** 951-973.

WOODCOCK, C. E. & GOPAL, S. 2000. Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science,* 14**,** 153-172.

WOODCOCK, C. E. & STRAHLER, A. H. 1987. The factor of scale in remote sensing. *Remote sensing of Environment,* 21**,** 311-332.

WOODCOCK, C. E., STRAHLER, A. H. & JUPP, D. L. 1988. The use of variograms in remote sensing: II. Real digital images. *Remote Sensing of Environment,* 25**,** 349-379.

WULDER, M. & BOOTS, B. 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing,* 19**,** 2223-2231.

WULDER, M. & BOOTS, B. 2001. Local spatial autocorrelation characteristics of Landsat TM imagery of a managed forest area. *Canadian journal of remote sensing,* 27**,** 67-75.

XIE, Y., SHA, Z. & YU, M. 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology,* 1**,** 9-23.

XU, H., DING, H., LI, M., QIANG, S., GUO, J., HAN, Z., HUANG, Z., SUN, H., HE, S. & WU, H. 2006. The distribution and economic losses of alien species invasion to China. *Biological Invasions,* 8**,** 1495-1500.

YANG, J. & PRINCE, S. D. 1997. A theoretical assessment of the relation between woody canopy cover and red reflectance. *Remote Sensing of Environment,* 59**,** 428-439.

APPENDICES

This chapter contains field data sheets used in the field, the schedule planned and pictures taken during some of the activities in the field.

Appendix 1: Field sheet

Date:…………………………

Name of farm:……………….....Site No:…………..…..GPS:..……………………..

Name of Farmer:…………………….……….Contacts:……………………………………..

Type of enterprise (farming system and/or natural environment):….…………………………………………………………………….

Bankrupt bush (*Seriphium plumosum*) treatment control, Last date of treatment:………………………………….………………………………………

Density of Bankrupt bush per 20 x 20 meter plot

Plot No	Canopy No	Length	Height	Width	Canopy spacing	GPS coordinates	Soil type	Soil color	Photo ID	Other (observation)

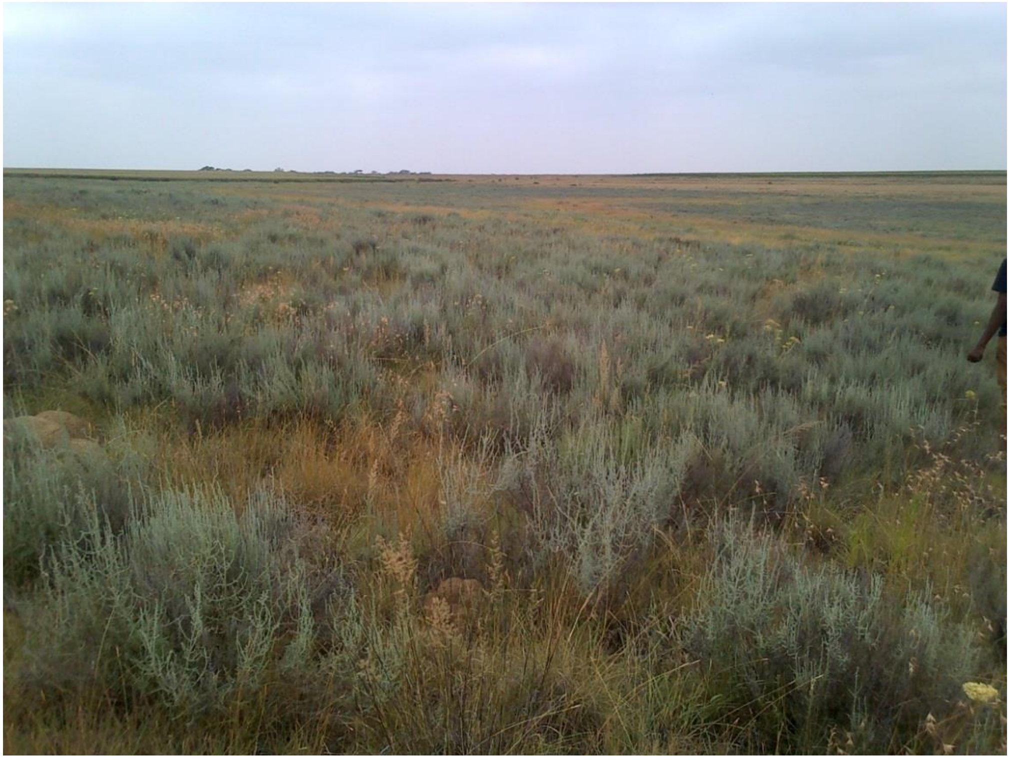
Notes:……………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

Appendix 2: Field work schedule

The field work will comprise locating, measuring and assessing the density of Bankrupt bush (*Seriphium plumosum*). We have selected plots in each property on which to carry out these tasks to save us time. The work is scheduled to take place as follows:

Date	**Properties to visit**
23/02/2015	Zyfer fontein; Hillyside; DeWildt; Rondebos; Nelville; Witsand; Witzand; Nieuwejaarsfontein; Aurora; Koekemoers rekwest;
24/02/2015	Jocador; Virginia; Virginia; Arundel; Bidsulphsberg; Canada; Cyferkuil; Liebenbergs bult.
25/02/2015	Astorea; Maclear; Frieden; Eenzaamheid; Groot Taaiboschfontein; Schurvekop
26/02/2015	Melsetter; Gethsemane; Waterval; Dupreezpoort; Langverwacht; Weiveld; Kalkoenkrans; Stoffelina; De la Harpe; Kranskop
27/02/2015	Deelfontein; Brakwater; Makwera; Liebenbergs bult.

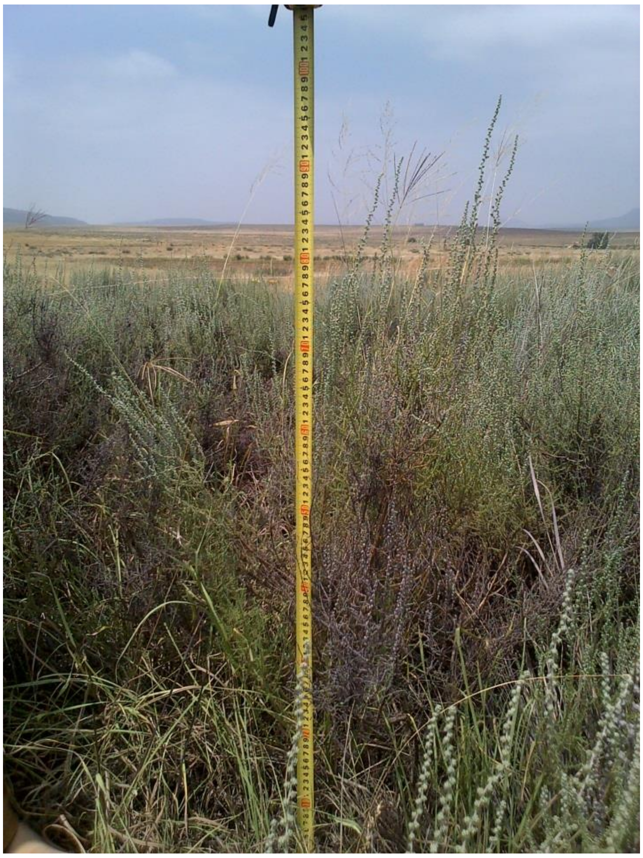
Appendix 3: Field pictures



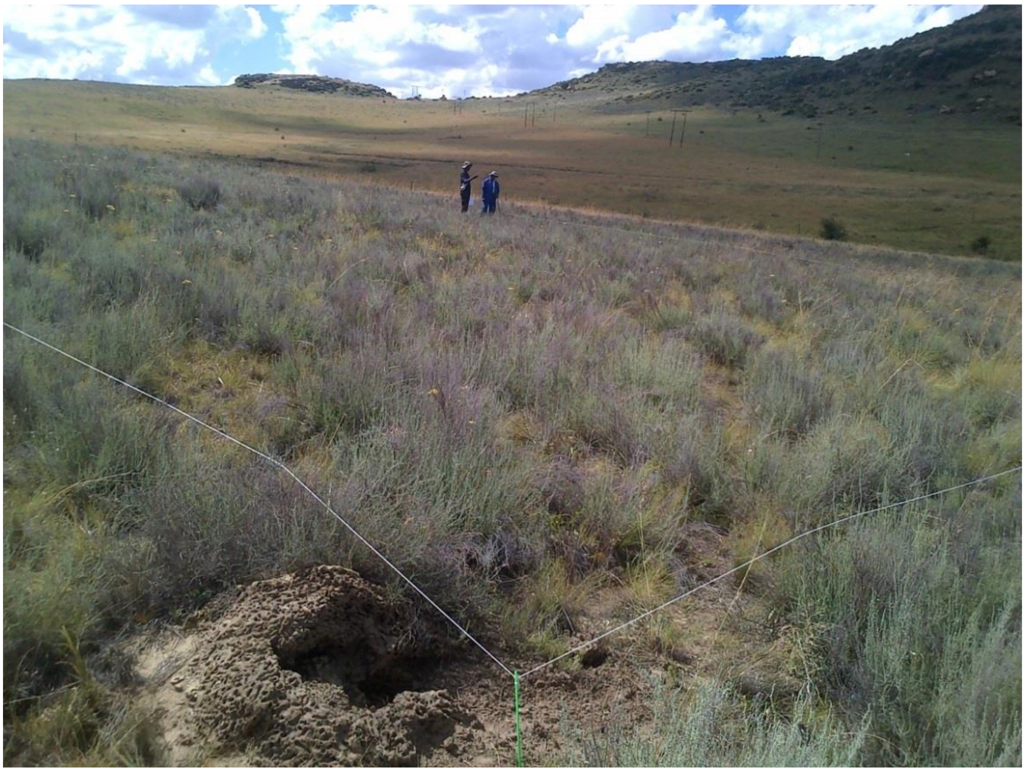
Picture 1: *Seriphium plumosum* infestation in a grassland.

Picture 2: Width/Length measurement of *Seriphium plumosum* canopy.

Picture 3: *Seriphium plumosum* canopy spacing measurement.



Picture 4: Height measurements of *Seriphium plumosum* canopy.



Picture 5: Demonstration of the constructed 20x20 meter plot in the field.

Picture 6: Biological control of *Seriphium plumosum.*

Picture 7: Fully dense *Seriphium plumosum* infested area.